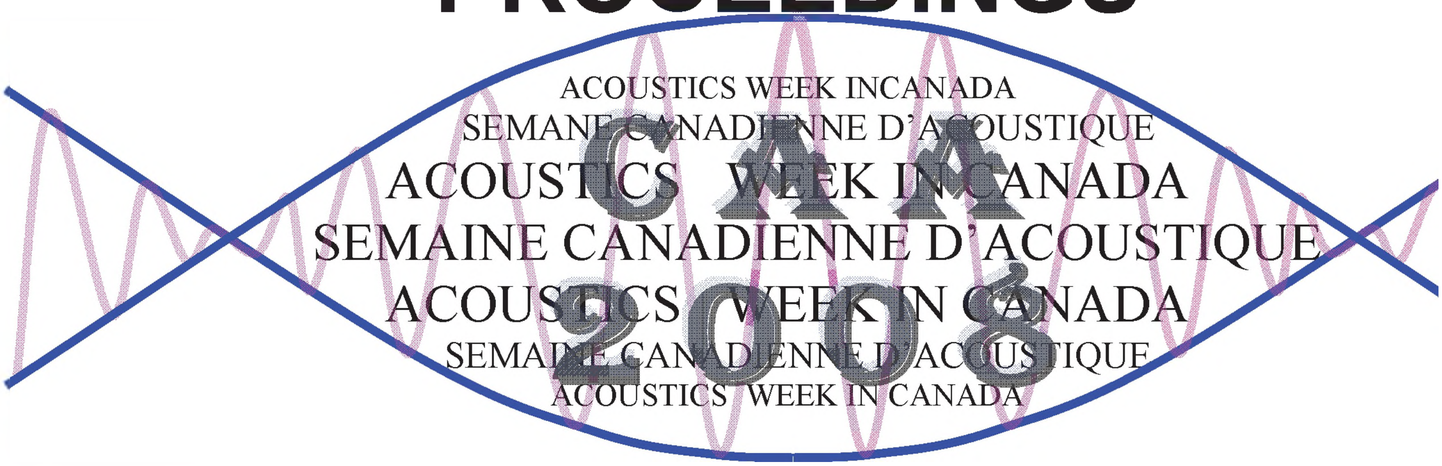
ISSN 0711-6659

canadian acoustics
acoustique canadienne

Journal of the Canadian Acoustical Association - Journal de l'Association Canadienne d'Acoustique

SEPTEMBER 2008 | SEPTEMBRE 2008
Volume 36 -- Number 3 | Volume 36 -- Numéro 3

PROCEEDINGS



COMPTES RENDUS

canadian acoustics

acoustique canadienne

THE CANADIAN ACOUSTICAL ASSOCIATION
P.O. BOX 1351, STATION "F"
TORONTO, ONTARIO M4Y 2V9

L'ASSOCIATION CANADIENNE D'ACOUSTIQUE
C.P. 1351, SUCCURSALE "F"
TORONTO, ONTARIO M4Y 2V9

CANADIAN ACOUSTICS publishes refereed articles and news items on all aspects of acoustics and vibration. Articles reporting new research or applications, as well as review or tutorial papers and shorter technical notes are welcomed, in English or in French. Submissions should be sent directly to the Editor-in-Chief. Complete instructions to authors concerning the required camera-ready copy are presented at the end of this issue.

ACOUSTIQUE CANADIENNE publie des articles arbitrés et des informations sur tous les domaines de l'acoustique et des vibrations. On invite les auteurs à soumettre des manuscrits, rédigés en français ou en anglais, concernant des travaux inédits, des états de question ou des notes techniques. Les soumissions doivent être envoyées au rédacteur en chef. Les instructions pour la présentation des textes sont exposées à la fin de cette publication.

CANADIAN ACOUSTICS is published four times a year - in March, June, September and December. The deadline for submission of material is the first day of the month preceeding the issue month. Copyright on articles is held by the author(s), who should be contacted regarding reproduction. Annual subscription: $20 (student); $60 (individual, institution); $300 (sustaining - see back cover). Back issues (when available) may be obtained from the CAA Secretary - price $10 including postage. Advertisement prices: $600 (centre spread); $300 (full page); $175 (half page); $125 (quarter page). Contact the Associate Editor (advertising) to place advertisements. Canadian Publication Mail Product Sales Agreement No. 0557188.

ACOUSTIQUE CANADIENNE est publiée quatre fois par année - en mars, juin, septembre et décembre. La date de tombée pour la soumission de matériel est fixée au premier jour du mois précédant la publication d'un numéro donné. Les droits d'auteur d'un article appartiennent à (aux) auteur(s). Toute demande de reproduction doit leur être acheminée. Abonnement annuel: $20 (étudiant); $60 (individuel, société); $300 (soutien - voir la couverture arrière). D'anciens numéros (non-épuisés) peuvent être obtenus du Secrétaire de l'ACA - prix: $10 (affranchissement inclus). Prix d'annonces publicitaires: $600 (page double); $300 (page pleine); $175 (demi page); $125 (quart de page). Contacter le rédacteur associé (publicité) afin de placer des annonces. Société canadienne des postes - Envois de publications canadiennes - Numéro de convention 0557188.

Acoustics Week in Canada
Semaine canadienne d'acoustique

Vancouver, British Columbia, 6 - 8 October / octobre 2008

SPONSORS • SPONSORS • SPONSORS

MAXXON

Dalimar Instruments Inc.

WorkSafe BC — Working to make a difference

Vibro-Acoustics / Control

Durisol

Can-Cell Industries Inc.

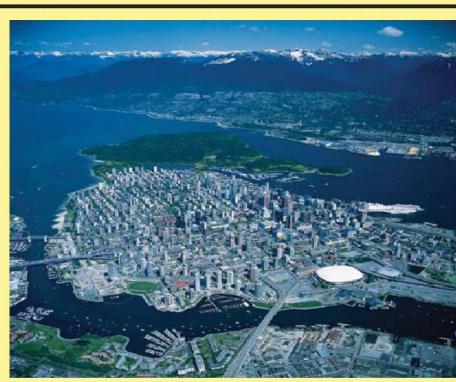
LMS — Engineering Innovation

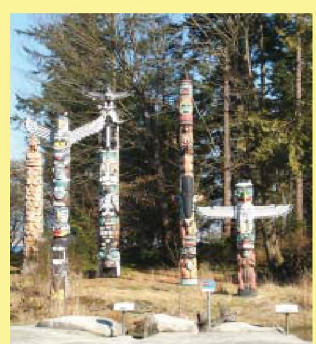
Blachford

SoundSorb

Brüel & Kjær

SPONSORS • SPONSORS • SPONSORS





Conference Chair / Président: Murray Hodgson [murray.hodgson@ubc.ca]
Technical Co-Chair / Directeur scientifique: Kimary Shahin [kns3@sfu.ca]

Conference Website at / Site Web du congrès à: www.caa-aca.ca/

GUEST EDITORIAL

En ma qualité d'organisateur, j'ai le plaisir de vous annoncer notre programme pour la Semaine canadienne d'acoustique 2008, qui prendra place à Vancouver en octobre. Grâce à l'aide enthousiaste de mon comité locale d'organisation, et au vif intérêt que vous avez témoigné au sein de la communauté canadienne d'acoustique, nous avons mis sur pied un formidable programme pour cette conférence. Vancouver est une ville très agréable à visiter au mois d'octobre. L'hôtel de la conférence est situé dans un emplacement idéal dans le quartier 'West End', à quelques pas de la plage de la Baie des Anglais ('English Bay') et proche du parc Stanley. Les participants seront accueillis par un aîné de la bande Premières Nations Musqueam. Le programme technique inclut des sessions plénières sur la production et troubles du débit, acoustique sous-marine, et paysages sonores. Il comprend 122 commun-ications portant sur plus de 20 sujets de recherche en acoustique et vibrations, allant de l'acoustique architecturale au débit à l'acoustique sous-marine. Cette publication contient les sommaires de deux pages de bon nombre de ces communications, ainsi que le programme détaillé de la conférence. L'exposition technique aura plus de 20 exhibitions. En plus du programme technique, nous avons prévu des activités sociales, incluant une visite guidée et récital d'orgue à la cathédrale Christ Church, une réception avant le banquet avec le groupe Coral and Em, le banquet de la conférence avec la présentation des prix suivi d'un concert par 4 musiciennes du groupe August. En bref, nous vous proposons 2½ journées d'activités acoustiques intéressantes et agréables. Je vous invite à nous rejoindre à Vancouver en octobre et espère avoir le plaisir de vous recevoir.

Murray Hodgson
UBC

As the general organizer, may I introduce you to our plans for Acoustics Week in Canada 2008, to be held in Vancouver in October. With the enthusiastic help of my local organizing committee, and the enthusiastic response of you in the Canadian acoustical community, we have put together a terrific conference program for your enjoyment. Vancouver is a great place to visit in October. The conference hotel is in a great location in the West End, steps from the beach at English Bay and close to Stanley Park. Delegates will be welcomed by an elder of the Musqueam First Nation. The exciting technical program includes plenary lectures on speech generation, underwater acoustics and soundscapes. It includes 122 technical papers in over 20 subject areas in acoustics and vibration, from architectural to speech to underwater. This proceedings issue contains two-page summaries of many of them, along with the detailed conference calendar. The equipment exhibition will have over 20 exhibitors. Besides the technical program, we plan a number of social activities, including a visit, talk and organ recital at Christ Church Cathedral, a pre-banquet reception to the background music of Coral and Em, and the conference banquet and award presentation, followed by a concert by the four singer-musicians of August. In short, we propose 2½ days of interesting and enjoyable acoustical activities. I invite you to join us in Vancouver in October, and look forward to welcoming you there.

Murray Hodgson
UBC

WHAT'S NEW ??

Promotions
Deaths
New jobs
Moves
Retirements
Degrees awarded
Distinctions
Other news

Do you have any news that you would like to share with Canadian Acoustics readers? If so, send it to:

QUOI DE NEUF ?

Promotions
Décès
Offre d'emploi
Déménagements
Retraites
Obtention de diplômes
Distinctions
Autres nouvelles

Avez-vous des nouvelles que vous aimeriez partager avec les lecteurs de l'Acoustique Canadienne? Si oui, écrivez-les et envoyer à:

Steven Bilawchuk, aci Acoustical Consultants Inc., Edmonton, Alberta, Email: stevenb@aciacoustical.com

Table of Contents/Tables des matières

Environmental Noise/Vibration

First Nations Languages Acoustics

'Green'-Building Acoustics

Measurement, Analysis and Musical Acoustics

Occupational Noise / Standards

Psychological Acoustics

Second Language Acquisition Acoustics

Sound Absorbing Materials

Speech Perception and Automatic Speech Recognition

Speech Production and Speech Disorders

Thermoacoustics

Underwater Acoustics

ACOUSTICS WEEK IN CANADA, Vancouver, 5-8 October 2008

--- Conference Technical Program ---

Sunday, October 5, 2008

4:30 PM – 9:30 PM	**CAA Board of Directors Meeting (Mountain Suite)**

Monday, October 6, 2008

7:30 AM – 4:00 PM	**Registration (Conference Foyer)**
8:15 AM – 8:30 AM	**Opening Ceremony & Welcome by Musqueam Elder Larry Grant (Nelson/Denman Rooms)**
8:30 AM – 9:30 AM	**Plenary Session (Nelson/Denman Rooms): The Origins of Constricted Voice Quality: The First Sounds of Speech; John H. Esling**
9:30 AM – 10:00 AM	**Coffee Break (Ballroom Foyer)**

	Nelson Room	Denman Room	Comox Room
	Second Language Acquisition Acoustics I Session Chair: Murray Munro	**First Nations Languages Acoustics I Session Chair: Kimary Shahin**	**Vibroacoustics Session Chair: Noureddine Atalla**
10:00 AM – 10:20 AM	'*I said made* (mate?)': Catalan/Spanish bilinguals' production of English word-final obstruents – Fullana & MacKay	A perception study of glottalization in Gitksan resonants – Lyon	Influence of dilatational propagating motion on diffuse field transmission loss of orthotropic sandwich composite panels – Ghinet & Atalla
10:20 AM – 10:40 AM	Effects of training modality on audio-visual perception of place of articulation and voicing in nonnative speech – Y. Wang et al.	Perturbations of pitch by ejectives in Upriver Halkomelem – Brown & Thompson	FE modeling of aircraft fuselage structure treated with viscoelastic material – Cintosun et al.
			Noise & Vibration Session Chair: Bill Gastmeier
10:40 AM – 11:00 AM	Learning Mandarin tones at sentence level through training: a pilot study – X. Wang	Tonogenesis in two Mayan languages: a contextual acoustic analysis – Arellanes et al.	Noise control in hydraulic system driven by Swash plate pump by optimizing control unit – Chikhalsouk & Bhat
11:00 AM – 11:20 AM	Acoustic realization and perception of English lexical stress by Mandarin learners – Lai et al.	The acoustic correlates of the unparsed: why we need more than a strong-weak distinction – Caldecott	Case study: Development of a high performance acoustical pipe lagging system – Kinart
11:20 AM – 11:40 AM	An acoustic study of stress in L2 production of German and Spanish – Miglio & Chun	Vowel quality and duration in Deg Xinag – Hargus	The evaluation of noise level in hand-held pneumatic tools (rock drill) by "PENEUROP CAGI TEST CODE" method – Majd & Nassiri
11:40 AM – 12:00 PM	The transfer of L1 acoustic cues in the perception of L2 lexical stress – Wang & Yoon	Conflict resolution in SENĆOŦEN: uvulars – /e/ sequences – Bird & Leonard	The evaluation of whole-body vibration level in hand-held pneumatic tools (rock drill) by "PENEUROP CAGI TEST CODE" method – Majd & Nassiri

12:00 PM – 1:00 PM	**Buffet Lunch (Ballroom Foyer)**

Monday, October 6, 2008 (continued)

	Nelson Room	Denman Room	Comox Room
	Environmental Noise I Session Chair: Bill Gastmeier	**Speech Production & Speech Disorders Session Chairs: Bryan Gick & Linda Rammage**	**Occupational Noise Standards Session Chair: Stephen Bly**
1:00 PM – 1:20 PM	An investigation into wind generated aero-acoustic tones – Ibbotson et ai.	Analysing coarticulation in Scottish English children and adults: an ultrasound study – Zharkova et al.	An overview of the standards and guidelines influencing the implementation of Noise Emission Declarations for Machinery – Bly & Haider
1:20 PM – 1:40 PM	Using a change in percent highly annoyed with noise as a potential health effect measure for projects under the Canadian Environmental Assessment Act: application to wind turbine noise – Keith et al.	A format frequency estimator for noisy speech based on correlation and cepstrum – Fattah et al.	ISO/IEC GUM applied to estimation of sound power measurement uncertainties – Keith
1:40 PM – 2:00 PM	Recent developments in environmental assessment methods for wind turbine noise – Howe & McCabe	Acoustic testing for phonologization – Shahin	Testing of Health Canada's acoustic chamber at the Consumer and Clinical Radiation Protection Bureau, based on ISO Standard 3745:2003 – Tsang et al.
2:00 PM – 2:20 PM	The role of source motion on noise generated by wind turbines – Richarz	Quantitative analysis of subphonemic flap / tap variation in NAE – Derrick & Gick	The ANSI Standard Standard S12.68-2007 method of estimating effective A-weighted sound pressure levels when hearing protectors are worn: A Canadian perspective – Voix et al
2:20 PM – 2:40 PM	Propagation of wind turbine noise in a boundary layer – Richarz	Tongue root retraction and tongue tip recoil in Xhosa alveolar click releases – Miller	CSA Appendix on measurement of noise exposure from headsets – Behar et al.
2:40 PM – 3:00 PM	Recent studies of infrasound from industrial sources – Gastmeier & Howe	**Coffee Break (Ballroom Foyer)**	
		Second Language Acquisition Acoustics II Session Chair: Yue Wang	**Classroom Acoustics Session Chair: Linda Rammage**
3:00 PM – 3:20 PM	**Coffee Break (Ballroom Foyer)**	Production of English lexical stress by inexperienced and experienced learners of English – He et al.	Acoustical evaluation of UBC non-classroom learning spaces – Hodgson & Villareal
	Sound Absorbing Materials Session Chair: Raymond Panneton		
3:20 PM – 3:40 PM	Pressure / mass method to mesure open porosity of porous solids – Salissou & Panneton	Durational properties of stressed syllables as a cue for English-accented French? – Guilbault	Ideal maximum noise levels for elementary school classrooms – Bradley
3:40 PM – 4:00 PM	Establishing relationships between acoustic and physical properties of shoddy-based fibre absorbers for acoustic modeling – Manning & Panneton	Do Mandarin-English bilinguals have an accent in their L1 vowel production? – Jiang	Statistical analysis of classroom questionnaires and acoustical parameters – Steininger & Hodgson
			Green Building Acoustics Session Chair: Maureen Connelly
4:00 PM – 4:20 PM	Fabrication of acoustic absorbing topologies using rapid prototyping – Godbold et al.	L2 English vowel learning by Mandarin speakers: Does perception precede production? – Thomson	Acoustical evaluation of six 'green' office buildings – Hodgson et al.
4:20 PM – 4:40 PM	Sound absorption of a microperforated panel backed by an irregular-shaped cavity – Wang et al.	An acoustic study of the L2 VGN rime production – Li	Sound transmission loss of green roofs – Connelly & Hodgson

	Acoustical Consulting—Current Challenges/Future Opportunities Session Chair: Clair Wakefield		
4:40 PM – 5:00 PM	Addressing the effects of overtopping vegetation on the performance of highway noise barriers – Wakefield	Mandarin speakers' productions of English vowels in real and pseudo words – Wu & Munro	Relationship between ventilation, air quality and acoustics in 'green' and 'grown' buildings – Khaleghi et al.
5:00 PM – 5:20 PM	The role of acoustics in sustainable design – Razavi	Variability in Cantonese speakers' production of English vowels – Munro	The acoustics of sustainable buildings – Richter et al.
5:20 PM – 6:00 PM			**Panel Discussion – Embedding acoustic into the integrated design process**

6:30 PM – 8:30 PM	**Christ Church Cathedral Tour, Talk and Organ Recital**

Tuesday, October 7, 2008

7:30 AM – 4:00 PM	**Registration (Conference Foyer)**
8:30 AM – 9:30 AM	**Plenary Session (Nelson/Denman Rooms): Studying the Sea with Sound; Stan Dosso**
9:30 AM – 7:00 PM	**Exhibition (Conference Foyer)**
9:30 AM – 10:00 AM	**Coffee Break (Ballroom Foyer)**

	Nelson Room	Denman Room	Comox Room
	Biomedical Acoustics Session Chair: Tarek El-Bialy	**First Nations Languages Acoustics II Session Chair: Jason Brown**	**Psychological Acoustics I Session Chair: Kathy Pichora-Fuller**
10:00 AM – 10:20 AM	Low intensity pulsed ultrasound stimulates ostegenic differentiation of human gingival fibroblasts – Mostafa et al.	Using acoustics to resolve a place controversy in Deg Xinag fricatives – Wright et al.	Effect of frequency in virtual rumble strips – Russo et al.
10:20 AM – 10:40 AM	Differential effect of therapeutic ultrasound on dentoalveolar structures during orthodontic force applications in-vitro (tension vs. compression forces – Lam et al.	Definitely indefinites? Using acoustics as a diagnostic in St'át'imcets – Caldecott & Davis	Listening with ear and hand: cross-modal integration in music perception – Maksimowski
10:40 AM – 11:00 AM	Modelling high frequency acoustic backscatter response from non-nucleated biological specimens – Falou et al.	Dentals are grave – Flynn & Fulop	A case of superior auditory spatial attention – Quelch et al.
11:00 AM – 11:20 AM	Finite element model of a SAW sensor for hydrogen detection – El-Gowini & Moussa	On the phonetics of schwa in Sliammon (M. Comox Salish): implications of the representations of Salish vowels – Blake & Shahin	The effect of informational masking and word position on sentence recall – Ezzatian et al.
11:20 AM – 11:40 AM	Cell expansion genes expression by therapeutic ultrasound: pros and cons – El Bialy & Aldosary	Tsilhqut'in ejectives: a descriptive phonetic study – Ham	The effect of types of acoustical distortion on lexical access – Pelletier et al.
11:40 AM – 12:00 PM	In-vivo ultrasound assisted tissue engineered articular condyle – El Bialy et al.	Laryngealized vowels in Quiaviní Zapotec: from phonetic realization to phonological contrast – Chávez-Peón	The effect of the degree of acoustical distortion on lexical access by younger adults – Coletta et al.

12:00 PM – 1:00 PM	**Buffet Lunch (Ballroom Foyer)**

Tuesday, October 7, 2008 (continued)

	Nelson Room	Denman Room	Comox Room
	Biomedical Acoustics II **Session Chair: Tarek El-Bialy**	**Environmental Noise II** **Session Chair: Bill Gastmeier**	**Psychological Acoustics II** **Session Chair: Kathy Pichora-Fuller**
1:00 PM – 1:20 PM	Ultrasound transmission through time-varying phononic crystals – Wright & Cobbold	The Harmonoise/Imagine method – van Leeuwen & van Banda	Is there a 'ramp archetype' of intensity in electroacoustics music? – Dean & Bailes
	Speech Perception & Automatic Speech Recognition I **Session Chair: Terrance Neary**		
1:20 PM – 1:40 PM	The temporal window of audio-tactile integration in speech perception – Gick & Ikegami	Typical hourly traffic distribution for noise modelling – VanDelden & Penton	Effects of emotional content and emotional voice on speech intelligibility in younger and older adults – Dupuis & Pichora-Fuller
1:40 PM – 2:00 PM	Spectral peaks as acoustic correlates to speech perception – Kiefte & Collins	Low Noise Pavement Technologies in the Canadian Context – A Review of Recent Studies Conducted in Alberta and Ontario – VanDelden, Penton & Haniff	Variability of the consonant modulation spectrum across individual talkers – Souza & Gallun
			Underwater Acoustics I **Session Chair: Stan Dosso**
2:00 PM – 2:20 PM	Acoustic modeling in automatic speech recognition – Deng	Implementing noise prediction standards in calculation software – the various sources of uncertainty – van Banda & Stapelfeldt	A semi-analytical approach to the study of the transient acoustic response of cylindrical shells – Leblond et al.
		Occupational Noise **Session Chair: Hugh Davies**	
2:20 PM – 2:40 PM	The temporal coordination of recited and spontaneous speech – Barbosa et al.	Noise exposure from communication headsets: the effects of environmental noise, device attenuation and preferred SNR under the device – Giguère & Dajani	Shock wave reflection and focusing phenomena in fluid-interacting shell systems – Iakovlev et al.
2:40 PM – 3:00 PM	The perceptual basis of velar epenthesis in Costa Rican Spanish – McLeod	Occupational noise exposure assessment using perceived and quantitative measures – Neitzel et al.	An efficient 3D PE propagation model with tessellation – Austin et al.
3:00 PM – 3:20 PM	Analysis and classification of a vowel database – Assmann et al.	Adjusting historical noise estimates by accounting for hearing protection use: a probabilistic approach and validation – Sbihi et al.	Array element localization of a bottom-mounted hydrophone array using ship noise – Ebbeson & Thierion
3:20 PM – 3:40 PM	**Coffee Break (Ballroom Foyer)**		
	Speech Perception & Automatic Speech Recognition II **Session Chair: Michael Kiefte**	**Measurement & Analysis** **Session Chair: Murray Hodgson**	**Underwater Acoustics II** **Session Chair: Michael Wilmut**
3:40 PM – 4:00 PM	Pattern recognition in speech perception research – Nearey	The noise scales and their units – Dehra	Comparison of measured and modelled transmission loss in Emerald Basin – Maranda & Collison
4:00 PM – 4:20 PM	Mismatch between the production and perception of F0 in New Zealand English ethnolects – Szakay	Measuring the displacement of engines using acoustics – Peppin	Bayesian source tracking and ocean environmental inversion – Dosso
4:20 PM – 4:40 PM	Processing of Japanese pitch accent by native Japanese and English listeners – Tu et al.	A cepstral-domain algorithm for pitch estimation from noise-corrupted speech – Shahnaz	Source tracking in an unknown ocean environment – Wilmut & Dosso
4:40 PM – 5:00 PM	The McGurk effect affected by the Right Ear Advantage – Scott	FFT Tutor: A MatLab-based instructional tool for FFT-parameter exploration – Singh et al.	Geoacoustic inversion of noise from ships-of-opportunity with unknown position – Tollefsen & Dosso

Tuesday, October 7, 2008 (continued)

5:00 PM	**AGM (Nelson Room)**
6:00 PM	**Reception (Ballroom Foyer), with performance by Coral and Em**
7:00 PM	**Banquet and Award Presentations (Denman/Comox Rooms)**
8:30 PM	**Concert by August (Denman/Comox Rooms)**

Wednesday, October 8, 2008

7:30 AM – 8:30 AM	**Registration (Conference Foyer)**
8:30 AM – 9:30 AM	**Plenary Session (Nelson/Denman Rooms): From Micro to Macro: Composing Microsound and Soundscapes; Barry Truax**
9:30 AM – 10:00 AM	**Coffee Break (Ballroom Foyer)**

	Nelson Room	Denman Room	Comox Room
	Auditory Scene Analysis Session Chair: Claude Alain	**Thermoacoustics Session Chair: Luc Mongeau**	**Architectural Acoustics Session Chairs: John Bradley & Murray Hodgson**
10:00 AM – 10:20 AM	Auditory scene analysis as a system – Bregman	Performance characterization of a small-capacity thermoacoustic cooler for air-conditioning applications – Paek et al.	Fire resistance and noise control in multi-family buildings – Quirt & Nightingale
10:20 AM – 10:40 AM	The role of attention in the development of auditory scene analysis – Sussman	A perfectly matched layer technique for Lattice Boltzmann Method – Najafi-Yazdi & Mongeau	Reverberation rooms and spatial uniformity – Ramakrishnan & Grewal
10:40 AM – 11:00 AM	Psychophysical and neuroelectric evidence of ontext effects in auditory stream segregation – Snyder	Numerical model of a thermoacoustic engine – Hireche et al.	Ray-tracing prediction of sound-pressure fields in empty and fitted rooms – Hodgson et al.
		Musical Acoustics Session Chair: Chris Waltham	
11:00 AM – 11:20 AM	The perception of auditory continuity with interrupted masking sounds – Ciocca & Chang	Structural segmentation of popular music with partially supervised clustering – Graves & Pedrycz	Speech privacy classes for rating the speech security of meeting rooms – Bradley & Gover
11:20 AM – 11:40 AM	Organising by object: How auditory memory can be structured within complex scenes – Dyson	Vibrational characteristics of harp soundboards – Waltham	The effects of room acoustics on speech privacy of meeting rooms – Bradley et al.
11:40 AM – 12:00 PM	Aging and the cocktail party problem: where is the problem? – Alain	Measurement and calculation of the Santur parameters – Heydarian	Prediction of Speech Transmission Index in eating establishments – Nahid

12:00 PM – 1:00 PM	**Buffet Lunch and Student-Presentation Awards (Ballroom Foyer)**
1:30 PM – 4:30 PM	**CSA Meeting (English Bay Suite)**

THE ROLE OF ACOUSTICS IN SUSTAINABLE DESIGN

Zohreh Razavi, M.A.Sc. EIT, LEED AP
Stantec Consulting Ltd. 111 Dunsmuir Street, Vancouver, BC, V6B 6A3.

1. INTRODUCTION

Good acoustics contribute to a quality work environment but it can pose challenges for sustainable designers. In fact, some of the most common practices associated with green design can actually negatively impact the acoustical performance of our workspaces. The solution, however, could be right above us.

Some green buildings have insufficient sound-absorbing materials due to considerable use of radiant chilled and cooling slabs and transparent envelope. This may cause excessive reverberation, resulting in an acoustical environment which "feels" noisy and can result in impaired verbal communication.

As architects continually strive to incorporate sustainability into their designs, the need for integrated design of all systems, including acoustics, becomes increasingly important. Since it has been investigated that one of the most important factors in greening a building is to provide an environment in which people can perform at their optimum level, designers must acknowledge that acoustics is a fundamental concern that can greatly contribute to the overall comfort level of a space and employee productivity.

The main purpose of this paper is to discuss some acoustical challenges in sustainable designs.

2. METHOD

2.1 Overview

Four projects with sustainable concepts of open-plan offices with reflected or acoustical ceiling were studied. The complaints from the occupants were the main reason for the studies. The size of the offices varied from 500 f^2 to 2000 f^2 with noise levels of NC 40 to > NC 65 at occupied situation.

2.2 Study cases

The projects were:

1. Open plan office, T-bar acoustical ceilings, carpeted floor, floor area of 1500 f^2 and adjacent to a transformer room, located in Fort Collins;
2. Open plan office, T-bar acoustical ceilings, carpeted floor, floor area of 1200 f^2 and adjacent to a mechanical room, located in Vancouver;
3. Open plan office, reflected ceilings, hard floor, floor area of 2000 f^2 and exposed ducting, located in Phoenix;
4. Home office, reflecting surfaces, floor area of 500 f^2, located in Richmond.

2.3 Project procedures

Each job site was visited in order to investigate the subjective perception of acoustical quality within the spaces. Prior to visit, plan drawings were reviewed to save time on sites. The noise measurements at a typical sitting positions for employees and at various operation levels for the noise sources were conducted. The measured octave band sound pressure levels were plotted and the equivalent NC levels within each space were found. The results were compared with NC levels recommended by ANSI standards for office environment. Recommendations for acoustical improvement within each space were provided.

2.4 Acoustical analysis procedure

During the acoustical measurements, noise sources (e.g. mechanical or electrical systems) were operating at different loads. The final recommendations were based on the normal operation at working hours. The recommendations were made in order to improve the acoustical environment in the open-plan office areas and increase the job performance of the employees which is the main objective of sustainable designs.

3. RESULTS

3.1 Acoustical measurements

Figure 1 shows the octave band sound pressure levels at octave band frequency of 63 Hz to 8000 Hz. These sound pressure levels were plotted along with NC curves in order to evaluate the noise criteria within each space. The sound pressure level values are the average sound pressure levels that have been measured within each space. As can be seen:

1. The NC level in the Vancouver office was NC 46 with the highest noise level at 125 Hz. Due to locating the office area adjacent to the mechanical room and locating the return air in the mechanical room beside the fan, the highest noise level in the office area was measured at blade pass frequency of the fan.

2. The NC level in the Phoenix office was NC 50 with higher noise levels at lower frequencies. All surfaces in this office were reflective with exposed ducting. The ducting was not internally lined and the "hum" out of the mechanical system was clearly perceptible by the employees.

3. The NC level in the Richmond home office was NC 55 with almost linear sound pressure level at all octave band frequencies. The noise sources in this office were the owner's kids' activities which was perceptible in the home office through

stairways. All surfaces in the office and stairways were reflective.

4. The NC level in the Fort Collins office was exceeding NC 65 which is well over the ANSI standard for the open offices. The peak was measured at 250 Hz, which is the harmonic of the 60 Hz fundamental operating frequency of the transformer voltage. This was due to the relative location of the office area to the transformer room. The transformer room was adjacent to the open-plan office area with noise leaking a path around the perimeter of its door. No vibration isolations measures were provided underneath the transformers. Thus, low frequency vibration was perceptible by even touching the floors. Return air in the office area were terminated in the transformer room and provided another path for noise transmission into the office area.

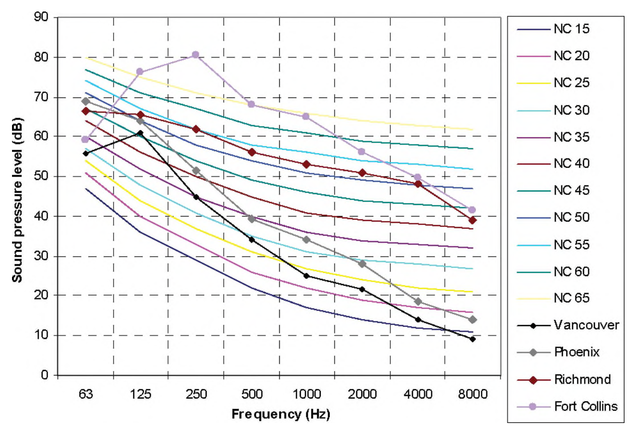


Figure 1. The sound pressure levels measured in dB, at octave band frequencies, in Vancouver, Phoenix, Richmond and Fort Collins open-plan offices.

3.2 Acoustical analysis

The high sound pressure level at 250 Hz in the Fort Collins office was an indication of noise transmission from the transformer room to the office area. This is an indication of not considering the basic principal of not locating noise sensitive areas adjacent to noisy areas, unless providing a buffer zone in between.

The same problem has been experienced in the Vancouver office by locating the mechanical room adjacent to the office areas. One common mistake would be considering mechanical noise for masking and ensuring speech privacy in open-plan offices. Mechanical systems have main components at lower frequencies; however, speech frequencies are at mid frequencies.

Eliminating absorptive panels from the office areas makes the offices echoy and noisy. In an echoy environment, concentration will be hard for employees which can negatively affect the employees' job performances. Mechanical ducting without internal lining could be sources of noise, especially at higher air velocities. These problems were experienced in Phoenix open-office and made the employees feel frustrated after a few hours of working within this environment.

Conserving energy and working from home at home offices is one of the sustainable solutions for the environment as long as all office requirements are being met. In office areas, employees who rated their workplace acoustics negatively are dissatisfied with their working environment. No acoustical treatment in the home office, instead of all of the reflective surfaces that the Richmond office has experienced, made the place ineffective. Optimal acoustical quality is a must in the office area which is located close to noise source and is being used for conference calls.

4. DISCUSSION

Open-plan offices are always acoustical challenges, especially with sustainable design goals of eliminating pours materials or minimizing them, and exposed ducting. However, one of the most important benefits to "greening" a building, aside from reduced environmental impacts, is the ability to provide an environment in which people can perform at their optimum level. Any gain in occupant productivity translates into enhanced building sustainability. Widespread dissatisfaction with the acoustics in workplaces can lead to costly errors in communication and reduced productivity due to distractions. Nevertheless, sustainable design does not have to occur at the expense of appropriate acoustics. Many acoustical products, such as ceiling tiles, insulation and carpeting, are recyclable or manufactured from recycled content and can in fact help projects meet sustainability goals. New sustainable acoustical products can solve acoustical problems in working areas. They can improve working environments by contributing to optimal performance and communication, thus helping to meet the objectives of sustainable construction.

REFERENCES

- ASHRAE Fundamentals Handbook (2007) Chapter 47, Sound an Vibration Control,
- ASTME E33.02 (1998), Open office acoustics, White Paper, ASTM Standardization News, 26 (12), 39-47,
- ASTME E 1374 (2006), Standard Guide for Open Office Acoustics and Applicable ASTM Standards,
- ANSI S12.60 (2002), Classroom Acoustics, Acoustical Performance Criteria, Design Requirements, and Guidelines for Schools.

ACOUSTICAL EVALUATION OF UBC NON-CLASSROOM LEARNING SPACES

Murray Hodgson and Jorge García Moreno Villarreal

Acoustics & Noise Research Group, SOEH-MECH, University of British Columbia, 3rd Floor, 2206 East Mall, Vancouver, BC, Canada V6T 1Z3 [murray.hodgson@ubc.ca]

1. INTRODUCTION

At the University of British Columbia (UBC), students share ideas, do homework, read, write, study, compute, discuss live or on cell-phones, or otherwise interact with fellow students to the benefit of learning in spaces such as lobbies, hallways, eateries, cafés, labs, libraries and common rooms. These areas are non-classroom learning spaces (NCLS). The main objective of this research project was to evaluate non-classroom learning spaces located in buildings at UBC, to determine the quality of their acoustical environments and how to improve their designs. Spaces were chosen for study based on the following criteria: be used by students for learning purposes; have capacity for at least 40 users; have a constant flow of users. In the end, 23 spaces in 11 buildings were studied. They were libraries, academic common areas, coffee shops, eateries and dedicated study spaces. The evaluation assessed the spaces by questionnaire survey and physical-acoustical measurement.

2. METHODOLOGY

2.1 Questionnaire

A questionnaire was developed for the study. Its length had to give a response time of 5-7 minutes. Both acoustical and non-acoustical questions were included, as follows:

- Non-acoustical: respondent demographics (sex, number in group, wearing of earplugs/headphones, current learning activities); perceived overall quality of the NCLS environment; perceived quality of non-acoustical aspects of the environment (lighting, air, temperature, furniture).
- Acoustical: perceived overall quality of the acoustical environment; effects of aspects of the acoustical environment; awareness of the acoustical environment; positive and negative consequences of the acoustical environment.

2.2 Physical-Acoustical Parameters

The following physical-acoustical parameters were measured:

- Noise levels. Equivalent-continuous noise levels in the 63-8000 Hz octave bands were measured. In the unoccupied spaces total, A-weighted and NC levels were determined. In the occupied spaces, total, A-weighted and NC(B) levels were determined;
- Reverberation time at mid-frequencies (**RT_{mid}**);
- Speech Intelligibility Index (**SII_n**), calculated at a receiver position from the normal-voice speech level, occupied noise levels and unoccupied RT_{mid}'s. SII_n at 1 m (SII_n1) was used to assess speech intelligibility. SII_n at 4 m (SII_n4) was used to assess speech privacy.

2.3 Acceptability Criteria

The evaluation criteria chosen for this study were adopted from various sources [1, 2, 3]:

- **RT_{mid}.** The spaces evaluated in this project are not classrooms, and are larger than those considered in classroom standards, sometimes considerably larger. Moreover, students experience the occupied space, and student absorption reduces the RT of the unoccupied space. Thus, for the non-classroom learning spaces in this study, RT<1.0 s was considered acceptable, and values less than 0.7 s were considered excellent;
- **SII_n.** Speech intelligibility was considered acceptable for SII_n1 values of 0.5-0.75; above 0.75, it was considered excellent. For speech privacy, SII_n4 values ranging from 0.10-0.20 were acceptable; values below 0.1 were considered excellent;
- **NC_u.** Continuous noise (mainly generated by mechanical services) in an unoccupied learning space should be in the range NC 25-30. Values below NC 25 were considered excellent;
- **NCB_o.** Continuous noise in an occupied learning space should not exceed NC(B) 40; values below NC(B) 35 were considered excellent;
- **$dBA_{u,o}$.** Values of total, A-weighted level up to 40 dBA were acceptable for unoccupied learning spaces, and up to 47 dBA for occupied learning spaces.

2.4 Test Protocol

Four visits were made to each space. For visits 1, 2 and 3 (occupied NCLS), noise levels were measured and the questionnaire administered in the periods 9:30-11:00, 12:00-14:00 and 14:00-16:00. For visit 4 (unoccupied NCLS), all physical measurements were performed.

3. RESULTS

3.1 Questionnaires

850 questionnaires were analyzed. The average results for each space were calculated, and spaces with better or worse quality identified. Average responses for each question were also calculated. Following are the main results: the learning activities reported most often were thinking and reading; lighting, air, temperature and furniture comfort generally enhanced learning—the acoustical

environment interfered with it; people moving and talking was the aspect of the acoustical environment that most impaired learning, followed by intermittent noise; distraction was the most reported negative consequence of the acoustical environment, followed by annoyance; difficulty hearing and talking were reported least; feeling relaxed was the most reported positive consequence of the acoustical environment, followed by feeling productive; conversational privacy was reported least; 22% of respondents reported that they chose their study location because of the acoustical environment; in most cases they chose a quiet location.

3.2 Statistical Analysis

Correlation

In order to observe if there were any apparent relationships between the questionnaire responses and the measured physical-acoustical parameters, Pearson's correlation coefficients between all data pairs were calculated. Values >0.2 in absolute value were considered significant and their apparent implications deduced.

Considering first only the questionnaire responses (note: 'satisfaction' refers to the perceived extent to which learning was interfered with or enhanced): no responses were correlated with the time of day or respondent sex; overall satisfaction with the learning environment was associated with increased satisfaction with people talking and moving, continuous noise and intermittent noise; overall satisfaction with the learning environment was associated with increased experiencing relaxed, energized and productive; overall satisfaction with the learning environment was associated with decreased distraction; satisfaction with lighting was associated with feeling productive satisfaction with air quality was associated with feeling relaxed; satisfaction with furniture comfort was associated with feeling relaxed and productive; satisfaction with the acoustical environment was associated with increased satisfaction with people talking and moving, continuous and intermittent noise and reverberation; satisfaction with the acoustical environment was associated with decreased annoyance, distraction, stress and difficulty hearing; satisfaction with people talking and moving, continuous and intermittent noise and reverberation were mutually correlated; satisfaction with people talking and moving, continuous and intermittent noise and reverberation were associated with decreased annoyance, distraction and stress; experiences of annoyance, distraction, stress, fatigue, difficulty hearing and difficulty talking were correlated; experiences of conversational privacy, and of feeling relaxed, energized and productive were correlated.

Second, considering only the physical-acoustical parameters: all noise levels and SII_n values were correlated; RT_{mid} was only correlated with SII_n4.

Finally, considering both the questionnaire responses and the physical-acoustical parameters: when noise levels were lower, students were more likely to be involved in reading; when noise levels were higher and SII_n's lower, people were less satisfied with the overall learning environment and with furniture comfort, were more likely to be involved in discussion, to work in groups, to report more difficulty hearing, slightly more difficulty talking, and to feel less productive, and were more likely to choose their study location because of the acoustical environment.

Regression analysis

Various multivariable linear-regression models were developed to predict the response to the question, "How well does the environment *in general* in this learning space interfere with or enhance your ability to use this space for your activities?" on a scale from -3 (interferes a lot) to +3 (enhances a lot) (variable *env_gen*). First, using only the other questionnaire responses as predictors, an optimal model which had an adjusted-R^2 of 0.48 was found. Second, using only the physical parameters as predictors, an optimal model which had an adjusted-R^2 of 0.19 was found. The best model, with an adjusted-R^2 of 0.53, was developed using both the questionnaire responses and the physical-acoustical parameters: $env\_gen = 0.151\ light + 0.126\ furn + 0.264\ people + 0.174\ prod + 0.401\ acoust - 0.014\ BNA_u + 0.348\ RT_{mid} + 2.188\ SII_n4$, in which *light* quantifies the perceived quality of the lighting, *furn* quantifies the perceived comfort of the furniture and *people* quantifies satisfaction with people talking and moving (on the same scales as *env_gen*); *prod* quantifies the reported feeling of productivity on a scale from 0 (not at all) to 5 (a lot), *acoust* = 1 if respondents chose their study location because of the acoustical environment and 0 if not. BNA_u is the total, A-weighted unoccupied noise level, RT_{mid} is the unoccupied mid-frequency RT, and SII_n4 is the normal-voice SII at 4 m.

4. DISCUSSION

According to the above regression model, occupant satisfaction with overall environmental quality can be improved by improving the lighting and furniture comfort, ensuring that people talking and moving are not a disturbance, decreasing noise levels and increasing speech privacy. The positive coefficient of the RT_{mid} term suggests that environmental quality can be increased by increasing the mid-frequency RT; however, it also increases with SIIn4 which decreases with increase RT, so the effect of reverberation is not simple.

REFERENCES

[1] ANSI S12.2-1995. American National Standard Criteria for Evaluating Room Noise. Acoustical Society of America, New York, 1995.

[2] Interim Sound and Vibration Design Guidelines for Hospital and Healthcare Facilities, Acoustical Society of America, New York, 2006.

[3] Noise and Vibration Control Engineering – 2nd edition, I. Ver, ed. (John Wiley & Sons, New York, 2005).

STATISTICAL ANALYSIS OF CLASSROOM QUESTIONNAIRES AND ACOUSTICAL PARAMETERS

Gavin Steininger and Murray Hodgson
Dept. of Mechanical Engineering, University of British Columbia, 2054-6250 Applied Science Lane, Vancouver, V6T 1Z4

1. INTRODUCTION

The acoustical conditions of a classroom may have a profound impact on the ability of occupants of that space to teach and learn. Within the activities of teaching and learning there seem to be several sub-categories of activities; these are lecturing, group work and independent work such as writing exams. For the first two, being able to communicate verbally is important, whereas for individual work a distraction-free environment is more important. This report will focus on the issue of effective verbal communication.

University classrooms have several acoustical attributes that are not common to other spaces. First, in a classroom the talker is, in general, farther from his or her audience than the distance at which he or she would normally endeavor to communicate in most other settings. Second, the communication is normally in the form of a monologue; consequently most signals generated by the audience can be classified as noise. Third, the noise sources of the classroom are commonly closer to the receiver than the talker. Finally, university classrooms tend to be large spaces.

On account of the unusual needs of classrooms it follows to develop methods to evaluate and predict the acoustical performance and optimize their designs. In order to accomplish this, it is necessary to have an objective measure of the acoustical quality of a classroom. The method used in this work to create an objective measure of the quality of a classroom was to first consider the subjective quality of the classroom as reported by its users. This was done through a survey-based metric, the Perceived Listening Ease (PLE) questionnaire [1]. The details of both the administration of the PLE survey form and the calculation of the PLE score will be described below. For the purposes of this summary it is only necessary to understand that a PLE score is calculated from questionnaire responses is believed to represent the best estimate of an occupant's subjective assessment of the acoustical quality of a space. The average PLE score of a space is consequently used as an estimate of its objective quality. This implicitly assumes that an individual PLE score is drawn randomly from a distribution whose mean is dependent on the classroom for which the PLE score was calculated. The primary objective of this report is to explore the relationship between the physical-acoustical attributes of a classroom and its mean PLE score, in order to evaluate the usefulness of PLE as a measure of classroom quality and, given that PLE is useful, to evaluate the performance of classrooms. It is also of interest to find a forward prediction model for the performance of a classroom which can be used to optimize the design.

2. METHOD

2.1 Data collection

The information used in this report was gathered from questionnaires distributed to students attending classes, and physical measurements taken in the classrooms.

The student questionnaire consisted of 73 questions. These questions were divided into three sections. These were the demographic information of the student, the student's evaluation of the listening environment in the classroom, and the student's evaluation of the course and its presentation. The questionnaires were completed during class time in the classroom that each questionnaire evaluated.

The physical measurements were taken in the classrooms during their unoccupied state. Simple physical measurements such as surface area and volume were derived from building schematics where possible and measured otherwise. The direct acoustical measures included background noise level (BNL), early decay time (EDT), reverberation time (RT-20) and early-late energy fraction (C50); these measurements where taken for each octave from 125 Hz to 8000 Hz. A list of the definitions of the physical measurements is given in Table 1. The measurements were made using the calibrated MLSSA or Win MLS systems and an omni-directional loudspeaker source. These measurements were repeated at nine locations on a 3-by-3 grid in each classroom.

2.2 Participants

Seventeen classrooms were considered in the study. Fourteen of these were seven pairs of pre- and post-renovated rooms. Two were identical rooms, one of which had been renovated; however only the renovated information is available. The final room was a newly constructed classroom. From the instructors of the seventeen classrooms contacted, 65 agreed to participate in the study. Each instructor participated with at least one class; however several instructors gave permission for surveys to be administered to multiple classes that they instructed. The result is that 82 class sections, with 4882 students, participated in the study.

The average age of students participating in this study at time of survey was 20.9 years of age. The number of student participants who identified themselves as male was 2847 (58.4%), the number of females was 1937 (39.8%); 98 students did not answer the question. Of the students surveyed, 1804 (37%) reported being in the Arts and 1555 (31.9%) reported being in the Sciences; the remaining 1522 students reported various specializations. The average year of study for undergraduate students at time of survey was 2.4 years. There were also 32 graduate students who responded to the survey. The number of ESL students in the study was 2076 (42.6%); the number of native English speakers was 2656 (54.5%). The students reported their usual location in the classroom as one of 30 possible locations on a 6-by-5 grid. For reasons of convenience and integrity it was necessary to truncate this down to four possible locations. These were in the front third of the room, the middle third of the room, the back third of the room, or when a respondent stated that they had no preferred location in the classroom. Students were distributed in these locations as follows: 1965 (40.3%) students sat in the front, 1767 (36.2%) in the middle, 716 (14.7%) in the back and 434 (8.9%) gave no preference. Only 136 (2.8%) students reported having a hearing impairment.

The global average PLE score was 62.6; the global standard deviation of PLE scores was 17. The average score for male respondents was 63.5, for female respondents it was 62.0. The average response from hearing-impaired respondents was 59.6, from non-hearing-impaired respondents it was 62.7. Similarly, students who suffered from repeated ear infections also rated the PLEs of classrooms lower; these students had an average PLE score of 60.6; students who did not have multiple ear infections had an average PLE of 62.7. For ESL students the average PLE response was 60.6; for native English speakers the average score was 64.4.

3. STATISTICAL ANALYSIS

3.1 Covariate selection

On account of the large impact of the demographic information on PLE score it was necessary to block for the demographic information of the respondents, and then to model the residuals against the physical-acoustical measurements of the classrooms.

An iterative holdout procedure was used to select the covariates. The selected variables were, the student's assessment of the air quality, lighting, seating and temperature of the classroom; the student's assessment of the instructor's accent, articulation and volume; whether the student is a native English speaker; the student's assessment of the course's difficulty and interest and; does the student do the pre-assigned work for the course.

3.2 Physical attributes selection

The multivariate model for the residuals of the demographic information and the physical attributes of the classrooms was selected using a step AIC procedure. The selected variables are shown in Table 2. A ".2" indicates that a variable was squared before the linear modeling process was preformed. The ".V" and ".VS" indicate an interaction with volume and the two way interaction between volume and surface area respectively.

4. SUMMARY OF FINDINGS

The first significant finding is that the effect of renovations on the PLE scores of the classrooms is almost always positive (increased acoustical quality). Another important observation is that rooms that have good acoustical qualities also have good non-acoustical qualities. This seems to imply that a person's perception of the acoustics of a space is linked to their perception of non-acoustical attributes.

With regards to comparisons that can be made between the PLE scores of the different classrooms, the distribution of the PLE scores for each of the classrooms appear to only differ in median values. Notably, the shape of the distributions, and the variances of the PLE scores for the different classrooms are similar.

In consideration of the relationships between PLE and the physical attributes of the classrooms, amongst the most important findings is that mid to high frequency (between 2000 and 4000 Hz) ventilation noise has a negative linear relation with PLE. Another important point is that most C50 octave band vales have a negative quadratic relation with PLE. This indicates that some reverberation is preferable in classrooms. It is only the low-frequency octaves (between 125 and 250 Hz) that have a linear relation; for these frequencies, the slope of the regression line is positive. The total amount of sound absorption in a classroom has a positive relation with PLE when considered as an interaction of surface area. There is some discord between this result and the one for C50 values, as they seem to indicate two different conclusions. It should be noted that this is not a result of there being an insufficiently large distribution of sound absorption across the classrooms to observe an optimal amount; if this were the case, then it would be unlikely that the effect would be observable in the case of the C50 values. A more likely explanation is that a student's evaluation of the reverberation of a classroom is linked to other attributes of a classroom, such as shape, which were not blocked for in this study. This theory is supported by the relationship of EDT with PLE; EDT is only important in small classrooms and has an optimal value of approximately 0.9 seconds.

In consideration of the physical attributes that were adjusted for the occupied state of the classrooms, some of the more important results are: Total background noise level has a negative linear relation with PLE; this is pleasantly consistent with the effects observed for ventilation noise. The component of speech intelligibility that is not explained by occupied EDT predicts PLE score; strangely this component is in some way linked to signal-to-noise; however signal-to-noise does not appear to be a strong predictor. The contradiction created by these two points is a strong indicator that the method that was used to adjust the signal-to-noise values for the presence of a speech reinforcement system was not adequate.

The component of PLE that is explainable by physical attributes of the classrooms is explainable by many physical attributes of the classrooms. This is unfortunately not useful for predicting the performance of classrooms; however it does imply that a careful study of the acoustically relevant physical attributes of spaces share some underlying components that are perceived by people and impact their assessments of the spaces.

Table 1. A list of the physical acoustical variables and their descriptions.

Variable	Description	Variable	Description
AlphaEDT	The total number of square meters of sound absorption based on EDT.	Total.BGNL	Total background noise level of occupied classroom.
Ao	Total sound absorption. adjusted for number of occupants.	SRS	Is a Speech reinforcement system used in the classroom.
Attendance	Estimated number of students attending the class at time of survey.	V	The classroom volume
f.m	Is the instructor male of female?	VN	Ventilation noise level.
Floor	The classroom floor area	**Extention**	**Description**
Registered	Number of students registered in the class.	.o	Indicates that the variable has bean adjusted for the occupied stat of the classroom.
S	The classroom surface area	.Mid.Freq	The average of the 500. 1000 and 2000 Hz values.
SI	Speech intelligibility	.Average	The average of all values.
SNA	Signal to noise ratio	.Global	The average over frequancys of all values.
Upseat	Are the seats upholstered	.1000.Hz	The ocatave band for which a given measument was taken
STI	Sound transmission index	.A.Weight	The total A weighted noise level

Table 2. A list of the physical-acoustical variables and their coefficients used in the PLE prediction model.

Variable	Coefficients	95%CI(LB	UB)	Standard Error	T-value	P-value	
(Intercept)	4.3965	-1.718	10.511	3.11959	1.409	0.159	
Attendance	-0.0137	-0.032	0.004	0.00912	-1.505	0.132	
C50.500.Hz	0.7903	0.469	1.111	0.16381	4.825	0.000	***
C50.500.Hz.2	0.1151	0.062	0.168	0.02706	4.255	0.000	***
EDT.125.Hz.VS	-27.2742	-32.874	-21.674	2.85713	-9.546	0.000	***
EDT.2000.Hz	-6.1873	-10.240	-2.135	2.06748	-2.993	0.003	**
EDT.2000.Hz.2.VS	4.8229	3.017	6.629	0.92131	5.235	0.000	***
EDT.250.Hz.2.V	3.2022	2.346	4.058	0.43684	7.330	0.000	***
EDT.500.Hz	6.8294	3.859	9.800	1.51566	4.506	0.000	***
f.m	1.6055	0.585	2.626	0.52066	3.084	0.002	**
VN.2000.Hz	-0.2867	-0.490	-0.083	0.10389	-2.760	0.006	**

REFERENCES

[1] Kennedy, L. Dillon Edgett, M. Hodgson, N. Lamb and R. Rempel, "Subjective assessment of listening environments in university classrooms", J. Acoust. Soc. Am. 119(1) 299-309 (2006).

SPEECH PRIVACY CLASS FOR RATING THE SPEECH PRIVACY OF MEETING ROOMS

John S. Bradley and Bradford N. Gover

Institute for Research in Construction, National Research Council, Montreal Rd., Ottawa, K1A 0R6

Introduction

Enclosed offices and meeting rooms are often required to provide speech privacy against eavesdroppers. That is, it should be difficult for people outside the room to understand or in some cases to even hear speech from the meeting room. Measurements of speech privacy must be capable of detecting weaknesses in individual components of the sound insulation of an office or meeting room and should indicate whether an eavesdropper would be able to understand or hear speech from the room. The degree of privacy is related to the probability of a speech privacy lapse.

The Uniform Weighted Signal-to-Noise Ratio

Initial listening tests [1] found a uniform-weighted signal-to-noise ratio (SNR_{UNI32}) to be the best measure for predicting the audibility and intelligibility of transmitted speech in noise. Equation (1) shows SNR_{UNI32} to be the speech level (L_{ts}) – noise level (L_b) differences averaged over the 1/3-octave bands from 160 to 5k Hz. These differences are clipped so that they cannot be less than –32 dB, for which cases speech would be inaudible.

$$SNR_{UNI32} = \frac{1}{16}\sum_{f=160}^{5000}\{L_{ts}(f) - L_b(f)\}_{-32} \tag{1}$$

Criteria for speech privacy goals were obtained by determining the thresholds of audibility and intelligibility of speech sounds transmitted through walls. These were the SNR_{UNI32} values for which 50% of a panel of good listeners could: just understand one word of a sentence, or just find the speech to be barely audible. These are given in terms of SNR_{UNI32} values in Table 1. The 'intelligibility threshold in rooms' applies to typical meeting rooms. The 'intelligibility threshold in free field conditions' applies to conditions with spatially separated speech and noise sources without significant reflected sounds. The threshold of audibility is not affected by these factors.

Threshold	SNR_{UNI32}
Intelligibility in free field	-16 dB
Intelligibility in rooms	-11 dB
Audibility speech sounds	-22 dB

Table 1. Audibility and intelligibility criteria for speech privacy designs.

New Measurement Procedure

Conventional sound transmission tests measure differences between average levels in the source and receiving spaces. The new speech privacy measurement procedure similarly measures the average levels in the source room because the talker could typically be anywhere in the source room. However, in the receiving space, transmitted levels are measured at spot locations usually 0.25 m from the separating wall/door as illustrated in Figure 1. Speech privacy is evaluated from the level differences between source room average levels and spot receiver levels in the adjacent space.

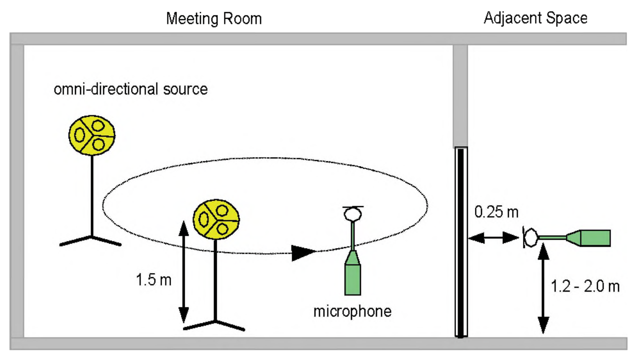


Figure 1. Measurement of level differences from a source room average to spot receiver positions in the adjacent space.

There are 3 reasons for this new approach: (1) the conventional approach assumes diffuse sound fields, but adjacent spaces will often not be at all diffuse (e.g. storage closet, open-plan office), (2) an eavesdropper would be more effective close to the separating partition where higher speech levels would occur, and (3) it is desired to evaluate the weaker components of the sound insulation rather than just an overall average.

Statistics of Speech Levels

The audibility and intelligibility of transmitted speech will depend on the speech levels in the meeting room as well as the sound transmission characteristics to the adjacent spaces. However, speech levels in meeting rooms are statistical in nature and will vary from moment to moment. To characterize these statistical properties, speech levels were recorded over 10 s intervals for 79 meetings in 39 rooms [2]. Speech levels were not related to measures of the room size or the number of occupants, and were combined into one cumulative probability distribution shown in Figure 2. From this figure it is possible to determine how often particular speech levels are exceeded and hence how often there may be speech privacy problems for a particular meeting room.

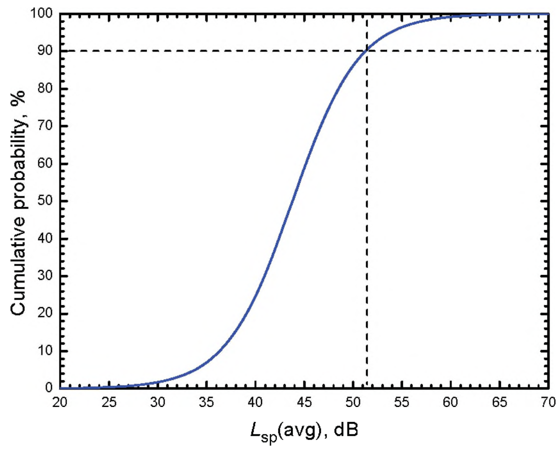


Figure 2. Cumulative probability distribution of measured speech levels in meeting rooms showing example that a 51.5 dB speech level is exceeded 10% of the time.

Speech Privacy Class

The speech privacy of a meeting room depends on both the level of the transmitted speech from the meeting room and the level of the ambient noise outside the meeting room. The Speech Privacy Class (*SPC*) is derived from the definition of SNR_{UNI32} in equation (1). If the –32 dB clipping in equation (1) is ignored, the equation can be re-written as,

$$SNR_{UNI32} \approx L_{ts}(avg) - L_b(avg) \quad (2)$$

where L_{ts}(avg) is the transmitted speech level and L_b(avg) is the ambient noise level at the receiver, both averaged over frequency from 160 to 5k Hz. L_{ts}(avg) can be replaced by the meeting room speech level L_{sp}(avg) – the measured meeting-room-to-spot-receiver level difference *LD*(avg). In addition, we assign the value –11 dB to SNR_{UNI32} to correspond to meeting the threshold of intelligibility, and obtain,

$$LD(avg) + L_b(avg) \approx L_{sp}(avg) + 11 \quad (3)$$

$LD(avg)+L_b(avg)$ is the Speech Privacy Class. Using Figure 2, the probability of the speech level L_{sp}(avg) in equation (3) being exceeded and a speech privacy problem occurring can be determined. That is, if $L_{sp}(avg) \geq$ SPC –11, the transmitted speech level will exceed the threshold of intelligibility and speech privacy will be compromised.

Table 2 lists *SPC* values and describes the related probability of speech privacy problems. As it extends from minimal to extremely high speech privacy, this process can be applied to a wide range of enclosed rooms.

Conclusions

SPC, determined from measurements of sound insulation and ambient noise, is useful for describing the speech privacy of an enclosed room in terms of the likelihood of a privacy lapse.

Acknowledgement

This work was jointly funded by the RCMP, PWGSC and NRC.

References

[1] Gover, B.N., and Bradley, J.S., J. Acoust. Soc. Am., 116, (6), 3480-3490 (2004).

[2] Bradley, J.S. and Gover, B.N., "Speech and Noise Levels Associated with Meeting Rooms", IRC Research Report RR-170, March (2004), revised December (2004).

Category	*SPC*	Description
	< 60	Speech expected to be frequently intelligible (more than once a minute) and almost always audible.
Standard Speech Privacy	60–65	Brief phrases expected to be occasionally intelligible (at most once every minute); speech sounds usually audible.
Enhanced Speech Privacy	65–70	Brief phrases expected to be rarely intelligible (at most once every 3.5 minutes); speech sounds frequently audible.
Standard Speech Security	70–75	Speech expected to be essentially unintelligible (brief phrases intelligible at most once every 15 minutes); speech sounds occasionally audible (at most once every minute).
Enhanced Speech Security	75–80	Speech expected to be unintelligible (brief phrases intelligible at most once every hour); speech sounds rarely audible (at most once every 3.5 minutes).
High Speech Security	80–85	Speech unintelligible (brief phrases intelligible at most once every 4.5 hours); speech sounds essentially inaudible (audible at most once every 15 minutes).
Top Speech Security	> 85	Speech unintelligible (brief phrases expected to be intelligible at most once every 20 hours); speech sounds inaudible (audible at most once every hour).

Table 3. Descriptions of the likelihood of transmitted speech being intelligible or audible for a range of SPC categories.

The Effects of Room Acoustics on the Speech Privacy of Meeting Rooms

John S. Bradley, Marina Apfel and Bradford N. Gover
Institute for Research in Construction, National Research Council, Montreal Rd., Ottawa, K1A 0R6

Introduction

Initial listening tests [1] established a uniform weighted signal-to-noise ratio (SNR_{UNI32}) as a good predictor of both the intelligibility and audibility of speech transmitted from an adjacent room. The same work established criteria for the speech privacy of meeting rooms in terms of the thresholds of audibility and intelligibility of speech from the room. However, subsequent validation tests in actual rooms [2] suggested that the criterion values were influenced by spatial and temporal room acoustics effects. Further listening tests reported here can quantitatively explain the differences.

The Previous Listening Tests

The initial listening tests [1] were carried out in approximately free-field conditions. As shown in Figure 1, speech sounds, modified to represent transmission through various walls, were reproduced by loudspeakers in front of the listener. Simulated ventilation noise was radiated from loudspeakers above the subject. The realistic separation of the speech and noise sources was intended to represent a worst-case condition for speech privacy in which the subject could understand more of the speech sounds.

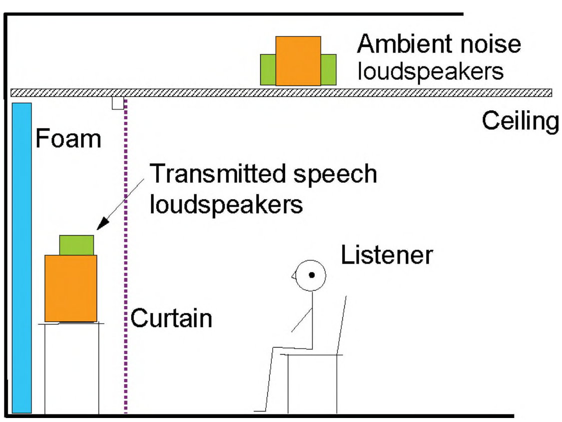


Figure 1. Initial listening test [1] set up in approximately free-field conditions.

The subsequent validation tests [2] involved radiating speech into one room from where it was transmitted through actual walls into a second room as shown in Figure 2. Both rooms had moderate reverberation times (0.8 and 0.64s) and simulated ventilation noise in the receiving room arrived diffusely at the listener who was located 0.25 m from the test wall. Although exactly the same speech tests were carried out, some results were quite different than in the first free-field experiment. Ratings of the audibility of speech sounds were very similar in both experiments. However, the intelligibility of the transmitted speech and the threshold of the intelligibility of the speech were different and the differences were equivalent to over a 5 dB change in signal-to-noise ratio. Ignoring these effects could lead to a costly over-design of the meeting room sound insulation.

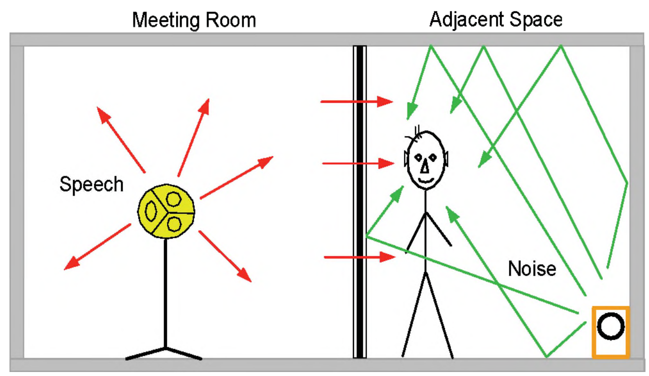


Figure 2. Two-room validation test setup in which subjects heard speech from an adjacent room in the presence of diffuse simulated ventilation noise.

New Listening Tests

New listening tests were carried out to understand the differences between the two previous tests, and how spatial and temporal room acoustics effects influence speech privacy. The new tests were carried out in simulated conditions in an anechoic room. Speech and noise test sounds could arrive from one or more of the 8 loudspeakers shown in Figure 3. Speech sounds from each loudspeaker could include: (a) direct sound only, (b) direct sound plus early reflections, or (c) direct sound, early reflections and reverberant sound. Simulated ventilation noise arrived from one loudspeaker or incoherently from all loudspeakers. The subjects repeated back test sentences so that the intelligibility of the speech could be determined.

Summary of Listening Test Results

It is well known that separating speech and noise sources in free-field conditions leads to a Spatial Release from Masking (SRM). SRM means that spatially separating speech and noise sources reduces the masking effect of the noise on the speech and hence leads to increased intelligibility scores. The magnitude of the release from masking can be described by the equivalent signal-to-noise ratio change relative to the case of coincident speech and noise sources.

In these new results these effects were sometimes further complicated by the addition of diffuse noise rather than uni-directional noise as well as varied amounts of early reflections and reverberant speech energy.

Figure 3. Eight-channel room acoustics simulation system in anechoic room.

The following key results were obtained:

Horizontally separating speech and noise sources by 32 degrees in free-field conditions reduced the masking effects on the speech by over 5 dB (in terms of equivalent signal-to-noise ratios) relative to the case of coincident speech and noise sources.

When diffuse noise was used instead of uni-directional noise, the results were similar to the case of coincident speech and noise sources for which there is no spatial release from masking.

A 90-degree vertical separation of speech and noise sources reduced the masking effects on speech by approximately 2 dB relative to the case of coincident speech and noise sources.

As expected [3], added early reflections of speech sounds that arrived within 50 ms after the direct sound had no effect on the intelligibility of the speech when the total speech level was kept constant.

When reverberant speech was added, while maintaining a constant overall speech level, the increase in the masking effect of the noise (in terms of equivalent signal-to-noise ratio) was proportional to the logarithm of the reverberation time (T_{60}) above T_{60} values of 0.5 s.

Diffuse noise and reverberant speech had additive independent effects that both increased the masking of the speech relative to spatially separated speech and noise sources.

The masking effects of the speech were the same for both natural speech and speech modified to represent the change in spectrum shape after transmission through a wall when using SNR_{UNI32} values to describe the conditions. For other signal-to-noise measures transmission through a wall changed the magnitude of the effects.

When simulated ventilation noise was radiated predominantly from groups of 3 loudspeakers, results were intermediate to those for completely diffuse noise and uni-directional noise.

Differences Between the Two Previous Experiments

The differences between the initial free-field and two - room experiments described in Figures 1 and 2 can be explained as due to two factors:

(1) The difference between the vertically separated speech and noise in the free-field experiment and the diffuse ambient noise in the two-room experiment changes the masking of the speech equivalent to about a 3.2 dB shift in signal-to-noise ratio.

(2) The room reverberation in the two-room experiment led to a further 3 dB increase in the masking of the speech, equivalent to a 3 dB increase in noise level.

The combined 6.2 dB effect is a reasonable estimate of the observed effects on intelligibility scores and intelligibility thresholds. In other rooms the differences could be a little larger or smaller depending on the actual reverberation times and spatial characteristics of the sounds.

Conclusions

The new results give a good estimate of the differences between the free-field and two-room experiments. They also give a quantitative indication of the importance of spatial and temporal room acoustics factors on the speech privacy of enclosed rooms. This improved understanding will make it possible to estimate the likely effects for other situations as they arise and to avoid costly over-design of the sound insulation of rooms.

Acknowledgements

This work was jointly funded by PWGSC, RCMP and NRC.

References

[1] Gover, B.N., and Bradley, J.S., "Measures for assessing architectural speech security (privacy) of closed offices and meeting rooms", J. Acoust. Soc. Am., 116, (6), 3480-3490 (2004).

[2] Bradley, J.S. and Gover, B.N., "Validation of Architectural Speech Security Results", IRC Research Report, RR-221, March 2006.

[3] Bradley, J.S., Sato, H. and Picard, M., "On the Importance of Early Reflections for Speech in Rooms", J. Acoust. Soc. Am. 113 (6) 3233-3244 (2003).

PREDICTION OF SPEECH TRANSMISSION INDEX IN EATING ESTABLISHMENTS

Musarrat Nahid and Murray Hodgson

School of Environmental Health, University of British Columbia, 2206 East Mall, Vancouver, BC, V6T 1Z3, Canada

1. INTRODUCTION

Eating establishments (EE: restaurants, cafes etc.) can produce noisy environments which can be an impediment to a comfortable conversation between customers [1]. As the noise level increases, talkers increase their voice levels in order to be heard over this noise—this is the Lombard effect [2]. This phenomenon can further increase the total noise level in an EE. Few studies have focused on controlling the noise level in eating establishments. None of those except one [3] has considered the Lombard effect, but that study didn't explore control measures to a considerable extent. The objective here is to model existing eating establishments (EEs) in the CATT Acoustics software and then, taking into account the Lombard effect, predict the acoustical conditions for speech (i.e., speech transmission index (STI) and, therefore, speech intelligibility (SI) and speech privacy (SP) in EEs without and with sound-control measures and, therefore, to determine how to design the EE to optimize the acoustical conditions.

2. METHOD

The primary talker (PT) and primary listener (PL) indicate the pair of customers between whom we want good SI. Other talkers/listeners are called secondary talkers (ST) and listeners (SL). The primary talker and the primary listener sit around a table facing each other so that the primary listener is in the direct field of PT.

The EE Voice Level Model [2] was used to take into account the Lombard effect, as CATT cannot do it automatically. It is hypothesized here that, according to this model [2], the Lombard effect occurs such that the voice output level $L_{pff1,n}$ varies with the background noise level L_n as follows:

$$L_{pff1,n} = L_{pff1,q} + asym/\{1 + \exp[(xmid - L)/scale]\} \text{ dBA}$$

in which *asym*, *xmid* and *scale* are Lombard-effect parameters, assumed unknown *a priori*, as is $L_{pff1,q}$ which is the voice level in the absence of noise. This model assumes that the room sound field is diffuse; however, a reverberant-field correction factor is used to correct for non-diffuseness of the sound field [2].

First, the talker voice level $L_{pff1,n}$ is predicted using the EE Voice Level Model. This value is put in CATT to model talkers (the sound source in CATT). Second, the SPLs at the primary listener (PL) due to secondary talkers are predicted by CATT to calculate the total noise level at PL. Finally, STI is predicted at PL (for SI) and SL (for SP). This is done by inputting noise levels calculated as the decibel sum of BNL (the noise level due to kitchen equipment, or music) and the total secondary-talker speech levels from Step 2.

In each EE configuration, predictions are done for different occupancies, namely low (LO) and high (HO) occupancies. The following configurations were evaluated: R (untreated configuration), PTC (PT seating at the corner of the room), STSS (STs are far away from PT), DV (volume is decreased by lowering the ceiling), IA-ceiling (increasing the absorption of the ceiling by applying a suspended acoustical-tile ceiling), IA-all surfaces (floor, ceiling, walls are made highly absorptive by applying thick carpet, suspended acoustical-tile ceiling, 15% perforated metal on 30 mm thick porous material), AB-2 (putting barriers around all tables, height of the barriers is 2 m, both sides of the barriers are made highly absorptive by applying wood-wool slab), AB-1 (same as AB-2 but the height of the barriers is now 1 m), RB-2 (same as AB-2 but the both sides of the barriers are highly reflective: glass, 6 mm), RB-1 (same as RB-2 but the height of the barriers is now 1 m). IA-DV (combining DV and IA-all surfaces), IA-AB-2 (combining IA-all surfaces and AB-2), DV-AB-2 (combining DV and AB-2), IA-DV-AB-2 (combining IA-all surfaces, DV and AB-2). The predictions were done in three EEs (Table 1) of different sizes.

Table1: Physical and acoustical characteristics of EEs.

EE name	Dimensions (m)	Customer density [#/floor area (m^2)]	α_{avg} [1]	N_t [2]		BNL (dBA)
				LO	HO	
MM	L_{avg}=14 W_{avg}=8.5 H=5	0.36	0.17	6	12	60.0
SS	L=12 W=5 H=4	0.75	0.16	5	10	57.4
LL	L=18 W=18 H=3.5	0.33	0.15	12	24	49.1

[1] Averaged over all octave-band frequencies (125-4k Hz)
[2] Number of talkers

CATT assigned the following quality ratings for STI: "Bad" (STI < 0.30), "Poor" (0.30<=STI<0.45), "Fair" (0.45 <=STI <0.60), "Good" (0.60<=STI<0.75), "Excellent" (0.75<=STI). Only those design-factor changes which resulted in 'fair' STI between the PT and PL in MM, both at LO and HO conditions, were incorporated in SS. Based on the results of these two types of model, a further attempt was made to obtain 'fair' STI in LL: it was divided in four compartments by using highly absorptive barriers rising up to ceiling ("Subdiv").

3. RESULTS AND DISCUSSION

Table 2: Predicted acoustical values at MM.

Case	Occu-pancy	STI at PL (for SI)	STI at SLs (for SP)
R	LO	0.39 (poor)	0.06-0.28
	HO	0.36 (poor)	0.04-0.25
PTC	LO	0.41 (poor)	0.05-0.33
	HO	0.44 (poor)	0.05-0.35
STSS	LO	0.42 (poor)	0.09-0.30
	HO	0.39 (poor)	0.05-0.27
DV	LO	0.40 (poor)	0.08-0.2
	HO	0.32 (poor)	0.01-0.21
IA-only ceiling	LO	0.36 (poor)	0.05-028
	HO	0.31 (poor)	0.02-0.23
IA-all surfaces	LO	0.40 (poor)	0-0.27
	HO	0.38 (poor)	0-0.25
AB-2	LO	0.50 (fair)	0.01-0.24
	HO	0.57 (fair)	0.01-0.29
AB-1	LO	0.48 (fair)	0.02-0.22
	HO	0.49 (fair)	0.01-0.22
RB-2	LO	0.49 (fair)	0.03-0.32
	HO	0.52 (fair)	0.04-0.36
RB-1	LO	0.42 (fair)	0.02-0.26
	HO	0.47 (fair)	0.01-0.3
IA-DV	LO	0.39 (poor)	0-0.26
	HO	0.37 (poor)	0-0.24
IA-AB-2	LO	0.48 (fair)	0-0.16
	HO	0.51 (fair)	0-0.15
DV-AB-2	LO	0.47 (fair)	0-0.21
	HO	0.49 (fair)	0-0.22
IA-DV-AB-2	LO	0.54 (fair)	0-0.21
	HO	0.50 (fair)	0-0.15

Table 3: Predicted acoustical values at SS.

Case	Occu-pancy	STI at PL (for SI)	STI at SLs (for SP)
R	LO	0.43 (poor)	0.16-0.26
	HO	0.36 (poor)	0.08-0.18
AB-2	LO	0.53 (fair)	0.02-0.22
	HO	0.55 (fair)	0-0.23
IA-AB-2	LO	0.54 (fair)	0-0.13
	HO	0.6 (fair)	0-0.18

Table 4: Predicted acoustical values at LL.

Case	Occu-pancy	STI at PL (for SI)	STI at SLs (for SP)
R	LO	0.45 (fair)	0.03-0.20
	HO	0.37 (poor)	0-0.11
Subdiv	LO	0.57 (fair)	0-0.26
	HO	0.53 (fair)	0-0.23

It seems that putting high and absorptive barriers (AB-2) around all tables of EEs may provide 'fair' STI in MM. Increasing the absorption of the room surfaces along with those barriers (IA-AB-2) may provide even better results, with the added benefit of a decreased noise level at PL compared to other configurations. These findings were also true for SS. Prediction results for the "Subdiv" configuration in LL also support these findings. Lowering the ceiling and using absorptive barriers together (DV-AB-2), or combining those two with increased absorption of the room surfaces (IA-DV-AB-2), provided good results in MM but the improvement was not significantly greater compared to IA-AB-2. Moreover, DV-AB-2 resulted in higher noise levels at PL than IA-AB-2 and IA-DV-AB-2 produced the same level of noise at PL as IA-AB-2.

REFERENCES

1. Christie, L. H. (2004). Psycho-to-building acoustics: are bars, cafes and restaurants acceptable acoustic environments? Research Report, Victoria University of Wellington.

2. Hodgson, M., Steininger, G., Razavi, Z. (April 2007). Measurement and prediction of speech and noise levels and the Lombard effect in eating establishments. *J. Acoust. Soc. Am.*, Vol. 121 (4).

3. Astolfi, A., Filippi, M. (2003). Good acoustical quality in restaurants: a comparison between speech intelligibility and privacy. *Proc. EuroNoise*, Naples, Italy, S102.

REVERBERATION ROOMS AND SPATIAL UNIFORMITY

Ramani Ramakrishnan[1,2], and Anant Grewal[3]
[1]Department of Architectural Science, Ryerson University, Toronto, Ontario
[2]Aiolos Engineering Corporation, Toronto, Ontario
[3]Institute for Aerospace Research, National Research Council of Canada, Ottawa, Ontario, K1A- 0R6

1. INTRODUCTION

Reverberation rooms are special test rooms used to evaluate the sound power level of sources as well as to qualify space bound hardware such as antennae and satellites to a high intensity noise environment with levels and spectral content representative of the acoustic environment present during launch. Combinations of reverberation rooms are used to evaluate transmission properties of building materials as well as absorption characteristics of noise control products. A number of standards are available that prescribe minimum requirements of reverberation rooms [1, 2].

The main characteristics of the reverberation rooms are: i) Adequate volume; ii) Suitable shape or diffusing elements or both; iii) Suitably small sound absorption over the frequency of interest; and iv) Sufficiently low background noise levels. [1, 2].

The volume of the chamber needs to be adequate as it determines the low-frequency limit of the room. Above the low-frequency limit, the room responds to bands of noise uniformly thus assuring spatial constancy of the sound levels. There are different methods to determine the low-frequency limit. One such limit is the Schroeder frequency and is given by [3],

$$f_c = 2000\sqrt{\frac{T_{60}}{V}} \qquad (1)$$

where, T_{60} is the chamber's reverberation time, sec. And V is the volume of the chamber in cubic meters.

The above limit is quite restrictive and when the sound levels are bands of noise, the volume can be lower and one can still maintain adequate spatial uniformity. The results of sound levels, from both single sinusoidal tones as well as bands of noise, measured in two reverberation chambers are presented in this paper to determine the adequacy of chamber volume.

2. CHAMBER VOLUME

Eq. (1) has provided a low-frequency limit which has been adopted by many standards and based on that requirement, the volume of the chamber has to be determined. As mentioned earlier, Schroeder requirement is quite restrictive.

Another empirical approach is to impose a norm of at least 20 modes per octave for acceptable uniformity. Slingerland, Elfstorm and Grün applied 20 modes/octave criterion and derived the following relationship for the cut-off frequency [4],

$$f_c = \frac{c}{\sqrt[3]{V}} \qquad (2)$$

where, c is the speed of sound.

The two different approaches produce different limits and the most commonly used Schroeder limit is too restrictive. Field measurements were conducted in two different chambers to determine the most reasonable limit that is practical and can be easily implemented. The description of the two chambers is presented next.

3. THE REVERBERATION CHAMBERS

Two chambers were used to determine the spatial uniformity of the chamber as well as the low-frequency cut-off limit of the chamber. The two chambers are located in Montreal and Ottawa respectively.

2.1 Chamber 1 – Concordia University

The smaller of the two chambers is located in the engineering building of Concordia University, Montreal and is used by the Building, Civil and Environmental Engineering Department (BCEE). The characteristics of the chamber are:
Length, L = 6.13 m; Width = 6.96 m; Height = 3.56 m; Chamber Volume = 152.3 cu.m.

The RT_{60} varied between 0.8 sec to 3 sec. across the frequency band. The cut-off frequency as per Eq. (1) is 188 Hz and as per Eq. 2 is 64 Hz.

2.2 Chamber 2 – National Research Council of Canada

The larger of the two chambers is located at the Structures, and Material Performance Laboratory (SMPL) of the National Research Council's Institute for Aerospace Research in Ottawa, Canada. The characteristics of the chamber are:
Length, L = 7.01 m; Width = 7.93 m; Height = 9.75 m; Chamber Volume = 542 cu.m.

The RT_{60} varied between 5 sec to 10 sec. across the frequency band. The cut-off frequency as per Eq. (1) is 272 Hz and as per Eq. 2 is 42 Hz.

2.3 Modal Compositions of the two chambers

The two chambers are rectangular in shape and the standing wave frequencies can easily be determined from basic descriptions [5] and are given by,

$$f_n = \frac{c}{2}\sqrt{\left[\frac{n_x}{L_x}\right]^2 + \left[\frac{n_y}{L_y}\right]^2 + \left[\frac{n_z}{L_z}\right]^2} \qquad (3)$$

The number of modes in each octave band was enumerated from the above equation and the results for the two chambers are given in Tables 1 and 2 respectively.

The results of Table 1 show that that the Chamber 1 can be comfortably used from the 125 Hz octave band to achieve acceptable spatial uniformity. This is borne out by the cut-off frequency of 64 Hz calculated from Eq. 2. The Schroeder limit for Chamber 1 is 188 Hz (from Eq. 1) which is very restrictive.

Table 1. Modal Composition of Chamber 1

Band No.	Lower Limit	Centre Frequency	Upper Limit	Number of Modes
1	22	31.5	44	2
2	44	63	88	12
3	88	125	177	44
4	177	250	355	165

Table 2. Modal Composition of Chamber 2

Band No.	Lower Limit	Centre Frequency	Upper Limit	Number of Modes
1	22	31.5	44	5
2	44	63	88	38
3	88	125	177	210
4	177	250	355	340

The results of Table 2 show that that the Chamber 2 can be comfortably used from the 63 Hz octave band to achieve acceptable spatial uniformity. This is borne out by the cut-off frequency of 42 Hz calculated from Eq. 2. The Schroeder limit for Chamber 1 is 272 Hz (from Eq. 1) which is very restrictive.

The validity of these limiting frequencies is confirmed through measurements and is presented next.

4. THE EXPERIMENT

Two chambers were used to determine the spatial uniformity of the chamber as well as the low-frequency cut-off limit of the chamber. Simple speakers (both low-frequency speakers and a bank of high frequency tweeters) were used to generate the sound. Both pink noise and sinusoidal tones (100, 150, 200, 250, 300, 400, 500 Hz) were generated and the resulting noise levels were measured at a number of locations, - between 48 and 54. The locations were chosen randomly at two different heights.

High intense sound were generated in the second chamber through hydraulically powered airstream modulators driven by high pressure air (150 to 200 psi) and connected to two exponential horns (25 Hz and 100 Hz). The resulting sound pressure levels were measured at eight locations at three different heights inside the chamber.

5. RESULTS

The results for Chamber 1 are presented for both broad-bands and sinusoids in Tables 3 and 4 below. Similarly, the results for the broad-band sound sources for Chamber 2 are shown in Table 5 below.

The ISO Standard 3741 [2] requires a minimum of 200 cu. m. as per the Schroeder limit of 125 Hz Octave band and the maximum allowable standard deviation is 1.5 dB. The results of Tables 4 and 6 show that even if one cannot meet the minimum volume requirement, the spatial uniformity of the chamber sound levels can be satisfied for broadband sound levels. For pure sinusoids, even though the volume requirements are satisfied, the results of Table 5 indicate that the spatial uniformity cannot be assured.

Table 4. Sound Levels in Chamber 1 (Broadband, 36 Locations)

1/3 Octave Band Centre Frequency, Hz	Average SPL, dB	Range dB	Standard Deviation, dB
50	87.7	12.4	3.7
63	89.3	10.7	2.8
80	93.9	6.6	1.7
100	99.7	5.0	1.1
125	97.1	5.5	1.2

Table 5. Sound Levels in Chamber 1 (Tones, 36 Locations)

Tones, Hz	No. of modes ± 5 Hz	Average SPL, dB	Range dB	Standard Deviation, dB
300	9	94.4	31.4	7.4
400	11	96.3	29.8	7.8
500	11	100.7	32.6	8.4

Table 6. Sound Levels in Chamber 2 (Broadband, 28 Samples)

1/3 Octave Band Centre Frequency, Hz	Average SPL, dB	Range dB	Standard Deviation, dB
50	145.4	10.0	3.8
63	146.1	8.3	2.6
80	144.9	7.1	2.6
100	144.2	3.9	1.4
125	144.2	4.0	1.4

REFERENCES

[1] American National Standard Methods for the Determination of Sound Power Levels of Small Sources in Reverberation Rooms. S1.21-1972 (Superceded by ANSI S1.31-1980 and S1.32-1980).

[2] Acoustics- Determination of Sound Power Levels of Noise Source- Precision Methods for Broad-Band Sources in Reverberation Rooms, ISO 3741.

[3] M. R. Schroeder, "Frequency-Correlation Functions of Frequency Responses in Rooms," J. Acoust. Soc. Am. Vol.34, pp. 1819, 1962. Also J. Acoust. Soc. Am. Vol.46, pp. 277-283, 1969.

[4] F. Slingerland, G.M. Elfstorm and W.E. Grün, "Performance and Operational Capabilities of the Large European Acoustic Facility (LEAF)," Proceedings of the International Symposium on Environmental Testing of Space Programmes – (ESA SP-304, September 1990).

[5] D. A. Bies and C.H. Hansen, "Engineering Noise Control – Theory and Practice," 3rd Edition, Spon Press (2003).

ACKNOWLEDGMENTS

The small reverberation room tests were conducted while the first author spent his sabbatical year qualifying the room at Concordia University, Montreal. The kind assistance provided by the BCEE (Building, Civil and Environmental Engineering) Department is duly acknowledged.

AUDITORY SCENE ANALYSIS AS A SYSTEM

Albert S. Bregman
Dept. of Psychology, McGill University, 1204 Penfield Ave., Montreal, QC, Canada H3A 1B1. al.bregman@mcgill.ca

1. INTRODUCTION

A serious problem faced by any listener is that the ears receive a mixture of all the environmental sounds that are present at a given moment. Yet in order to generate appropriate responses, the auditory system must be able to build representations of the individual sounds that have created the mixture; this accomplishment is called auditory scene analysis (ASA). It seems to be accomplished by a two-stage process. First it analyses the incoming signal into its frequency components, both at a given time and extending over time. Then bottom-up processes of perceptual grouping use various acoustic relations among the components to build up evidence favouring the grouping of certain subsets of components, each subset representing a single environmental sound, with its own spectral and temporal properties (Bregman, 1990). Top-down processes use these "grouping recommendations" in building a representation of the streams.

1.1 Laboratory phenomena related to ASA.

There are a number of phenomena, including stream segregation, illusory continuity, fusion and decomposition of complex sounds, which are best viewed as glimpses of a single, coherent ASA system. Stream segregation is typically studied by alternating two types of tones (call them A and B). When the difference in the feature that distinguishes A from B tones is large enough and the speed of the sequence is fast enough, the listener perceives two parallel but independent streams of sound, one restricted to the A tones and the other to the B tones. Differences between A and B tones can be in terms of frequency (for pure tones), or in pitch and timbre (for complex tones), or in how the properties of a tone change over time, or for separation in the spectrum (for band-limited noises), or where they seem to come from in space (for all tones).

Illusory continuity is the apparent continuity of a sound, A, through a loud interruption, B, despite the fact that A is turned off during the interruption. It is stronger when the interrupting sound, B is shorter, and when it would have masked A even if A had actually been present during the interruption. It is viewed as a perceptual compensation for masking, a way of dealing with a sonic environment composed of many sounds, where one can temporarily mask another.

Fusion and decomposition of complex sounds occurs when many different frequency components are detected at the same time in a sensory input. Fusion into a single sound is favoured by harmonic relations among the components, a common spatial origin, their proximity in frequency and correlation in how they change over time.

These phenomena are best viewed as glimpses of a single, coherent system. The alternative view would see them as distinct, each with its own physiological basis.

2. PHYSIOLOGICAL INVESTIGATION

From a physiological perspective, we may well discover analysis mechanisms for the different acoustic features that favour the grouping of components. For example, Fishman, Arezzo, and Steinschneider (2004) presented an ABAB... sequence to awake monkeys while neural activity was recorded in primary auditory cortex (A1). Using pure tones, they varied the frequency separation of the A and B tones, the tone presentation rate, and the duration of each tone. Recordings were made at the cortical site which responded maximally to tone A (the "A-site").

In human psychophysics it has been found that as the speed of the sequence is increased, the segregation of the ABAB… sequence into A and B streams becomes stronger. In the cortex of monkeys, at slow speeds the A-site responded also to the B tones (but to a lesser degree). The greater the frequency difference between A and B, the less the A-site responded to B. But more interestingly, as the tone rate increased, while the response at the A-site to the A tones was somewhat reduced, the A-site's response to B was even more greatly reduced, so that the A-site yielded a neural response pattern dominated by responses to the A tone, occurring at half the alternation rate. In other words it becomes more selective in favouring the A stimuli.

2.1 Interpretation of the results

These neural activities were taken by the. authors as an important cause of perceptual segregation. In effect the A-site "sees" the A stream only (a B-site would see only the B stimuli). They also pointed out that since no response was required of the monkeys, the research was studying an automatic or obligatory process.

One is tempted to conclude that the physiological mechanism of stream segregation has been discovered. One of the difficulties with such a straightforward interpretation is that the results depend on the use of pure-tone stimuli. Because the earliest research used pure tones as A and B, and manipulated their similarity be making them different in frequency, most of the physiological and animal research has also used pure tones. As a consequence, the explanations are often specific to pure-tone stimuli.

However, A and B can be made different in many other ways. For example they can appear to emanate from different spatial locations, e.g., by the manipulation of interaural differences in time of arrival, or be different in loudness, factors that will promote segregation. If A and B are complex periodic tones, a difference in their pitches can also favour segregation. Also for these tones, differences in the placement of formants (peaks in their spectra) can favour their segregation. It is even true that A and B tones can have the same pitch and loudness, come from the same location, have components located in the same spectral region, with identical amplitude spectra and still be made to segregate – by having the phase relations among the components of each tone be different, altering their timbres.

One might argue that for each feature there are separate cortical sites that respond to different values of that feature, each one working in the same way as the frequency sites described by Fishman et al. (2004). The actions of any or all of these sites could produce stream segregation. But this argument does not take into account the fact that the various sorts of differences between A and B tones tend to interact. Any particular acoustic difference only promotes, rather than directly determining, the grouping of components to represent individual environmental sounds. The more factors that favour a particular grouping, the stronger the total evidence for that grouping will be. The various acoustic differences can also compete with one another, some favouring one grouping, and others favouring different ones. It seems that in such a case the grouping with the most evidence in its favour will be the one that is perceived.

In addition to the interaction of the various bottom-up acoustic analyses, these also interact with top-down processes, such as knowledge of the class of signal involved. Such stored knowledge can influence the attentional processes that participate in the building of the perceptual representations of individual streams of sound. The overriding role of top-down processes is particularly evident in the case of speech, where sine-wave-analog speech, despite being a stripped-down cartoon of speech, can still be understood, despite the fact that some of the bottom-up acoustic cues favouring integration of the components of the "speech" signal are missing.

So one cannot say that any particular acoustic relation, such as the frequency difference studied by Fishman et al. (2004) is *the* physiological cause of grouping, even in the simple laboratory example of the alternation of two tones.

3. ASA AS A COHERENT SYSTEM

In the introduction, we described three different phenomena as being glimpses of the action of the ASA system: stream segregation, perceptual fusion of simultaneous components, and illusory continuity. We can consider the argument that each of these has its own distinct physiological basis.

One argument against it is that these three phenomena respond to the same variables. For example, two narrow-band noise bursts, A and B, can be created where A has a higher pass band than B, and these can be alternated, each burst separated from the next by a wideband noise (W), in the pattern AWBWAWBW…. (Bregman, Colantonio & Ahad, 1999), We can use this stimulus to look at both stream segregation and illusory continuity. When the centre frequencies of the A and B bands are close together they will not segregate, and will also appear to connect up behind the wide-band interruption, yielding illusory continuity. When their centre frequencies are further apart, both segregation and continuity are affected. A and B are heard as separate streams, and also fail to connect up behind the noise. We have argued that both stream segregation and the continuity illusion are products of a single ASA system. We have also argued that the process called the "old-plus-new heuristic" can explain both illusory continuity and the decomposition of the spectrum into separate sounds (Bregman, 1990)

The question is whether the phenomena described in this paper result from the activity of an integrated ASA system or are the result of a haphazard set of physiological processes, each perhaps having evolved independently to favour the correct parsing of the incoming sound. There are a number of arguments for preferring the view of ASA as a coherent system. They include (1) the desire for parsimonious explanation: The alternative is to believe that a bundle of unrelated phenomena just happen to be that way for idiosyncratic physiological reasons. This is an unparsimonious explanation – a bristly hypothesis that needs a shave with Occam's razor; (2) the fact that these phenomena result from processes that serve a common function; (3) the fact that they interact in ways that are predictable under the assumption of a unified system; (4) that they respond to many of the same variables; and (5) the heuristic value, for researchers, in finding the factors that may affect all of the various grouping phenomena..

REFERENCES

Bregman, A.S. (1990). *Auditory scene analysis: the perceptual organization of sound.* Cambridge, Mass.: The MIT Press, 1990 (Paperback 1994).

Fishman, Y.I., Arezzo, J.C., & Steinschneider, M. (2004). Auditory stream segregation in monkey auditory cortex: effects of frequency separation, presentation rate, and tone duration. *J. Acoust. Soc. Am. 116*, 1656-1670.

Bregman, A.S. Colantonio, C. & Ahad P.A. (1999). Is a common grouping mechanism involved in the phenomena of illusory continuity and stream segregation? *Perception & Psychophysics, 61 (2),* 195-205.

ORGANISING BY OBJECT : HOW AUDITORY MEMORY CAN BE STRUCTURED WITHIN COMPLEX SCENES

Benjamin J. Dyson[1]
[1]Department of Psychology, Ryerson University, 350 Victoria Street, Toronto, Ontario, Canada, M5B 2K3

1. INTRODUCTION

Objects are considered important units of short-term memory in vision [1,2]. This experimental series was an attempt to detail the extent to which analogous memorial organisation can take place in audition. The initial design was based on [1] in which participants were presented with two objects differing along two separate attributes (i.e., line: orientation and texture; box: size and gap location). Participants were better at remembering two attributes when they were derived from the same object (e.g., orientation and texture) relative to retrieving two attributes derived from different objects (e.g., orientation and size).

1. METHOD

2.1 Stimuli

Variations in two 500 ms sounds (tone and noise) were created, with each sound having a 10 ms linear onset and offset. Noise was low-pass filtered to sound like *wind* (+20 dB under 510 Hz, 0 dB 510-1200 Hz, -20 dB over 1200 Hz) or high-pass filtered to sound like *rain* (-20 dB under 510 Hz, 0 dB 510-1200 Hz, +20 dB over 1200 Hz). Amplitudes for both types of noise were then increased or decreased linearly in intensity to give the impression of moving *towards* or *away* from the listener. The tone was either *high* (1122 Hz) or *low* (750 Hz) in pitch and was frequency modulated (modulation frequency: 5 Hz, deviation frequency: 10%) or not to give the *presence* or *absence* of a warble. All variations of tone and noise were mixed and calibrated binaurally at 70 dB SPL(A) using Sennhesier HD580 headphones, and a Brüel & Kjœr sound level meter (Type 2610) and artificial ear (Type 4153).

2.2 Design and procedure

At each trial, participants were presented with a blank screen for 1000 ms and then two cue dots for 1500 ms, after which the combined noise/tone was played. Two of four possible prompts followed. Each prompt presented a to-be-remembered attribute centre screen with the two possible values of the attribute left and right of centre: 'Wind [CLIMATE] Rain' and 'Towards [DIRECTION] Away' for the noise, 'Low [PITCH] High' and 'No [WARBLE] Yes' for the tone. For both prompts, participants pressed a left or right button in accordance with their memory of the acoustic attribute. All possible combinations of tone and noise were presented equally and each attribute was interrogated equally across same- and different-object responses, and across first and second prompts. All experimental analyses were based around repeated-measures ANOVAs with the factors of object (same, different) and response (first, second). Since the distinction between same and different object was essentially meaningless during the first response, RT advantages and/or error rate reductions for same-object responding relative to different-object responding were expected only for the second response in a pair.

2. RESULTS

3.1 Experiment 1

Experiment 1 set out to establish the basic effect in that participants should be better at retrieving multiple attributes when they originate from the same auditory object relative to different auditory objects. Table 1 supports the critical object x response interaction ($F[1,10] = 6.03$, $p = .034$), with Tukey's HSD test ($p < .05$) revealing faster RTs for the second response in a pair when the to-be-remembered attribute was derived from the same object as the first (see [3] for further details).

Table 1. Summary statistics for Experiment 1

	RT (ms)	RTSE (ms)	Error (%)
1st Response Same	1273	93	12.89
1st Response Diff	1276	95	13.14
2nd Response Same	992	57	14.28
2nd Response Diff	1091	40	14.80

3.2 Experiment 2

The flexibility of object-based organisation was tested in Experiment 2 via the use of a cue [4]. If such organisation was purely strategic, then it should be possible to abolish the effect by alerting participants to the fact that they will have to retrieve information from different objects prior to acoustic stimulation. For half of the same-object trials and half of the different-object trials, participants were cued to the attributes they would have to respond to. The ANOVA revealed an average benefit of over 200 ms for cueing ($F[1,17] = 45.86$, $p < .001$). However, the effect of cue failed to modulate the object x response interaction ($F[1,17] = 4.86$, $p = .041$; see Table 2). Tukey's HSD test ($p < .05$) confirmed the same-object advantage for the second response in a pair for both cued and uncued conditions.

Table 2. Summary statistics for Experiment 2

UNCUED	RT (ms)	SE (ms)	Error (%)
1st Response Same	1526	71	11.11
1st Response Diff	1526	77	12.58
2nd Response Same	1264	47	14.84
2nd Response Diff	1292	56	15.45

CUED	RT (ms)	SE (ms)	Error (%)
1st Response Same	1306	80	9.38
1st Response Diff	1296	65	12.93
2nd Response Same	996	77	14.59
2nd Response Diff	1073	73	12.41

3.3 Experiment 3

Object-based organisation is often put into competition with other forms of organisation such as space [2] when assessing memorial structure. Experiment 3 was identical to the uncued condition of Experiment 2, save for that stimuli were now presented monaurally and for half of the same- and different-object trials, sounds were presented to the same or different ear. If space and object make additive contributions to the organisation of auditory memory, then the benefit accrued for same-object responding should be larger during different ear conditions. As shown in Table 3, ear of delivery failed to impact upon the standard object x response interaction ($F[1,17] = 4.97$, $p = .039$), with Tukey's HSD test ($p < .05$) in which the same-object advantage was revealed for the second response in a pair. The data are consistent with the idea that space works as a useful organisational factor in auditory scene analysis only when other grouping mechanisms are ambiguous [5].

Table 3. Summary statistics for Experiment 3

SAME EAR	RT (ms)	SE (ms)	Error (%)
1st Response Same	1260	77	12.15
1st Response Diff	1288	77	11.20
2nd Response Same	1005	45	14.93
2nd Response Diff	1122	44	14.14

DIFFERENT EAR	RT (ms)	SE (ms)	Error (%)
1st Response Same	1265	74	10.76
1st Response Diff	1245	54	12.24
2nd Response Same	1036	45	15.71
2nd Response Diff	1113	47	15.53

3.4 Experiment 4

By introducing time lags into the paradigm, it should be possible to assess the temporal limits of auditory object-based organisation. Time lags of 0ms or 2000ms were introduced inbetween the end of the sound and start of the first prompt, and/or, the end of the first feedback and start of the second prompt (appearing as *x–y* Lag in Table 4). Here, the standard object x response interaction was subsumed by a further interaction with the second kind of time lag ($F[1,17] = 14.88$, $p = .001$). Tukey's HSD test ($p < .05$) revealed that the same-object advantage was abolished with long delays inbetween first and second prompt. As shown in Table 4, the data are consistent with the characterisation of object-based organisation in terms of mutual interference during different-object trials rather than mutual facilitation during same-object trials [1], in that different-object trials show significant speeding as time lag increases.

Table 4. Summary statistics for Experiment 4

0–0 LAG	RT (ms)	SE (ms)	Error (%)
1st Response Same	1320	68	11.11
1st Response Diff	1309	67	11.46
2nd Response Same	1089	78	11.46
2nd Response Diff	1169	68	9.90

0–2000 LAG	RT (ms)	SE (ms)	Error (%)
1st Response Same	1287	57	10.42
1st Response Diff	1288	72	13.37
2nd Response Same	1112	74	13.37
2nd Response Diff	1077	58	13.19

2000–0 LAG	RT (ms)	SE (ms)	Error (%)
1st Response Same	1085	53	10.24
1st Response Diff	1092	54	11.63
2nd Response Same	1001	76	11.63
2nd Response Diff	1156	56	10.76

2000–2000 LAG	RT (ms)	SE (ms)	Error (%)
1st Response Same	1123	59	14.58
1st Response Diff	1120	67	14.93
2nd Response Same	1027	68	13.37
2nd Response Diff	1062	53	14.93

3.5 Experiment 5

Experiment 5 refuted an alternative explanation of the data based on different attribute association strengths between conditions. In previous experiments, an attribute (i.e., PITCH) predicted only one same-object attribute (i.e., WARBLE) but two different-object attributes (i.e., DIRECTION or CLIMATE). However, an object x response interaction was revealed ($F[1,27] = 4.50$, $p = .043$) when the first attribute equally predicted its sibling attribute or only one of the attributes from the other object (see Table 5).

Table 5. Summary statistics for Experiment 5

	RT (ms)	SE (ms)	Error (%)
1st Response Same	1449	69	9.82
1st Response Diff	1483	77	10.38
2nd Response Same	1169	50	13.76
2nd Response Diff	1272	50	13.00

3. DISCUSSION

The data across Experiments 1-5 support the case for object-based organisation in auditory memory. This kind of organisation appears to be exogenous (Experiments 1 and 2), does not interact with spatial organisation (Experiment 3) but is sensitive to temporal delay (Experiment 4). Despite initial transduction difference between the senses, there is an extent to which the phenomenology of everyday life is represented in a rich, multi-modal code. These current observations suggest that associated features (i.e., objects) serve as important cognitive representations for both visual and auditory stimuli [6, 7] and that as a shared representational code, objects may be a useful construct in thinking about how disparate sensory information is eventually bound together. Indeed, a recent revision of one influential model of memory posits additional mechanisms that attempt to account for exactly these kinds of concern (e.g., episodic buffer; [8]). Further research examining the same-object advantage using multi-modal stimuli should reveal additional insights into how we organise the collection of not only sounds, but also sights and sensations that define our previous experiences.

REFERENCES

[1] Duncan, J. (1984). Selective attention and the organization of visual information. *Journal of Experimental Psychology: General, 113*, 501-517.

[2] Lee, D. & Chun, M. M. (2001). What are the units of visual short-term memory, objects or spatial locations? *Perception & Psychophysics*, *63*, 253-257.

[3] Dyson, B. J. & Ishfaq, F. (2008). Auditory memory can be object-based. *Psychonomic Bulletin & Review*, *15*, 409-412.

[4] Awh, E., Dhaliwal, H., Christensen, S., & Matsukura, M. (2001). Evidence for two components of object-based selection. *Psychological Science*, *12*, 329-334.

[5] McDonald, K. L. & Alain, C. (2005). Contribution of harmonicity and location to auditory object formation in free field: evidence from event-related brain potentials. *Journal of the Acoustical Society of America, 118*, 1593-1604.

[6] Kubovy, M. & Van Valkenburg, D. (2001). Auditory and visual objects. *Cognition*, *80*, 97-126.

[7] Dyson, B. J. (in press). Perceptual organisation. In C. Plack (Ed.) *The Oxford Handbook of Auditory Science: Auditory Perception*. Oxford University Press: Oxford.

[8] Baddeley, A.D. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences*, *4*, 417-423.

LOW INTENSITY PULSED ULTRASOUND STIMULATES OSTEOGENIC DIFFERENTIATION OF HUMAN GINGIVAL FIBROBLASTS

Nesrine Mostafa [1], Paul Scott [2], Douglas N Dederich [3], Michael Doschak [4], Tarek El-Bialy [1]

[1] Departments of Dentistry and Biomedical Engineering, University of Alberta, Edmonton, AB., Canada

[2] Department of Biochemistry, University of Alberta, Edmonton, AB., Canada.

[3] Department of Dentistry, University of Alberta, Edmonton, AB., Canada.

[4] Faculty of Pharmacy and Pharmaceutical Sciences, University of Alberta, Edmonton, AB., Canada

1. INTRODUCTION:

Low intensity pulsed ultrasound (LIPUS) is widely used as a therapeutic tool in medicine. Several cell types have been reported to be sensitive to LIPUS exposure, including bone cells [1], cementoblasts [2], and odontoblast-like cells [3]. Certain clinical applications of LIPUS treatment were also studied. LIPUS was shown to enhance healing of periodontal defects in animals [4], and the repair of orthodontically induced tooth root resorption in humans[5]. However, little is known about the effects of LIPUS on the proliferation and differentiation of human gingival fibroblasts (HGF) and its possible application in periodontal therapy. Only two studies showed significant increases in cell proliferation and collagen production by HGF with LIPUS [6,7]. However, the effects of LIPUS on HGF differentiation were not examined. Therefore, our research tested the in-vitro effects of LIPUS on HGF osteogenic differentiation as a new tool for periodontal therapy.

2. METHODS:

HGF cells were cultured in 48-well plates at 2.5×10^3 cells/well. HGF received LIPUS treatment for 5 or 10 min/day for 28 days (1.5-MHz frequency and 30mW/cm2 intensity). LIPUS treated and untreated HGF were analyzed for different cell activities at weeks 1, 2, 3 and 4.

2.1. Cell viability assay:

The MTT (4,5-dimethyl thiazol-2-yl)-2,5-diphenyl tetrazolium bromide) assay was used as a measure of cell viability. Yellow MTT is reduced to purple formazan in the mitochondria of living cells, which is then quantified at $\lambda_{absorbance}$: 570nm. Cell survival after LIPUS treatment was expressed relative to untreated control HGF [8].

2.2. Differentiation assay (Alkaline phosphatase "ALP"):

ALP enzyme cleaves the phosphate group from ALP substrate p-nitro phenol phosphate to produce p-nitrophenol, which is measured at $\lambda_{absorbance}$: 405. ALP activity was reported in terms of the p-nitrophenol/ well and normalized to the DNA content (µg DNA/well) in each lysate to obtain specific ALP activity (ALP/DNA) [9].

2.3. Proliferation assay (total DNA content):

The remaining cell lysates from the ALP assay were subsequently used to measure DNA content. They were analyzed with the CyQUANT DNA kit following the manufacturer's instructions using a fluorescent plate reader ($\lambda_{excitation}$ at 480 nm, $\lambda_{emission}$ at 527 nm). [9]

2.4. Reverse-transcriptase polymerase chain reaction (RT-PCR):

Gingival fibroblasts were harvested using Trizol® Reagent. The total RNA was extracted using the RNeasy Mini Kit following the manufacturer's instructions. Then, RNA was quantified fluorometrically ($\lambda_{excitation}$: 480 nm, $\lambda_{emission}$: 527 nm). 0.3 µg of RNA was then used for the RT reaction using the Omniscript kit [9]. The resulting cDNA was used as a template for PCR amplification of osteopontin (OPN) and the housekeeping gene glyceraldehyde-3-phosphate dehydrogenase (GAPDH) was used as the internal control.

3. RESULTS:

3.1. Cell viability:

Both LIPUS treatments did not affect cell viability at any time point.

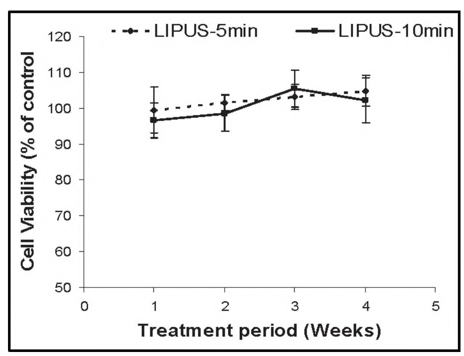


Figure (1): Cell viability for LIPUS treated HGF. Each point represents the percentage of the value to untreated control HGF.

3.2. Specific ALP activity (ALP\DNA):

ALP activity increased with 5min of LIPUS treatment compared to the other groups ($p<0.05$) by week 3. After week 4, ALP activity was increased with both LIPUS treatments 5 min ($p<0.005$) and 10 min ($p<0.05$) when compared to untreated HGF.

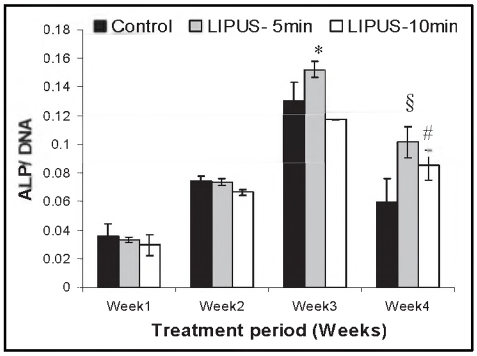


Figure (2): Changes in the specific ALP activity (*: p<0.05 as compared to other groups, §: p<0.005 and #: p<0.05, compared to the control).

3.3. Total DNA Content:

All treatments demonstrated an increase in total DNA content as a function of time, yet there were no significant differences between different treatments at any time point

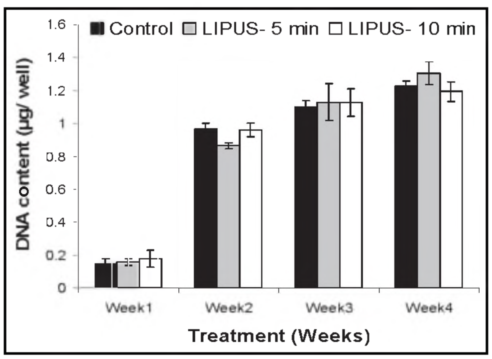


Figure (3): Changes in DNA content.

3.4. RT-PCR:

LIPUS treatment (5min/day) consistently induced significant upregulation of OPN gene expression compared to other groups starting from week 2 (p<0.05), with the highest stimulation observed at week 3 and 4 (P<0.005).

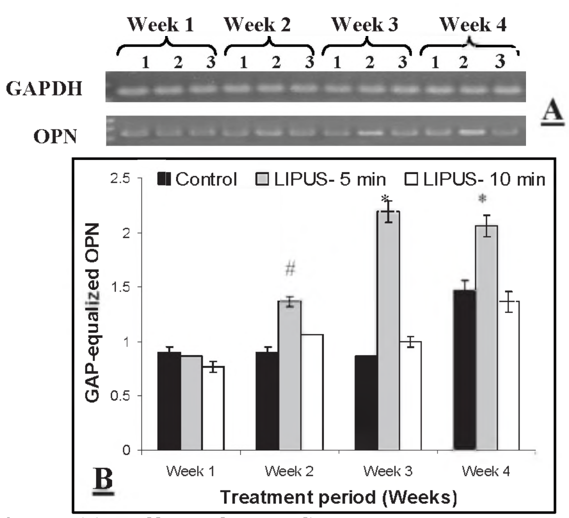


Figure (4): Effect of LIPUS treatment on expression of OPN in HGF, as analyzed by RT-PCR. A. cDNA levels of OPN for 1: control, 2: LIPUS- 5min, and 3: LIPUS-10min group at 1, 2, 3 and 4 weeks. B. Densitometric analyses of OPN expression for LIPUS treated and untreated HGF after equalization with GAPDH. (#: p<0.05 and *: p<0.005 as compared to other groups).

4. DISCUSSION

Our study is the first to examine the effects of LIPUS on the differentiation of HGF. LIPUS stimulation (5 min/day) enhanced the differentiation of HGF, as evidenced by the significant increase of ALP activity and OPN expression. ALP is considered an early marker for osteoblast [1] and cementoblast [2] differentiation. Interestingly, our results are consistent with other studies, which reported significant increases in ALP activity with LIPUS treatment in osteoblasts [1], cemetoblasts [2] and odontoblast-like cells [3]. OPN is an extracellular matrix protein that is expressed in the early stages of mineralized tissue development. LIPUS stimulation of odontoblast–like cells resulted in the significant increase of OPN expression [3]. Our results suggest that furthering the therapeutic usage of LIPUS stimulation would be beneficial for periodontal therapy.

REFERENCES

1. Maddi, A., et al., Long wave ultrasound may enhance bone regeneration by altering OPG/RANKL ratio in human osteoblast-like cells. Bone, 2006. 39(2): p. 283-8.
2. Dalla-Bona, D.A., et al., Cementoblast response to low- and high-intensity ultrasound. Arch Oral Biol, 2008. 53(4): p. 318-23.
3. Scheven, B.A., et al., Short-term in vitro effects of low frequency ultrasound on odontoblast-like cells. Ultrasound Med Biol, 2007. 33(9): p. 1475-82.
4. Ikai, H., et al., Low-intensity pulsed ultrasound accelerates periodontal wound healing after flap surgery. J Periodontal Res, 2008. 43(2): p. 212-6.
5. El-Bialy, T., I. El-Shamy, and T.M. Graber, Repair of orthodontically induced root resorption by ultrasound in humans. Am J Orthod Dentofacial Orthop, 2004. 126(2): p. 186-93.
6. Doan, N., et al., In vitro effects of therapeutic ultrasound on cell proliferation, protein synthesis, and cytokine production by human fibroblasts, osteoblasts, and monocytes. J Oral Maxillofac Surg, 1999. 57(4): p. 409-19; discussion 420.
7. Reher, P., et al., Therapeutic ultrasound for osteoradionecrosis: an in vitro comparison between 1 MHz and 45 kHz machines. Eur J Cancer, 1998. 34(12): p. 1962-8.
8. Issa, Y., et al., Cytotoxicity of metal ions to human oligodendroglial cells and human gingival fibroblasts assessed by mitochondrial dehydrogenase activity. Dent Mater, 2008. 24(2): p. 281-7.
9. Varkey, M., et al., Osteogenic response of bone marrow stromal cells from normal and ovariectomized rats treated with a low dose of basic fibroblast growth factor. Tissue Eng, 2007. 13(4): p. 809-1

ACKNOWLEDGMENT

This project was supported by Dr. Tarek EL-Bialy Start up fund (Faculty of MEdicine and Density, University of Alberta)and NSERC I2I phase 1 grant.

HISTOMORPHOMETRIC ANALYSIS OF THE EFFECT OF THERAPEUTIC ULTRASOUND ON THE DENTOALVEOLAR STRUCTURES DURING ORTHODONTIC FORCE APPLICATION IN-VITRO

B. Lam[2], S. Aldaghreer[1], and A.J. Sloan.[3] T. El-Bialy[1]*,
[1]Division of Orthodontics, Faculty of Medicine & Dentistry; [2]Faculty of Science, University of Alberta, Edmonton, Alberta, Canada and [3]School of Dentistry, Cardiff University, Wales, UK.
*corresponding author, telbialy@ualberta.ca

1. INTRODUCTION

Contact between odontoblasts and the dentin matrix is required to maintain the phenotypic morphology and secretory activity of odontoblasts. [1] To accurately assess the phenotypic morphology and secretory activity of odontoblasts, a longer term model of the dentin-pulp complex is required. Magloire et al. were able to culture human tooth slices such that the cell viability was maintained for a significant period; however the model allowed limited experimental manipulation [2]. In 1998 Sloan *et al* studied the process of dentinogenesis in rat incisor for up to 14 days *in vitro*. They demonstrated that the dentin-pulp complex from mature rodent tissues can be cultured successfully for substantial periods of time and would provide a useful model for the study of dentinogenesis and tissue repairDhopatkar introduced a variation of the tooth slice model which allowed investigation of the dentin pulp complex undergoing orthodontic force in vitro [4].

Low intensity pulsed ultrasound (LIPUS) has been found to stimulate angiogenesis during wound healing and to enhance bone growth into titanium porous-coated implants [5] to enhance bone healing after fractures [6, 7] and bone formation after mandibular osteodistraction [8, 9]. LIPUS can enhance mandibular growth in growing patients with hemifacial microsomia. The expression of bone proteins: osteonectin, osteopontin, and bone sialoprotein has been found to respond to therapeutic ultrasound in a dose dependent manner.

El-Bialy et al. reported that LIPUS minimized orthodontically-induced root resorption and accelerated healing by reparative cementum in 12 patients over a 4 week period of simultaneous tooth movement and LIPUS application. It was reported that LIPUS exposure affected cementoblasts by regulating mRNA expression of alkaline phosphatase, which plays a role in the mineralization process but had no effect on cell proliferation [10]. In addition, it has been reported that LIPUS stimulates odontoblast matrix production and gene production in vitro [11]. This suggests that LIPUS may have distinct effects of cell viability, cell adhesion and gene expression of odontoblasts in culture.

We hypothesize LIPUS will enhance dentine and cementum matrix modelling formation as well enhance alveolar bone and periodontal ligament remodelling.

2. METHODS

2.1 Sample Collection and Preparation

The mandibles were dissected from 28-day-old male SD rats euthanized by cervical dislocation. Transverse sections of approximately 1.5 mm thickness were cut with a 0.006” diamond wafer saw (ISOMET® Wafering Blade, Series 15HC, 3" x 0.006" x 1/2").

The sections were washed several times in washing medium at 37°C immediately after cutting. Slices were placed in culture medium (100 ml), containing DMEM, vitamin C (0.15 mg/ml), 10% heat inactivated fetal calf serum, L-glutamine (200mM) and 1% penicillin/ streptomycin solution. The mandible slices were cultured at 37°C in an atmosphere of 5% CO_2 in air, in a humidified incubator for 24 hours. After changing the media, springs consisting of 0.016×0.025 were applied to each slice. This is what is referred to as the mandible slice organ culture (MSOC). The springs were calibrated to deliver 50 grams of force to each slice across the slice passing through the tooth and periodontal ligament inside the mandible slice. Upon application of the compression force by the spring, the mandible slice organ culture (MSOC) were deformed and resulted in two areas of periodontal ligament compression as well as two areas of tension.

2.2 Application of Ultrasound

After the application of the springs one hour LIPUS was applied using a 2.5 transducer producing incident intensity of 30 mW/cm^2 of the transducer’s surface area to the slices in the treatment group. The sample was divided into three groups Control (n=7), 5 minutes US (n=10), 10 minutes US (n=10) for one week.

2.3 Assessment

Once the treatment period was completed, histomorphometric analysis was performed. The images were captured using a digital camera [(CCD), Leica, Wetzlar, Germany)] with 40X magnification lenses. Field of view for each image was standardized using a calibration ruler of 2mm in length. Histomorphometric analysis was performed using Meta-Morph software (Molecular Devices Corporation, Sunnyvale, CA , U.S.A.). Predentine and cementum thicknesses and cell counts for the odontoblastic and the preodontoblastic layers as well as for the periodontal ligament spaces were calculated for the compression sides and the resultant tension sides. All data collected was analyzed by Multi-Variate Analysis Of Variance (MANOVA) test with Bonferroni test as the sample size was relatively small. Data analysis was performed using SPSS statistical package (Version 15, Chicago, IL, USA).

3. RESULTS

It was observed that both cementum and predentin thickness were increased in the ultrasound group. [Figure 2] In addition, the number of odontoblast, preodontoblast, and periodontal ligament (PDL) cells were increased in the ultrasound groups. The observed cell counts and thickness measurements may be dependent on the amount of ultrasound exposure.

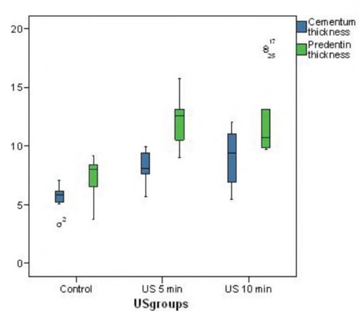


Figure 1

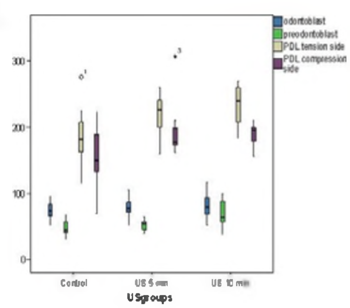


Figure 2

Figure 1. Graphical presentation of cementum and predentine thicknesses

Figure 2. Graphical presentation of cell count in the odontoblast, preodontoblast, PDL in both tension and compression sides.

4. DISCUSSION

Mandible slice organ culture is an effective means of investigating the dentin-pulp complex while allowing experimental manipulation. The application of the springs allowed compression and tension of the complex to replicate effects of orthodontic movements. LIPUS can be applied to this model in such that the entire complex received even distribution of the treatment.

The increase in cementum thickness may be a result of altered mRNA expression of alkaline phosphatase or increase in matrix production by cementoblasts [10]. As stated in the introduction, LIPUS may have stimulated odontoblast matrix production resulting in an increase in preodontoblast and odontoblast cells. Previous studies suggest that LIPUS stimulates PDL cell proliferation [12] which was observed in our results.

From a molecular and cellular level, it can be observed that LIPUS may be beneficial to the restructuring of the mineralized tissue along the dentin-pulp complex when undergoing orthodontic force.

References

1. Heywood, B.R. and J. Appleton. The ultrastructure of the rat incisor odontoblast in organ culture. Arch Oral Biol, 1984. 29(4): p. 327-9.
2. Magloire, H., A. Joffre, and F. Bleicher, An in vitro model of human dental pulp repair. J Dent Res, 1996. 75(12): p. 1971-8.
3. Sloan, A.J., et al., An in vitro approach for the study of dentinogenesis by organ culture of the dentine pulp complex from rat incisor teeth. Archives of Oral Biology, 1998. 43: p. 421-430.
4. Dhopatkar, A.A., et al., Anovel in vitro culture model to investigate the reaction of the dentine–pulp complex to orthodontic force. Journal of Orthodontics, 2005. 32: p. 122-132.
5. Tanzer, M., et al., Effect of Noninvasive Low Intensity Ultrasound on Bone Growth into Porous-Coated Implants. Journal of Orthopaedic Research, 1996. 14: p. 901-906.
6. Abramovich, A., Effect of ultrasound on the tibia of the young rat. J Dent Res, 1970. 49(5): p. 1182.
7. Heckman, J.D., et al., Acceleration of tibial fracture-healing by non-invasive, low-intensity pulsed ultrasound. J Bone Joint Surg Am, 1994. 76(1): p. 26-34.
8. El-Bialy, T.H., et al., The Effect of Pulsed Ultrasound on Mandibular Distraction. Annals of Biomedical Engineering, 2002. 30: p. 1251-1261.
9. Shimazaki, A., et al., Low-intensity pulsed ultrasound accelerates bone maturation in distraction osteogenesis in rabbits. The Journal of Bone and Joint Surgery, 2000. 82-B: p.1077-1082.
10. Dalla-Bona, D.A., et al., Effects of ultrasound on cementoblast metabolism in vitro. Ultrasound in Medicine and Biology, 2006. 32(6): p. 943-948.
11. Scheven BA, Millard JL, Cooper PR, Lea SC, Walmsley AD, Smith AJ. Short-term in vitro effects of low frequency ultrasound on odontoblast-like cells. Ultrasound Med Biol. 2007 Sep;33(9):1475-82. Epub 2007 May 24.
12. Tomofuji T, Kusano H, Azuma T, Ekuni D, Yamamoto T, Watanabe T, Kishimoto T.: Gingival cell proliferation induced by use of a sonic toothbrush with warmed silicone rubber bristles.J Periodontol. 2004 Dec;75(12):1636-9.

Acknowledgement

This research was supported by Dr. Tarek El-Bialy Start-up fund, Faculty of Medicine and Dentistry, University of Alberta, Edmonton, Alberta Canada.

MODELLING HIGH FREQUENCY ACOUSTIC BACKSCATTER RESPONSE FROM NON-NUCLEATED BIOLOGICAL SPECIMENS

Omar Falou[1], J. Carl Kumaradas[2], and Michael C. Kolios[1,2]
[1]Dept. of Electrical and Computer Engineering, Ryerson University, 350 Victoria St., ON, Canada, M5B 2K3
ofalou@ryerson.ca
[2]Dept. of Physics, Ryerson University, 350 Victoria St., ON, Canada, M5B 2K3

1. INTRODUCTION

It has been shown that high frequency ultrasound (20MHz - 60MHz) can be used to detect structural and physical changes in cell ensembles during apoptosis [1]. Apoptosis, or programmed cell death, was originally defined by Kerr *et al.* in 1972 [2] and is characterized by: nuclear condensation and DNA degradation, cytoplasm shrinkage, and fragmentation of the cell into membrane-bound bodies [3]. Ultrasonic backscatter from cell ensembles treated with the chemotherapeutic cisplatin (which induces apoptosis) increased by 9-13dB resulting in much brighter images in the areas the cells respond to the treatment. However, the mechanism that causes this increase in the backscatter is not known. Theoretical models of ultrasound scattering at the cellular level are needed in order to develop methods for using backscatter measurements to determine cell response to treatment. The development of such models requires an understanding of the various mechanical properties of the cell components.

Baddour *et al.* [4] performed successful measurement of high-frequency (10-65 MHz) backscatter frequency responses from a single eukaryotic cell. A recent study by the same group [5] showed that for prostate carcinoma (PC-3) cells whose nucleus to cell volume ratio equals 0.33, the backscatter response could be modeled as a fluid sphere. However, for human acute myeloid leukemia (OCI-AML-5) cells whose nucleus to cell volume ratio equals to 0.50, the backscatter response could not accurately be modeled as a fluid sphere.

This work attempts to investigate the response of some non-nucleated biological specimens by measuring the backscatter response from sea urchin oocytes and comparing it to theoretical predictions from a fluid sphere model to further provide evidence of their fluid-like nature.

2. METHODS

A very sparse suspension of oocytes from purple sea urchin were prepared in artificial seawater ($\rho \sim 1000\ Kg/m^3$, $c \sim 1500\ m/s$) at room temperature. These oocytes were selected because of their spherical shape and narrow size distribution (mean oocyte diameter 74 µm). A VisualSonics VS40B (VisualSonics Inc., Toronto, Ontario, Canada) ultrasound imaging device was used to acquired data using three transducers, with different resonant frequencies, f number, and focal lengths (20 MHz polyvinylidene fluoride: f2.35, 20 mm; 40 MHz polyvinylidene fluoride: f3, 9mm; 80 MHz lithium niobate: f3, 6 mm). These transducers were excited at 19, 40, and 55 MHz pulses, respectively. Only data from the -6-dB bandwidth of each transducer were used in the analysis which gave an overall bandwidth spanning 10-62 MHz. Ten independent acquisitions rf signals were performed using each transducer. The acquired rf lines were then thresholded by discarding all lines not containing any data greater than 90% of the maximum value found in all rf lines for a given transducer in order to remove empty rf lines and indirect oocyte hits. A Hamming window of width of 2 µs was applied to all remaining lines. In addition, visual inspection was performed to eliminate lines which exhibited abnormal pattern due to the presence of more than one oocyte in the resolution volume of the transducer. For each transducer, the remaining rf lines were superimposed, then averaged to obtain a single rf line corresponding to the average backscatter from a single oocyte. The backscatter transfer function ($\mathrm{BSTF}_{\mathrm{expr}}(\omega)$) was calculated as:

$$\mathrm{BSTF}_{\mathrm{expr}}(\omega) = \frac{R_{\mathrm{expr}}(\omega)}{R_{\mathrm{ref}}(\omega)}$$

where $R_{\mathrm{expr}}(\omega)$ is the Fourier transform of the average backscatter signal from a single oocyte. $R_{\mathrm{ref}}(\omega)$ is the Fourier transform of the reference signal (the reflection from a flat S_1O_2 crystal placed at the transducer focus in artificial sea water at room temperature). The values of the backscatter transfer function are shown in the form of plots expressed in relative decibels (dBr) to the backscatter intensity from the reference. The least squares analysis was used to determine the theoretical frequency responses that best agree with their corresponding experimental backscatter transfer functions. This was performed by minimizing R^2, the sum of the squares of the offsets of the theoretical responses from their corresponding experimental BSTFs. Theoretical frequency responses were calculated for a fluid sphere using the Faran scattering model [6] by the letting the Poisson's ratio approach 0.5.

3. RESULTS

Figure 1 shows the theoretical (Faran model: oocyte density=$1168\ Kg/m^3$, speed of sound=$1554\ m/s$) and measured backscatter frequency responses of a single sea urchin oocyte in sea water at room temperature subject to three incident pulses (19 MHz, 40 MHz, and 55 MHz). A contour plot of the least square fit coefficient, R^2, of experimental and theoretical backscatter transfer functions as a function of the oocyte density and speed of sound is shown in figure 2. Experimental data for the 20 MHz transducer (which generated the 19 MHz pulse) were not used in the calculation of R^2 due to the presence of ripples in the experimental backscatter frequency responses of the oocyte (see figure 1). Densities and speeds of sound for the oocyte ranged from $1100\,Kg/m^3$ to $1250\,Kg/m^3$ and from $1520\,m/s$ to $1580\,m/s$, respectively.

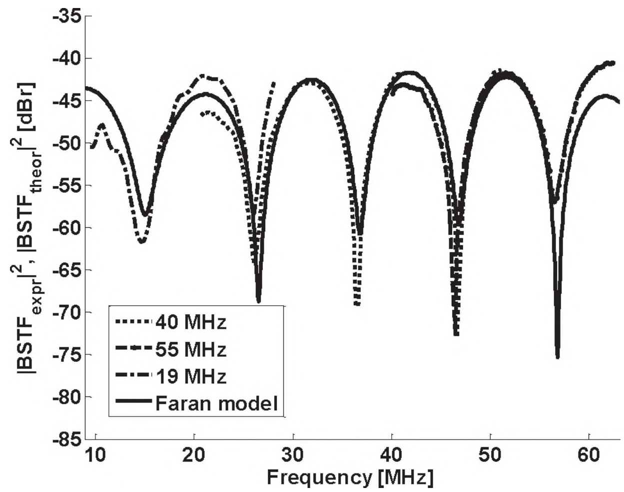


Fig. 1. Theoretical (Faran model) and measured backscatter frequency responses of single sea urchin oocytes in sea water at room temperature subject to three incident pulses: 19 MHz, 40 MHz, and 55 MHz.

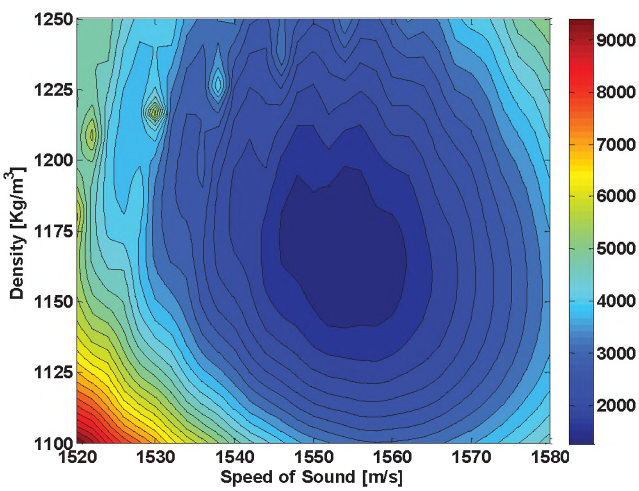


Fig. 2. Contour plot of the least square fit coefficient, R^2, of the experimental and theoretical backscatter transfer functions as a function of the sphere density and speed of sound.

4. DISCUSSION

The absence of the sharp resonance peaks in the frequency responses from a single sea urchin oocyte (figure 1) reveal the absence of detectable shear waves propagation inside the oocyte. This implies that oocytes behave more like fluid scatterers at high ultrasonic frequencies. For the 20 MHz transducer, the presence of ripples in the experimental backscatter transfer functions may be due to the transducer's low performance. This will be further investigated in future work. The least square fit analysis reveals that a density and speed of $1168\,Kg/m^3$ and $1554\,m/s$, respectively, provide the best fit of the theoretical frequency responses with their corresponding experimental backscatter transfer functions. A value of $1168\,Kg/m^3$ for the density is not surprising since sea urchin oocytes tend to settle at the bottom of the container when suspended in sea water. Similarly, the speed of sound in the sea urchin was found to be $1554\,m/s$. This value is very reasonable since it is close to that of biological tissue ($1540\,m/s$).

In conclusion, this study shows that the high frequency ultrasound backscatter from some non-nucleated biological specimens, such as sea urchin oocytes, are best modelled using the fluid sphere scattering model. Future work include the application of this methodology to investigate scattering from AML cells using a finite-element layered model, in which the cell is represented by a fluid shell (cytoplasm) surrounding an elastic sphere (nucleus).

REFERENCES

[1] G. J. Czarnota, M. C. Kolios, H. Vaziri, S. Benchimol, F. P. Ottensmeyer, M. D. Sherar, and J. W. Hunt, "Ultrasonic biomicroscopy of viable, dead and apoptotic cells," *Ultrasound in Medicine and Biology,* vol. 23, pp. 961-965, 1997.

[2] J. F. R. Kerr, A. H. Wyllie, and A. R. Currie, "Apoptosis - Basic Biological Phenomenon with Wide-Ranging Implications in Tissue Kinetics," *British Journal of Cancer,* vol. 26, pp. 239-257, 1972.

[3] G. Hacker, "The morphology of apoptosis," *Cell and Tissue Research,* vol. 301, pp. 5-17, Jul 2000.

[4] R. E. Baddour, M. D. Sherar, J. W. Hunt, G. J. Czarnota, and M. C. Kolios, "High-frequency ultrasound scattering from microspheres and single cells," *Journal of the Acoustical Society of America,* vol. 117, pp. 934-943, 2005.

[5] R. E. Baddour and M. C. Kolios, "The fluid and elastic nature of nucleated cells: Implications from the cellular backscatter response," *Journal of the Acoustical Society of America,* vol. 121, pp. EL16-EL22, Jan 2007.

[6] J. J. Faran, "Sound Scattering by Solid Cylinders and Spheres," *Journal of the Acoustical Society of America,* vol. 23, pp. 405-418, 1951.

ACKNOWLEDGMENTS

This project was funded by the Canadian Institute of Health Research (grant #79447). This research was undertaken, in part, thanks to funding from the Canada Research Chairs Program awarded to Michael Kolios. Omar Falou is the recipient of an Ontario Graduate Scholarship. Funding to purchase the equipment was provided by the Canada Foundation for Innovation, the Ontario Ministry of Research and Innovation and Ryerson University.

CELL EXPANSION GENES EXPRESSION BY THERAPEUTIC ULTRASOUND. PROS AND CONS.

Tarek El-Bialy[1], Taghreed Aldosary[2]

[1]. Departments of Dentistry and Biomedical Engineering, University of Alberta, T6G 2N8; [2]Departments of Dentistry, University of Alberta, T6G 2N8

1. INTRODUCTION

Mesenchymal stem cells (MSCs) have the capabilities for self-renewal and differentiation into cells with the phenotypes of, among others, bone, cartilage, and fat. Because bone marrow derived cells, which are a main source of MSCs, are not always acceptable due to a significant drop in their cell number and proliferative/differentiation capacity with age, there is a need for a technique that is capable of expansion of these cells or finding an alternative stem cell source.

Nucleostemin is a newly discovered nucleolar protein present in both embryonic and adult rat central nervous system stem cells, and several human cancer cell lines [1]. It has been reported also that Nucleostemin gene is a marker of proliferating stromal stem cells in adult human bone marrow [2].

Ultrasound, an acoustic pressure wave at frequencies above the limit of human hearing, is transmitted into and through biological tissues and is being used widely in medicine as a therapeutic, operative, and diagnostic tool.[3, 4]. It has been reported that Low Intensity Pulsed Ultrasound (LIPUS) enhances different types of cell proliferation that includes chondrocytes [5]; skin fibroblasts [6] and increase VEGF vascular endothelial growth factor (VEGF), a growth factor associated with endothelial cell proliferation and migration [7]. The effect of LIPUS on stem cells is not fully studied or understood. It also has been shown that ultrasound stimulates tendon cell proliferation and upregulate of proliferating cell nuclear antigen (PCNA) [8]. Also, therapeutic ultrasound upregulate c-myc, a proto-oncogene, binds DNA through its basic helix-loop-helix/leucine zipper domain, which is essential to cell proliferation and differentiation [9].

Human Umbilical Cord PeriVascular (HUCPV) cells are a potential substitute for BMCs due to the immaturity of newborn cells [10]. Umbilical cord blood transplantation (UCBT) has become an established haematopoietic stem cell therapy for patients with no reported immune-rejection. [10]. Due to donor number limitation is a major constraint to bone marrow mesenchymal stem cell therapy, there is a high need for alternative cell sources for such cell-based therapies [11].

Taking the aforementioned information, we hypothesized that LIPUS can up-regulate the following genes Nucleostemin gene in rapidly proliferative and self renewal cells (Bone marrow stem cells, Human Umbilical perivascular endothelial cells..

2. METHOD

2.1 Cells

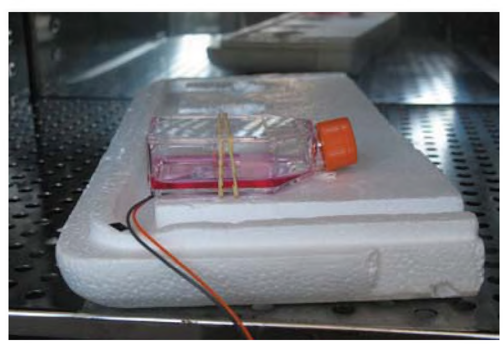
Bone marrow stem cells (BMSCs) were isolated from tibiae and femora of -3 month old male Sprague-Dawley rats (200gms). The femora will be extracted, and the bone marrow cells were flushed with basic media using 10-ml syringe from both sides. Then the cells were centrifuged for 6 min at 600 rpms and were resuspend in tissue culture media, and plated at 5 x 107 cell/mm culture dish (6-well plate), then were incubated in 95% and CO2 at 37° C, with fresh media every 3 days. Cells were expanded and trypsenized and passaged after 2 weeks. Cells were seeded in 75 mm flasks. Cells were expanded in culture medium that contained (Dulbecco's Modified Eagle's Medmium-low Glucose medium (DMEM-LG), 10% fetal bovine serum (FBS), 1% streptomycin/penicillin, and ascorbic acid (50 μg/ml).

Human Umbilical Cord PeriVascular (HUCPV) were donated by Dr. JE Davis, UT, Toronto, ON, Canada. Details about isolation and behavior of these cells are reported previously [11].

2.2 LIPUS

LIPUS was applied for twenty minutes per day to the base of the cell culture flasks. LIPUS was applied using a 2.5 cm2 transducer that delivers 1.5 MHz with a duty cycle of ¼ and delivers SATA intensity of 30 mW/cm2. The power output was calibrated weekly to ensure that the output is maintained throughout the experiment. Ultrasound was applied to the bottom of each cell flask for 20 minutes per day for 7 and 20 days (Figure 1).

Fig. 1. Application of LIPUS to the cell culture flasks.



2.3 RT-PCR

Primers for the Nucleostemin were as follows: Forward:TCCGAAGTCCAGCAAGTATTG; and reverse: AATGAGGCACCTGTCCACTC. The housekeeping gene glyceraldehyde-3-phosphate dehydrogenase (GAPDH) was used as a control. Samples were amplified for 22 - 30 cycles in an Eppendorf Master-cycler thermal cycler. PCR reactions were electrophoresed on 2% agarose gels and visualized under UV light after staining with Ethidium Bromide using image analyzer.

3. RESULTS

3.1 Cell response to LIPUS.

Figures 1 and 2 show PCR reactions gel electrophoresis. It appears that LIPUS has increased Nucleostemin expression in both types of stem cells.

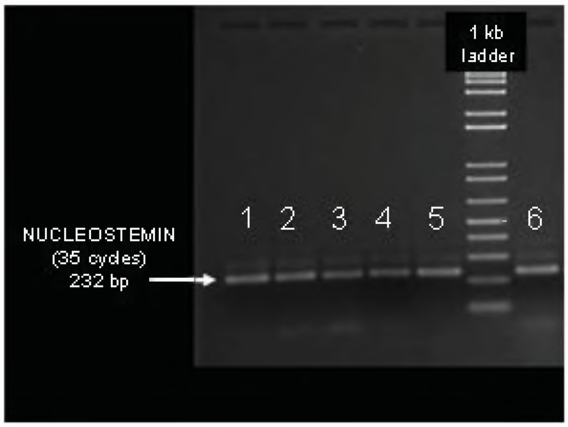


Fig. 1. Nucletemin up-regulation by LIPUS after 7 days in rat BMSCs. RT-PCR of mRNA from BMCs in Culture. 1,2 (T0), 3,4 (Control), 5,6 (LIPUS).

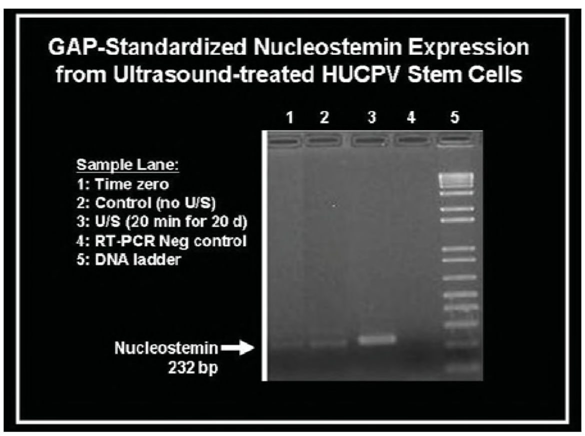


Fig. 2. Nucletemin up-regulation by LIPUS after 20 days (3) compared to control (2) lanes in HUCPVSCs.

4. DISCUSSION

Our results support previous findings that LIPUS can increase gene expression by different cells. The increased gene expression of Nucleostemin is in accordance with increased cell proliferation. Rat bone marrow stem cells responded positively to LIPUS by increased Nucleostemin expression after 7 days, while HUCPVSCs responded to LIPUS by increased expression of Nucleostemin expression by LIPUS after 20 days. This preliminary study suggests that optimization experiment to test expression of Nucleostemin by each type of stem cells be conducted.

Our data suggests that Therapeutic ultrasound might be beneficial in enhancing proliferation of stem cells and maintaining their pluripotent characteristic by continuing expression of Nucleostemin. However this might be a negative effect on neoplastic cells. Since Nucleostemin is also a marker of cancer cells [1], until a study is conducted on the effect of LIPUS on cancer cell proliferation and associated gene expression, it should be recommended that caution must be taken before application of therapeutic application should cancerous cells are expected to be in the vicinity of the application area until further investigation might shows it is safe in such a condition.

REFERENCES

[1] Tsai RY, McKay RD. A (2002) Nucleolar mechanism controlling cell proliferation in stem cells and cancer cells. Genes Dev;16:2991–3003.

[2] WAEL KAFIENAH, SANJAY MISTRY, CHRISTOPHER WILLIAMS, ANTHONY P. HOLLANDER (2006). Nucleostemin Is a Marker of Proliferating Stromal Stem Cells in Adult Human Bone Marrow. STEMCELLS;24:1113–1120.

[3] Ziskin, M., Applications of ultrasound in medicine-comparison with other modalities, in Ultrasound: medical applications, biological effects, and hazard potential, M.H. Repacholi, M. Grandolfo, and A. Rindi, Editors. 1987, Plenum Press: New York. p. 49-59.

[4] Dyson, M., Therapeutic applications of ultrasound., in Biological effects of ultrasound, W.L. Nyborg and M.C. Ziskin, Editors. 1985, Churchill Livingstone: New York. p. 121-133.

[5] Mukai S, Ito H, Nakagawa Y, Akiyama H, Miyamoto M, Nakamura T.: Transforming growth factor-beta1 mediates the effects of low-intensity pulsed ultrasound in chondrocytes..Ultrasound in Medicine and Biology, 2005. 31(12): p. 1713-1721.

[6] Doan, N., et al., In vitro effects of therapeutic ultrasound on cell proliferation, protein synthesis, and cytokine production by human fibroblasts, osteoblasts, and monocytes. J Oral Maxillofac Surg, 1999. 57(4): p. 409-19; discussion 420.

[7] Reher, P., et al., Effect of ultrasound on the production of IL-8, basic FGF and VEGF. Cytokine, 1999. 11(6): p. 416-23.

[8] Tsai WC, Hsu CC, Tang FT, Chou SW, Chen YJ, Pang JH.Ultrasound stimulation of tendon cell proliferation and upregulation of proliferating cell nuclear antigen.J Orthop Res. 2005 Jul;23(4):970-6. Epub 2005 Jan 26.PMID: 16023014.

[9] Park, K. S., Lee, Y. S., & Kang, K. S. (2006). In vitro neuronal and osteogenic differentiation of mesenchymal stem cells from human umbilical cord blood. Journal of Veterinary Science, 7(4), 343-348.

[10] Hofmeister, C. C., Zhang, J., Knight, K. L., Le, P., & Stiff, P. J. (2007). Ex vivo expansion of umbilical cord blood stem cells for transplantation: Growing knowledge from the hematopoietic niche. Bone marrow transplantation, 39(1), 11-23.

[11] Sarugaser, R., Lickorish, D., Baksh, D., Hosseini, M. M., & Davies, J. E. (2005). Human umbilical cord perivascular (HUCPV) cells: A source of mesenchymal progenitors. Stem cells, 23(2), 220-229.

ACKNOWLEDGMENTS

This project was funded by University of Alberta, Faculty of Medicine and Dentistry, Start-up fund and University of Alberta Hospital Fund and Saudi Cultural and Educational Bureau, Canada. The authors would like to thank Dr. Jie Chen and Ying Tsui from the department of Electrical Engineering, University of Alberta, Edmonton, Canada for their help in providing the LIPUS devices.

IN-VIVO ULTRASOUND ASSISTED TISSUE ENGINEERED ARTICULAR CONDYLE

Tarek El-Bialy[1], Hasan Uludag[2], Walied Mousa[3], Saranjeev Lalh[4], Nadr Jomha[5] and Stephen Badylak[6].
[1]Dept. of Dentistry and Biomedical Engineering, University of Alberta, Canada, T6G 2N8, [2]Dept. of Chemical and Material Engineering, University of Alberta, Canada, T6G 2N8, [3]Department of Mechanical Engineering, University of Alberta, Canada, T6G 2N8, [4]Renew Oral and maxillofacial Surgery, Edmonton, Canada, T6G 2N8, [5]Surgery Department, Orthopedic surgery, University of Alberta, Canada, T6G 2N8; [6]McGowan Institute for Regenerative Medicine, Pittsburgh, USA,15219.

1. INTRODUCTION

The increased prevalence of joint degeneration has led to incredible progress in the field of bone and cartilage tissue engineering [1,2]. Regardless of the current applications of cell and tissue-based therapies, such as chondrocyte transplantation [3] for the treatment of articular cartilage defects, prosthetic replacement of the entire articular surface continues to be the foremost practice for total joint replacement [4,5]. In-vitro and in-vivo experiments for tissue engineering articular condyles were proposed to replace synovial joints such as the temporomandibular joint (TMJ) [6,8]. However these trials did not provide convincing evidence of functional integration of the tissue engineered articular condyles [8]. Low Intensity Pulsed Ultrasound (LIPUS) has been shown to enhance osteogenic and chondrogenic differentiation [9-10]. Also, it has been shown that LIPUS enhances bone formation and growth by bone and cartilage matrix production [11-12]. In addition, it has been shown that LIPUS enhances tissue engineering of bone construcst in-vitro [12]. The aim of this study was to evaluate the effect of LIPUS on in-vivo functional integration of autologus tissue engineered mandibular condyle (TEMC) for replacement of excised temporomandibular articular condyles.

2. METHOD

2.1 Animals.

This research was approved by the University of Alberta Animal Care Committee. Eight skeletally mature rabbits were selected and divided into three groups. Group 1 [3 rabbits] (Ultrasound with tissue engineered mandibular condyle), group 2 [3 rabbits] (Ultrasound with empty scaffold), and group 3 [2 rabbits] (empty scaffold, no ultrasound).

2.2 Stem cell isolation and conditioning.

All rabbits had bone marrow stem cells (BMSCs) isolated from their femur bones according to the previously established protocol [13]. BMSCs were differentiated into chondrogenic or osteogenic cells. Chondrogenic medium contained (DMEM+10% FBS + 1% pen-strep*fungizone [as a base medium] \ in addition to 0.1mM None essential amino acid + 50ug/ml ascorbic acid-2-phosphate + 10nm dexamethazone + 5ug/ml insulin + 5ng/ml TGF-B1) while osteogenic medium contained (base medium in addition to 50ug/ml ascorbic acid + 2-phosphate + 10nm dexamethazone + 7mM + 1ug/ml BMP-2). LIPUS was applied for 20 minutes /day for four weeks to the cells to enhance their differentiation.

2.3 Experiment procedure

The chondrogenic and osteogenic differentiated cells were seeded into collagen sponges (Halistate, Dental Implant Technology, Scottsdale, AZ, USA) that were housed into biodegradable extra-cellular matrix (ECM) scaffold to form TEMCs. The biodegradable scaffold was prepared as follows. The urinary bladder of market weight (6 months of age) pigs were harvested following euthanasia and cleaned of excess connective tissues and immersed in 1.0N saline to remove the urothelial cells on the luminal surface. The resulting material was referred to as urinary bladder matrix (UBM). The UBM ECM was then lyophilized. Six sheets of the UBM ECM material were laminated by vacuum pressing with placement of a central accumulation of comminuted (powdered) ECM. This local accumulation of powdered ECM was placed between the third and fourth layers, effectively creating a "pillow". The construct was then sterilized with ethylene oxide (ETO). The central pillow region of the construct served as the meniscal substrate site while the adjacent multi-laminate sheets served as suture anchoring points.

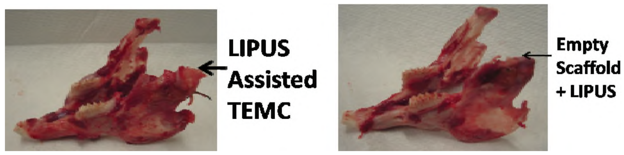
2.3 In-vivo TMJ condylectomy and implantation of the TEMC.

The left TMJ condyle in each rabbit was excised under general anesthesia (Ketamine (70 mg/mL) and Xylazine (10 mg/mL) at a rate of 1 mL/kg IM) according to the previously published technique [14]. Post-surgical care included antibiotic (Baytril 5mg/kg) twice a day SC for about 3 - 4 days and analgesic (Metacam 0.2 mg/kg) once a day for two days SC then 0.1 mg/kg once a day as needed. Rabbits were fed a soft diet for few days then they resumed normal food. The tissue engineered condyles were implanted into the amputated temporomandibular joint (TMJ) articular condyle and were fixed in place using a hardening calcium sulfate/phosphate material (ProDense, Wright Medical Technology Canada Ltd., Mississauga, ON, Canada). In groups 2 and 3, the amputated TMJ articular condyles were replaced by empty scaffolds. Groups 1 and 2 were treated daily for 20 minutes by an ultrasound device that delivers a power of 30 mW/cm2 with pulse frequency of 1.5 M Hz, pulse repetition frequency of 1 K Hz. Four weeks after implantation of the TEMC or empty scaffolds, rabbits were euthanized and evaluated by microCT scanning as well as by histological examination.

3. RESULTS

3.1 Gross anatomy of the dissected mandibles.

Figure 1 shows the gross pictures of the dissected mandibles in each group. It can be seen that group 1 mandible shows comparable morphological and dimensional features of the TEMC and normal condyle. However in all other groups there was no noticeable growth of the excised mandibular condyles.



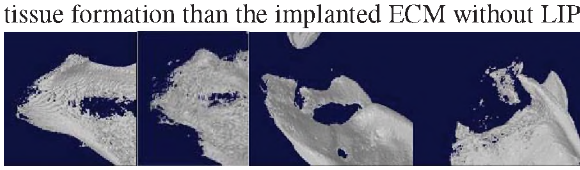
A B

C

Fig. 1. A) Group 1, (B) Group 2 and (C) group 3. Right condyles are normal condyles while the left condyles are the experimental ones.

3.2 MicroCT evaluation.

Figure 2 reveals that the MicroCT evaluation of the TEMC in groups 1 has similar features of the normal condyle, while the empty ECM scaffold with LIPUS shows more hard tissue formation than the implanted ECM without LIPUS.

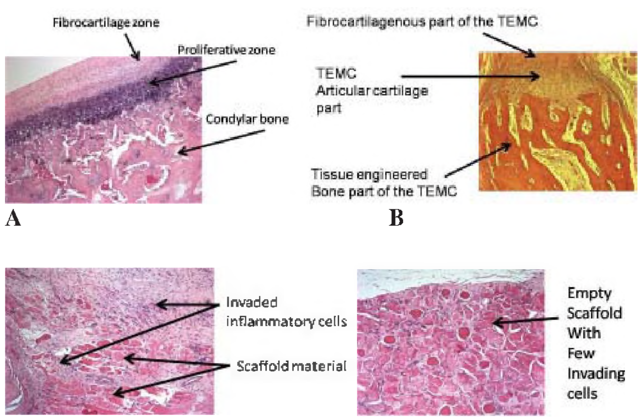


A B C D

Figure 2. MicroCT scanning of the normal (A); LIPUS assisted TEMC (B); Empty ECM scaffold + LIPUS (C) and site of implanted empty ECM scaffolds without LIPUS (D).

3.3. Histological examination

Figure 3 shows the photomicrographs of the histological sections of normal (right) and left TEMC in group 1 as well as the site of implanted ECM scaffold with or without LIPUS.



A B

C D

Figure 3. Photomicrograph of the histological examination of normal condyle (A); LIPUS assisted TEMC in group 1 (B); Empty scaffold with LIPUS (C) and empty scaffold without LIPUS (D).

It can be seen that the LIPUS assisted TEMC had similar morphological features of the normal control mandibular condyle. The implanted ECM with LIPUS showed inflammatory cell invasion of the empty scaffold with LIPUS and empty scaffold TEMCs without LIPUS showed minimal cell invasion of the ECM scaffold.

4. DISCUSSION

To our knowledge, this is the first report that TEMC can successfully replace excised articular joint condyle. Our results support previous findings that LIPUS enhances bone and cartilage matrix formation and integration. However, this preliminary result requires further investigation regarding gene expression in each group as well as mechanical testing of the TEMC. One limitation of this study is that we did not test a positive control of the in-vivo TEMC without LIPUS. This was based on our in-vitro study that showed LIPUS enhanced matrix formation by the differentiated chondrogenic and osteogenic BMSCs.

REFERENCES

[1] Langer RS, Vacanti JP. (1999). Tissue engineering: the challenges ahead. Sci Am.;280:86-9.

[2] Petit-Zeman S. Regenerative medicine. Nat Biotechnol. 2001;19:201-6.

[3] Brittberg M, Peterson L, Sjogren-Jansson E, Tallheden T, Lindahl A. (2003). Articular cartilage engineering with autologous chondrocyte transplantation. A review of recent developments. J Bone Joint Surg Am.;85:109-15.

[4] Cohen J. (1998). Current concepts review. Corrosion of metal orthopaedic implants [letter]. J Bone Joint Surg Am.;80:1554.

[5] Schmalzried TP, Callaghan JJ. Wear in total hip and knee replacements. J Bone Joint Surg Am. 1999;81:115-36.

[6] Abukawa H, Terai H, Hannouche D, Vacanti JP, Kaban LB, Troulis MJ. (2003). Formation of a mandibular condyle in vitro by tissue engineering. J Oral Maxillofac Surg.;61:94-100.

[7] Weng Y, Cao Y, Silva CA, Vacanti MP, Vacanti CA. (2001). Tissue-engineered composites of bone and cartilage for mandible condylar reconstruction. J Oral Maxillofac Surg.;59:185-90.

[8] Alhadlaq, A., and Mao, J..J.: (2005). Tissue-Engineered Osteochondral Constructs in the Shape of an Articular Condyle.J Bone Joint Surg Am.;87-(5): 936-944.

[9] Takayama T, Suzuki N, Ikeda K, Shimada T, Suzuki A, Maeno M, Otsuka K, Ito K.. (2006). Low-intensity pulsed ultrasound stimulates osteogenic differentiation in ROS 17/2.8 cells. Life Sci. 2007 Feb 13;80(10):965-71. Epub Nov 29.

[10] Unsworth J, Kaneez S, Harris S, Ridgway J, Fenwick S, Chenery D, Harrison A.(2007). Pulsed low intensity ultrasound enhances mineralisation in preosteoblast cells. Ultrasound Med Biol.;33(9):1468-74. Epub 2007 Aug 8.

[11] Miyamoto K, An HS, Sah RL, Akeda K, Okuma M, Otten L. (2005). Thonar EJ, Masuda K.: Exposure to pulsed low intensity ultrasound stimulates extracellular matrix metabolism of bovine intervertebral disc cells cultured in alginate beads. Spine. 1;30(21):2398-405.

[12] Moinnes JJ, Vidula N, Halim N, Othman SF. (2006). Ultrasound accelerated bone tissue engineering monitored with magnetic resonance microscopy. Conf Proc IEEE Eng Med Biol Soc.;1:484-8.

[13] Powsner, E.R. and Fly M.N. (1962). Aseptic aspiration of bone marrow from the living rabbit. J Appl. Physiol.;17:1021-2

[14] Dimitroulis G, Slavin J. (2006). The effects of unilateral discectomy and condylectomy on the contralateral intact rabbit craniomandibular joint. J Oral Maxillofac Surg.;64(8):1261-6.

ACKNOWLEDGMENTS

This project was funded by U of A Hospital Fund grant number G533000198 and Dr. Tarek El-Bialy Start-up fund, Faculty of Medicine and Dentistry, U of A, Canada. The authors would like to thank Dr. Jie Chen and Ying Tsui Dept Electrical Engineering, U of A for their kind support in providing the LIPUS devices and Dr. Michael Doschak Faculty of Pharmacy, U of A for his support with the Micro CT scanning of the study samples.

ULTRASOUND TRANSMISSION THROUGH TIME-VARYING PHONONIC CRYSTALS

Derek W. Wright[1] and Richard S.C. Cobbold[1]
[1]Institute of Biomaterials and Biomedical Engineering, University of Toronto, 164 College St, Ontario, Canada M5S3G9

1. INTRODUCTION

Over the past few years there has been a substantial growth of interest into new engineered materials called metamaterials, which include photonic crystals. These materials are designed to manipulate electromagnetic waves, such as microwaves and light, in new ways that enable properties and effects not found in natural materials. This has stimulated major interest in the development and understanding of their acoustic wave analogue, namely phononic crystals [1][2]. Phononic crystals manipulate acoustic waves through scattering to realize numerous unique effects. In particular, it may be possible to achieve sub-wavelength imaging, which has already been predicted and experimentally demonstrated for electromagnetic metamaterials [3][4], and has been predicted for certain photonic crystals [5].

Phononic crystals are artificial crystal arrangements of materials with differing acoustic properties. The incident acoustic waves scatter off of the material interfaces present in the crystal lattice, interfering constructively or destructively. Accordingly, the wavelengths of interest are those comparable to the spacing of the crystal lattice. Some of the effects already experimentally demonstrated in certain phononic crystals are transmission band-gaps [6], negative refraction [7], ultrasound focusing in 2D [1] and 3D [8], and phonon tunnelling [9].

The acoustic properties of a phononic crystal are fixed by its design parameters, such as the crystal lattice configuration and spacing, and the materials used. Once a crystal has been fabricated it is not possible to change its acoustic properties, resulting in a static crystal whose usefulness is limited to its designed operating parameters.

We seek to control the acoustic properties of a phononic crystal by changing its material parameters as a function of time, which will dynamically affect the acoustic wave scattering within the crystal. This may enable time-variant phononic crystals which would have adjustable acoustic properties depending on the nature of the material parameter variation. In order to determine how material parameter variation will affect the acoustic properties of a phononic crystal, we have extended the existing static scattering theory to handle time-varying material parameters. We wish to examine the resulting governing equations to gain insight into the factors that affect acoustic transmission through time-varying phononic crystals.

2. METHOD

2.1. Our new method

We have used the 1D transmission matrix method (TMM) as a starting point since it gives a closed-form solution to wave propagation in periodic media [10]. The existing TMM theory was then expanded to accommodate time-varying material parameters, which we call the *time-varying transmission matrix method* (TV-TMM). Incident waves within a time-varying phononic crystal are modulated at every scattering interface. This continuous interaction between the propagating waves and the time-varying parameters dramatically affects the nature of wave transmission through a phononic crystal. The resulting closed-form expressions can provide insight into how the time-varying material parameters affect wave propagation.

The finite-difference time-domain (FDTD) method [11] can also simulate time-varying phononic crystals, however, it gives no analytical insight into *why* a particular result occurs – only *what* occurs. Moreover, because the accuracy depends on the resolution in the space and time domains, better results take longer to generate. The FDTD method was thus used as a benchmark to verify our new method.

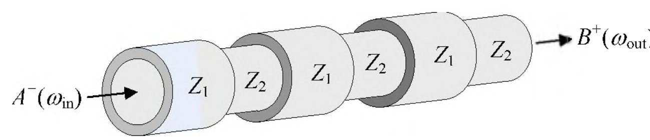


Fig 1. A corrugated tube waveguide. The incident waves are scattered at every impedance discontinuity, which occur between each tube section. $A^+(\omega_{in})$ and $B^+(\omega_{out})$ are the input and transmitted spectra, respectively.

2.2. A representative 1D system

A corrugated tube waveguide is a tube in which the diameter of the tube changes periodically along its principle axis as shown in Figure 1. Acoustic wave propagation within such a corrugated tube waveguide is analogous to acoustic wave propagation with a 1D phononic crystal consisting of alternating layers of differing media. Each tube section of a given diameter has particular acoustic impedance, Z. Thus, the alternating diameter tube shown in Figure 1 has two acoustic impedances: Z_1 and Z_2.

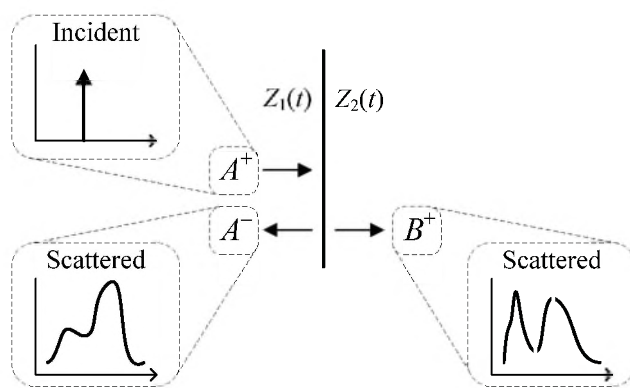


Fig 2. An illustration of how an incident monotone (A^+) is scattered and modulated (A^- and B^+) at every time-varying impedance discontinuity. The inset plots represent frequency-domain spectra.

In a corrugated tube waveguide, when an incident acoustic wave meets an impedance discontinuity the wave is scattered, meaning that some of the energy is reflected and the remainder is transmitted. In the static case, the transmission and reflection coefficients are simply constants, so that the scattered waves have the same frequency content as the incident wave. However, in the dynamic case the impedances are time-varying due to material parameter variation, thus the transmission and reflection coefficients are also time-varying. The incident wave is modulated by the time-varying transmission and reflection coefficients resulting in a spectrum of scattered waves. Such a situation is illustrated in figure 2 for an incident monotone.

3. RESULTS

3.1. Verification

Using the 1D tube waveguide from [12] as a representative model of a 1D phononic crystal, transmission equations for both continuous (including non-periodic) and periodic material parameter variations were derived. The new TV-TMM formulation matches the predictions of our FDTD simulator. Furthermore, the TV-TMM simulation execution times were four to five orders of magnitude faster than the FDTD simulations used for comparison and provide exact solutions at the frequencies of interest.

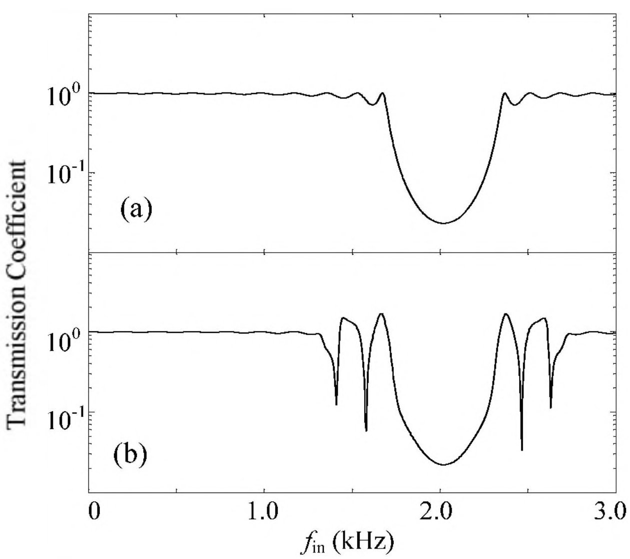


Fig 3. Fundamental frequency transmission coefficients for ten (a) static and (b) time-varying segments. In (b), the parameter variation frequency is 500 Hz and the second tube cross-sectional area is varying by ± 5%.

3.2. Results

The fundamental frequency transmission coefficient for the ten tube segments is shown for both the static and time-varying cases in figure 3, generated by varying the input frequency and plotting the output magnitude of that same frequency. Figure 3a shows the transmission band-gap characteristics centred at 2 kHz in the static case. In figure 3b, the transmission spectrum now includes energy that has been modulated to and from the harmonics by the time-varying transmission and reflection coefficients. Notice that the shapes of the band-gap and band-gap edges have been significantly altered when compared to the static case. This raises the possibility of changing the band-gap characteristics using material parameter variation as the controlling mechanism. One could envision a controllable band-pass or notch filter using such an arrangement.

4. DISCUSSION

Our new method for predicting the transmission properties of time-varying phononic crystals gives closed-form solutions to the transmission through single and multiple time-varying phononic crystal segments. Using this method, we predict that the transmission properties of a time-varying phononic crystal can be made dramatically different from a similar static phononic crystal.

Currently, we are analyzing the resulting equations for acoustic wave transmission through time-varying phononic crystals provided by our new method. We hope to derive a better understanding of the relationship between the nature of the material parameter variation and the resulting output spectra. Furthermore, two and three dimensional phononic crystals have the additional property of acoustic wave refraction, a fact which leads us to ask what effect material parameter variation may have on imaging through these crystals.

To help us answer these questions, we intend to extend our 1D TV-TMM formulation into 2D by applying a similar approach to that used in 2D multiple scattering theory [13]. The extension from 2D to 3D should also be possible since 2D and 3D multiple scattering theories are quite similar in form.

REFERENCES

[1] A. Sukhovich, L. Jing & J. H. Page. (2008). Negative refraction and focusing of ultrasound in two-dimensional phononic crystal, Phys. Rev. B 77 014301

[2] J. H. Page, A. Sukhovich, S. Yang, M. L. Cowan, F. Van Der Biest, A. Tourin, M. Fink, Z. Liu, C. T. Chan & P. Sheng. (2004). Phononic crystals Phys. Stat. Sol. (b) 241 3454–62

[3] J. B. Pendry. (2000). Negative refraction makes a perfect lens Phys. Rev. Lett. 85 3966–9

[4] A. Grbic & G. V. Eleftheriades. (2004). Overcoming the diffraction limit with a planar left-handed transmission-line lens Phys. Rev. Lett. 92 117403

[5] F. AbdelMalek, W. Belhadj & H. Bouchriha. (2005). FDTD study of subwavelength imaging by a photonic crystal slab Photonics Nanostruct. Fundam. Appl. 3 19–24

[6] M. S. Kushwaha & P. Halevi. (1994). Band-gap engineering in periodic elastic composites Appl. Phys. Lett. 64 1085–7

[7] X. Zhang & Z. Liu. (2004). Negative refraction of acoustic waves in two-dimensional phononic crystals Appl. Phys. Lett. 85 341–3

[8] Yang S., Page J. H., Liu Z., Cowan M. L., Chan C. T. & Sheng P. (2004). Focusing of sound in a 3D phononic crystal Phys. Rev. Lett. 93 024301

[9] S. Yang, J. H. Page, Z. Liu, M. L. Cowan, C. T. Chan & P. Sheng. (2002). Ultrasound tunneling through 3D phononic crystals Phys. Rev. Lett. 88 104301

[10] C. E. Bradley. (1994). Time harmonic acoustic Bloch wave propagation in periodic waveguides. Part I. Theory J. Acoust. Soc. Am. 96 1844–53

[11] Y. Tanaka, Y. Tomoyasu & S-i Tamura. (2000). Band structure of acoustic waves in phononic lattices: Two-dimensional composites with large acoustic mismatch Phys. Rev. B 62 7387–92

[12] X. Hu, Z. Hang, J. Li, J. Zi & C. T. Chan. (2006). Anomalous Doppler effects in phononic band gaps Phys. Rev. E 73 015602(R)

[13] P. Martin. (2006). Multiple Scattering (New York: Cambridge University Press)

ACKNOWLEDGMENTS

We wish to acknowledge financial support from the Natural Sciences and Engineering Research Council of Canada and an Ontario Graduate Scholarship to DWW.

NOISE REDUCTION IN HYDRAULIC SYSTEM DRIVEN BY SWASH PLATE PUMP BY OPTIMIZING THE CONTROL UNIT

Molham Hasan Chikhalsouk and Rama B. Bhat
Department of Mechanical and Industrial Engineering –Concordia University, 1455 de Maisonneuve Blvd. West, Montreal, Quebec, H3G 1M8, Canada

1. INTRODUCTION

Swash plate pumps are used in wide range of mobile and stationary applications, where they exhibit certain advantages in comparison to the other modes of power transmission. Some of these advantages are: high specific power, enhanced static and dynamic characteristics, and lower power requirements, when equipped with control unit to deliver specific flow rate that matches the load. However, the pump generates excessive noise when equipped with a double feedback control unit implementing a PD controller. Many approaches were followed to reduce the noise at the design stage. Some of these are implementing a pair of silencing grooves. However, none of the previous studies investigated the effect of the control unit on the quietness of the pump. In the present paper, the effect of using a novel control strategy with a single PID controller on noise levels will be evaluated and compared with the current control strategy.

2. PUMP DESCRIPTION

Swash plate pump mainly consists of a swash plate mechanically connected to a finite number of pistons, which are distributed in a circular array. The pistons are reciprocating in their cylinders and they are nested in a cylinder block. The fluid traffic from cylinders are controlled by means of a port plate. The length of the piston stroke is determined by the angle of inclination of the swash plate, which is varied by a control cylinder.

3. CONTROL UNIT

In order to save on the pump power and to keep the fluid properties within specifications, the pump is equipped with a control unit. The control unit determines the swiveling angle of the swash plate and in turn the piston stroke. In order to push the swash plate and generate the required angle, there is a need for two inputs which are a hydraulic input and an electronic input. The hydraulic input is generated by the hydraulic unit which consists of a secondary pump to generate the pressurized fluid on the control cylinder, and it is connected to the inlet of a hydraulic proportional valve. The valve has a moving part (spool), its housing (sleeve), and two restoring springs that push the spool against a solenoid. The solenoid is activated by the control current, which is generated by the electronic part of the control unit.

4. CONTROL UNIT AND NOISE

There is a relationship between the noise and the control unit, where the type of the controller improves or worsens the quietness of the pump. The control unit follows different strategies such as PD, PID, or PI. It is known that the operation of the pump causes noise, and that noise can be expressed in sinusoidal form, where it has small amplitude and high frequency, as follows:

$$y(t) = A \sin(\omega t) \tag{1}$$

where y is the noise, A is the amplitude of the noise, ω is the noise frequency, and t is time.

Accordingly, in PD controller, the derivative gain amplifies the noise and PID suppresses the noise.

5. CONTROL STRATEGIES

The current model of the pump comes with a double negative feedback controller (the inner loop with a PID controller to control the position of the spool, and the outer loop with a PD controller). A single PD controller proposed by Khalil et al [2] causes excessive noise. Hence, it is proposed to investigate the improvement in the pump performance that can be achieved by replacing the PD by a PID controller.

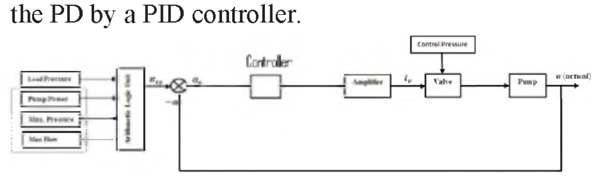


Figure 1: Swash plate pump with single feedback control loop

Figure 1 shows the control strategies with two different controllers (PD/ PID).The two strategies have almost the same structure, and the only difference is the kind of the controller.

6. RESULTS

The noise levels are investigated for a control strategy with a single PD controller proposed by Bhat and Khalil (2002), where the parameters of the controller are the proportional gain of unity and the derivative gain equals 0.02. Also, the noise levels of a control strategy with a

PID controller proposed by Chikhalsouk and Bhat (2008) are studied. The PID is parameterized to have the following gains: The proportional gain equals1.9, the integral gain equals 8.4 and the derivative gain is equal to 0.01. The measured noise levels are shown in Figure 3.

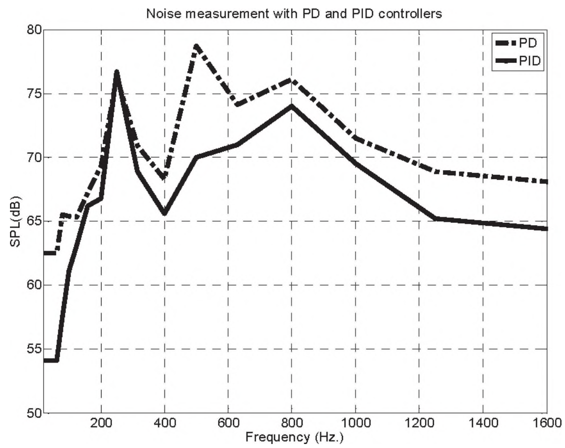


Figure 2: Noise measurement with PD and PID controllers

Figure (2) shows the noise levels measured when the pump is equipped with two different controllers (PD and PID). It can be noticed that the noise levels decrease by replacing the PD controller by the PID controller, particularly at the higher frequencies. The only exception happens at 270 Hz. where they have the same value regardless of the implemented controller.

7. CONCLUSIONS

Swash plate pump must be equipped with a control strategy to save on the pump power and maintain the pump in good quality. The kind of the controller determines the quietness of the pump, where the current model implements a PD controller which amplifies the noise and makes the pump noisy. Using a suitable PID controller with the optimum gain reduces the noise remarkably.

REFERENCES

[1] Chikhalsouk, M. H., and Bhat, R. B., (2008).” Optimizing the Performance of the Swash Plate Pump with Conical Arrangement by Implementing Single Loop Feedback Controller”. Proceedings of CSME forum, Ottawa, Canada, June 6-9, 2008.

[2] Khalil, M. K. B., Yurkevich, V., Svoboda, J., and Bhat, R. B., (2002).”Implementation of Single Feedback Control Loop for Constant Power Regulated Swash Plate Axial Piston Pumps”. International Journal of Fluid Power, Vol. 3, No. 3, pp. 27-36.

THE EVALUATION OF NOISE LEVEL IN HAND-HELD PNEUMATIC TOOLS (ROCK DRILL) BY "PNEUROP CAGI TEST CODE" METHOD

Farhad Forouhar Majd[1,2], Parvin Nassiri[2]
[1]Dept of Occupational Health, School of Public Health, Isfahan University of Medical Sciences and Health Services, Hezar Jerib Ave, Isfahan, Iran, forouhar@hlth.mui.ac.ir
[2]Dept of Occupational Health, School of Public Health and Institute of Public Health Research, Tehran University of Medical Sciences, Enghelab Ave. Tehran, Iran, nassiri@sina.tums.ac.ir

1. INTRODUCTION

The objective of this study is evaluation of noise in hand–held pneumatic tools (Rock drills) used in Lashotor stone mines in Isfahan by the method of PNEUROP CAGI TEST CODE specifies the procedures which can be applied to certain items of construction equipment.

Extremely loud noise damages hearing. In rare cases, the damage is momentary and caused by shock waves from, for instance, an explosion [1]. Damage to hearing normally occurs after many years of exposure to loud noise. Such damage is not sudden and total, but a gradual deterioration of the hearing which is most pronounced around the frequency of 4000 Hz [2]. Compressed air driven hand-tools are often used as examples of noisy machines. This is not surprising considering that air tools are extremely common and many old, compressed air tools are noisy [3].

This method is designed to evaluate noise and vibration propagated by the hand–held pneumatic tools [4]. Rock drill is one of the four product categories specially identified as a major source noise and vibration in the federal noise control Act of 1972 [5]. There are three different types of noise sources including: 1. Process noise, produced by contact between the machine and a surface, 2. Exhaust air noise, partly caused by the flow variation as the compressed air passes the motor, partly by the aerodynamic sound generated in the exhaust air channel, and 3. Vibration radiated noise, from the surface of the machine produced by the moving parts of machine and the flow of air inside the machine [6, 7].

In this study the rock drill produced noise in above specify test method, its control procedures are also recommended.

2. METHOD

2.1 PNEUROP CAGI TEST

The method is based on ANSI (1971) for measuring hand-held pneumatic tools overall noise such as rock drill, paving breaker, and is also famous to "PNEUROP CAGI TEST CODE" [8, 9]. This paper will deal with the methods required for determining the amount of noise produced by a rock drill. The aim of these measurements is either to obtain a value of the overall noise emitted by the tool, or to understand why a tool emits noise and therefore provide a basic for making noise reducing modifications.

2.2 The equivalent value and noise dose

High sound levels over long periods of time damage hearing. For this reason, major efforts are being made to reduce sound at workplaces to harmless levels. In the first step it is possible to deal with methods and criteria for determining the amount of noise a worker is exposed to. Rock drill produced a continuous noise based on ISO standard and therefore we can calculate the total dose in regarding to duration and level of the various noises which the worker is exposed with the aid of the following formula:

$$Leq = 10\,.\,Log\left[\frac{1}{T}\sum t_i\,.\,10^{0.1\,.\,L_{pi}}\right]$$

2.3 measurement conditions

In the second step, the measurements shall be performed with the center of the machine, 1 meter above a reflecting floor [10]. In fact, the noise assessment will be done on the spherical area so that geometric center of drill should be located in the spherical center and also in a 1 meter distance of land. All values are measured in terms of sound pressure level in a situation of "A" for frequency weighting and "slow" for time weighting by B&K 2230 sound level meter and 1625 analyzer. Table 1 shows the allowable noise levels of construction equipment including rock drill given by ANSI S5.1-1971.

Table 1. Allowable values of construction equipment sound level, ANSI S5.1-1971.

Equipment	A-Weighted Sound Level, dB(A)
Earthmoving :	
front loader	75
backhoes	75
dozer	75
tractor	75
scraper	80
grader	75
truck	75
paver	80
Material handling:	
concrete mixer	75
concrete pump	75
crane	75
derrick	75
Stationary:	
Pumps	75
generators	75
compressors	75
Impact:	
pile drivers	95
jack hammers	75
rock drills	80

2.4 Sound power level

In order to calculate the sound power level of a drill by use of sound pressure level [11] so that the drill is located in the center of a hypothetical sphere with the surface of 'S' and the reference surface of 1square meter, the following formula can be used:

$$Lw_A = L_p + 10Log_{10}(S/S_0)$$

3. RESULTS

As can be seen, SPL values are shown in Table 2. The background values are defined for two situations while compressor was turned on or off. The overall sound pressure level based on ANSI S5.1-1971 has been equal to 94.63 dB (A).

Table 2. Exposure levels of measured SPL in a rock drill

Sound pressure levels(dB)	Max sound levels(dB)	Background sound levels(dB)
95.5	105	55.7*
94.5	103	37.3**
94	101	
93.8	102	
95.5	104	
$SPL_{(f)}$=94.63	$Max_{(f)}$=103	Leq=88 (dB)
*The compressor turned on **The compressor turned off		

The graph below shows an increase gradually in sound pressure level generated by the rock drill from frequency of 31.5 Hz to 4000 Hz with a fall in 1000 Hz frequency gotten by sound level meter in a frequency analysis of 1/1band octave.

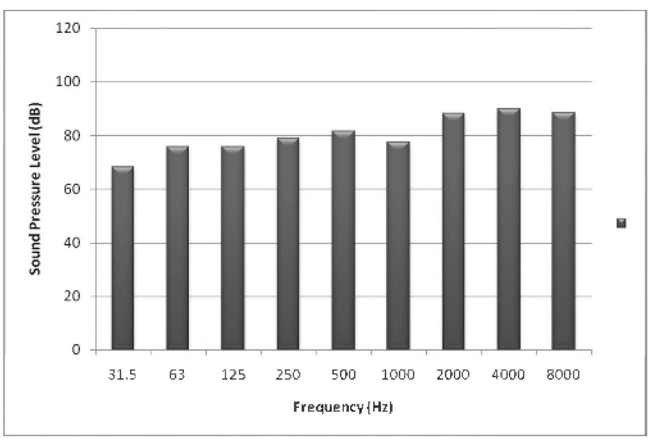


Fig. 1. 1/1 octave band frequency distribution of the rock drill radiated noise from 31.5 Hz to 8000 Hz using by miners in stone mines.

The critical frequency was 4000 Hz with a value of 90 dB to 105 dB depending largely on the design of the work piece and given load to drill. Figure 1 shows the values of a rock drill SPL applied in the stone mine measured in a distance of two meter of rock drill.

A wide range of noise frequencies produced by the rock drill are shown in Figure 2. There is a slightly rise from 20 Hz to 8000 Hz frequencies with a peak of 90.1 dB at 4000 Hz. The graph shows higher levels of noise in high frequency limits. It can be important while using sound absorbing material to control noise emitted from the exhaust system of pneumatic tools. Perhaps achieving control approaches look complicated but the above assessment can help us to apply one or more of control methods given below.

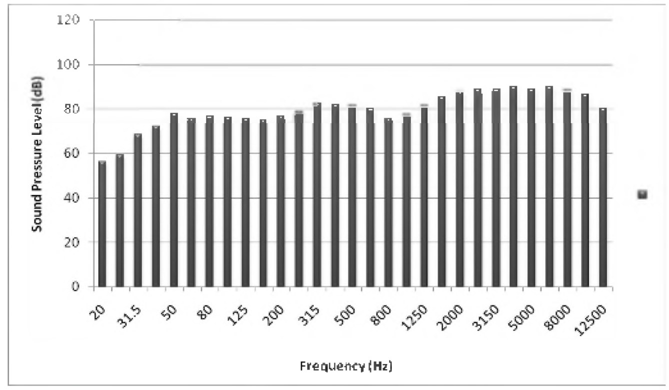


Fig. 2. 1/3 octave band frequency distribution of the rock drill radiated noise using by miners in stone mines from 20 Hz to 12500 Hz.

4. DISCUSSION

The overall SPL will get from the values and evaluated levels around the drill so that we can find the sound power level of drill. This is necessary factor to apply control approaches on machine.

However, there are three different methods for control of a rock drill noise applied in stone mines [12, 13].

- Reduction of process noise
- Exhaust air noise reduction
- Reducing vibration-induced noise

The machine casing of hand-held pneumatic tools is a source of noise although the levels are usually not large compared with, for instance, exhaust air noise.

REFERENCES

[1] Handbook of noise and vibration control, *Malcolm J. Crocker*, 2007.
[2] Industrial health, *M Takamatsu*, Journal, 1982.
[3] Measurement of airborne noise emitted by construction equipment intended for outdoor use, *ISO*, 1976.
[4] Handbook of noise and vibration control, *Antony Barber*, 1992.
[5] Surface rock drills, *Atlas Copco*, 1991.
[6] Handbook of noise control Chap 13-construction equipment *M C. GrawHill* – Book company, *Harris* .C 1973.
[7] Sound and vibration, VOL 10, page 6-33, Purcell, 1976.
[8] Test code for the measurement of sound from pneumatic equipment *ANSI*, 1971.
[9] Test – site measurement of maximum noise emitted by engine powered equipment, *ASA*, 3, 1975.
[10] Ergonomic tools in our time, *Atlas Copco tools*, 1988.
[11] Method for designation of sound power emitted by machinery and equipment by machinery and equipment, *ASA*, 5, 1976 *ANSI*, 1976.
[12] Sound and vibration, Vol5. Page 4, *Anonymous*, 1991.
[13] Noise control foundation, *LEPOR – M*, 1983.

THE EVALUATION OF WHOLE-BODY VIBRATION LEVEL IN HAND-HELD PNEUMATIC TOOLS (ROCK DRILL) BY "PNEUROP CAGI TEST CODE" METHOD

Farhad Forouhar Majd[1,2], Parvin Nassiri[2]
[1]Dept. of Occupational Health, School of Public Health, Isfahan University of Medical Sciences and Health Services, Hezar Jerib Ave, Isfahan, Iran, forouhar@hlth.mui.ac.ir
[2]Dept. of Occupational Health, School of Public Health and Institute of Public Health Research, Tehran University of Medical Sciences, Enghelab Ave, Tehran, Iran, nassiri@sina.tums.ac.ir

1. INTRODUCTION

The evaluation of whole-body vibration in hand–held pneumatic tools (Rock drills) used in lashotor stone mines in Isfahan by the method of PNEUROP CAGI TEST CODE [1, 2] shows the methods which are provided how to use of construction equipment in stone mines under work circumstances.

All hand-held machines transmit vibrations to our hands while we are working with them [3]. Vibrations consist of a reciprocal motion in handles which arises when a number of forces with varying directions and magnitude influence the machine and set it in motion. Work with intensely vibrating machines for a long period of time may give rise to different types of injury. Main types of injuries likely to occur are largely vascular injury, nerve injury, skeletal injury and joint injuries. Perhaps, the most famous injury is vibration induced white fingers or Raynaud's phenomena [4].

This method is designed to evaluate noise and vibration propagated by the hand–held pneumatic tools [5]. Rock drills are categorized as the major sources of noise and vibration by the federal noise control Act of 1972 [5, 6].

The paper demonstrates vibration values generated by rock drill in three directions of X, Y and Z axes and of course a comparison of three vital indices for workers while working with equipment with the standard graphs related to the above specify test method.

2. METHOD

2.1 Physical quantities

It has been found that measurement of vibrations on hand-held tools is far more difficult than we originally envisaged. When measuring vibrations, it is very difficult to decide subjectively whether or not the results of our measurements are reasonable. A motion can be described in terms of displacement, velocity or acceleration [7]. Between these three units, displacement is the most obvious unit for engineer and the unit of acceleration is particular useful in making human vibration measurements [8, 9].

2.2 Whole-body indices

This paper focuses on measuring and predicting of created vibrations as whole–body indices in a rock drill used in stone mines by B & K vibration meter 2512 model. In order to evaluate them we had to locate the whole–body accelerator on the ground where the operator was standing and processing his work [10]. The indices include Reduced Comport (RC), Fatigue Decreased Proficiency (FDP), and Exposure limit (EL) for a frequency response of 1-80 Hz.

3. RESULTS

As can be seen in table 1, obtained conclusions have been shown considerable values of vibration acceleration in Z, X and Y axes for exposure limit and fatigue decreased proficiency respectively. All values are considered for a 5 hour time of exposure.

Table 1. Exposure levels of whole-body transmitted vibrations from a rock drill in X, Y, and Z axes.

Vibration levels	X,Y (dB)	X,Y(m/s^2)	Z (dB)	Z(m/s^2)
Exposure Limit				
Peak	110	0.31	142	12.5
Leq	101	0.11	130	3.2
rms	103	0.14	127	2.2
Fatigue Decreased Proficiency				
Peak	123	1.4	137	7
Leq	101	0.25	121	1.1
rms	112	0.39	125	1.77

This is clear from figure 1 that the emitted vibrations from rock drill are strongly issued in Z direction of whole-body vibration axes with a peak of 12.5 m/s^2 for E.L.

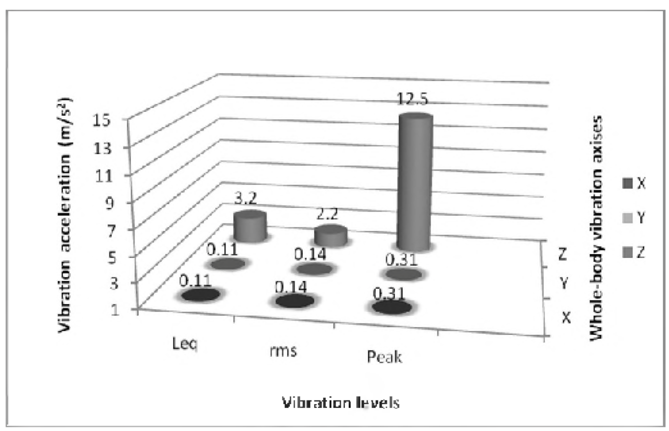


Fig. 1. Evaluated measures of Exposure Limit index of whole-body vibration acceleration produced by a rock drill used in stone mine in three different axes (X, Y, Z) for leq, rms and peak vibration levels.

The obtained results for fatigue decreased proficiency index demonstrate a peak of 7 m/s^2 acceleration in Z direction compared to acceleration values of other vibration axes.

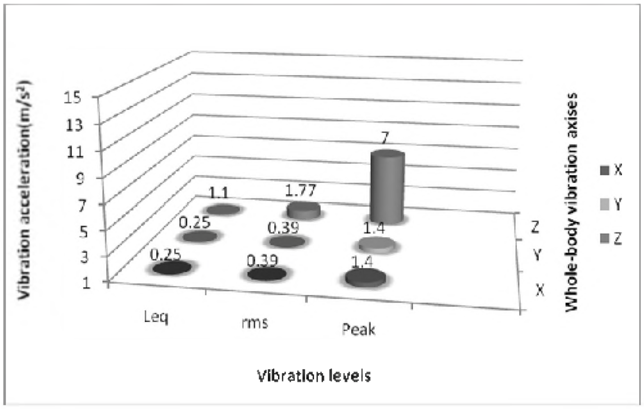


Fig. 2. Evaluated measures of Fatigue Decreased Proficiency index of whole-body vibration acceleration produced by a rock drill used in stone mine in three different axes (X, Y, Z) for leq, rms and peak vibration levels.

The given values in table 2 are extracted from the ISO 2631, 1985 graphs to describe permissible exposure levels of whole-body vibrations in low frequencies of a hand-held pneumatic tool [11].

We have concluded that the operator of rock drill should work only for a period of 1.5 hour per day when receiving all band of frequencies from 1 to 80 Hz, otherwise we should protect him against vibration low frequencies.

The situation for fatigue decreased proficiency index is different and more limited than exposure limit index. In the table, the maximum permitted period of exposure to vibration is 30 min in a day. In other words, the FDP index says that operators should be protected against 1-30Hz frequencies so that they are able to work for a long time with the rock drill.

Table 2. Allowable exposure time to rock drill whole-body vibration.

Criteria	Allowable exposure time(hr)	Frequency rate(Hz)
Exposure Limit	8	25-80
	5	18-80
	4	15-80
	2.5	11-80
	1.5	1-80
Fatigue Decreased Proficiency	8	40-80
	5	30-80
	2.5	16-80
	1	11-80
	0.5	1-80

4. DISCUSSION

Although we were faced to high measures of vibrations in different axes, but the most important of index in comparison with its recommended limits can be exposure limit, because it is a better weighted index for human vibration in workplaces [12]. Reduced comport Index is not also recommended for occupational and industrial jobs because of having more limitations.

With regarding to the above results the whole-body vibration frequencies from the specified rock drill and in low frequencies can be considered as a hazardous vibration source for human body especially for circulatory system, nerves and skeleton. We must pay attention to this subject that many of these workers prefer to do their duties while sitting on the handles of drill because of long period of time for drilling.

However, reduction of vibrations usually means that a thorough analysis of dynamic forces in the machine and process must be performed. This force analysis is followed by a study in different mass spring systems. Acceleration can be reduced by decreasing the force of excitation or increasing the effective mass. Acceleration can also be reduced through attenuation via a mass spring system [13, 14].

In each case, reduction of emitted vibration can influence the vibration induced noise depending on vibration frequencies and therefore one of the best approach of control noise and vibration of a handle tool is vibrating surface elimination through:

- increasing the mass of drill body
- isolating the outer shell of motor
- balancing power transmission system

REFERENCES

[1] Handbook of noise control Chap 13-construction equipment *M.C GrawHill* – Book company, *Harris* .C 1973.
[2] Ergonomic tools in our time, *Atlas Copco tools*, 1988.
[3] Handbook of human vibration, *M.J Griffin*, 1990.
[4] Industrial health, *M Takamatsu*, Journal, 1982.
[5] Test code for the measurement of sound from pneumatic equipment *ANSI,* 1971.
[6] Surface rock drills, *Atlas Copco,*1991.
[7] Handbook of noise and vibration control, *Malcolm J. Crocker*, 2007.
[8] Handbook of noise and vibration control, *Antony Barber*, 1992.
[9] Noise control foundation, *LEPOR – M*, 1983.

[10] Sound and vibration, vol. 5. Page 4, *Anonymous,* 1991.
[11] Sound and vibration, vol. 10, page 6-33, *Purcell*, 1976.
[12] Test – site measurement of maximum noise emitted by engine powered equipment, *ASA*, 3, 1975.
[13] Measurement of airborne noise emitted by construction equipment intended for outdoor use, *ISO*, 1976.
[14] Method for designation of sound power emitted by machinery and equipment by machinery and equipment, *ASA*, 5, 1976 *ANSI*, 1976.

An Investigation into Wind Generated Aero-Acoustic Tones

Peter Alexander Ibbotson[1], Jason Cambridge[2] and John Pearse[2]
[1]Marshall Day Acoustics, Christchurch, New Zealand, peter.ibbotson@marshallday.co.nz
[2]School of Mechanical Engineering, University of Canterbury, Christchurch, New Zealand

1. INTRODUCTION

In early 2008, a newly constructed six level car parking building at a private hospital in Christchurch, New Zealand was the cause of complaint due to a tonal noise audible during windy conditions. Subjectively the noise was "like someone rubbing their finger around a crystal wine glass". A review of literature suggested that the noise could be generated by an aero-acoustic tone generated by wind flowing either over a thin obstruction or into a slot [1]. The building had two different types of balustrades and a number of thin slots, however the external balustrades were considered to be the most likely cause of the tone, due to their location where they were subjected to high wind velocities. The external balustrades were constructed from 36 steel balusters of dimensions 50mm x 6mm x 1050mm. The balusters were arranged vertically and spaced at 93mm centres. 70 balustrades were located in the building.

2. RESULTS OF TESTING

Testing was performed during windy conditions using FFT analysis to allow a narrow band assessment of the tone to be performed. The following graph shows the results of one measurement:

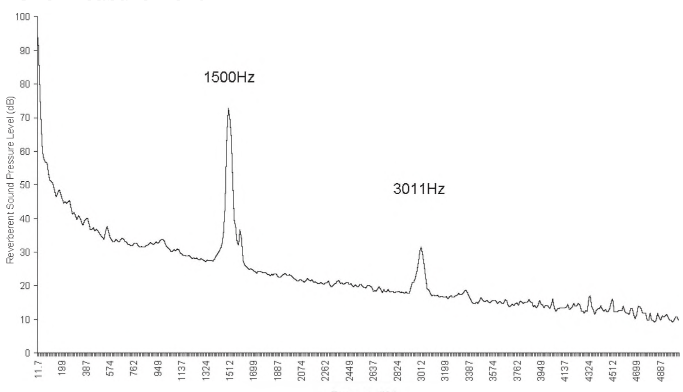


Figure 2.1: Reverberant sound pressure level in car park during tone generation. A distinct tone at 1500 Hz with a secondary tone at 3011Hz can clearly be seen. A smaller peak at 1550Hz is also shown.

A forced vibration response of the balusters was measured by exciting the fundamental mode natural frequency with a hammer and measuring the response with an accelerometer. The results show that the external baluster does have a natural frequency response at 1550 and 2753Hz in bending. Although a natural frequency response was not observed at 1500Hz it is possible that the baluster has a 1500Hz natural frequency of vibration in another axis that was not measured by the single axial accelerometer used for the measurement. Figure 2.2 shows the results of these measurements:

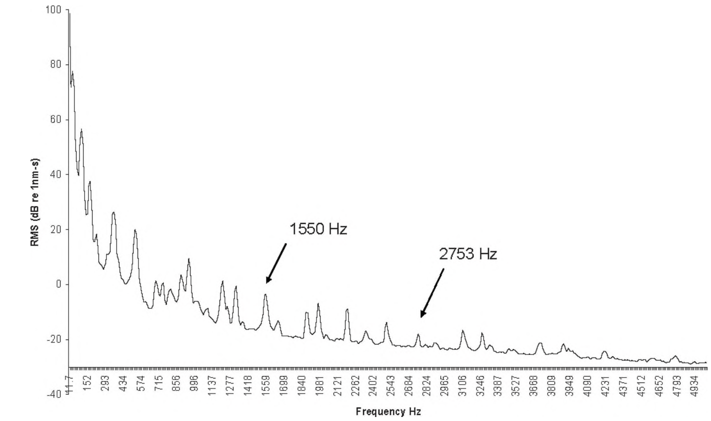


Figure 2.2: Forced frequency response of the external balustrade. The graph shows a natural frequency response at 1550Hz and 2753Hz.

Vibration monitoring of balustrades was also performed during windy conditions when the tone was being generated. Figure 2.3 shows the results of the monitoring of the external balustrade:

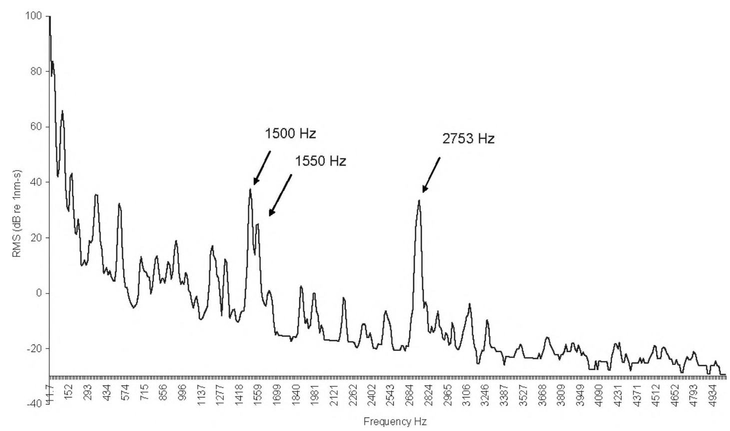


Figure 2.3: Vibration measurement of external balustrade during tone generation. A response at 1500Hz, 1550Hz and 2753Hz was observed.

The above results clearly show a vibration at around the same frequencies as the measured airborne tone. Given that the measurements of forced vibration showed a natural response at these frequencies it was considered the airborne tones were generated by the vibrating balusters. The mechanism by which the balusters are forced to vibrate was considered likely to be vortex shedding or a standing wave effect caused by the spacing of the balusters.

The standing wave frequency between the balusters was predicted using the formula f = V/D where f is the frequency of the standing wave, V is the speed of sound in air and D is the wavelength of which half corresponds to the distance between the balusters. Given the balusters are spaced at

93mm the standing wave frequency was predicted to be 1814Hz. This does not correspond with any of the frequencies measured.

An attempt to predict the vortex shedding frequency was performed using the equation for the Strouhal number $S(c) = f_v c/U$ where f_v is the dominant shedding frequency, U is the velocity of the free stream air and c is the length of the baluster parallel to the wind direction. The balusters were effectively flat plates with a blunt edge and a chord/thickness ratio of 8.3. Research has shown that Strouhal numbers for these types of plate vary stepwise with chord/thickness ratio and are related by integer multiples of 0.6 [2]. However complicating this prediction was the fact that the wind direction was observed to be at an angle to the balustrade during tone generation, and that the balusters did not have a true blunt edge but a slight rounding. This made the correct Strouhal number difficult to determine. In any event, even assuming a Strouhal number of 1.8 [2] the observed tones did not line up well with the predicted natural Strouhal frequency. In spite of this it was considered that vortex shedding was the most likely cause of the vibration of the balusters.

3. WIND TUNNEL TESTING

In order confirm that the external balustrade was responsible for the tone generation, wind tunnel testing was performed. Testing at a range of frequencies and balustrade angles showed that the tone under investigation was generated only between 10.7 – 13.6 m/s at an angle of 24° ± approximately 3° from the wind flow direction. Figure 4.1 shows the generated sound pressure level and frequency as a function of wind speed.

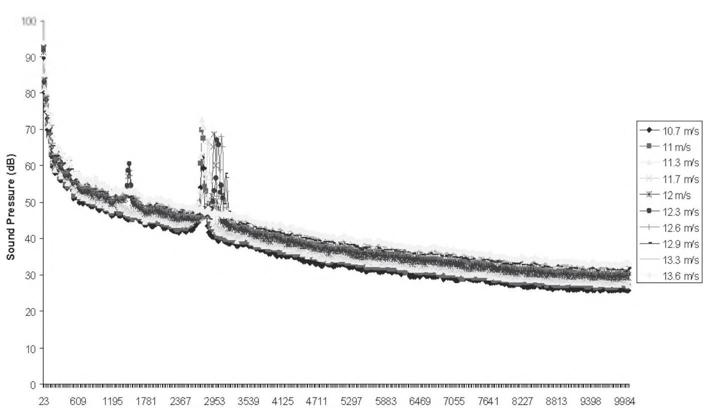


Figure 3.1 Sound pressure level and frequency as a function of wind speed.

It can be seen that both tones were successfully generated, however the tone at around 3000Hz was louder in the wind tunnel whereas the 1500Hz tone was louder in-situ. It can be seen that the tone varies with wind speed from 2742Hz to 3140Hz for wind speeds between 10.7 to 13.6 m/s.

The vibration response of the baluster was also measured during tone generation. Vibration responses between 1500 and 1550Hz and between 2742 to 3140 Hz were observed. The largest vibration response by a significant margin was observed at 11 m/s at a frequency around 2750Hz which is likely due to the natural mode of vibration of the baluster at this frequency.

3.1 Treatment Testing

The following treatments were tested in an attempt to eliminate the tone;

1. 14mm x 35mm expanded metal;
2. 50 x 50mm coarse square wire mesh;
3. 19mm x 19mm fine square wire mesh, and
4. Plastic stripping approximately 12mm in diameter.

All treatments were affixed to the leading edge of the balustrade. It was predicted that by affixing the above treatments the airflow over the balusters would be disrupted and the Strouhal vortex shedding frequency would be altered. All treatments were successful in suppressing the tone.

Vibration measurements were performed during testing of each treatment which showed that although the tone was suppressed by the treatments, some vibration remained for all treatments other than the plastic stripping. It is noted that in similar experiments by others, certain trailing edge treatments, such as perforated sheeting, have caused the amount of noise generated to increase [3]. On this basis the plastic stripping was considered the most suitable treatment.

Additional experiments showed that where plastic stripping was applied to each alternate baluster this reduced the level of noise considerably, however the tone still remained. Turning the balustrade around so that it was affixed at the leading edge rather than the trailing edge also reduced, but did not eliminate, the tone.

3.2 Hotwire anemometry testing

Hotwire anemometer testing was performed during conditions where the tone was being generated. This testing showed that the amount of turbulence from the balustrade reduced significantly when the leading edge treatment was affixed

REFERENCES

[1] Parker, R (1997). Aeroacoustics, *International Journal of Fluid Dynamics*, 1, 1.
[2] Nakamura, Y. Ohya, Y. and Tsuruta, H (1991). Experiments on vortex shedding from flat plates with square leading and trailing edges. *J Fluid Mech*, 222, 437 – 447.
[3] Scott, C. Girdzlusz, E (2008) Whistling Gratings. *Acoustics Bulletin*, Jan/Feb, 18-20

A JUSTIFICATION FOR USING A 45 dBA SOUND LEVEL CRITERION FOR WIND TURBINE PROJECTS

Stephen E. Keith, David S. Michaud, Stephen H.P. Bly
Healthy Environments and Consumer Safety Branch, Product Safety Directorate, Consumer and Clinical Radiation Protection Bureau, Acoustics Division, 775 Brookfield Rd. 6301B, Ottawa, Ontario Canada, K1A 1C1 skeith@hc-sc.gc.ca

1. INTRODUCTION

When required, Health Canada provides advice based on well-accepted scientific evidence for a link between noise exposure and health. Such advice may be requested by Responsible Authorities designated under the *Canadian Environmental Assessment Act* (CEAA) [1] to determine whether it is likely that noise related to a project will cause significant adverse effects. The intent of the CEAA is to ensure that actions are taken to promote sustainable development without causing significant effects. A change in percentage highly annoyed with noise (%HA_n) has been used as one of the measures to determine health impacts in environmental assessments for noise [2], including noise generated by wind farms.

Wind energy is projected to increase to 10000 MW by 2015 in Canada, placing a growing demand on Health Canada to provide health effects advice for proposed wind turbine projects (see refs in [3]).

This paper summarizes how Health Canada derived a noise criterion of 45dBA as the level at which mitigation is recommended for wind turbines operating in quiet rural areas. This criterion is intended to avoid noticeable rattles, sleep disturbance and an increase in %HA_n greater than 6.5%.

2. HISTORY OF %HA_n AND ITS CURRENT USE BY HEALTH CANADA IN ENVIRONMENTAL ASSESSMENTS

In 1978, Schultz published a synthesis of international research on community reaction to transportation noise that provided a relationship between the %HA_n as a function of day-night sound level (Ldn) of the transportation noise source under study. Updates to the Schultz curve (reviewed in [2]) have included the ISO 1996-1:2003 standard [4] (adopted without change by CSA in 2005) where the relationship between the rating level (RL) and %HA_n is given by:

$$\%HA_n = 100/[1+\exp(10.4-0.132*RL)] \qquad \text{Eq.1}$$

The RL in Eq. 1 is typically an adjusted Ldn, with adjustments made depending on the type of noise source and source characteristics. ISO 1996-1:2003 notes that research has shown that there is a greater expectation for and value placed on "peace and quiet" in quiet rural areas. This may be equivalent to a RL adjustment of up to 10 dB. If, as is usual, a wind farm is situated in a quiet rural community, then a quiet rural area adjustment is a basis of Health Canada's proposed criterion.

Based on the characterization of a "severe" noise impact in a report developed by Hanson et al [5] for the U.S. Department of Transportation, Health Canada has recommended that noise mitigation be considered when a *project-related*, long-term increase (from baseline), in the calculated %HA_n, exceeds 6.5%. Eq. 1 is used, with RL derived from all potential adjustments for source type, source characteristics, time of day and if the impacted community is a quiet rural area. Additional rationales for using a change of 6.5% in the %HA_n have been reviewed by the authors elsewhere [2]. Essentially, in the scenario where a project-related increase in noise level increases baseline levels from a normally acceptable urban living environment (i.e. 60 Ldn) to a normally unacceptable urban living environment (i.e. 65 Ldn) the corresponding calculated increase in %HA_n is nearly 6.5%. It is then assumed that the same change in %HA_n can be used to define the change from normally acceptable to normally unacceptable living environments for a broader range of initial noise environments.

The non-linear nature of the dose-response relationship for %HA_n between 43 Ldn and 77 Ldn, makes the threshold for the increase in sound levels to achieve a severe noise impact (i.e. a 6.5% increase in %HA_n) smaller as the baseline sound levels increase.

3. USING A CHANGE IN %HA_n TO EVALUATE THE POTENTIAL HEALTH EFFECTS OF WIND TURBINE NOISE

Keith et al [3] described how Health Canada's proposed sound level criterion is based, in part, on the project-related changes in %HA_n, using Eq.1, where RL for wind turbines is taken to be Ldn, the same as for other industrial noise sources and road traffic noise. ISO 1996-1:2003 recommends RL = Ldn for these latter two sources. These changes are evaluated in terms of changes in $\%HA_n$ from anthropogenic sources without the wind turbine(s), to the noise environment with the wind turbines, as per Eq. 1. For quiet rural areas, the Ldn is adjusted by +10 dB in Eq.1.

Examples of quiet rural areas include, but are not limited to, those with a dwelling density of less than 8 dwellings per square kilometre with day and night background levels less than 45 dBA and 35 dBA, respectively. In these areas, Health Canada has proposed that mitigation be considered, if, at a height of 1.5 m, at the most exposed façade of a noise sensitive receptor, the predicted sound level produced by wind turbine operations exceeds 45 dBA.

This proposal represents a cautious one because the predicted noise levels are meant to be evaluated at the wind speed that produces the *highest* wind turbine sound power level, while background noise is evaluated in calm winds. This accounts for sheltering by obstructions as well as pronounced wind shear effects that have been observed under stable atmospheric conditions. Based on the assumption that the turbines operate continuously at approximately their maximum sound power output, the same criterion value is applied to day and night time Leq values.

Table 1 shows how, using Eq. 1, the proposed criterion level compares to a level that guards against an increase in the $\%HA_n$ of more than 6.5%.

Table 1. The change in $\%HA_n$ in quiet rural areas with operational wind turbine noise levels set at 43.5dBA (rounded to 45 dBA in text).

	Lday	Lnight	Ldn	Adj.	%HA
Baseline	45.0	35.0	45.0	--	1.1
Operation	43.5	43.5	49.9	+10	7.6

Operational levels of 45 dBA in quiet rural areas should adequately protect against low frequency noise impacts from wind turbines. ANSI specifies that in the 63 Hz octave band, moderately noticeable vibrations are associated with a sound level of 70 dBZ, or 44 dBA (reviewed in [3]). Also, a 45 dBA Leq for constant noise is below all specified World Health Organization guideline levels for effects from sleep disturbance, speech disturbance, moderate annoyance, or hearing impairment [6].

It could be argued that it would be more appropriate to use published dose response relationships for wind turbine noise [7] [8], rather than Eq.1, as described above. However, the first published curve [7] was relatively small in its scope. Also, the second publication [8] showed a dose response relationship only for percentage annoyed, not $\%HA_n$. As discussed in Keith et al [3], the data for wind turbines is not convincingly different from Eq.1 with a +10 dB quiet rural area adjustment for Ldn.

4. COMPARISON TO EXISTING PROVINCIAL GUIDELINES, POLICIES & LEGISLATION

Ontario, Alberta and British Columbia (BC) are the only Canadian provinces with guidance specific to wind turbines. For Ontario, in quiet areas, for wind speeds below 6 m/s, the noise limit is 40 dBA, or the minimum hourly background, and at 11 m/s the noise limit rises to 53 dBA. In a quiet rural area, application of Alberta's Energy Utilities Board Directive 038 would yield a criterion with a night time Leq of 40 dBA for wind speeds between 6-9 m/sec, the only speeds for which the Directive prescribes predictions. The province of BC adopted a "*Land use operational policy, wind power projects on Crown land*", which specifies the maximum sound levels from wind turbines is 40 dBA determined at constant wind speeds (e.g. 8-10 m/sec) (see refs in [3]). Therefore, under most situations, the Alberta and BC limits are both more restrictive than the current proposal. Only under conditions where wind speeds are high enough to increase sound levels by 5 dBA would this generalization change.

REFERENCES

1. Canadian Environmental Assessment Act. 1992 Amended December, 2003 http://www.ceaa.gc.ca/012/002/CEAA-Overview_e.pdf (accessed 07/09/2008).
2. Michaud DS, Bly, S, and Keith, SE (2008) Using a change in percent highly annoyed with noise as a potential health effect measure for projects under the Canadian Environmental Assessment Act. *Canadian Acoustics*, 36: 13-28.
3. Keith SE, Michaud, DS, and Bly, SHP (2008) A proposal for evaluating the potential health effects of wind turbine noise for projects under the Canadian Environmental Assessment Act. *Journal of Low Frequency Noise, Vibration & Active Control*, (submitted).
4. ISO (2003) Acoustics-Description, measurement and assessment of environmental noise-Part 1: Basic quantities and assessment procedures. ISO 1996-1:2003(E).
5. Hanson C et al (2006) Transit Noise and Vibration Impact Assessment. FTA-VA-90-1003-06 May: 1-274.
6. World Health Organization (1999) Guidelines for Community Noise. WHO, Geneva.
7. Pedersen E and Waye, KP (2004) Perception and annoyance due to wind turbine noise--a dose-response relationship. *J Acoust. Soc Am*, 116: 3460-3470.
8. Pedersen E and Persson, WK (2007) Wind turbine noise, annoyance and self-reported health and wellbeing in different living environments. *Occup. Environ. Med.*, 64: 480-486.

Recent Developments in Environmental Assessment Methods for Wind Turbine Noise

Brian Howe and Nick McCabe
HGC Engineering, 2000 Argentia Rd., Plaza 1, Suite 203, Mississauga, Ontario, Canada L5N 1P7

1. BACKGROUND

The mechanism of noise generation by large wind turbine generators (WTGs), and the sound level impact can vary dramatically over time at a given receptor location, and to a degree greater than many other types of industrial noise. It is widely known that WTGs tend to emit greater sound power as wind speed increases, but it is perhaps less predictable how other environmental factors influence the radiated sound power for a given reference-height wind speed and the propagation of sound to a receptor.

To ensure acceptable noise impacts at residential neighbors, it is appropriate for regulators to require that the acoustic impact of proposed WTGs be assessed using standardized methods and against standardized criteria. Yet, because of the wide range in actual impact at a receptor from one observation period to another, selecting a standardized condition covering all possible situations is difficult.

In 2007, HGC Engineering prepared for the Canadian Wind Energy Association (CanWEA) a document outlining best practices for the developers of wind farm projects, and summarizing the assessment methodologies in the different jurisdictions in Canada [1]. The various guides in use at that time all provided general limits for sound levels, but less guidance on the prediction methods, leading assessors to rely on general purpose methods, and to develop their own assumptions. For these reasons, the acoustic assessment of wind farms in many jurisdictions tended to have a somewhat arbitrary nature; different assessors would produce somewhat different results.

To deal with this issue, the Ontario Ministry of Environment (MOE) published a guideline document which required the use of ISO 9613 for calculations related to the propagation of sound, and IEC 61400-11 to establish sound power. However, there remained variability between assumptions used in the analysis by different assessors, and therefore variability in the results and recommended setback distances from residential receptors. Accordingly, the MOE published a new draft document [2] which makes changes aimed at reducing the variability amongst assessors by specifying certain assumptions to be used in the analysis and mandates the consideration of wind profile effects.

Correctly applying the MOE procedures will generate for each reference height wind speed a predicted sound level at any given receptor. The recent changes will reduce the variability of results amongst assessors, but the question remains as to how well the predicted results will mirror the real-world sound level impact. A discussion of the actual extent of this latter variability is important, as there is a perception amongst some in the public that a sound level measurement made under any arbitrary atmospheric condition should reflect the prediction.

2. FACTORS CONSIDERED

Some factors governing the observed sound levels affect the sound power emitted by a turbine, some affect the propagation from a WTG to a receptor, and some do both. This paper focuses on factors which change with time; factors like distance and site topography also influence sound propagation but are not discussed here.

2.1 Air absorption

The effect of variations in air temperature and humidity are considered by noise propagation models. Using ISO 9613 it is generally possible to see a change of 1 to 2 dB over typical source to receiver distances by modifying air properties. The draft MOE document mandates that air absorption be considered at 10°C and 70% relative humidity. The practical result of this change is unlikely to reduce the variability in predicted results by more than 1 dB.

In practice, changes in air density or humidity with elevation may also cause some other interesting effects. Fog has been subjectively identified as a factor increasing the apparent sound by some residents near wind farms. In addition to the influence on propagation, low-lying fog layers may result in high wind shear coefficients effectively resulting in little or no background sound concurrent with high sound power emissions from the WTGs.

2.2 Ground absorption

The propagation of sound varies with ground type, or, for a given receptor, with seasonal variations in the ground condition (snow covered ground, hard ice, grasses, etc). In the case of WTGs, where the sound source is very high, it is less clear what effect seasonal variation may have. In typical modeling methods implementing ISO 9613, the difference between fully absorptive ground and fully reflective ground is generally about 3 dB for second-storey receptors, although somewhat larger at lower elevations.

The draft MOE interpretation document suggests that a global value of G=0.7 be used. The practical difference in modeling between this value and fully absorptive ground appears to be about 1 dB. An alternate method described by the MOE makes use of specific G values for the source, middle and receiver regions.

2.3 Wind profile

Wind profile relates to the variation in wind speed with height above grade. The term "wind shear" is used to describe the same thing, generally assuming a logarithmic profile. The wind shear exponent quantifies the wind shear.

IEC 61400-11 deals with wind profile through consideration of "roughness length". It attempts to normalize actual wind profiles encountered during the measurements by defining a fixed reference "roughness length", and transforming measured sound levels to theoretical sound levels under the standard condition. A standard roughness length of 0.05 is equivalent to a wind shear exponent of about 0.16.

From projects assessed by HGC Engineering, wind shear exponents approaching 0.5 are common under nighttime conditions. To put this in terms of wind speeds, a 10 m wind speed of 7 m/s with a wind shear coefficient of 0.16 results in an 80 m wind speed of 9.8 m/s, whereas the same 10 m wind speed with a coefficient of 0.5 results in an 80 m wind speed of 19.5 m/s.

As required by IEC 61400-11, most manufacturers list the sound power output of their turbines as a range, correlated with 10 m wind speeds under the reference wind profile condition. In practice, then, where the wind shear exponent might vary from 0.05 to 0.45 through a given day, the sound power output from the turbine might vary over the entire range (which could be 5 to 10 dBA), even while the speed at the reference height remains constant.

The MOE document mandates that wind profiles be considered by adjusting the manufacturer's stated sound emissions in consideration of the site specific wind profile. While this is clearly important, the document does not provide assistance as to how the site specific wind profile is to be developed. Such practical considerations as to how many site-specific wind measurements are to be made, at what elevations, for what duration, and – perhaps most importantly – what value amongst the tremendous range that will be calculated over the monitoring period should be selected as the governing wind profile, are not discussed.

2.4 Wind direction

For many wind farms, typical setback distances are currently in the range of 400 to 500 metres for the closest residences. This is larger than in many industrial noise impact situations making the influence of wind direction much more significant. This is particularly true in the case where many WTGs are located on only one side of a residence. At the distances in question, it is not unreasonable to expect that there would be times when the WTGs can barely be heard, and others when they are the dominant source of sound. Wind turbines are also directional in their acoustic radiation; as changes in wind direction change the orientation of a WTG with respect to a receptor, further changes in observed sound can occur.

The MOE document does not address wind direction, and assessors will continue to assume the "moderate downwind condition" or long term average described by ISO 9613. This minimizes variation between assessors, but is a major factor in the variability of actual measured sound levels and their deviation from predictions.

2.5 Summary

Of the four factors discussed above, the most important in governing the variability in actual measured sound levels are considered to be wind profile and wind direction. The MOE interpretation document now requires the consideration of wind profile, buts lacks specific details as to how this is to be done. Wind profile will continue to be an important factor leading to variability of predictions amongst assessors in Ontario. Ground and air absorption will now be handled in the same way by all assessors, but these factors are considered less important in practice.

It is useful to recognize that ISO 9613 does not purport to predict sound pressure levels under all conditions, and limits the applicability of the stated accuracy of +/- 3 dB to relatively low source heights and modest distances, both of which tend to be exceeded in the case of WTGs.

3. ACOUSTIC AUDIT RESULTS

To demonstrate the typical variability of sound levels over time, Figure 1 show the range of average (L_{EQ}) sound pressure levels measured by HGC Engineering at wind farms - one over 7 days in Nova Scotia, and one over 9 days in Ontario. In both instances the measurement locations are about 350 metres from the nearest WTG. In the chart, data for intervals during which there was negligible electrical power produced by the WTGs have been removed, and the data has been plotted against the wind speed.

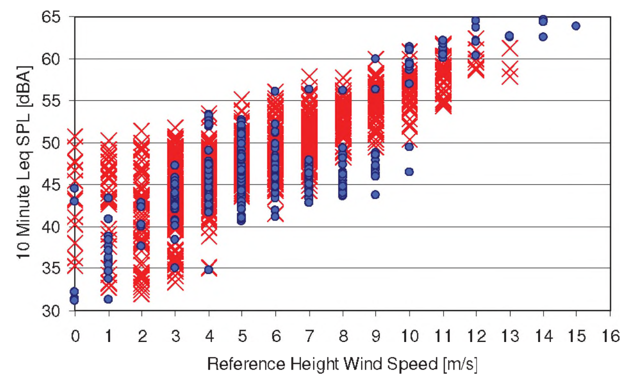


Fig. 1. Sound levels versus reference height wind speeds measured at wind farms in Nova Scotia (blue circles) and Ontario (red crosses).

The above results show a wide range of sound levels for each reference wind speed. The variability is typically +/- 5 dB or more, and is greater at low reference-height wind speeds where the influence of wind profile is greatest.

4. CONCLUSION

Assessing the environmental noise impact from WTGs is necessary to ensure compatibility with nearby residential properties. Standardized methods for doing so have been evolving and the MOE for example, recently revised their guidelines to minimize the variability of results between assessors by prescribing certain assumptions. The importance of wind profile has been acknowledged in the guideline, but this alone is unlikely to reduce variability between assessors until a standard definition of worst-case site specific wind shear has been agreed upon.

Acoustic audits show temporal variability of sound levels of at least +/- 5 dB for a given reference height wind speed. This degree of variability exceeds any likely differences expected between assessors, and highlights that actual sound levels will potentially exceed those predicted by ISO 9613 at times. Thus, it is important to realize that while standardized assessment methods provide a useful and consistent basis for assessing WTGs, they do not necessarily reflect the range of actual sound levels expected at receptors.

REFERENCES

[1] HGC Engineering, (2007). Wind turbines and sound: Review and best practice guidelines. Available at: http://www.canwea.ca/municipalities/municipalities_bestpractices_e.php

[2] Ontario Ministry of Environment, (2008). Noise guidelines for wind turbines. Interpretation for applying MOE NPC publications to wind power generation facilities. Draft May 2008.

RECENT STUDIES OF INFRASOUND FROM INDUSTRIAL SOURCES

William J. Gastmeier, MASc, PEng and Brian Howe, MEng, MBA, PEng
HGC Engineering, 2000 Argentia Rd., Plaza 1, Suite 203, Mississauga, Ontario, Canada L5N 1P7

1. BACKGROUND

Infrasound from industrial equipment and its potential effects on residential neighbours has recently been the subject of discussion in the media, with infrasound from wind turbines being a particularly contentious issue on occasion. In 2007, HGC Engineering investigated infrasound at several wind power projects and prepared a summary document for the Canadian Wind Energy Association (CanWEA) [1] indicating that infrasound was not an issue at modern wind turbine installations. Following on from that study, HGC Engineering has investigated three additional situations involving the perception of infrasound by individuals. These involved the measurement of infrasound sources using refined measurement methods, determining the physical effects of infrasound at neighbouring receptors and the manner in which infrasound was perceived by the neighbours. The sources included wind turbines, a large reciprocating engine used for power generation and an industrial sieve used to sift material. The results are summarized in this paper.

2. STUDIES

Infrasound is defined as "a wave phenomenon of the same physical nature as sound but with frequencies below the range of human hearing". Generally, hearing via the auditory nervous system is considered to occur at frequencies above 20 Hz. However, although infrasound may not be "heard" based on the normal meaning of the word, under certain circumstances it can be perceived by humans. There is some degree of auditory perception below frequencies of 20 Hz and there are non-auditory mechanisms such as the vestibular balance system and the resonant excitation of body cavities by which humans can sense infrasound. As determined in these studies, humans can also perceive the effects of infrasonic excitation on structures, such as perceptible vibration in walls and windows and secondary effects such as the rattling of lightweight building components.

2.1 Wind Turbine Generators

Measuring infrasonic sound levels produced by wind turbine generators is difficult in that it is hard to separate the effect of the turbines from infrasound naturally occurring due to local wind and distant atmospheric effects. Earlier attempts by HGC Engineering measured infrasound outside at various locations near and far from wind turbines and concluded that there was no material or consistent increases noted near the wind turbines.

Additional outdoor measurements have now been conducted using refined methods, incorporating a 1" free field microphone and a large multi-layered windscreen. Infrasound measurements were also conducted inside a residence near a wind power project to eliminate the direct contribution of infrasound due to wind at the microphone, and investigate the influence in occupied spaces.

Several sets of measurements were sequentially completed with and without the wind turbine generators operating. The residence is located about 325 m from the closest of several 1.8 MW wind turbines. During the measurements the wind was blowing towards the residence from a concentration of wind turbines at a speed of approximately 5 m/s.

Figure 1 presents the two minute average sound level in the dining room of the residence with and without the operation of the wind farm. The levels of infrasound are compared to perception thresholds suggested by Watanabe and Møller [2], which is shown as a hatched line.

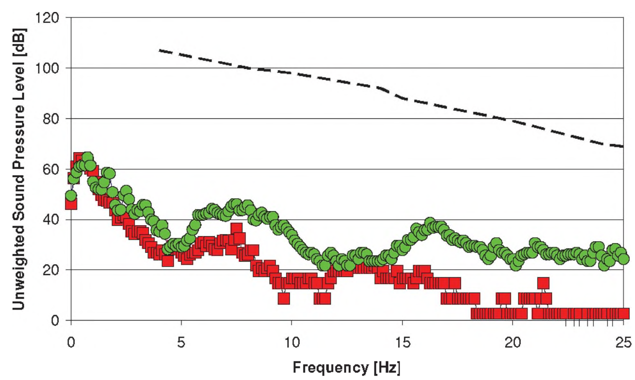


Fig. 1. Infrasonic sound levels measured in a residence, with (●) and without (■) wind turbine generators in operation.

The measurements indicate that the operation of the wind farm increases the sound levels by approximately 10 dB at selected frequencies. Regardless, if the wind turbine generators were operating, the sound levels are at least 30 dB less than the perception threshold suggested by Watanabe and Møller.

Vibration measurements were conducted on the foundation of the residence and on the foundation of several of the closest turbines to confirm that ground borne vibration was not a factor. These measurements demonstrated that in the both instances, the vibration levels are well below the perception limits discussed in ISO 2631/2 "Evaluation of human exposure to whole-body vibration Part 2".

The overall conclusion is that there is no evidence to suggest that infrasound should be a source of complaints by the occupants.

2.2 Low Speed Diesel Engine

HGC Engineering had an opportunity to investigate infrasound produced by two 30 MW low speed diesel engines used to provide electrical power for a Caribbean island. The 9 cylinder engines have an operating speed of 107 rpm. They were found to cause off-site excitation at a frequency of 10.7 Hz due to a 6th order (6 times the fundamental operation speed) "X" mode vibration of the engines. Essentially, the engines have a resonance whereby the top of the engine twists back-and-forth. The large surfaces of the engine efficiently radiate pressure pulsations which pass out through the building envelope with virtually no attenuation.

The infrasound was creating complaints at a neighbouring office building and hotel. The principal complains related to obvious rattling and visual motion of ceiling tiles and various lightweight fixtures. There were no complaints about physical perception per se.

To address the vibration, it would have been prohibitively expensive for the manufacturer to modify the engine to avoid the 6th order excitation. Rather, tuned mass dampers (TMDs) were originally installed at the ends of the engine block to reduce the vibration. Unfortunately, the tuning of the original TMDs was found to be unreliable; the spring elements failed due to fatigue, shifting the resonant frequency dramatically. The situation was finally solved by installing powered vibration compensators at the top of each end of the engine. These compensators are rotating eccentric masses which are adjusted to provide a force out of phase and equal to the inertia force of the vibrating engine.

Figure 2 presents the frequency spectra of the sound pressure level measured in the conference room of the office building located approximately 80 m from the building that houses the two engines. The levels were measured while both engines were running at full load, with and without the powered vibration compensators operating.

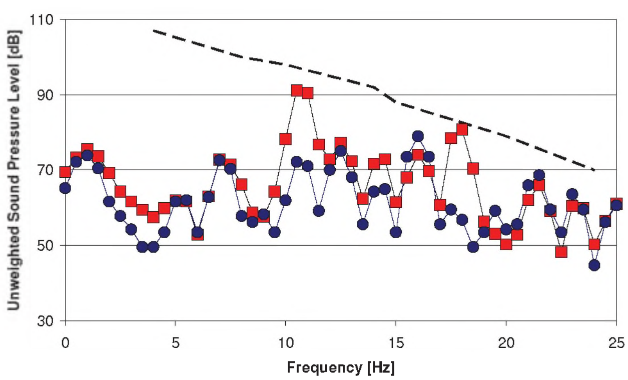


Fig. 2. Infrasonic sound levels measured indoors at 80 m from a 30 MW low speed diesel engine, with a powered vibration compensator (■) and without (●).

The compensators are extremely efficient at controlling the infrasound, reducing the sound pressure levels in the office space by 19 dB at a frequency of 10.5 Hz. Subjectively, the rattling of the ceiling and various fixtures within the offices was essentially eliminated as a result of the compensators, although there remains some residual low frequency air-borne excitation created by the firing of the engines at a frequency of 16 Hz.

It is interesting to note that the infrasound levels due to the engines were initially almost identical to the perception limits defined by Watanabe and Møller, and yet the complainants did not express concerns related to perception.

2.3 Vibrating Screens and Sieves

Infrasound proved to be a potential issue for a manufacturing plant that produces a foamed glass bead insulation product. The manufacturing process utilizes a number of large vibrating screens and sieves to classify the product by size. HGC Engineering completed extensive measurements at a plant in Germany, in order to aid in the design of a similar plant in Canada.

The screens operated at a frequency of 16 Hz. Vibration transmitted via the structure to the building shell was the principal source of infrasound noted off the property. Occupants of an office located approximately 75 m from the main processing building experienced visual vibration of its windows and light building components. One-third octave sound levels outside the windows were 87 dB at 16 Hz. Physically, the infrasound was not otherwise noticeable by the occupants in the building.

There were residential uses at approximately 150 m from the process. At this distance, the 16 Hz sound levels were reduced to 77 dBA. With these levels measured in one-third octave bands, it may not be precise to compare them to the aforementioned studies, but nonetheless: no complaints from the residents were noted even though the levels are approaching the perception limits as defined by Watanabe and Møller.

In designing the new plant in Canada, the frequency of the vibrating sieves was increased to 20 Hz, not to simply move the frequency of excitation above the infrasonic region, but to increase the efficiency of the vibration isolation systems supporting the sieves. In addition, the sieve building was designed to provide additional noise reduction at 20 Hz. This reduced the 20 Hz sound pressure level to 70 dB at the equivalent distance of 75 m.

3. CONCLUSION

HGC Engineering has recently conducted several infrasound studies using refined measurement methods to isolate the infrasound energy produced by industrial sources from naturally occurring infrasound in the environment. The results confirm a previous study prepared for the Canadian Wind Energy Association, concluding that infrasound from wind turbine generators is well below any realistic human perception limits.

The results also show that the effects of infrasound generally manifest themselves as the visual movement of lightweight building components and fixtures before human perception becomes an issue. This is particularly evident in instances when the infrasound is produced in a narrow frequency range where the probability of matching the resonant frequency of a component increases.

REFERENCES

[1] HGC Engineering, Canadian Wind Energy Association, (2007). Wind Turbines and Infrasound. Available at: http://www.canwea.ca/images/uploads/File/CanWEA_Infrasound_Study_Final.pdf

[2] Watanabe and Møller, "Low frequency hearing thresholds in pressure field and free field", Journal of Low Frequency Noise and Vibration, 1990b

TYPICAL HOURLY TRAFFIC DISTRIBUTION FOR NOISE MODELLING

Peter VanDelden[1], Scott Penton[2], and Aaron Haniff[3]
RWDI AIR Inc., 650 Woodlawn Rd West, Guelph, Ontario, Canada N1K 1B8
[1]peter.vandelden@rwdi.com [2]scott.penton@rwdi.com [3]aaron.haniff@rwdi.com

1. INTRODUCTION

A prominent feature of the environmental noise in our western world is the sound of vehicles travelling on our roads. Noise from vehicles is estimated using models that incorporate many factors, including traffic volume. However, the ability to characterize the variation in noise level is limited by the temporal resolution of the traffic volume that is available. This study provides input to evaluating the variation of noise level over the course of a day by examining variation in the number of vehicles.

Noise levels are usually represented on the basis of hourly, day/night, or 24-hour time periods. 24-hour time periods usually require no more than the Average Annual Daily Traffic (AADT) volume for a road segment. When breaking the evaluation period into daytime hours and night-time hours or further resolving it to a one hour time period, it becomes necessary to have information about how the traffic is distributed.

Traffic volumes distributed over one hour time periods are not always available for the specific road segments that are being modelled. Since traffic data is used primarily for transportation planning, the traffic volumes are frequently counted only during the morning peak and afternoon peak hours, or during an eight hour daytime period. Data for the evening and night-time periods is therefore missing.

The problem of determining how low the traffic volumes and sound levels fall during evening and night-time periods is particularly real in jurisdictions such as Ontario, where the lowest sound levels are used in determining the applicable limits for permitting of industrial noise. When the data is not available on an hourly basis for a desired road, a comparable road in the vicinity is often used as a proxy. This study addresses the situation where a suitable proxy is not available, by developing a typical hourly distribution.

2. DATA SET

2.1 Data Sources

The typical hourly traffic distribution presented in this study is empirically derived. Data was requested from the provincial and local levels of government across Ontario. Care was taken to represent the widest possible variety of roads by soliciting data from cities, towns, and municipalities in all regions of the province. Each was requested to send data sets reflecting the diversity of road types within the jurisdiction. Rural, collector, arterial, and controlled access roads are all represented.

Each of the data sets was accumulated using automated counters. A large number of the data sets included multiple days of counting. A few of the counts included complete breakdowns by vehicle type and size. However, the amount of this type of data was not sufficient to warrant further examination.

Traffic counting for the data sets obtained was conducted from March through November, with the majority occurring between April and October. One of the counts included hourly averages for each month of the year.

In general, the traffic distributions follow a common pattern. Between 6 am and 8 am there is a rapid rise in volume. The volume is drops somewhat, but rises through the late afternoon. After peaking at the end of the afternoon, it drops off into the evening and night-time hours. The lowest volume occurs between the hours of 2 am and 4 am.

2.2 Data Selection

Most of the data sets that were received were suitable for this analysis. Data sets that were significantly influenced by local features, had low traffic volumes, or were only part of a road segment are excluded in the analysis. Local features such as a dominant industry with shift changes or a nearby school, showed the ability to bias the hourly distribution on a road segment. While this is normal for certain road segments, this is clearly not universal. Data sets with low total volumes were also poorly behaved. They included hours where no vehicles were recorded, and showed comparatively large impacts for small changes in traffic volumes. Upon reviewing the distribution of traffic in the uni-directional data sets, it was apparent that they could not be considered together with the bi-directional data sets. These data sets were typically missing the rapid morning rise in traffic, or the large volumes of the late afternoon and early evening traffic. These data sets would skew the results away from normal conditions, and were therefore not included.

After removing the unsuitable data sets, just over 100 data sets remained for analysis. The distribution of AADT's is shown in Figure 1.

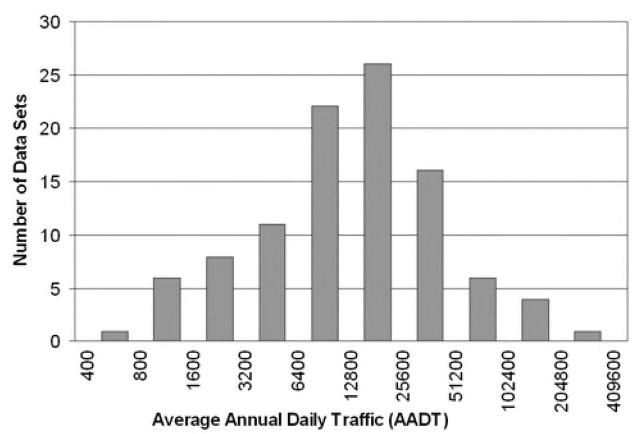


Figure 1. Distribution of AADTs in Data Set

3. ANALYSIS AND RESULTS

3.1 Hourly Traffic Volume

The accumulated suitable data sets were analyzed using the percentage of the AADT occurring in each hour of the day. A single data set was used for each road segment.

Where a number of days were counted for a segment, an average was calculated over the measurement period. The typical distribution was calculated as the mean of each hour over the data set, as shown in Figure 2. The range of values and standard deviation are shown in Table 1.

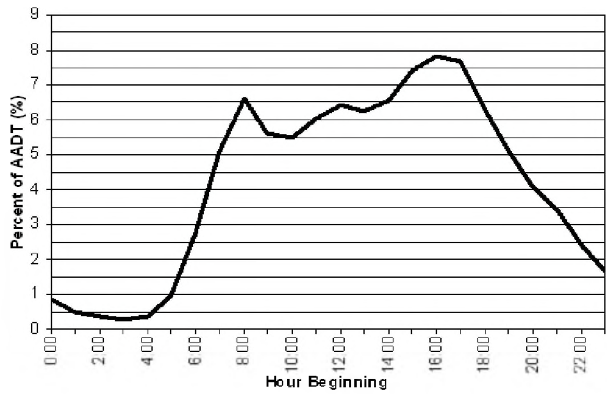


Figure 2: Typical Hourly Traffic Distribution in Percent AADT

Table 1. Range of Values and Standard Deviation for the Supporting Data Set (in Percent)

Hour Beginning	Typical	Maximum	Minimum	Standard Deviation
0:00	**0.87**	1.88	0.18	0.44
1:00	**0.49**	1.21	0.09	0.27
2:00	**0.36**	0.86	0.07	0.21
3:00	**0.30**	0.76	0.05	0.17
4:00	**0.36**	0.87	0.07	0.21
5:00	**0.95**	2.68	0.37	0.54
6:00	**2.75**	5.18	1.19	1.43
7:00	**5.05**	8.59	2.13	2.30
8:00	**6.55**	11.08	3.30	2.81
9:00	**5.62**	7.70	3.96	2.24
10:00	**5.50**	7.73	3.81	2.21
11:00	**6.04**	9.76	4.19	2.48
12:00	**6.48**	9.78	4.45	2.65
13:00	**6.26**	9.75	4.24	2.56
14:00	**6.60**	9.62	4.44	2.63
15:00	**7.41**	10.40	5.51	2.91
16:00	**7.82**	10.34	5.83	3.06
17:00	**7.65**	9.30	5.58	3.01
18:00	**6.27**	8.72	4.42	2.50
19:00	**5.12**	7.44	3.52	2.06
20:00	**4.09**	6.30	2.18	1.69
21:00	**3.41**	5.21	1.30	1.44
22:00	**2.41**	4.09	0.78	1.08
23:00	**1.67**	3.79	0.46	0.86

The significance of the statistical data becomes apparent when the hourly distribution is translated into sound levels. For illustration purposes an L_{eq}(24 hour) of 60 dBA was used for comparison with the L_{eq}(1 hour) sound levels as shown in Table 2. The respective increase or decrease due one standard deviation increase or decrease in traffic volume is also presented.

Table 2. Hourly Sound Levels and Variation Based on an L_{eq}(24hr) of 60 dBA.

Hour Beginning	Sound Level (dBA)	+1 std. dev.	-1 std. dev.
0:00	53	+2	-3
1:00	51	+2	-3
2:00	49	+2	-4
3:00	49	+2	-4
4:00	49	+2	-4
5:00	54	+2	-4
6:00	58	+2	-3
7:00	61	+2	-3
8:00	62	+2	-2
9:00	61	+1	-2
10:00	61	+1	-2
11:00	62	+1	-2
12:00	62	+1	-2
13:00	62	+1	-2
14:00	62	+1	-2
15:00	62	+1	-2
16:00	63	+1	-2
17:00	63	+1	-2
18:00	62	+1	-2
19:00	61	+1	-2
20:00	60	+1	-2
21:00	59	+2	-2
22:00	58	+2	-3
23:00	56	+2	-3

3.2 Day/Night Split

The amount of traffic occurring in each part of the 16/8 and 15/9 hour splits was calculated for each distribution. Table 3 presents the average and range of these values.

Table 3. Day/Night Split Ratios

	16:8 hr	15:9 hr
Mean	92:8	90:10
Maximum Range	97:3	96:4
Minimum Range	87:13	83:17

4. DISCUSSION

To understand the value of this typical traffic distribution, it is necessary to know its limitations. The first indication of limitations is derived from the input data. Low volume roads, unidirectional roads, or those with significant local influences are not covered in this typical hourly traffic distribution.

Regulatory parameters put a practical lower limit on the AADT values where this distribution can be applied. Conditions, including AADT, that are sufficient to generate L_{eq}(24 hour) values of 57 dBA and 52 dBA would be needed to raise the lowest hourly sound levels above the minimum limit values of 45 dBA and 40 dBA respectively. More significantly, the models that use this data also provide limitations. For example, the algorithms currently used in Ontario set a minimum limit of 40 vehicles per hour. An AADT of over 13,000 would be necessary to satisfy this criterion during the early morning hours.

This study proposes a typical hourly traffic distribution together with some indication of its range of variability. It provides a method of estimating minimum hourly sound levels when actual data is missing or incomplete and good proxy data is not available.

THE ACOUSTIC CORRELATES OF THE UNPARSED : WHY WE NEED MORE THAN A STRONG-WEAK DISTINCTION

Marion Caldecott[1]
[1]Dept. of Linguistics, Totem Field Studios, University of British Columbia, Vancouver, BC, Canada V6T 1Z4

1. INTRODUCTION

This paper reports results from 2 experiments on St'át'imcets (Lillooet Salish) that test the prediction that phonologically distinct domains in the Prosodic Hierarchy are also acoustically distinct. In particular, that when a phonological distinction between non-prominent syllables exists, an acoustic distinction should also produced by speakers, contra the traditionally assumed strong-weak dichotomy. These acoustic differences should be reflected in traditional prominence cues as well as boundary effects.

The current model of the Prosodic Hierarchy permits a 3-way syllable distinction at Pword level [1]: 1) stressed head of foot, 2) unstressed non-head of foot; and 3) unstressed, unparsed at foot level, or 'extrapod'. Stressed syllables are accepted to be more prominent or stronger (have higher F0, duration and intensity) than unstressed [2]. However, the acoustic characteristics of this third type of syllable have been ignored. It has generally been assumed that extrapods and unstressed syllables are indistinguishably non-prominent or 'weak' [3][4].

Another underlying principle of the current model is that the prosodic constituents which are the domains of phonological processes are acoustically distinct [5]. These distinctions are generally discussed in terms of prominence, as above, or Prosodic Strengthening [6].

If the model permits a ternary syllable distinction at the Pword level, then given a systematic mapping, there should also be a ternary acoustic distinction. This predicts that if footed, unstressed syllables and extrapods are shown to be phonologically distinct domains, they should also be acoustically distinct. Testing this prediction requires a language that makes such a phonological distinction.

St'át'imcets presents a case in which some word-final suffix vowels are subject to phonological reduction when they are parsed as footed, unstressed syllables but not extrapods [7]. If the hypothesis above is correct, this predicts St'át'imcets word-final unstressed syllables and extrapods should show distinct acoustic characteristics.

Experiment 1 compares the prominence cues of Pword-final unstressed syllables and extrapods. The prediction tested is that extrapods will be more prominent in terms of F0, duration and intensity than unstressed syllables.

Experiment 2 focuses on boundary effects. It has been shown that a vowel's position on the F1/F2 plane is more peripheral at higher boundaries in the Prosodic Hierarchy [6]. The prediction tested here is that, given that Pword-final extrapods lack a foot boundary, they will be less peripheral on the F1/F2 plane than Pword-final footed, unstressed syllables.

2. METHOD

2.1 Participants

Participants in both experiments were 4 fluent St'át'imcets speakers. AP: female, 64 yrs, Northern dialect and LT: female, 77 yrs, Southern dialect, took part in both experiments. RW: female, 77 yrs, Northern dialect and CS: 85 yrs, Northern dialect participated in Experiment 1. CA: male, 67 yrs, Northern dialect; and HD: male, 70yrs, Southern dialect, participated in experiment 2.

2.2 Stimuli

Stimuli in both experiments were word-final suffixes containing /a,i,u/. Tokens in which the suffix was parsed as footed, unstressed syllables or extrapods were created and checked with speakers. In experiment 1, 1sg.subj. /-ɬkan/, 3pl.subj. /-wit/ and 3pl.obj. /-tumuɬ/ were selected. In experiment 2, possessive suffixes were selected: 2sg /-šu/, 3 sg. /-ša/ and 3pl. /-i/. In both cases, tokens were placed in contextual target sentences that controlled for intonation. Due to dialect differences, speakers in experiment 2 were presented with 2 vowels each. 22 repetitions of each token were recorded. For experiment 1: 3 vowels x 2 syllable types x 22 repetitions = 132 tokens per speakers. For experiment 2: 2 vowels x 2 syllable types x 22 repetitions = 88 tokens per speaker.

2.3 Experiment procedure

In both experiments, speakers were shown MS PPt slides with pictoral scenarios to elicit the target sentence. In experiment 1, speakers answered a St'at'imcets yes/no question based on what they saw on the slide. In experiment 2, the target sentence was an imperative of the type "Give /his/her/their/your object to me/Henry", with a slide showing the object and the recipient. Speakers could choose to be prompted in English. All speakers were recorded using a Marantz PMD660 solid-state recorder and Audio-technica ATM75 head-mounted cardioid condenser microphone in a private home.

2.4 Acoustical analysis procedure

Measurements for both experiments were done in PRAAT [8]. Vowels were segmented and a script was used to extract duration and F0 and intensity at midpoint of the vowel. Within subject comparisons were done using independent sample t-tests with conservative p-factors P=.000 =*** p≤.01=**, p≤.05* [9].

3. RESULTS

All speakers make a prominence distinction between Pword-final unstressed syllables and extrapods. In addition, 3 of 4 speakers show boundary effects as well.

3.1 Prominence Effects

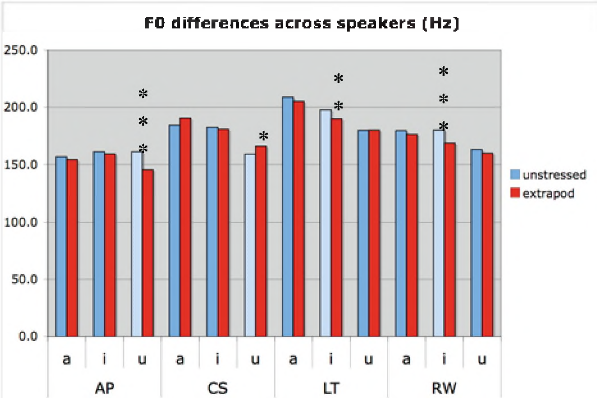


Figure 1 Mean F0 values for vowels across speakers.

Figure 1 shows that all speakers make a distinction in F0 in either /i/ or /u. For 3 of the four speakers, unstressed vowels had higher pitch than extrapods.

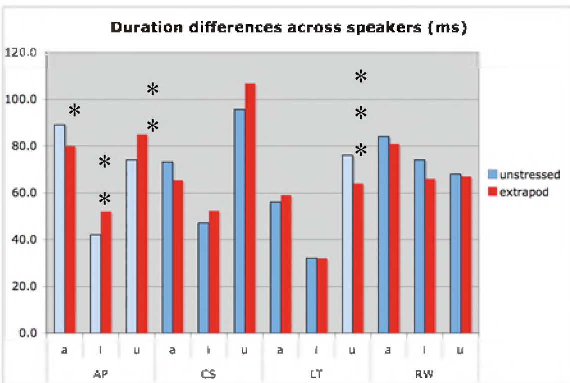


Figure 2 Mean duration values for vowels across speakers

Figure 2 shows only 2 speakers make a duration distinction, in half of which, extrapods were longer than unstressed syllables.

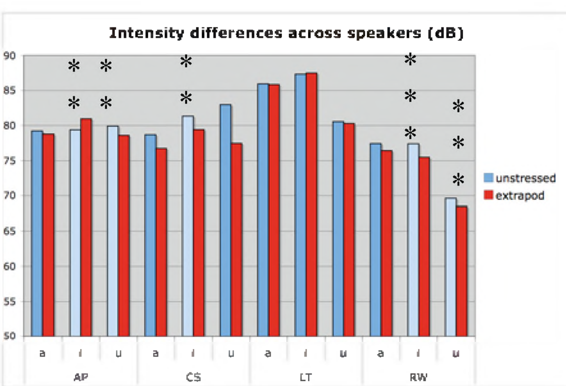


Figure 3 Mean intensity values for vowels across speakers

Figure 3 shows that 3 speakers show significantly greater intensity in unstressed syllables than extrapods.

3.2 Boundary Effects

3 speakers make a distinction at Pword level in F1/F2 values: AP makes no distinction. CA's unstressed /a/ is * higher than extrapod, and unstressed /u/ is **backer. HD's unstressed /i/ is **fronter than extrapod, and LT's unstressed /i/ is **higher and ***fronter than extrapod.

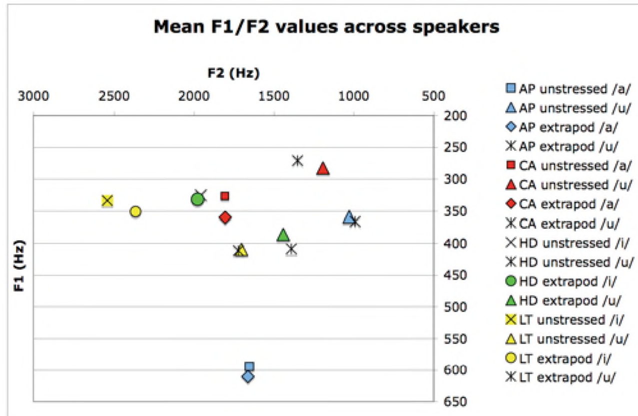


Figure 4 Mean F1/F2 values

4. DISCUSSION

These results support the prediction that a phonological distinction at Pword level is reflected in an acoustic distinction. The direction of the difference was not as predicted for the prominence results—in the majority of cases, unstressed syllables were more prominent than extrapods. The direction of boundary effects were as predicted, with unstressed syllables being more peripheral than extrapods. This supports a version of Prosodic Strengthening that is sensitive to all boundaries, not just the outermost one.

The differences between footed, unstressed vowels and extrapods cannot be accounted for in terms of stress, or positional effects: both are unstressed (weak) and occur word finally. The difference is that one is a constituent of a foot, while the other is not. Parsing, in addition to headedness, or strength, must be taken into account when characterising syllables at the Pword level.

REFERENCES

[1] Selkirk, L. (1995). The prosodic structure of function words. UMass Occasional Papers in Linguistics: 18. 439-469.

[2] Lieberman, P. (1959). Some acoustic correlates of word stress in American English. JASA 32: 451-454.

[3] Halle, M. & J. Vergnaud. (1987). An Essay on Stress. Current Studies in Linguistics. MIT Press.

[4] Nespor, M. & I. Vogel (1986). Prosodic Phonology. Foris: Dordrecht.

[5] Beckman, M. & J. Pierrehumbert. (1986) Intonational structure in Japanese and English. Phonol.Yearbook 3, 255-309.

[6] Cho, T. (2005). Prosodic strengthening and featural enhancement: Evidence from acoustic and articulatory realization of /a,i/ in English. JASA 117 (6): 3867-3878.

[7] Davis, H. (in prep). A Teaching Grammar of St'át'imcets.

[8] P. Boersma and D. Weenik, "Praat: Doing phonetics by computer (version 4.5.08)" [Computer program], 2006 Apr 4 [cited 2007 Jul 22], Available HTTP: http://www.praat.org/

[9] Koch, K. (2008). Intonation and Focus in Nlhe7kmxcin. PhD thesis. UBC.

ACKNOWLEDGMENTS

Kukwstum'ckálap Aggie Patrick, Carl Alexander, Ceda Scotchman, Herman Dan, Laura Thevarge and Rose Whitley. This project was funded by SSHRC grant #752-2003-0330 and the Jacobs Research Fund.

VOWEL QUALITY AND DURATION IN DEG XINAG

Sharon Hargus
Dept. of Linguistics, University of Washington, Box 354340, Seattle, WA USA 98195-4340

1. INTRODUCTION

Deg Xinag, an Athabaskan language spoken in western Alaska, has been described as containing the following vowel inventory: /e a o ə ʊ/ (Krauss 1962, Leer 1979). The vowel "/ʊ/" is restricted in distribution in Deg Xinag, only occurring adjacent to uvulars. Deg Xinag "/ʊ/" is a reflex of Proto-Athabaskan *ʊ, but /ʊ/ in Deg Xinag gives the auditory impression of being a short version of /o/. If true, then Deg Xinag would be typologically unusual in having no high vowel phonemes. This study investigated the following questions: what are the spectral properties of the Deg Xinag vowels? Which vowels are significantly different in duration?

2. METHOD

2.1 Participants

Participants are 8 adult native speakers (3 male, 5 female) of Deg Xinag, between the ages of approximately 68-76 at the time of recording. English is a second language for all speakers, but proficiency varies.

2.2 Recording materials and recording procedure

Two word lists were constructed. In a word list constructed for measuring vowel duration, words were recorded in a sentence context. Vowels were recorded in one of two consonantal contexts: surrounded by coronal stops, or preceded by a uvular stop and followed by a coronal stop. In another word list constructed for measuring vowel quality, words were recorded in isolation. The vowel quality word list recorded the vowels in several consonantal contexts: preceded and followed by coronal stops; preceded by a uvular consonant; followed by a uvular consonant.

On both the vowel quality and vowel duration word lists, words or sentences were recorded in random order (the same random order for each speaker). Four repetitions were elicited. Some speakers voluntarily produced more than four repetitions. Some repetitions were later excluded due to excess background noise or some similar reason. Because no speaker was literate in Deg Xinag, the words were elicited through a combination of translation from English and/or prompting in Deg Xinag.

Recordings were made using a professional CD recorder or compact flash recorder with an AT 4041 microphone attached. Data was recorded at 44,100 Hz and downsampled to 11,025 Hz.

2.3 Acoustic analysis

The first four formants of vowels on the vowel quality word list were measured using PRAAT (version 4.3.27), with Maximum Formant set at 5000 Hz for men and 5500 Hz for women.

Vowel duration was measured using Multi-Speech 2.5. In a first pass through all the data, tags were placed at the onset and offset of F2. In a second pass, duration was measured between tags.

2.4 Statistical analysis

The vowel duration measures were subjected to repeated measures analysis of variance, with the dependent variable being each speaker's mean vowel duration for each vowel category. Vowel was the independent variable. Post hoc analysis was performed with Fisher's PLSD.

3. RESULTS

3.1 Spectral properties of vowels

For vowels in the coronal context, /ə/ has higher F2 than /e/, /o/, impressionalistically [ɪ]. In Figure 1, a typical plot of the contrasting vowels in this context, notice that the F1 of /ə/ is higher F1 than that of either /e/ or /o/.

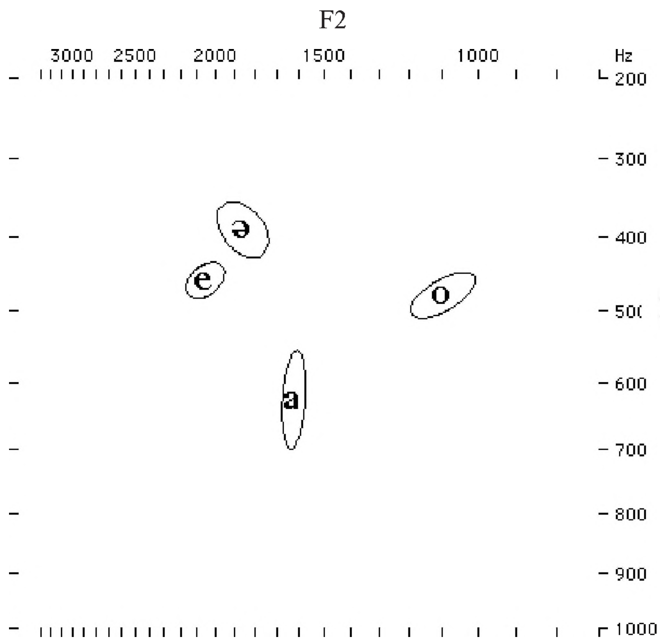


Figure 1. F1 x F2 plot of vowels in coronal context (for HM, a female speaker)

In a uvular context, /ə/ has a lower F1 than /e/ or /o/. The vowel /ʊ/ also occurs in this context. For all speakers, there is overlap in F1 and F2 between /o/ and /ʊ/, with greater degrees of overlap when a uvular follows. Figure 2 shows a typical vowel plot for vowels before a uvular consonant.

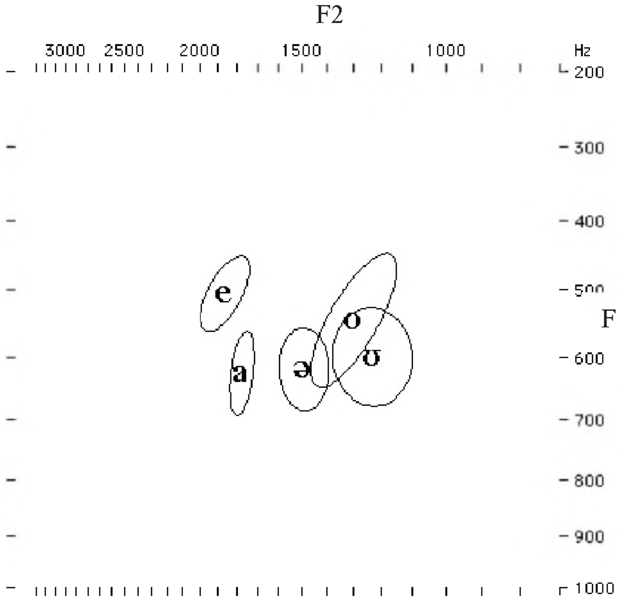


Figure 2. F1 x F2 plot of vowels following a coronal consonant and before a uvular consonant (for HM, a female speaker)

3.2 Vowel duration

Vowel duration results for the four vowels that can occur surrounded by coronal stops are shown graphically in Figure 3.

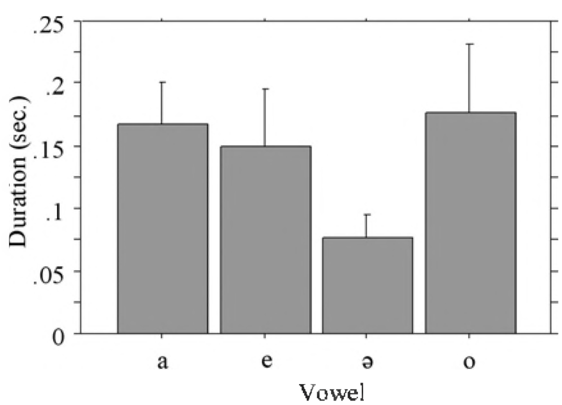


Figure 3. Bar graph showing means for /a e ə o/ preceded and followed by coronal stops. Error bars represent one standard deviation.

The vowels differ significantly in length ($F[7,21] = 21.455$, $p < .0001$). Post hoc analysis indicates that /ə/ is significantly shorter than each of /e o a/, but /e o a/ do not differ in length from each other.

Vowel duration results for the five vowels that can occur after a uvular stop are shown graphically in Figure 4.

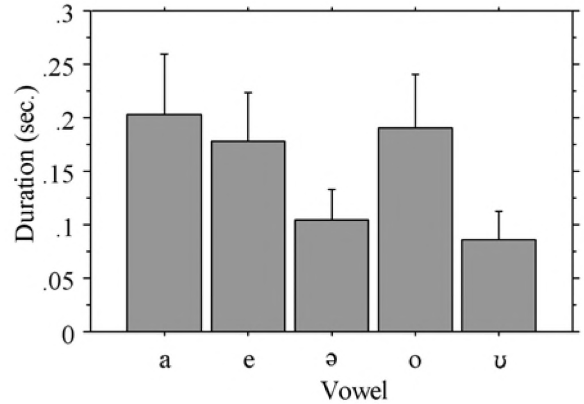


Figure 4. Bar graph showing means for /a e ə o U/ preceded and followed by coronal stops. Error bars represent one standard deviation.

The vowels differ significantly in length ($F[7,28] = 34.633$, $p < .0001$). Post hoc analysis indicates that /ə ʊ/ are significantly shorter than each of /e o a/, but not from each other. /e o a/ do not differ in length from each other.

4. DISCUSSION

Qualitative inspection of the spectral properties of the Deg Xinag vowels indicates two points of interest: "/ʊ/" is similar to /o/, and /ə/ has a high vowel allophone [ɪ] which occurs surrounded by coronal consonants. Thus while Deg Xinag has no high vowel phonemes, there are phonetic high vowels.

Statistical analysis indicates that the vowels of Deg Xinag can be divided into two sets, a short set consisting of /ə ʊ/ and a long set consisting of /e o a/. Durational differences between the longer and shorter vowels of Deg Xinag are comparable to normative vowel duration data available for other Athabaskan languages (Witsuwit'en (Hargus 2007) and Tsek'ene (Hargus in preparation)), and support the reconstruction of Proto-Athabaskan as having full (*i: *e: *a: *u:) and reduced vowels (*ə *ʊ *"α") (Krauss 1964).

REFERENCES

Hargus, Sharon. 2007. *Witsuwit'en Grammar: Phonetics, Phonology and Morphology*. Vancouver: UBC Press.

Hargus, Sharon. in preparation. Kwadacha (Fort Ware Sekani) Grammar, Seattle and Fort Ware, Ms.

Krauss, Michael. 1962. Ingalik, Ms.

Krauss, Michael. 1964. 'Proto-Athapaskan-Eyak and the Problem of Na-Dene I: The Phonology.' *International Journal of American Linguistics* 30:118-131.

Leer, Jeff. 1979. *Proto-Athabaskan Verb Stem Variation: I. Phonology*. Fairbanks: Alaska Native Language Center, University of Alaska Fairbanks.

ACKNOWLEDGMENTS

This project was made possible by funds provided by NSF (OPP-0337704 and DEL-0651853) and by the University of Alaska (2006 grant to Michael Krauss).

USING ACOUSTICS TO RESOLVE PLACE CONTROVERSIES IN DEG XINAG FRICATIVES

Richard Wright, Sharon Hargus, Julia Colleen Miller
Dept of Linguistics, University of Washington, Seattle WA 98195-4340 USA

1. INTRODUCTION

Field linguists have disagreed about whether the Deg Xinag reflexes of the Proto-Athabaskan third person plural subject prefix *χ- (Leer 2000) and areal prefix *χʊ - (Leer 2005) contain the uvular fricative [χ] or the glottal fricative [h]. In an earlier acoustic study of differences among the eight voiceless fricatives of Deg Xinag that can occur in stems (Wright, Hargus, and Miller 2005, in prep.) we found that /χ/ and /h/ differ significantly in skew and kurtosis but not center of gravity, lowest spectral peak, or standard deviation. In this study we investigate the identity of the prefixal fricative ("x"): specifically, whether it patterns with stem /χ/ or /h/.

2. METHOD

2.1 Participants

Participants are 8 adult native speakers (3 male, 5 female) of Deg Xinag, between the ages of approximately 68-76 at the time of recording. English is a second language for all speakers, but proficiency varies.

2.2 Recording materials and recording procedure

Words on a word list were recorded in isolation in random order (the same random order for each speaker). The word list illustrated three types of fricatives: stem-initial voiceless uvular [χ], stem-initial voiceless glottal [h], and the prefixal fricative whose place of articulation is at issue ("x"). Four lexical items per fricative category were recorded. In two of the words the following vowel was rounded and in two of the words the following vowel was unrounded. Four repetitions of each word were elicited. Some speakers voluntarily produced more than four repetitions. Some repetitions were later discarded due to background noise. Because no speaker was literate in Deg Xinag, the words were elicited through a combination of translation from English and/or prompting in Deg Xinag by the second author.

The recordings were made using a professional CD recorder or compact flash recorder. Participants wore a head mounted microphone (Shure SM-10) to control for source distance. Recordings were made at 44,100 Hz and downsampled to 22,050 Hz.

2.3 Acoustic analysis

Four spectral measures (center of gravity, standard deviation, skew and kurtosis) over a 30 ms. window at the midpoint of the fricative were made using PRAAT (version 4.3.27 and previous). The lowest main spectral peak was identified using FFT spectra (512 points, or approximately 25 ms. window) generated with Multi-Speech (2.5 and previous). The lowest main peak was defined for this study as the frequency above 500 Hz after which there was a sustained drop in frequency. Intensity was also measured over a 25 ms. window at the fricative midpoint using Praat.

2.4 Statistical analysis

A series of two-factor repeated measure analyses of variance and Bonferroni-Dunn post-hoc tests were conducted to examine the effects of fricative Place (/χ/, /h/ or "x") and vowel Rounding on each measure.

3. RESULTS

3.1 Spectral differences

For center of gravity Place had a significant effect ($F[2,14] = 13.484$, $p = .0005$) but Rounding of the following vowel did not. There was no significant Place by Rounding interaction for this measure. Figure 1 illustrates the center of gravity results. Stem /h/ is significantly different from both prefixal "x" and stem /χ/ in center of gravity, but "x" and /χ/ do not differ by this measure.

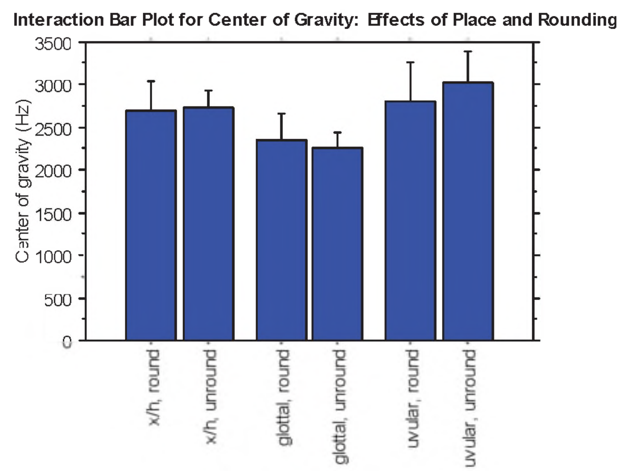


Figure 1. Effects of Place and Rounding on center of gravity

Place and Rounding had significant effects on standard deviation (Place, $F[2,14] = 6.149$, $p = .0121$; Rounding, $F[1,14] = 35.492$, $p = .0006$). The glottal and uvular fricatives differed significantly from each other but not from "x". Place and Rounding also had significant effects on kurtosis (Place, $F[2,14] = 26.628$, $p < .0001$; Rounding, $F[1,14] = 36.498$, $p = .0006$). For this measure /h/ differed significantly from both the uvular fricative and prefixal "x", but the latter two did not differ from each other. Place and Rounding also had a significant interact effect for kurtosis ($F[2,14] = 5.189$, $p = .0206$).

There were no significant effects of Place or Rounding on skew.

Rounding had a significant effect on lowest peak ($F[1,14] = 59.141$, $p = .0001$), but Place did not.

3.2 Intensity differences[a]

For intensity Place and Rounding both had significant effects (Place, $F[2,12] = 5.201$, $p = .0236$; Rounding, $F[1,12] = .0042$, $p = 20.106$). For this measure, the uvular and prefixal "x" fricatives are significantly different, but "x" and /h/ do not differ from each other in intensity, nor do the uvular and glottal fricatives. Figure 2 illustrates the intensity results. There was no significant interaction effect of Place and Rounding.

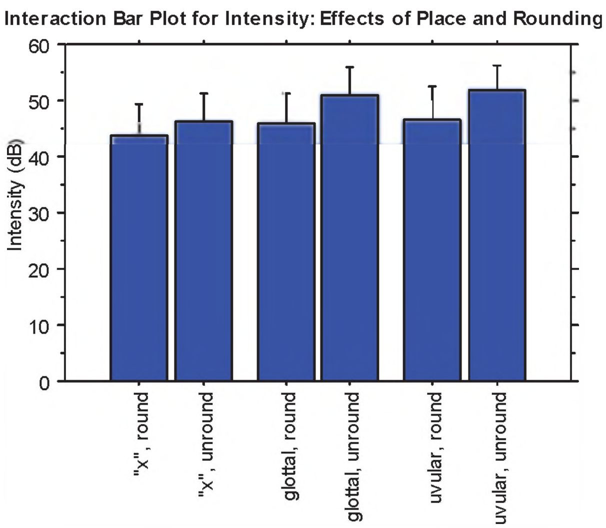


Figure 1. Effects of Place and Rounding on intensity

3.3 Summary

The significant differences between prefixal "x" and the glottal and uvular fricatives in stems for the measures presented in 3.1 and 3.2 are summarized in Table 1:

	"x"
center of gravity	= χ, ≠ h
standard deviation	--
kurtosis	= χ, ≠ /h/
skew	--
lowest peak	--
intensity	≠ χ, = h

Table 1. Significant differences between prefixal "x" and stem /χ h/ (Bonferroni-Dunn test)

4. DISCUSSION

Prefixal "x" patterns with the uvular and not the glottal fricative for two of the spectral measures but with the glottal fricative and not the uvular fricative for the intensity measure. We suggest that the conflict between the two types of results may be a result of the fact that prefixal "x" occurs in unstressed syllables on our word list, as opposed to /χ/ and /h/, which occur in stems, which are stressed. Although stress in Deg Xinag has not been instrumentally studied, there are well known differences in stress between prefixes and stems in Athabaskan languages, with stems consistently reported as receiving stress in a variety of languages (Rice and Hargus 2005). We are therefore inclined to give greater weight to the spectral measures, and identify "x" as the uvular fricative.

The Deg Xinag third person plural subject prefix "x"- (now equated with χ- on the basis of our results) has been reconstructed for Proto-Athabaskan as *χ- (Leer 2000). Assuming that this reconstruction is correct, and valid also for the other instances of Deg Xinag prefixal /χ/-, Deg Xinag can be seen as conservative, but perhaps headed to a stage in which prefixal /χ/- is being replaced with the glottal fricative, as it has been in e.g. Witsuwit'en (Hargus 2007), a related Athabaskan language. Although Witsuwit'en has a contrast between uvular and glottal fricatives in stems, like Deg Xinag, the third person plural prefix is /h/- and the areal prefix is /ho/-~/w/-. The change in Witsuwit'en (and incipient change in Deg Xinag) may be essentially prefixal lenition of stricture, perhaps due to lack of consistent stress on prefix syllables. We also note that since there is no contrast between prefixal /χ/ and /h/, there may be a decreased functional load on "x"-, thereby increasing its variability and reduction (Lindblom 1990).

REFERENCES

Hargus, Sharon. 2007. *Witsuwit'en Grammar: Phonetics, Phonology and Morphology*. Vancouver: UBC Press.

Leer, Jeff. 2000. 'The Negative/Irrealis Category in Athabaskan-Eyak-Tlingit.' In *The Athabaskan Languages: Perspectives on a Native American Language Family*, ed. by Theodore B. Fernald and Paul R. Platero. Oxford: Oxford University Press. 51-72.

Leer, Jeff. 2005. 'How stress shapes the stem-suffix complex in Athabaskan.' In *Athabaskan Prosody*, ed. by Sharon Hargus and Keren Rice. Amsterdam and Philadelphia: John Benjamins. 278-318.

Lindblom, Bjorn. 1990. 'Explaining phonetic variation: a sketch of the H&H theory.' In *Speech Production and Speech Modeling*, ed. by William J. Hardcastle and Alain Marchal. Dordrecht: Kluwer Academic Publishers. pp. 403-439.

Rice, Keren, and Sharon Hargus. 2005. 'Introduction.' In *Athabaskan Prosody*, ed. by Sharon Hargus and Keren Rice. Amsterdam: John Benjamins. 1-45.

Wright, Richard, Sharon Hargus, and Julia C. Miller. 2005. 'An acoustic study of Deg Xinag fricatives.' *Journal of the Acoustical Society of America* 117:2491.

ACKNOWLEDGMENTS

This project was made possible by funds provided by NSF (OPP-0337704 and DEL-0651853) and by the University of Alaska (2006 grant to Michael Krauss).

[a]Intensity results are currently available only for 7 of 8 speakers.

DEFINITELY INDEFINITES? USING ACOUSTICS AS A DIAGNOSTIC IN ST'ÁT'IMCETS

Marion Caldecott[1], Henry Davis[1]
[1]Dept. of Linguistics, Totem Field Studios, University of British Columbia, Vancouver, BC, Canada V6T 1Z4

1. INTRODUCTION

In St'át'imcets (Lillooet Salish) questions such as "Swat ku ats'xentáli ku swat?" are apparently ambiguous between a multiple WH reading ("Who saw whom?") and an indefinite object reading ("Who saw someone")[1]. Given that WH phrases are freely construed as indefinites in non-question contexts, the question arises as to whether the final 'swat' above is ambiguous between a WH phrase and an indefinite object, or whether it is unambiguously indefinite. A similar ambiguity occurs in embedded questions such as "I don't know who saw someone" and "I don't know who saw whom".

This experiment tests the hypothesis that St'át'imcets speakers, like speakers of German [2], will use prosody to distinguish between the WH and indefinite readings of the WH phrases above. If WH phrases in multiple WH questions are ambiguous between two readings, we predict there will be an acoustic distinction between the WH reading and the indefinite reading. If they are unambiguously indefinites, we predict no acoustic distinctions.

2. METHOD

2.1 Participants

Participants were 3 fluent St'át'imcets speakers: AP: female, 64 yrs, Northern dialect; CA: (brother of AP), male, 67 yrs Northern dialect; HD: male, 70yrs, Southern dialect.

2.2 Stimuli

Tokens were 3 sentences with 2 WH phrases per sentence: a control with an indefinite object only reading (1), a matrix WH question (2), and an embedded WH question (3).

(1) Cw7aoz t'u7 ku swat ku tsew'entáli ku swat.
'Nobody kicked anybody.'
(2) Swat ku tsew'entáli ku swat?
'Who kicked someone/who?'
(3) Cw7aoz kwens zwáten lhswátas ku tsew'entáli ku swat.
'I don't know who kicked someone/who.'

2.3 Experiment procedure

Speakers were shown 4 different scenarios using 4 transitive verbs. Each scenario had 5 subscenes and was repeated twice. Each subscene contained one character who always asked "Who xed someone?" and one who always asked "Who xed who?". Each scene had 5 indefinite questions, 3 multiple WH questions, 2 control answers, 2 indefinite answers (embedded questions), and 2 WH answers (embedded questions). 2 filler answers per subsection were also elicited. In total, each speaker recorded 128 tokens (64 matrix questions, 32 embedded questions, 16 control (32 for HD) and 16 fillers). AP and CA were recorded together, while HD and LT were recorded separately. Speakers took turns asking and answering questions. All speakers were recorded in a private home using a Marantz PMD660 solid-state recorder. CA, HD and LT were recorded using a SHURE headworn condenser WH30XLR microphone and AP using a SHURE dynamic LOZSM63LB microphone.

2.4 Acoustical analysis procedure

Measurements were done in PRAAT [3]. Final WH words were marked and measurements for maximum F0, pitch peak alignment, duration and mean intensity were measured for each glide+vowel sequence.

3. RESULTS

Results were analysed in SPSS through a series of independent sample t-tests with a conservative significant p value of $p=.000$ and a marginal $p\leq.01$ [4]. Tokens were excluded due to misspeak, deletion of target and mis-measurement.

Table 1. Mean values for embedded questions

		AP	CA	HD
Control	N	10	11	30
	F0	193.7	150.65	136.21
	dur	203.77	168.93	195.89
	peak%	56	0	0
	intens	73.22	75.98	75.57
WH	N	17	12	15
	f0	208.87	145.51	138.97
	dur	192.65	176.31	197.56
	peak%	83	0	3
	intens	71.58	75.29	74.72
Indefinite	N	16	17	12
	F0	209.36	144.54	136.86
	dur	201.69	170.28	194.1
	peak%	81	0	0
	intens	71.61	74.89	74.24
WH-Indef	F0	NS	NS	NS
	dur	NS	NS	NS
	peak%	NS	NS	NS
	intens	NS	NS	NS
Control-WH	F0	NS	NS	NS
	dur	NS	NS	NS
	peak%	NS	NS	NS
	intens	NS	NS	NS
Control-Indef	F0	NS	NS	NS
	dur	NS	NS	NS
	peak%	NS	NS	NS
	intens	NS	NS	**p=.01**

Table 1 shows that no speaker made a distinction between indefinites and WH phrases in embedded questions. Speakers also make no distinction between the control and WH phrase, and only 1 speaker makes a marginal distinction in intensity between the control and the indefinite. The control N was particularly low for AP and CA because their dialect permits a construction that omits the WH phrase.

Table 2. Mean values for matrix questions

		AP	CA	HD
Control	N	10	11	30
	F0	193.7	150.65	136.21
	dur	203.77	168.93	195.89
	peak%	56	0	0
	intens	73.22	75.98	75.57
WH	N	17	24	15
	f0	257.32	147.23	155.47
	dur	191.79	195.14	194.46
	peak%	92	34	33
	intens	76.83	75.61	86.79
Indefinite	N	34	43	16
	F0	169.68	143.16	151.41
	dur	215.08	185.79	195.86
	peak%	15	21	13
	intens	72.82	75.91	87.17
WH-Indef	F0	**p=.00**	NS	NS
	dur	**p=.00**	NS	NS
	peak%	**p=.00**	NS	**p=.003**
	intens	**p=.00**	NS	NS
Control-WH	F0	**p=.00**	NS	**p=.00**
	dur	NS	**p=.003**	NS
	peak%	**p=.005**	**p=.001**	**p=.00**
	intens	**p=.00**	NS	**p=.00**
Control-Indef	F0	**p=.01**	**p=.002**	**p=.00**
	dur	NS	**p=.01**	NS
	peak%	**p=.002**	**p=.001**	**p=.003**
	intens	NS	NS	**p=.00**

Table 2 shows that results for matrix questions are considerably different than for embedded questions. Two speakers make a distinction between WH phrases and indefinites: AP makes a distinction in every cue, and HD makes one in peak percent. All speakers make a distinction between the control and matrix question WH and indefinites in peak percent, and all but CA in pitch as well. Duration was a strong cue only for CA, while HD had strong intensity differences.

4. DISCUSSION

Our results support the hypothesis that St'át'imcets WH-phrases are unambiguously indefinites. Speakers did not distinguish between them in embedded questions, and the differences found in matrix questions can be seen as a factor of methodological/metalinguistic complications. AP and HD both found the task challenging, in that they were aware that the St'át'imcets translation of both English questions was the same. AP, in particular, made a conscious effort to distinguish the St'át'imcets questions using an English style raised pitch contour, which was quite distinct from CA's pronunciation. Both AP and HD seemed less aware of this distinction in the embedded questions, as reflected in their results.

The differences between the control and matrix WH and indefinites can be explained by an intonational difference between questions and declaratives. In St'át'imcets matrix questions, initial WH words and non-WH words differ precisely in peak percentage [5]. St'at'imcets yes/no questions are also produced at a significantly higher register than declaratives [6]. It appears that matrix WH questions in St'at'imcets, unlike in English [7], are produced with different intonation than declaratives.

If WH phrases in St'át'imcets are indefinites, as shown by these results, this raises the question of why they are subject to superiority effects, like other WH phrases [7]. Research including more speakers and a more opaque methodology is needed to confirm these results.

REFERENCES

[1] Davis, H. 2008. WH-in situ: Japanese meets Salish. In Michiko Takeuch (ed.) Kanagawa University Studies in Language Special Issue: Individual Languages and Language Variation. KU Center for Language Studies. 17-58
[2] The Oxford-Duden German Dictionary. Dudenredaktion & German Section of the Oxford University Press Dictionary Department (eds). Clarendon Press. Oxford.
[3] P. Boersma and D. Weenik (2008), "Praat: Doing phonetics by computer (version 5.0.20)" [Computer program], Retrieved 2008 July 2 from: http://www.praat.org/
[4] Koch, K. (2008). Intonation and Focus in Nlhe7kepmxcin. PhD. Thesis. UBC.
[5] Oberg, M. (2007). Intonation Contours in St'át'imcets. UBCWPL 20. Papers for ICSNL 42.
[6] Syrdal, A. & M. Jilka. *(2003). To rise or to fall: That is the question. Paper presented at the 146 ASA meeting. Austin, Texas.
[7] Caldecott, M. (2007). Some facts about intonation in St'át'imcets. Paper presented at ICNSL 42. UBC Kelowna
[8] Davis, H. 2006. St'át'imcets Superiority. Paper presented at ICSNL 41. University of Victoria. Victoria, B.C.

ACKNOWLEDGMENTS

Thank you to our speakers Aggie Patrick, Carl Alexander and Herman Dan.

This project was funded by SSHRC #410-2003-1138.

Dentals are Grave

Darin Flynn[1] and Sean Fulop[2]
[1]Dept. of Linguistics, University of Calgary, 2500 University Dr NW, Calgary, Alberta, Canada T2N 1N4
[2]Dept. of Linguistics, California State University, 5245 N. Backer Ave. PB92, Fresno, CA, USA 93740-8001

1. INTRODUCTION

Phonetic features require either an articulatory or an acoustic basis. Defining a feature in an optimal fashion can lead to improved explanatory force concerning, e.g., phonetically motivated sound change. This paper will highlight the increased explanation of certain auditorily based sound changes and assimilations, obtained by adjusting the definition of the feature [grave], and concomitant adjustments to the classification of segments. In particular, non-sibilant dentals must be [grave]. Like all coronals, dentals are considered [acute] in Jakobsonian taxonomy [1] et sequentes. However, their noise energy and their involvement in [flat] enhancement and assimilation suggest instead that they are [grave], like labials and velars.

2. CLASSIFICATION

With noise measurements alone, it is notoriously difficult to discriminate reliably between non-sibilant dental and labial consonants, as both present generally level spectra with no significant peaks.

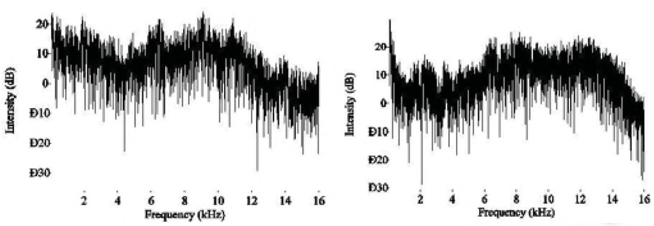


Fig. 1. Simple Fourier power spectra of [θ] in Slavey [t̪͡θʰah] ‘carrot’ and [f] in English ‘fan’ (rendered in Praat)

Indeed their noise energies are so similar that labials commonly substitute for dentals across languages. Table 1 showcases such substitution in Shihgot’ine, a North Slavey (Athabascan) dialect centered in Tulít’a, NT; cf. dentals in South Slavey (NT, AB).

Table 1. Dentals > labials in Shihgot’ine Slavey [2]

Slavey	*Tulít’a*		*Sl.*	*Tul.*	
ʔeht̪͡θáa	ehpaː	‘dryfish’	θɛ̃́	fɛ̃ʔ	‘star’
-t̪͡θʰiʔ	-pʰiʔ	‘head’	θa	fa	‘sand’
t̪͡θ’ih	p’ih/p’ĩ́ɛ	‘mosquito’	-ðáʔ	-va	‘mouth’
-t̪͡θ’éhé	p’é/p’ɛ́h	‘sinew’	-ðeʔ	-ve	‘liver’

Since labials are always considered [grave], we can find no support at all for the claim that (inter)dentals are not [grave], let alone [acute]. Rather it seems clear that (inter)dentals can only be given the same value of [grave] as the labials from which they are so hard to distinguish.

Moreover, given their rather level spectrum, labials (and dentals) cannot be [grave] in the sense of “having predominantly low frequency energy” [1]. Rather, given that an acoustic feature must really be an *auditory* feature, we propose to redefine [grave] as the *audible* presence of significant low frequency noise in a sound. In particular this means that the low frequency noise (< 2.5 kHz) must not be overshadowed by predominant high frequency noise (as in sibilants).

On this definition, [grave] applies equally to labials and dentals as it does to velars, which present a preponderance of noise in lower frequencies. This allows us to rationalize shifts not only between velars and labiodentals as in Table 2, but also between velars and dentals as in Table 3.

Table 2. [f] ~ [x] in K’áshogot’ine Slavey (Radı̨lı̨h Kóé, NT) [2]

fori ~ xori	‘quickly’	lifuʃé ~ lixuʃé	‘fork’
lifótõ ~ lixótõ	‘nine’	fawéhgewe ~ xawéhgewe	‘Old Baldy’

Table 3. [θ, ð] > [x, ɣ] in South Slavey (Tthedzéh Kǫ́ę́, NT) [2]

Standard	*Tthedzéh*		*St.*	*Tth.*	
θe-	xe-	PERF.	-ðáʔ	-ɣáʔ	‘mouth’
θẽ́	xẽʔ	‘star’	-ðéh	-ɣéʔ	‘skin’

3. ENHANCEMENT

According to Jakobson et al. [1] [grave] is *enhanced* (cf. [4]) by another “low tonality” feature of vocoids, [flat], characterized by a downward shift of formants—particularly F_2. (Similarly, consonantal [acute] is enhanced by vocalic [sharp], an upward shift of formants.) Indeed, across languages, F_2 transitions tend to be lower or equal in dentals vs. alveolars [5]. This pattern is shown for American English in Table 4.

Table 4. Starting F2 values (Hz) for alveolars vs. dentals [6]

si	2050	zi	1950	di	2000	θi	1950	ði	1950
sæ	1700	zæ	1700	dæ	1750	θæ	1650	ðæ	1650
sɔ	1150	zɔ	1200	dɔ	1350	θɔ	1050	ðɔ	1150
su	1600	zu	1550	du	1700	θu	1600	ðu	1500

Like its consonantal counterpart [grave], [flat] has diverse articulatory exponents in speech: labialization, velarization, pharyngealization, and retroflexion. We present diachronic evidence that dentals—as [grave]—are enhanced acoustically by all such incongruent articulations.

3.1 Labialization and velarization

Table 5 illustrates that dental consonants, which remain in South Slavey, have evolved into labiovelars in the North Slavey dialect Sahtúgot’ine, centered in Délı̨ne, NT. This sound change also occurred in K’áshogot’ine, another North Slavey dialect (NT), in Tłı̨chǫ Yatii (NT), and word-finally in Gwich’in (YK). As predicted, the [grave] feature of dentals was enhanced by the [flat] feature of labialization and velarization (and the dental gesture was eventually lost).

Table 5. Dentals > labiovelars in Sahtúgot'ine Slavey [2]

Slavey	*Déline*		*Slavey*	*Dél.*	
t̪͡θʰe	kʷʰe	'rock'	-t̪͡θ'ɛ́hɛ́	-kʷ'é	'sinew'
-t̪͡θ'iəne	kʷ'ɛnɛ́	'bone'	-ðé	-wéʔ	'liver'

3.2 Pharyngealization

Table 6 illustrates that dental consonants, which remain intact in Dëne Sųłiné (among other northern Athabascan languages), have evolved into pharyngealized sibilants ("emphatics") in Tsilhqot'in (BC). As predicted, the [grave] tonality of the dental gesture was enhanced (and eventually replaced) by the [flat] tonality of tongue root retraction.

Table 6. Dentals > emphatics in Tsilhqot'in [2]

D. Sųł.	*Tsilh.*		*D. Sųł.*	*Tsilh.*	
t̪͡θʰɛ̃ɬ	t̪͡s̪ʰĩ́ɬ	'axe'	θɛ-	s̪ɛ-	perf. conj.
-t̪͡θʰí	-t̪͡s̪ʰí	'head'	jaθ	jəs̪	'snow'
-t̪͡θ'i	-t̪͡s̪'i	'stay (pl)'	-ðá	-z̪í	'mouth'

Interestingly, Tsilhqot'in's neighbor St'at'imcets Salish has pharyngealized coronal approximants /z̪, z̪'/ which are phonetically dental or interdental [7]. (Arabic has a similar voiced continuant, called ðˤaːʔ.)

3.3 Retroflexion

Retroflexion cannot enhance dentalization, as these gestures are incompatible. Revealingly, however, an interdental approximant /ð̞/ which occurs in disparate Philippine languages has evolved into a retroflex lateral /ɭ/ in Southern Kalinga, and a retroflex rhotic /ɻ/ in Madukayang Kalinga, Balangao, Mansaka and Upper Tanudan Kalinga. We assume that retroflexion came to substitute for interdentalization on the basis of a shared "low tonality": [flat] in /ɭ, ɻ/ and [grave] in /ð̞/. (A recent study of Kagayanen /ð̞/ confirms that it is not [flat]; its F_2 and F_3 are similar to those of an alveolar liquid [3].)

4. ASSIMILATION

That dentals are [grave] predicts that coronal consonants may become dental when released into a [flat] vowel or approximant. This is because "low tonality" in an approximant or vowel, viz. [flat], can be mistaken for "low tonality" in a preceding consonant, viz. [grave]—a kind of acoustic assimilation. This prediction is confirmed in the subsections below.

4.1 Back vowels

Table 7 illustrates that in the Australian language Lardil /t/ is realized as dental before /u, a/ (and as laminal-postalveolar before /i/). On our interpretation, the "low tonality" of [flat] in /u, a/ is assimilated into /t/ as [grave]/dental (and the "high tonality" of [sharp] in /i/ is assimilated into /t/ as [acute]/laminal-postalveolar).

Table 7. Coronal allophony in Lardil [4]

nom.	*fut.*	*nonfut.*	*Acc.*	
kalt̪it	kalt̪it̪-uɽ	kalt̪it̪-at	kalt̪it̪-in	'urine'
jaɽput	jaɽput̪-uɽ	jaɽput̪-at	jaɽput̪-in	'snake, bird'

4.2 Retracted vowels

A palatographic study of Kamwe (Afro-Asiatic) reveals that coronal consonants are alveolar or postalveolar when adjacent to advanced tongue root vowels, but dental when adjacent to retracted tongue root vowels [8]. A similar pattern occurs in Kalenjin (Nilo-Saharan) [9]. In our view, the "low tonality" of [flat] in retracted vowels is assimilated into coronal consonants as [grave]/dental.

4.3 /ɻ, ɝ/

In Irish English, alveolar consonants can be realized as dental before /ɻ, ɝ/, which are retroflexed (and perhaps rounded) [10]. For instance, /t, d, n, l/ are dental in e.g. *train, spider, manner, pillar.* Again, on our interpretation, the "low tonality" of [flat] in retroflex/rounded /ɻ, ɝ/ is assimilated into coronal consonants as [grave]/dental (cf. [10]).

5. CONCLUSION

We have argued that the Jakobsonian feature [grave] does not require a predominance of low-frequency noise, but rather requires that the noise below 2.5 kHz is "sufficiently audible" owing to a lack of predominance of high-frequency noise. This effectively extends the reach of the feature, since all the noisy sounds which were classed as [grave] under the original definition still are—notably labials and velars. We have argued that non-sibilant dentals, too, are [grave]. On the one hand, their noise energy is very similar to that of labials. On the other hand, their interaction with the vocalic feature [flat] across languages strongly suggests that they bear the consonantal counterpart [grave].

REFERENCES

[1] R. Jakobson, G. Fant & M. Halle. (1952). *Preliminaries to speech analysis.* Cambridge, MA: MIT Press.

[2] D. Flynn & S. Fulop. (2005). Acoustic features in Athabaskan. Paper presented at the Annual Meeting of the Linguistic Society of America, Oakland, CA, Jan. 9, 2005.

[3] K. Olson & J. Mielke. (2008). Acoustic properties of the interdental approximant in Kagayanen. *J Acoust Soc Am* 123(5), 3460.

[4] K. N. Stevens, S. J. Keyser, & H. Kawasaki. (1986). Toward a phonetic and phonological theory of redundant features. In J. S. Perkell & D. H. Klatt (eds.), *Invariance and variability in speech processes*, 426-449. Hillsdale, NJ: Lawrence Erlbaum.

[5] S. N. Dart & P. Nihalani. (1999). The articulation of Malayalam coronal stops and nasals. *JIPA* 29, 129-142.

[6] J. Olive, A. Greenwood & J. Coleman. (1993). *Acoustics of American English Speech: A dynamic approach.* New York: Springer-Verlag.

[7] D. Flynn. (2008). Emphasis in the Pacific Northwest Plateau. Paper presented at the 43rd Annual International Conference on Salish and Neighboring Languages, North Vancouver, BC, Jul. 25.

[8] J. Laver. (1965). Some observations on alveolar and dental consonant-articulations in Higi. *Journal of West African Languages*, 2, 59-61.

[9] J. Local & K. Lodge. (2004). Some auditory and acoustic observations on the phonetics of [ATR] harmony in a speaker of a dialect of Kalenjin. *JIPA*, 34, 1-16.

[10] J. Harris. (2002). Dentals are dark: coronals in Irish English. Paper presented at the Ninth International Phonology Meeting, Vienna, Nov. 1-3, 2002.

This project was funded by SSHRC (CE101-3663).

Acoustical Evaluation of Six 'Green' Office Buildings

Murray Hodgson[1], **Rosamund Hyde**[2], **Blair Fulton**[2] **and Catherine Taylor-Hell**[2]

[1] Acoustics & Noise Research Group, SOEH-MECH, University of British Columbia, 3rd Floor, 2206 East Mall, Vancouver, BC, Canada V6T 1Z3 [murray.hodgson@ubc.ca]

[2] Stantec Consulting Ltd., 1100 – 111 Dunsmuir Street, Vancouver, BC, Canada V6B 6A3

1. OBJECTIVES AND METHODOLOGY

The objective of this work was to evaluate six 'green' office buildings acoustically, to learn design lessons. It involved a meeting with designers, performing an occupant satisfaction survey (using a web-based survey developed by the Center for the Built Environment at the University of California at Berkley), analyzing the acoustical responses, walking through the building, planning acoustical measurements, performing and analyzing the acoustical measurements and considering the design implications of the results.

The study involved six very different nominally-'green' office buildings, all designed to prevailing sustainable-development principles, evaluated 1-5 years after occupancy. Descriptions can be found at www.sbtc.ca/index.cfm?bd=KBDet.cfm&id=60. All buildings had mainly glass façades for day-lighting, with sun shades and operable windows, and contained a mix of private and shared offices, and open-office cubicles.

2. MEASUREMENTS AND ACCEPTABILITY CRITERIA

The objective here was to use physical measurements to evaluate the acoustical environment, to explain the survey results which identified situations (workplaces and building conditions) of high and low occupant satisfaction. Workplaces at which measurements were performed were chosen to correspond to high and low occupant satisfaction. In general, these included desks in open-plan, shared and private offices, located in quiet and noisy areas, near and far from operable windows. Furthermore, measurements were made under building conditions expected to correspond to high and low satisfaction (windows or doors closed or open, quiet or noisy external source). Table 1 shows the four acoustical parameters that were measured. Also shown are the acceptability criteria used to evaluate each aspect of the acoustical environments in these office buildings.

Table 1. Acoustical measurement parameters and acceptability criteria.

Measurement parameter	Acceptability criterion
Background noise level, NC in dB	NC 30-35 in meeting, conference rooms NC 35-40 in workspaces
Reverberation time (mid-frequency), RT_{mid} in s	< 0.75 s for comfort, verbal communication
Speech Intelligibility Index, SII	> 0.75 for high speech intelligibility < 0.2 for high speech privacy
Noise Isolation, NIC in dB	NIC 35-40 for executive offices, conference rooms NIC 30-35 for general offices, meeting rooms

3. RESULTS

3.1 Designer meetings

Following are the main points relevant to acoustics learned from the designers at the meetings with them: LEED certification is often a goal that influences design; design often does not involve specialized acoustical expertise—acoustical consultants deal with 'special cases'; quantitative acoustical design targets are never set; designers are aware of acoustical issues; external noise (and pollution) concerns may rule out a fully-natural ventilation concept; 'green' buildings often have operable windows, which causes noise concerns if there's an external noise source; low noise levels resulting from absence of a forced-air system result in low speech privacy; client's wishes (e.g. for open-office design) may affect design; budget short-falls at the end of the project may affect acoustical quality; obtaining good noise isolation involves lined return-air ducts, upholstered furniture, acoustical ceilings, carpet, open-office partitions; some buildings are designed for any occupant; the internal 'fit-up' (e.g. acoustical treatments) is done later by contractors for tenants (on limited budgets); designers often believe their building is well designed, and is successful with occupants.

3.2 Occupant satisfaction surveys

The Berkley survey asks occupants to rate their general satisfaction with the building and with their workspace, with the office layout, with the office furnishings, with thermal comfort, air quality, lighting, acoustic quality and with the washrooms. Occupants rated quality on a scale of -3 (maximum dissatisfaction) to +3 (maximum satisfaction).

Figure 1 shows the results of the occupant satisfaction surveys done in five of the six buildings. Also shown (Ref) are the average scores from all buildings ('green' and non-'green') surveyed using the CBE survey. In general, satisfaction ratings were positive indicating satisfaction. Occupants were very satisfied with their buildings and workspaces, with the furnishings, office layouts, cleanliness and maintenance and with the washrooms. They were generally very satisfied with the lighting, and some-

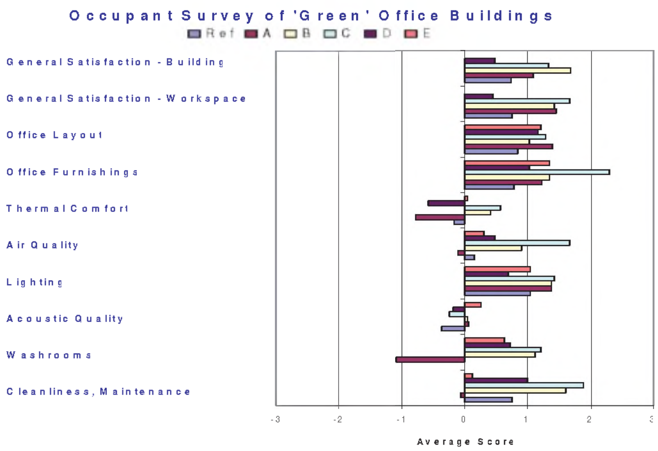


Figure 1. Occupant satisfaction survey results for five 'green' office buildings.

what satisfied with air quality. Satisfaction with thermal comfort varied from somewhat satisfied to somewhat dissatisfied. Occupants were generally dissatisfied with the acoustical environment, which often received the lowest rating. Speech privacy is the biggest acoustical issue.

3.3 Acoustical measurements

Following are the main results of the acoustical measurements:

• Background Noise Level: NC 26-34 (unoccupied, natural ventilation); NC 35-42 (unoccupied, forced-air ventilation); NC 45-60 (external noise, windows open); NC 40-60 (occupied);

• Reverberation Time: open-office areas: 0.6-1.0 s (low absorption); 0.2-0.4 s (high absorption); private offices: 0.4-0.7 s (low absorption); 0.2-0.4 s (high absorption); hallways, atriums: 0.9-2.4 s;

• Speech Intelligibility (private office, across desk, casual voice): 0.3-0.6 (forced-air ventilation, low absorption); 0.7 to 0.8 (natural ventilation, high absorption);

• Speech Privacy. Between open-office cubicles, casual voice): 0.3-0.6 (forced-air ventilation, low absorption); 0.7-0.8 (natural ventilation, high absorption). Outside-inside private office (door open, casual voice)=0.7;

• Noise Isolation: into closed offices = NIC 25-30 (door closed); = NIC 9-15 (door open); between work areas = NIC 7-20.

3.4 Design implications

The main acoustical design implications of the results related to low background noise levels, inadequate speech privacy, excessive reverberation, inadequate noise isolation between workplaces in open and shared work areas, and inadequate internal and external wall isolation. Following are details as they relate to 'green'-building issues:

• since LEED virtually ignores acoustics, a building designed to obtain LEED certification is unlikely to have adequate attention paid to the acoustical environment;

• 'green' buildings often are designed to have natural/displacement ventilation systems; these can affect the acoustical environment beneficially or detrimentally, resulting in low background-noise levels and low noise isolation; however, forced-air ventilation can figure in 'green'-building design;

• many 'green' buildings have little sound-absorption; this affects the acoustical environment detrimentally, resulting in excessive reverberation, low acoustical privacy and inadequate attenuation of sound propagating through the building; however, beneficial sound-absorbing materials can figure in 'green'-building design;

• if a 'green' building, designed with a ventilation system relying on operable windows, is located next to a significant noise source, noise problems are likely, especially if the windows open on the source side;

• a 'green' building designed to rely on a natural/displacement ventilation system, and with transparent envelope for day-lighting, may overheat on hot, sunny days, forcing occupants to open windows and office doors, resulting in excessive noise and low speech privacy;

• background-noise levels in a 'green' building with full or partial natural-ventilation system may be lower than as expected in a conventional building with a forced-air system. These low levels may make it more difficult to achieve adequate speech privacy;

• a 'green' building designed to rely on a displacement ventilation system usually involves air-transfer openings and/or ducts in partitions. These significantly reduce noise isolation between areas, even when treated acoustically.

4. DISCUSSION AND CONCLUSION

The acoustical environment is often judged the least satisfactory aspect of 'green' office buildings by the occupants. They are dissatisfied with excessive noise and poor speech privacy, and consider that the acoustical environment does not enhance their ability to work (i.e. productivity). Speech privacy is often the biggest concern.

The results of this study suggest that improving acoustical environments in 'green' buildings fundamentally requires good acoustical design – that is, the application in design of existing knowledge, with input from an acoustical specialist from the beginning of the design process. This knowledge relates to site selection and building orientation, to the design of the external envelope and penetrations in it, to the building layout and internal partitions, to the design of the HVAC system, to the appropriate dimensioning of spaces, and to the amount and location of sound-absorbing treatments. For a satisfactory acoustical environment, the advice of the acoustical specialist must be followed, and the budgetary resources made available for it to be implemented.

ACKNOWLEDGMENT

The authors gratefully acknowledge the support of the EcoSmart Foundation, Vancouver, which is supported by the Government of Canada.

SOUND TRANSMISSION LOSS OF EXTENSIVE GREEN ROOFS - FIELD TEST RESULTS

Maureen Connelly[1, 2], Murray Hodgson[2]

[1]Centre for the Advancement of Green Roof Technology, School of Construction and the Environment, British Columbia Institute of Technology, 555 Great Northern Way, Vancouver, BC, Canada V5T 1E2; [2]School of Environmental Health and Department of Mechanical Engineering, University of British Columbia, 2206 East Mall, Vancouver, BC, Canada V6T 1Z3

1. INTRODUCTION

Green roofs have the potential to provide excellent external/internal sound isolation due to their high mass, low stiffness and damping effect and, through surface absorption, to reduce noise pollution in the community from aircraft, elevated transit systems, industrial sites and noise build-up in urban areas. This paper reviews the acoustical characteristics and the potential contributions of green roofs to the acoustical environment, investigates applicable literature and sound transmission theory and reports on new empirical findings on the transmission loss of green roofs. A diffuse to free field intensity level measurement methodology was developed to obtain the presented results and for use in a future field test facility.

Current construction practices, driven in part by sustainable building rating programs, have led to an increased use of lightweight metal roof assemblies and a decreased use of ceilings. Green roofs can provide a higher transmission loss than the additional ceiling element and improve transmission loss throughout the full architectural frequency range, specifically desirable in residential occupancies developed below aircraft flight paths. The field testing conducted on two 33 m^2 low profile extensive green roofs indicated an increase of 5 to 13 dB in transmission loss over the low and mid frequency range (50 Hz to 2000 Hz), and 2 dB to 8 dB increase in transmission loss in the higher frequency range relative to the transmission loss of a reference roof.

2. GREEN ROOFS

2.1 Acoustical contribution and benefits

Green roofs have the potential to provide excellent external/internal sound isolation due to their high mass and low stiffness; through surface absorption, green roofs have the potential to reduce noise pollution in the community from elevated transit systems, industrial sites and noise build-up in urban areas. The sound transmission characteristics of a green roof are governed by the multiple layers of fluid, solid, and poro-elastic materials that comprise the full profile of the vegetated roof system. This research is initially focused on flat, nominally sloped (2% to 4%) extensive green roofs. Extensive green roof systems are comprised of the roof deck, vapour barrier, insulation, waterproofing membrane, root barrier, water reservoir/ drainage layer, filter fabric, substrate and drought-tolerant plant species. Extensive green roofs have a shallow substrate profile, 40 mm to 150 mm thick, and are installed on both conventional and protected membrane roof systems—often installed on buildings without significant cost for additional structural loading. Significant research has determined that green roofs can reduce stormwater runoff and lower a building's energy demand for cooling/heating through improved thermal performance (1). Green roofs can provide mitigation of unacceptable noise levels that affect the health, safety and well-being of the urban population; however, the acoustical benefits of green roof technologies have not yet been investigated.

2.2 Literature Review

The review of empirical findings on the TL of roofs highlights three summarizing concepts. First, the use of additional materials to mass load and add damping to the roof can virtually eliminate the coincidence effect and increase transmission loss at low frequencies (2, 3, 4). Second, in the absence of green roof technology, increased TL was achieved by the addition of a ceiling; this addition to the roof assembly increased TL only in the mid and high frequency ranges, not in the low frequency range of potentially disturbing noise from aircraft (5, 6). Third, two reports on pre-cultivated green roof mats provided evidence that the moisture content of the substrate and the water retention mat is a physical property that affects the acoustical characteristics (7,8).

It is suspected that transmission losses in the layer of the substrate will be of major importance. Findings support the hypothesis that soil texture affects the attenuation of sound as it passes through the depth of the soil (9). The physical properties most prevalent in the research include particle size distribution, bulk density and porosity, flow resistivity, and tortuosity; soil conditions of moisture and compaction are variables affecting the acoustic characteristics (10). The findings from the literature support the investigation of sound transmission loss as a function of the plant species on the green roof. The vegetated root interface with the soil has been identified as affecting the normal specific impedance (12).

3. METHOD

A reverse testing method initiated by Mulholand and Sharp (1978), proved to be very useful as a strategy for developing a methodology for this research and a future field test facility (12). The ISO 15186 series indoor to outdoor method (13), propagating sound from an interior diffuse field to an exterior free field was adopted. The ISO 15186 standard uses an intensity approach to evaluate the transmitted acoustic intensity radiated by the element under test while the incident intensity is deduced from the average sound pressure level in the source room. Sound transmission loss is calculated as:

$$TL = \left[L_{p1} - 6 + 10\lg\left(\frac{S}{S_0}\right)\right] - \left[\overline{L}_{In} + 10\lg\left(\frac{S_M}{S_0}\right)\right]$$

L_{p1} is the average sound pressure level in the source room
S is the area of the separating partition under test
L_{In} is the average normal sound intensity level over the measurement surfaces
S_M is the total area of the measurement surfaces
S_0 =1 m^2.

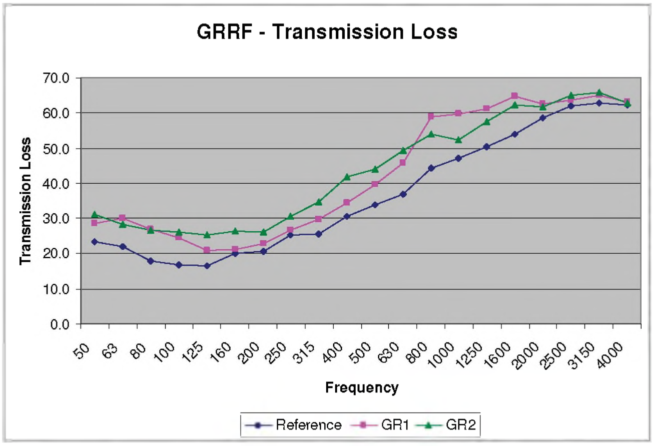


Fig. 1. Measured transmission losses of three roofs.

The Green Roof Research Facility at the BCIT Centre for the Advancement of Green Roof Technology has three independent research roofs originally commissioned in 2003 for the evaluation of stormwater runoff characteristics and thermal performance of green roofs. One roof is a conventional system which acts as a reference roof test specimen, the other two roofs (GR1 and GR2) have the same roof system to the top of the membrane as the reference roof, plus identical green roof components, differing only by the depth of substrate. GR1 has 75 mm and GR2 has 150 mm of substrate. The planting was consistent in its establishment at the time of sound transmission evaluation. Potential sound flanking paths through roof drains, which lead to internal meters, and the roof jack conduits, containing the thermal performance and weather station wiring, were eliminated with sand filled bags and 12 mm steel plates. There is no additional ceiling in the research facility.

An array of five loudspeakers was used in the GRRF interior to create an approximately diffuse sound field. The average sound pressure level was 93 dB generated in each 1/3 octave band. For calculation of the TL the space averaged sound pressure was measured below each of the three roofs. The radiating intensity was measured at 12 discrete points on each of the three roofs.

4. RESULTS

Figure 1 illustrates the resulting transmission loss/frequency curves. All three roofs exhibited an increase in TL with increasing frequency. There is a dip in TL of the reference roof at 125 Hz; however, it is not conclusive whether this is due to the coincidence effect. GR1 (75 mm substrate) exhibited inconsistent increases in TL over the frequency range. The consistent increase in TL of GR2 (150 mm substrate) is illustrated in the TL curve above that of the reference roof. The substrate and green roof materials increased the TL over the low-mid frequency range of 50 Hz to 2000 Hz, by 5 dB to 13 dB, and in the higher frequency range by 2 dB to 8 dB. This is a significant decrease in low and mid frequency sound level transmission; the green roof then provides the opportunity to eliminate the need for ceiling installations for the purpose of increase sound transmission loss. Comparable sound transmission loss through the addition of a ceiling and insulation may not be attainable at low frequencies up to 125 Hz.

5. CONCLUSION

Existing sound transmission algorithms do not adequately predict TL of light-weight roof system or green roofs, nor describe the potential effect of moisture content of the substrate. The sound energy is dissipated in the substrate and provides a mass loading and damping effect on to the light-weight roof deck. Current construction practices, driven in part by sustainable building rating programs, have led to an increased use of lightweight metal roof assemblies and a decreased use of ceilings. Green roofs will provide a higher TL than the additional ceiling element and improve TL throughout the full architectural frequency range, specifically desirable in residential and institutional occupancies developed below aircraft flight paths.

REFERENCES

[1] Connelly, M & Liu, K 2005, Green Roof Research in British Columbia - An Overview. Version of Greening Rooftops for Sustainable Communities, May 5-6, 2005, Washington, D.C. pp. 1-17.
[2] Sharp, BH 1973, A Study of Techniques to Increase the Sound Insulation of Building Elements. U.S. Department of Commerce PB-222 829, Washington.
[3] Friberg, R 1973, 'Transmission Loss and Absorption Factors for Corrugated Steel Roofs, Insulated on the Outside' Proceedings of Inter-Noise, Copenhagen, 1973, pp.213-217.
[4] Alexander, NJH & O'Connor, DE & Orlowski, RJ 1980. 'Industrial Claddings: Sound Absorption and Transmission. Department of Applied Acoustics', University of Salford, Salford.
[5] Bradley, JS 1998, 'Insulating Buildings Against Aircraft Noise: A Review'. National Research Council, Ottawa, Internal Report IRC-IR-760.
[6] Bradley, JS & Lay, K & Norcross, SG 2002, 'Measurements of the Sound Insulation of a Wood Frame House Exposed to Aircraft Noise'. National Research Council, Ottawa, Internal Report IRC-IR-831
[7] Gerhardt, HJ & Grundmann, R 1992, 'Flow mechanical and acoustic investigation on XERO FLOR roof planting system' Report.
[8] Ouis, D & Lagstrom, J 2004, 'The Noise Insulating Capacity of Green Roofs,' Malmo University, Sweden.
[9] Oelze, ML & O'Brien, WD & Darmondy, RG 2002, 'Measurement of Attenuation and Speed of Sound in Soils', Soil Science Society of America Journal , vol. 66, pp.788-796.
[10] Attenborough, K 1982, 'Acoustical Characteristics of Porous Materials', Physics Report, vol. 82, no.3, pp.179-227.
[11] van der Heijden, L.A.M & Claessen, V & de Cock, N 1983 'Influence of Vegetation on Acoustic Properties of Soils' Oecologia vol.56, pp.226-233.
[12] Mulholland, KA 1971, 'Method for measuring the sound insulation of facades: Factors to be considered', Applied Acoustics, vol.4, pp. 279-286.
[13] Anon, E1289-97(2003). Standard Specification for Reference Specimen for Sound Transmission Loss

ACKNOWLEDGMENTS

The authors would like to acknowledge the funding support of NSERC and the industry partners of the BCIT Centre for the Advancement of Green Roof Technology

These findings have been published in 2008 Conference Proceedings, Sixth Annual Greening Rooftops for Sustainable Communities, Connelly M. & Hodgson M, 'Sound Transmission of Green Roofs'

RELATIONSHIP BETWEEN VENTILATION, AIR QUALITY AND ACOUSTICS IN 'GREEN' AND 'BROWN' BUILDINGS

Alireza Khaleghi[1], Karen Bartlett[2], Murray Hodgson[2]
[1]Dept. of Mechanical Engineering, University of British Columbia, 2054-6250 Applied Science Lane, Vancouver, V6T 1Z4
[2]School of Environmental Health, University of British Columbia, 3rd Floor, 2206 East Mall, Vancouver, V6T 1Z3

INTRODUCTION

There is a close relationship between the three factors, ventilation (in general, the heating, ventilation and air conditioning (HVAC) concept/system), indoor air quality (IAQ) and acoustics, and the design of each influences the performance of the others significantly. The application of various standards, methods and criteria in building design results in different conditions and qualities of the resulting ventilation, air quality and acoustical environment in different buildings. This provides an opportunity for investigation and evaluation of the impacts of various design concepts on the internal environmental conditions of buildings.

The present study investigated the relationship between the building design concept and resulting environmental factors, and the relationship between these, in 'green' and conventional, non-'green' ('brown') buildings.

The reduction of energy consumption, using natural ventilation (if applicable) and by the design of the concept of 'green' buildings, is considered very seriously in building design nowadays. Of course, introducing a new type of HVAC system influences the ventilation, indoor-air and acoustical qualities directly. In order to evaluate the environmental factors comprehensively, spaces in both 'green' and 'brown' buildings were chosen such that different types of ventilation concepts/systems – forced-air, displacement and natural – were involved. This paper presents observations about the results and the relationships between them.

METHODOLOGY

The factors investigated in this study were acoustics, ventilation and IAQ. Following are details of the environmental factors measured:

- Acoustics [1,2]:
  background-noise levels (unweighted and A-weighted octave-band and total, NC, RC II); mid-frequency reverberation times (RT_{mid}); noise isolation (octave-band, NIC).
- Ventilation [3]:
  air-exchange rate (ACH = air changes per hour); ventilation rate was quantified by measuring the air flow entering the spaces, and using the SF_6 tracer-decay method.
- IAQ [4,5,6]:
  (glass)fibre dust (fibre concentration) ; ultrafine particulates (ratio of indoor-to-outdoor PC, inside PC – 20% of outdoor PC; PC = particulate concentration, UPC = ultra-fine particulate concentration) ; Volatile Organic Compounds (VOC concentration);

In order to investigate the above components more comprehensively, spaces were chosen which were easy to access, and which included a broad range of HVAC systems, furnishings and acoustical treatments. Hence, three buildings on the UBC campus, with rooms which contained the types of spaces required for this study, and which were adjacent (and, thus has similar outdoor environmental conditions), were selected for monitoring. These three buildings were: 1) BUILD_K, a 'green' building with LEED Silver ranking, 2) BUILD_C ('brown') and 3) BUILD_M ('brown'). The buildings were located next to campus roads, but were generally in quiet external environments.

In view of the fact that the acoustical conditions in any space are a function of its geometry, furnishings and furniture density, the rooms investigated were divided into four main groups:

- Office spaces – these were generally small rooms with a maximum of two occupants. The spaces had natural or forced-air ventilation, with carpets and acoustic tiles, and generally high or low furnishing density;
- Small classrooms – these spaces were generally larger than the spaces in the first group; they had forced-air or natural ventilation systems, and the major distinction that they possessed acoustic ceilings;
- Large spaces with substantial acoustic treatment – this category included large-volume spaces ventilated with displacement and natural ventilation systems; they had acoustic tiles and were carpeted, and contained a high furniture density;
- Large spaces with some acoustical treatment – these spaces covered a wide range of common, large educational spaces with different types of acoustical treatment and furnishings.

This pilot study involved 13 rooms; because of the statistical limitations associated with such a small sample size, the results are more indicative than definitive.

RESULTS

The two main building concepts investigated were construction style (i.e. 'brown', hybrid or 'green') and ventilation-system type (i.e. natural, displacement or forced-air). Environmental factors were correlated with ventilation-system type and construction style. The results with the highest correlation coefficients between the ventilation-system concept and environmental factors can be summarized as follows:

• In spaces with natural-ventilation systems:
the levels of unweighted low-frequency and total noise were lower; the number of air changes per hour was lower; the fibre concentration was lower (due to the type of furnishings in the spaces surveyed); the ratio of indoor-to-outdoor UPC was high (due to no filtration and inadequate control of outdoor air); the indoor temperature was lower at the time of monitoring (i.e. winter).
• In spaces with displacement ventilation systems:
the unweighted and A-weighted mid-frequency noise levels were high; total A-weighted sound-pressure levels were higher; the NC level of the noise was higher (because of nearby exhaust fans).
• In spaces with forced-air ventilation systems:
the unweighted low-frequency and total sound-pressure levels were high; the number of ACH was higher; the fiber concentration was higher (due to the type of furnishings); the ratio of indoor-to-outdoor PC was significantly lower; the indoor temperature was higher at the time of monitoring (i.e. winter).

The results with the highest correlations between the environmental factors and construction style can be summarized as follows:
• in 'brown' buildings:
the NC level was lower than in hybrid and 'green' buildings; this was mainly due to the presence of spaces with displacement ventilation and of naturally-ventilated spaces with open windows in other groups; the VOC concentration in rooms was lower; the ratio of indoor-to-outdoor PC was lower.
• in hybrid buildings:
unweighted and A-weighted low, mid and total sound-pressure levels were higher; rooms had greater NC levels; ventilation rates (ACH) were higher.
• 'green' buildings had:
lower unweighted low-frequency noise levels; lower total, unweighted sound-pressure levels; lower ventilation rates; lower fibre concentrations in the spaces (due to the absence of acoustic tile and carpets, and the low furniture density); higher ratios of indoor-to-outdoor PC; lower indoor temperatures (in winter).

As another part of the investigation, the impact of window status in naturally-ventilated rooms on the environmental conditions inside the spaces was investigated. Figure 1 shows the most significant correlations when the windows were open. In general:
• the unweighted and A-weighted levels of noise in all frequency bands and, therefore, total sound-pressure levels were noticeably increased;
• the magnitude of the indoor – 20% outdoor PC was greatly increased; this was due to the introduction of high volumes of unfiltered air into the spaces by opening the windows.

Moreover, it was observed that ventilation rate had a significant negative correlation with PC. In addition, VOC concentration was lower in spaces with lower ventilation rates (i.e. naturally-ventilated spaces, in general); this was further confirmed by a high positive correlation between ventilation rate and VOC concentration.

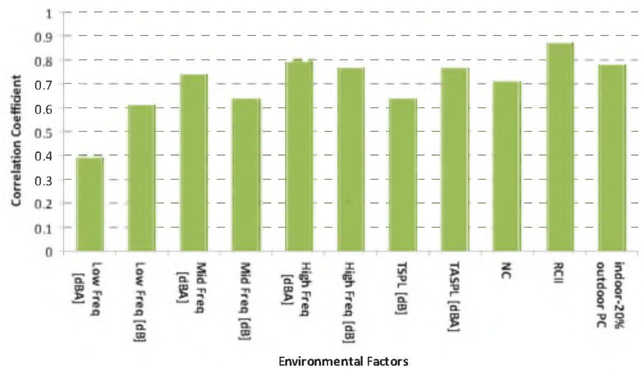


Figure 1. Influence of open windows on environmental factors in 'green' buildings

CONCLUSION

The main conclusions of this pilot study are as follows:
• forced-air ventilation gives better IAQ, but higher HVAC noise levels; IAQ and noise level are directly related;
• in naturally-ventilated spaces with radiant ceiling slabs, lack of acoustic treatment gives lower fibre concentrations, but worse acoustical conditions;
• naturally-ventilated spaces have unsatisfactory ventilation quality but acceptable noise levels with the windows closed, and satisfactory ventilation quality but excessive noise levels with the windows open, even without significant external noise sources;
• naturally-ventilated spaces with few furnishings or sound-absorbing materials have higher IAQ;
• acoustical treatment can enhance acoustic quality, but worsens IAQ.

The results suggest that the optimum building design would use a mechanical-ventilation system designed according to current standards, with a ventilation rate conforming to current standards (to dilute contaminants), and with a carefully selected amount, type and location of acoustical treatment (use materials which generate less contaminants such as fibres, VOCs, etc.; use wall absorption in combination with radiant ceiling slabs).

REFERENCES

[1] ASHRAE Handbook of HVAC Applications, *Sound and Vibration* (ASHRAE, 2003).
[2] ANSI/ASA S12.60-2002, *American National Standard Acoustical Performance Criteria, Design Requirements, and Guidelines for School* (ANSI, 2002).
[3] ASHRAE Standard 62.1, Ventilation for Acceptable Indoor Air Quality (ASHRAE, 2004).
[4] WCB Laboratory Services, *Fibres in Air* (Workers Compensation Board, 1991).
[5] D.W. Bearg, *Indoor Air Quality and HVAC Systems* (Lewis, 1993).
[6] R.M. Hester, *Volatile Organic Compounds in the Atmosphere* (Cambridge: Royal Society of Chemistry, 1995).

THE NOISE SCALES AND THEIR UNITS

Himanshu Dehra[1]
[1]American Institute-Industry Forum on Energy, 1-140 Av. Windsor, Lachine, Canada, H8R 1P7

1. INTRODUCTION

The standardization of noise scales is performed by presenting their sources, their definitions, their measurement equations and their units.

The interference of noise arises due to difference of power of two intensities. The intensity of power for any particle body is a function of development of various stresses. The phenomenon of acoustic resonance occurs when critical stress level, matches with the natural stress level necessary for oscillation of a particle body. The criteria for generation of acoustic resonance include waves propagated with transmission of light, sound, noise, heat, electricity, fluid and fire from a particle body [1, 2]. The psychological feeling of sensation and perception of noise from light, sound, heat, electricity, fluid and fire is a physiological response from the sensory organs of a standard (average) human body [1, 3].

The objective is to standardize the characterization of noise interference due to difference of power of two intensities, which can be due to transmission of light, sound, heat, electricity, fluid and fire into a particle body. The sources of noise, their definitions, their measurement equations and their units are presented in the subsequent sections.

2. SOURCES OF NOISE

The sources of noise are classified according to the type of wave of interference:

Light: The light is a visual sensation evaluated by an eye with seeing of a radiant energy in the wavelength band of electromagnetic radiation from approximately between 380 to 765 nm (nm = nanometer = $(1+10^9)^{-1}$ meter). The units of light are based on the physiological response of a standard (average) eye. The human eye does not have the same sensitivity to all wavelengths or colors. The solar energy spectrum in the visible region contributes in adding daylight as a visual sensation to the human body.

Sound: The sound is a hearing sensation evaluated by ear due to fluid pressure energy in the frequency band approximately between 20 Hz and 20,000 Hz. The units of sound are based on the physiological response of the standard (average) ear. The human ear does not have the same sensitivity to the whole frequency band.

Heat: The heat is a sensation of temperature evaluated by a radiant energy in the wavelength band of electromagnetic radiation from approximately between 0.1 μm to 100 μm (μm = micrometer = $(1+10^6)^{-1}$ meter). The units of heat are function of sensation of temperature. The sensation of temperature is a measure of hotness and coldness. Thermal comfort is an evaluation of comfort zone of temperature on the basis of physiological response of a standard (average) human body. The solar energy spectrum in the ultra violet radiation region contributes to sensation of discomfort of the human body.

Electricity: The electricity is a sensation of shock evaluated by skin of an observer due to an electromagnetic energy stored in a conductor short-circuited by a human body either due to pass of direct current or an alternating current.

Fluid: The fluid is a combined sensation of ventilation and breathing evaluated by the amount of fluid passed either externally or internally through a standard (average) human body.

Fire: The fire is a sensation of burning caused due to combined exposure of skin to radiation energy and fluid acting on a standard (average) human body.

3. DEFINITIONS

The criteria for definitions of noise are based on areas of energy stored in a wave due to interference, speed of wave and difference of power between two intensities of wave [4].

Noise of Sol: The noise of sol is noise occurring due to difference of intensities of power between two solar systems. The amplitude of a solar energy wave is defined as the power storage per unit area per unit time. The solar power is stored in a packet of solar energy wave of unit cross sectional area and of length s, the speed of light.

Noise of Therm: The noise of therm is noise due to difference of intensities of power between two heat power systems. The amplitude of a heat wave is defined as the power storage per unit area per unit time. The heat power is stored in a packet of heat wave of unit cross sectional area and of length s, the speed of light.

Noise of Photons: The noise of photons is noise due to difference of intensities of power between two lighting systems. The amplitude of a light beam is defined as the power storage per unit area per unit time. The light power is stored in a packet of light beam of unit cross sectional area and of length s, the speed of light.

Noise of Electrons: The noise of electrons is noise due to difference of intensities of power between two electrical power systems. The amplitude of an electricity wave is defined as the power storage per unit area per unit time. The electrical power is stored in a packet of an electricity wave of unit cross sectional area and of length s, the speed of light.

Noise of Scattering: The noise of scattering is noise due to difference of intensities of power between two fluid power systems. The amplitude of a fluid wave is defined as the power storage per unit area per unit time. The fluid power is stored in a packet of fluid energy wave of unit cross sectional area and of length s, the speed of fluid.

Noise of Scattering and Lightning: The noise of scattering and lightning is a noise due to difference of intensities of power between two fire power systems. The amplitude of a flash of fire is defined as the power storage per unit area per unit time. The fire power of light is stored in a packet of flash of fire of unit cross sectional area and of length s, the speed of light. The fire power of fluid is stored in a packet of flash of fire of unit cross sectional area and of length s, the speed of fluid.

Noise of Elasticity: The noise of elasticity is a noise due to difference of intensities of power between two sound systems. The amplitude of a sound wave is defined as the power storage per unit area per unit time. The sound power is stored in a packet of sound energy wave of unit cross sectional area and of length s, the speed of sound.

4. NOISE MEASUREMENT

Noise of Sol: The solar power intensity I is the product of total power storage capacity for a packet of solar energy wave and the speed of light. The logarithm of two solar power intensities, I_1 and I_2, gives power difference for two solar power intensities. It is mathematically expressed as [4, 6]:

$$\text{Sol} = \log \left(I_1\right) \left(I_2\right)^{-1} \tag{1}$$

Where, Sol is a dimensionless logarithmic unit for noise of sol. The decisol (dS) is more convenient for solar power systems. Since a decisol (dS) is $1/11^{th}$ unit of a Sol, it is mathematically expressed by the equation:

$$\text{dS} = 11 \log \left(I_1\right) \left(I_2\right)^{-1} \tag{2}$$

Noise of Therm: The heat power intensity I is the product of total power storage capacity for a packet of heat energy wave and the speed of light. The packet of solar energy wave and heat energy wave, have same energy areas, therefore their units of noise are same as Sol.

Noise of Photons: The light power intensity I is the product of total power storage capacity for a packet of light energy wave and the speed of light. The packet of solar energy wave and light energy wave, have same energy areas, therefore their units of noise are same as Sol.

Noise of Electrons: The electrical power intensity I is the product of total electrical storage capacity for a packet of electricity wave and the speed of light. The packet of solar energy wave and an electricity wave, have same energy areas, therefore their units of noise are same as Sol.

Noise of Scattering: The fluid power intensity I is the product of total power storage capacity for a packet of fluid energy wave and the speed of fluid. The logarithm of two fluid power intensities, I_1 and I_2, gives power difference for two fluid power intensities. It is mathematically expressed as:

$$\text{Sip} = \log \left(I_1\right) \left(I_2\right)^{-1} \tag{3}$$

Where, Sip is a dimensionless logarithmic unit for noise of scattering. The decisip (dS) is more convenient for fluid power systems. Since a decisip (dS) is $1/11^{th}$ unit of a Sip, it is mathematically expressed by the equation:

$$\text{dS} = 11 \log \left(I_1\right) \left(I_2\right)^{-1} \tag{4}$$

The water is a standard fluid used with a specific gravity of 1.0 for determining the energy area for a fluid wave.

Noise of Scattering and Lightning: The intensity, I, of flash of fire with power of light, is the product of total power storage capacity for a packet of fire wave and the speed of light. The intensity, I, of flash of fire with power of fluid, is the product of total power storage capacity for a packet of fire wave and speed of fluid. The combined effect of scattering and lightning for a noise due to flash of fire is to be determined by superimposition principle. The packet of solar energy wave and a flash of fire with power of light, have same energy areas, therefore their units of noise are same as Sol. The flash of fire with power of light may also include power of therm. The packet of fluid energy wave and a flash of fire with power of fluid, have same energy areas, therefore their units of noise are same as Sip. A multiplication factor of a specific gravity of fluid is used in determining the areas of energy for the case of fluids other than water.

Noise of Elasticity: The sound power intensity I is the product of total power storage capacity for a packet of sound energy wave and the speed of sound. The logarithm of two sound power intensities, I_1 and I_2, gives power difference for two sound power intensities. It is mathematically expressed as [5]:

$$\text{Bel} = \log \left(I_1\right) \left(I_2\right)^{-1} \tag{5}$$

Where, Bel is a dimensionless logarithmic unit for noise of elasticity. The decibel (dB) is more convenient for sound power systems. Since a decibel (dB) is $1/11^{th}$ unit of a Bel, it is mathematically expressed by the equation:

$$\text{dB} = 11 \log \left(I_1\right) \left(I_2\right)^{-1} \tag{6}$$

The units of noise scales and their limiting conditions are presented in Table 1.

TABLE I. NOISE SCALES

Reference[a] (I_2 = 1 Wm^{-2})	Noise Scales and limiting Conditions		
	Noise of Sol	Noise of Scattering	Noise of Elasticity
Units	Sol	Sip	Bel
I_1 = 1 Wm^{-2}	No Positive Solar Energy	No Positive Fluid Energy	No Positive Sound Energy
I_1 = 1+→0 Wm^{-2}	Deacreasing Solar Energy	Decreasing Fluid Energy	Decreasing Sound Energy
I_1= + ve	Increasing Solar Energy	Increasing Fluid Energy	Increasing Sound Energy
I_1 = -1 Wm^{-2}	Negative Solar Energy	Negative Fluid Energy	Negative Sound Energy
	Darkness	Low Pressure	Inaudible range
I_1 = -ve	Darkness incresaing, distance from point source of light increasing	Low pressure increasing, vacuum approaching	Inaudible range increasing, vacuum approaching
I_1 = -1+→0 Wm^{-2}	Negative Solar Energy	Negative Fluid Energy	Negative Sound Energy
	Decreasing Darkness	Decreasing Low Pressure	Decreasing inaudible range

[a] Reference value of I_2 = 1 Wm^{-2} signifies the limiting condition with areas of noise interference approaching to zero.

5. CONCLUSION

The noise scales and their units are presented.

REFERENCES

[1] Dehra, H. (2007). A Unified Theory for Stresses and Oscillations, Canadian Acoustics, vol. 35, pp. 132-133.

[2] Del Toro, V., (1986). Electrical Engineering Fundamentals, 2nd ed., Prentice-Hall of India, New Delhi.

[3] Goldstein, E. Bruce (1999). Sensation & Perception, Books/Cole Publishing Company, 5th edition.

[4] Dehra, H. (2008). Power Transfer and Inductance in a Star Connected 3-Phase RC Circuit Amplifier, in the proceedings of American Institute of Chemical Engineers - AIChE 2008 Spring Meeting, New Orleans, LA, USA, April 6-10, 2008, session 96a.

[5] Randall, Robert H. (1951). An Introduction to Acoustics, Dover Publications, republication of 1st edition.

[6] Dehra, H. (2007). A Theory of Acoustics in Solar Energy, Proceedings of the 1st International Conference on Advances in Energy Research, Indian Institute of Technology (IIT) Bombay, India, December 12-14, 2007.

A CEPSTRAL-DOMAIN ALGORITHM FOR PITCH ESTIMATION FROM NOISE-CORRUPTED SPEECH

Celia Shahnaz, Wei-Ping Zhu, and M. Omair Ahmad
Centre for Signal Processing and Communications, Dept. of Electrical and Computer Engineering
Concordia University, Montreal, Quebec, Canada H3G 1M8

1. INTRODUCTION

Reliable pitch estimation is one of the most significant problems in speech processing applications since any error in pitch detection has a deleterious effect on system performance. Many applications also need robust pitch estimation from noise-corrupted speech. Since noise obscures the periodic structure of speech, the pitch estimation in noise is an intricate task. Even though a large number of pitch estimation algorithms have been disclosed in literature for clean speech, it is rather surprising that only a few algorithms have been proposed for the pitch estimation in the presence of noise [1]-[2].

The main purpose of the present work was to develop an accurate algorithm for pitch estimation from noisy speech observations with an aim to substantially reduce the pitch-errors for a wide range of speakers. We propose to employ a Discrete Cosine Transform (DCT) based power spectral subtraction scheme for enhancing noisy speech prior to pitch estimation. Then, in order to remove the detrimental effect of formants, the de-noised speech is inverse filtered to yield an output referred to as the Linear Prediction (LP) residual. The kernel of the proposed method lies in the introduction of a DCT power cepstrum (DPC) of the LP residual that exhibits a more prominent peak at the true pitch relative to that demonstrated by the conventional cepstrum of noisy speech. Consequently, global maximization of the DPC results in a momentous improvement in the pitch estimation accuracy. Extensive simulation results confirm that for both low and high-pitched speakers, our algorithm consistently outperforms the state-of-the-art pitch estimation methods in the white or multi-talker babble environmental noise.

2. PROPOSED METHOD

2.1 Pre-processing

Assuming $x(n)$ and $d(n)$ as clean speech and additive noise signals, respectively, the observed noisy signal $y(n)$ is given by,

$$y(n) = x(n) + d(n) \tag{1}$$

$y(n)$ is segmented into frames with a frame size N by the application of a window function $w(n)$. As a pre-processing, one-dimensional DCT is performed on the windowed noisy frame $y'(n)$ which is given by the relation,

$$Y'(k) = c(k)\sum_{n=1}^{N} y'(n)\cos\left[\frac{\pi(2n-1)(k-1)}{2N}\right] \tag{2}$$

In (2), the co-efficients c are given as,

$$c(k) = \sqrt{\frac{1}{N}} \quad for\, k = 1,\; c(k) = \sqrt{\frac{2}{N}} \quad for \quad 2 \le k \le N \tag{3}$$

The DCT co-efficients corresponding up to the upper frequency limit of the first formant is retained and the rest of the co-efficients is set to zero. From the resulting DCT sequence denoted as, $Y_w(k)$, a time domain pre-processed noisy speech frame, $y_w(n)= x_w(n)+d_w(n)$, can be obtained through the inverse DCT operation. Since noise is additive both in signal and DCT domain, $Y_w(k)$ can be written as,

$$Y_w(k) = X_w(k) + D_w(k) \tag{4}$$

where, $X_w(k)$ and $D_w(k)$ are the DCTs of $x_w(n)$ and $d_w(n)$, respectively.

2.2 Noise reduction

The instantaneous power spectrum of $y_w(n)$ is approximated as follows,

$$|Y_w(k)|^2 \approx |X_w(k)|^2 + |D_w(k)|^2 \tag{5}$$

Prior to pitch estimation, in order to enhance speech, a DCT based modified power spectral subtraction scheme is derived from (5) as,

$$|\hat{X}_w(k)|^2 = \begin{cases} |Y_w(k)|^2 - \alpha|\hat{D}_w(k)|^2 & if\ |\hat{X}_w(k)|^2 > 0 \\ \beta|\hat{D}_w(k)|^2, otherwise \end{cases} \tag{6}$$

where, β is the spectral floor parameter, and α refers to the over-subtraction factor. The DCT noise power spectrum is estimated from the beginning silence frames and updated during the immediate past silence frames before the speech frame using the following averaging rule,

$$|\hat{D}_w^m(k)|^2 = \lambda|\hat{D}_w^{m-1}(k)|^2 + (1-\lambda)|Y_w^m(k)|^2 \tag{7}$$

where, m represents the frame index, and λ the forgetting factor. In order to compensate for the noise spectrum errors, the value of α is adequately adapted from frame to frame as a function of segmental noisy signal to noise ratio ($NSNR$) of the frame where, $NSNR$ and α are formulated as,

$$NSNR = \frac{\sum_k |Y_w(k)|^2}{\sum_k |\hat{D}_w(k)|^2}, \quad NSNR_{dB} = 10\log_{10} NSNR \tag{8}$$

$$\alpha = \alpha_0 - \frac{NSNR_{dB}}{s}, \quad -5 \le NSNR_{dB} \le 20 \tag{9}$$

where, $\alpha = 1$ for $NSNR_{dB} > 20$, $\alpha = \alpha_{max}$ for $NSNR_{dB} < -5$, with α_0, the desired over-subtraction factor at $NSNR_{dB} = 0$, α_{max}, the maximum allowable value of α, the constant s are chosen experimentally as 4, 5 and 20/3, respectively. Once the subtraction is performed in the DCT domain based on (6), an enhanced speech frame is obtained using the following relationship,

$$x_w^{en}(n) = IDCT\left\{|\hat{X}_w(k)| e^{j\arg(Y_w(k))}\right\} \tag{10}$$

where, $IDCT$ stands for the inverse DCT.

2.3 Pitch Estimation

Our idea is to translate the pitch estimation problem from $y_w(n)$ into a problem of estimating the pitch from $x_w^{en}(n)$ based on the knowledge of the instants of Glottal Closure (GCIs) derived from the Linear Prediction (LP) analysis. Representing τ as the lag variable and estimating the ACF of $x_w^{en}(n)$ as,

$$\phi_x(\tau) = \frac{1}{N}\sum_{n=0}^{N-1-\tau} x_w^{en}(n) x_w^{en}(n+\tau), \quad \tau \geq 0 \tag{11}$$

least squares solution of the Linear Prediction Co-efficients (LPCs) denoted as, $\{a_k, k=1,\ldots,p\}$ can be computed using the following set of equations,

$$\begin{bmatrix} \phi_x(p) & \phi_x(p-1) \ldots & \phi_x(1) \\ \phi_x(p+1) & \phi_x(p) \quad \ldots & \phi_x(2) \\ \vdots & \vdots & \vdots \\ \phi_x(p+S-1) \ldots & & \ldots \phi_x(S) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} = - \begin{bmatrix} \phi_x(p+1) \\ \phi_x(p+2) \\ \vdots \\ \phi_x(p+S) \end{bmatrix} \tag{12}$$

In (12), p is the order of prediction and $\tau = p+1, p+2, \ldots, p+S$ are considered, where, S governs the number of equations. In order to remove harmful vocal-tract (formants) effect from the extraction procedure of pitch, $x_w^{en}(n)$ is inverse filtered with the LP parameters $\{a_k\}$. The output of the inverse filter is referred to as the linear prediction (LP) residual given by,

$$\Lambda(n) = x_w^{en}(n) + \sum_{k=1}^{p} a_k x_w^{en}(n-k) \tag{13}$$

The LP residual $\Lambda(n)$ corresponds to an estimate of the excitation source of $x_w^{en}(n)$. In order to handle the heavy noise, a DCT power cepstrum (DPC) of the LP residual $\Lambda(n)$ is introduced as,

$$\Omega(n) = \left(IDCT(\log|DCT(\Lambda(n))|^2)\right)^2 \tag{14}$$

The DPC of the LP residual is more convenient in that it emphasizes the true pitch-peak compared to that revealed by the conventional cepstrum. If F_s is the sampling frequency (Hz), by searching for the global maximum of $\Omega(n)$, the desired pitch (F_0) is obtained as,

$$F_0 = \frac{F_s}{T_0}, T_0 = \arg\max_n[\Omega(n)] \tag{15}$$

3. SIMULATION RESULTS

3.1 Simulation conditions

The performance of the proposed method is evaluated using the *Keele* reference database [3]. This database provides a reference pitch at a frame rate of 100 Hz with 25.6 ms window. The *Keele* database has studio quality, sampled at 20 kHz with 16-bit resolution. In order to use this database, we have chosen the same analysis parameters (frame rate and basic window size). For simulation, white and multi-talker babble noises from the *NOISEX'92* database are used. The noisy speech with SNR varying from 5 dB to ∞ dB is considered for Simulations. For windowing operation, we have used a normalized hamming window. The parameter β was set to 0.002 and in the estimation of $\{a_k\}$ parameters, S is chosen as $5p$ with p=10.

Table 1. Percentage gross pitch-error for the white noise corrupted speech at SNR = 5dB

Speaker	Proposed Method	CEP Method	WAC Method
Male	5.05	26.80	9.38
Female	3.71	24.56	6.61

Table 2. Percentage gross pitch-error for the multi-talker babble-noise corrupted speech at SNR = 5dB

Speaker	Proposed Method	CEP Method	WAC Method
Male	10.59	40.2	19.35
Female	5.40	31.63	14.53

3.2 Performance comparison

We have evaluated and compared the performance of the proposed pitch estimation algorithm with the cepstrum (CEP) [2] and weighted autocorrelation (WAC) methods [4]. For a speaker group, the percentage gross pitch-error GPE (%) is calculated considering two male (or female) speakers. We have defined GPE (%) which is the ratio of the number of frames giving ''incorrect'' pitch values to the total number of frames. The estimated pitch is considered as ''incorrect'' if it falls outside 20% of the true pitch value. In Table 1. and Table 2., GPE(%) for male and female speaker group are summarized considering the white noise and babble noise-corrupted speech signals, respectively, at an SNR=5 dB. It is evident that in comparison to the CEP and WAC methods, a significant reduction in the GPEs(%) is achieved by the proposed algorithm for both female and male speakers in the presence of a white or a babble noise.

4. DISCUSSION

In this paper, a new pitch estimation algorithm for noisy speech preceded by a DCT domain noise reduction scheme is presented. In order to indicate accurately the approximate location of the GCIs to be used for pitch estimation, a DCT power cepstrum (DPC) of the LP residual is proposed that is capable of reducing the pitch-errors in a difficult noisy condition. We argue through simulation results that the proposed method is suitable for a wide range of speakers and significantly outperforms some of the reported methods in the present of a heavy noise in terms of percentage gross pitch-errors.

REFERENCES

[1] D. O'Shaughnessy, (2000). Speech communications: human and machine. *IEEE Press, NY*, second edition.

[2] W. J. Hess, (1993). Pitch Determination of Speech Signals. *New York: Springer*.

[3] G. Meyer, F. Plante, and W.A. Ainsworth, (1995). A pitch extraction reference database. *EUROSPEECH'95*, 827-840.

[4] T. Shimamura and H. Kobayashi, (2001). Weighted autocorrelation for pitch extraction of noisy speech. *IEEE Trans. Speech Audio Processing*, 9, 727-730.

FFT TUTOR: A MATLAB-BASED INSTRUCTIONAL TOOL FOR FFT PARAMETER EXPLORATION

Arvind Singh, Tom De Rybel, and José R. Martí
Dept. of Electrical and Computer Engineering, The University of British Columbia, 2332 Main Mall, Vancouver, British Columbia, Canada, V6T 1Z4 tomr@ece.ubc.ca

1. INTRODUCTION

Although the Fast Fourier Transform (FFT) has been the staple of signal processing [Oppenheim(1998)] for many years, it is still frequently misapplied. In many cases, the confusion stems from misconceptions regarding the relations between time and frequency-domain parameters. Also, spectral leakage due to mismatches between the sample rate and the harmonic contents, and the choice of windowing technique, are a frequent cause of trouble.

In this paper, we will first present a summary of how the various FFT parameters relate and can be chosen in a practical way, followed by a discussion on spectral leakage, windowing and zero-padding. Then, a MatLab-based tool is introduced to help visualize the relevant concepts.

The tool allows the user to graphically evaluate the influence of the analysis parameters on harmonic signals, as well as on a custom dataset, such as a sound recording. This, then, allows the user to experiment with, and optimize, the FFT analysis parameters to enhance the resulting FFT spectrum, as well as visually compare the inverse of the spectrum produced with the original time-domain signal.

2. FFT Parameters at a Glance

In order to use the FFT effectively, the parameters governing the time and frequency domain windows must be well chosen. It are only the samples in these windows which are of concern to the FFT transform and it has no further knowledge of the signal of interest. Errors in the choice of window size and transform parameters are a main cause of poor performance of the FFT.

To select the parameters, it must always be remembered that the FFT assumes that the input signal is periodical. Thus, the signal present in the time window should be precisely an integer number of periods of the signal so that, when the window is repeated periodically, the resulting time sequence should be a faithful reproduction of the periodical input.

Small mis-matches, such as "one sample off" introduce discontinuities and can produce very noticeable degradation of the spectrum. Windowing is then used to improve the end-to-end match.

The relation between the FFT parameters [Marti(2002)] is illustrated in Fig. 1. The rules can now be summarized as:

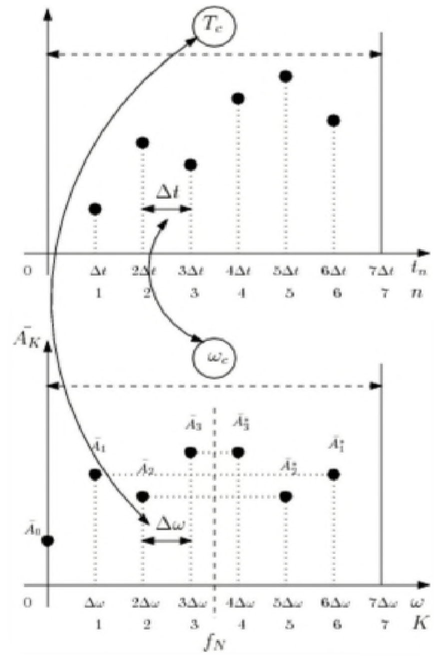


Fig.1: Relations between the time and Frequency-domain windows

$$T_c = N\,\Delta t \qquad \omega_c = \frac{2\pi}{\Delta t}$$

$$\Delta t = \frac{2\pi}{\omega_c} \qquad \Delta\omega = \frac{2\pi}{T_c}$$

$$f_N = \frac{1}{2\,\Delta t}$$

- The size of the time window T_c determines the lowest frequency that can be represented, and thus the frequency resolution $\Delta\omega$ of the transform.

- The sample rate Δt for a given time window determines the number of points N in the window. This determines the bandwidth ω_c. The Nyquist criterion allows 1/2 of this to avoid aliasing. In practice, 1/10 to 1/5 is used due to filter limitations.

- The frequency-domain components include one DC term A_0, and N complex conjugate frequency components An around the Nyquist frequency f_N.

3. FFT TUTOR

Fig. 2 shows a screen-shot of the MATLAB software tool developed, demonstrating a case where the time window is chosen incorrectly.

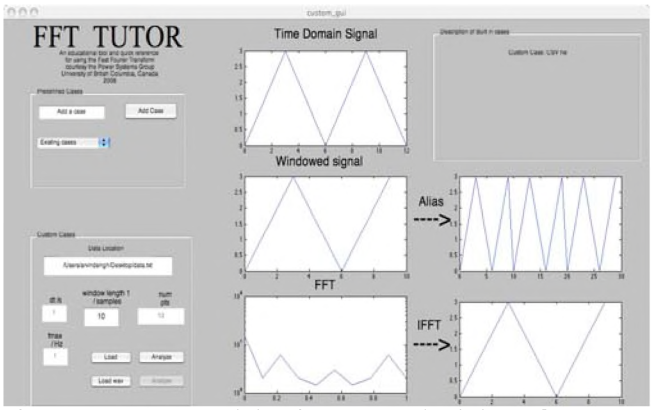


Fig. 2. FFT-TUTOR tool showing a custom loaded case from a CSV file.

The tool features four predefined cases:

- Single Frequency
- Double Frequency
- Zero padding (periodic)
- Zero padding (pulse)

For each results using proper windowing selection is illustrated, as well as results obtained with commonly encountered errors. Each case is furnished with a short explanation in the top right hand corner.

3.1 Single Frequency Case

Fig. 2 illustrates the effect of incorrectly choosing window lengths for periodic signals. An FFT frequency domain window spans from 0 to 2pi radians, the latter non-inclusive. This corresponds to one whole period of the periodical signal in time domain. A common mistake is that the time domain window often contain one point too many because both the 0 rad and 2pi rad values are included in the window. If sampling is done every 90 degrees, for example, then only the 0, 90, 180, and 270 degree values should be included in the window.

3.2 Double Frequency

When multiple known frequencies are present in the signal, proper windowing demands that integer numbers of periods of each frequency are present in the signal. If this condition can not be met, the periods of those frequency components do not match the time window, and spectral leakage will occur. This leakage is visible in the spectrum as additional components spread around the frequency bin that matches the component the closest. Energy that should be purely in the correct, but not available, spectral component thus “leaks” into neighboring bins. One way to think of this is that the various components interpolate the missing required frequency bin.

3.3 Zero padding (periodic)

Zero padding is often used for increasing frequency resolution, since it effectively lengthens the size of window. It is often misapplied, however, in use with periodic signals. In these cases, it results in a different time domain version of the signal as seen by the FFT, as shown for a sine wave in Fig. 3, and corrupts the frequency response by introducing artificial frequency components.

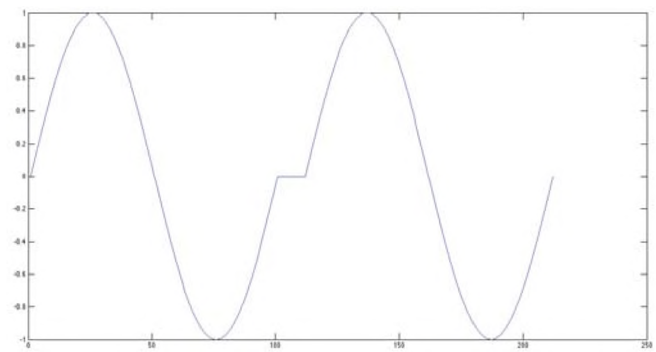


Fig. 3: Time-domain signal due to incorrect zero padding.

3.4 Zero padding (pulse)

Unlike periodic signals, zero padding can be used effectively to increase the frequency resolution for impulses. However this is not true in all cases. In fact, the term zero-padding is somewhat misleading. Padding a transient signal assumes that the padded values are approximately the same as the actual signal if a longer window was taken. Therefore, if a DC offset is present, the signal must be padded to the DC value to avoid corruption of the frequency spectrum.

3.5 Custom cases

In addition to the predefined cases outlined previously, the software tool can read CSV files as well as *.wav files. The user can specify the window length and observe the aliased signal, frequency spectrum and reconstructed signal, as shown in Fig. 2.

REFERENCES

Oppenheim, A. V. and Ronald, W. S. and Buck, J. R. (1998). Discrete-Time Signal Processing, 2nd edition. Prentice Hall.
Martí, J. R. (2002). EECE466 Digital Signal Processing Systems, Course notes.

AUTHOR NOTES

The FFT Tutor program can be obtained free of charge from the authors for private and commercial use.

STRUCTURAL SEGMENTATION OF MUSIC WITH FUZZY CLUSTERING

Daniel Graves[1], Witold Pedrycz[1]
[1]Dept. of Electrical & Comp. Eng., University of Alberta, Edmonton, 9107 – 116 St., AB, Canada T6G 2V4

1. INTRODUCTION

Often in music we talk about its structural components as in intro, verse, chorus, bridge, outro, etc. In this research our objective is to automatically break music into these basic musical structures found in most popular music. The motivation behind musical segmentation is the many potential applications such as music thumbnail generation for sampling music from a database, fast-forward mechanisms for jumping to the next musical structure, music information retrieval, music summarization, and searching or browsing a musical database.

The goal of this research is to segment a digital music file such as an MP3 file. Digital music files are widely available due to the popularity of Internet music retailers; hence, the structural segmentation can be applied to any song purchased or converted to this digital format.

Musical segmentation was performed by [1] using MPEG-7 features and constrained clustering based on K-Means. Goto, c.f. [2], develops a method called RefraiD that detects the chorus sections of music and can even detect key changes in choruses using a 12-dimensional chroma feature vector. Peeters, c.f. [3], investigates musical segmentation of structural components using Mel Frequency Cepstral Coefficients (MFCC) and compares the sequence approach of structural segmentation with the state approach (HMM) and conclude that the state approach is more robust and computationally efficient. Authors in [4] propose a method of musical segmentation by detecting boundaries first, followed some aggregation. Many features are used including MFCCs. Abdallah et. al., c.f. [5], build a musical segmentation architecture based on a Bayesian framework.

2. METHOD

The MPEG-7, a well known standard for description of media and its digital storage, is employed in the automatic structural segmentation of music. Much pre-processing is performed similar to that done in [1]. All music files were converted to mono and had a sampling rate of 44.1kHz. A band spacing of $1/8^{th}$ octave is used for the AudioSpectrumEnvelope audio descriptors outlined in the MPEG7 standard. The hop size described in the standard was set to the period of the beat in the song so that each sample in the spectrum corresponded to one beat. The beat was detected by the algorithm implemented in the software tool Matlab-XM (http://mpeg7.doc.gold.ac.uk/mirror/index.html). The spectrum is normalized by the L_2-norm and the dimensionality of the spectrum is reduced to 20 dimensions by principal components analysis (PCA) producing a feature vector of 21 dimensions called AudioSpectrumProjection in the MPEG-7 standard where the 21^{st} dimension is the relative power of the beat.

A hidden Markov model (HMM) is trained on the entire AudioSpectrumProjection sequence where the output of each state is modeled by a single Gaussian distribution. The Viterbi algorithm is used to determine the most likely sequence of states for the observed vector sequence. Since each sample in AudioSpectrumProjection corresponds to a beat in the music, every beat corresponds to a state produced from the HMM. As suggested by [1] the number of states denoted by N in the HMM is selected to be large, i.e. N=80, since the HMM is unable to capture structural information in terms of individual states; however, structural information can be automatically distinguished using the local distribution of states mathematically denoted by the vector $\mathbf{x}_t = [c_0, c_1, .. c_{N-1}]$ where N is the number of states in the HMM and c_0 is the count of states W=11 samples from time t where W is the size of the window for generating the local distributions.

Hence the feature space of each song consists of the tuplet of pairs $\mathbf{X} = \{\mathbf{z}_t = (t, \mathbf{x}_t) \mid 0 \le t < L\}$ where L is the length of the song. The features are clustered using a modified version of Fuzzy C-Means (FCM) clustering where time and the distributions are treated as semantically distinct features in clustering. Since the time t and the local distribution $\mathbf{x}_t$ at time t are semantically distinct features, they should be distinguished as two separate blocks of features in the clustering, namely by modifying the distance function

$$d(\mathbf{z}_1, \mathbf{z}_2) = \|\mathbf{x}_1 - \mathbf{x}_2\| + \alpha \|t_1 - t_2\| \qquad 1.$$

where $d(\mathbf{z}_1, \mathbf{z}_2)$ is the distance metric in standard FCM clustering. Just as with FCM, the number of clusters c must be specified beforehand. The update of membership values u_{ik} for i=1…c and k=1…L is identical to FCM except in the replacement of the metric $d(\mathbf{z}_1, \mathbf{z}_2)$. The update of the cluster centers is accomplished by the modified expression

$$\mathbf{v}_i = \frac{\sum_{k=1}^{L} u_{ik}^m \mathbf{z}_k}{\sum_{k=1}^{L} u_{ik}^m} \qquad 2.$$

where m is the fuzzification coefficient and $\mathbf{v}_i$ is the centroid distribution for each cluster i=1…c.

3. EXPERIMENTS

A number of experiences were conducted to evaluate the performance of the structural segmentation on some popular songs by the band Coldplay. Two evaluation criteria are used: classification rate (CR) and the f-measure (F). Precision and recall which are used to calculate the f-measure are denoted P and R respectively. Both measures are based on a sample-by-sample basis meaning that samples from each cluster are labelled according to the largest reoccurring class in each cluster and compared with the class label provided by a human expert. The number of classes was set to the number of different structural segments in the music and the number of clusters was set to the number of structural segments in the music (i.e. 1 intro, 3 verse, and 2 chorus segments gives c=6 clusters and k=3 classes). Tables 1 and 2 summarize the results.

Table 1. Results using reference distributions

Song	CR	F	P	R
Yellow (k=5)	89.0%	79.8%	80.2%	79.5%
A Message (k=4)	84.3%	74.0%	77.5%	70.9%
Fix You (k=6)	86.2%	77.2%	80.6%	74.0%
Swallowed in the Sea (k=4)	75.5%	62.6%	73.3%	54.7%
Talk (k=5)	77.4%	64.3%	60.8%	68.3%
A Rush of Blood to the Head (k=4)	89.4%	83.4%	92.4%	76.0%
In My Place (k=5)	86.0%	75.7%	77.5%	73.9%
Politik (k=6)	81.2%	68.9%	69.9%	68.0%
God Put A Smile Upon Your Face (k=6)	91.1%	83.0%	85.5%	80.6%
The Scientist (k=5)	83.1%	69.0%	67.4%	70.7%

Table 2. Results using FCM-DFS clustering

Song	CR	F	P	R
Yellow (c=10,m=1.8,a=0.1)	77.7%	70.2%	65.3%	75.9%
A Message (c=8,m=2,a=0.15)	89.2%	82.8%	88.1%	78.1%
Fix You (c=8,m=2,a=0.25)	79.8%	77.7%	71.1%	85.6%
Swallowed in the Sea (c=7,m=1.8,a=0.05)	89.7%	84.9%	79.2%	91.6%
Talk (c=8,m=1.4,a=0.7)	80.6%	68.4%	67.6%	69.1%
A Rush of Blood to the Head (c=8,m=1.5,a=0.1)	86.6%	81.0%	81.3%	80.6%
In My Place (c=9,m=1.4,a=0.1)	83.5%	71.2%	74.5%	68.1%
Politik (c=9,m=1.4,a=0.15)	88.4%	79.9%	79.0%	80.9%
God Put A Smile Upon Your Face (c=11,m=1.45,a=0.05)	71.3%	67.6%	62.6%	73.6%
The Scientist (c=8,m=2,a=1)	83.9%	73.5%	71.3%	75.9%

The results from the reference distributions are obtained by constructing reference distributions for each class that are the centroid or prototype distribution for each class as done by [1]. Distributions are labelled according to the reference distribution that is the closest in the Euclidean distance sense. The results produced by using reference distributions do not take advantage of time information. Hence, when clustering with reference distributions, the number of clusters (c) equals the number of classes (k).

With FCM-DFS, the number of clusters (c) can be greater then the number of classes (k) which accounts for the fact that there can be several chorus sections in a song. Since several choruses will be at different time instances and since we are making use of time information in clustering, each instance of the chorus is its own cluster. Varying the number of clusters (c) in clustering was not done in these experiments and will be the focus of future research. The number of clusters was fixed to be the number of segments in the song, i.e. total number of verses, choruses, etc.

4. DISCUSSION

The observations of the results are interesting as in a number of the test songs, the FCM-DFS performs better then the results from the reference distributions. Since the results from using the reference distributions do not employ the additional information about time, it is observed that FCM-DFS, which makes use of time information while clustering, is beneficial to the problem of musical segmentation.

The song "Yellow" and "God Put A Smile On Your Face" give relatively poor results compared with the other songs and compared with the technique of using reference distributions. A likely reason is that some of the structural segments are short and FCM is having trouble capturing these short segments.

The direction of future research will focus on looking at more songs and comparing FCM-DFS with current state of the art techniques in musical segmentation.

REFERENCES

[1] Levy, M., Sandler, M., (2008). Structural segmentation of music audio by constrained clustering. IEEE Trans. on Audio, Speech, and Language Processing, 16(2), 318-326.
[2] Goto, M. (2003). A chorus detecting method for musical audio signals. Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing, 437-440.
[3] Peeters, G. (2003). Deriving musical structures from signal analysis for music audio summary generation: "sequence" and state approach, in CMMR (LNCS2771) Lecture Notes in Computer Science, New York: Springer-Verlag, 142-165.
[4] Ong, B.S., Herrera, P. (2005). Semantic segmentation of music audio contents, Proc. Int. Conf. Computer Music.
[5] Abdallah, S. (2005). Theory and evaluation of a Bayesian music structure extractor, Proc. ISMIR, 420-425.

ACKNOWLEDGMENTS

Support from the NSERC and Canada Research Chair (CRC) Program (W. Pedrycz) is gratefully acknowledged. This research is supported by a post-graduate masters NSERC fellowship and an iCore fellowship (D. Graves).

MEASUREMENT AND CALCULATION OF THE PARAMETERS OF SANTUR

Peyman Heydarian and Lewis Jones
Sir John Cass Dept. of Art, Media and Design, London Metropolitan University, 41-44 Commercial Rd, London E1 1LA, UK

1. INTRODUCTION

This paper, concerns the instrument, Santur and measurement and calculation of its physical characteristics and acoustical properties such as pitch deviation and inharmonicity factor. These parameters provide a better understanding of the instrument.

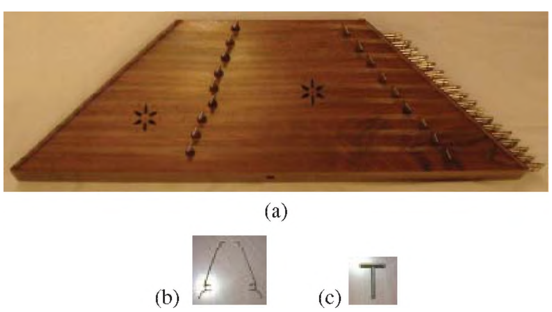
The Santur

The Santur (Fig.1-a), is a flat string instrument, which is played with a pair of hammer sticks. Santur originated in Iran and is known as a Hammered Dulcimer in English. It is a direct ancestor of Piano.

The pair of hammer sticks (Fig.1-b), are held between the index and the middle fingers and are used to hit the strings. When the notes are played, a small deflection of the strings creates a loud voice. Sticks are usually coated by a felt. The impact, makes it thinner and harder through time.

Four strings are vibrated for each note. They are stretched on the sound board, pulled between the string holders (Fig.2-a) and the tuning pegs (Fig. 2-b, Fig. 3-b), and sit on a bridge between these two ends (Fig.3-a).

The strings hinge on the left and right edges of the soundboard. The notes are adjusted by the tuning pegs, using a tuning key (Fig.1-c). It is also used as a hammer to hit the tuning pegs. The bridges are movable and can continuously change the pitch by several whole steps.

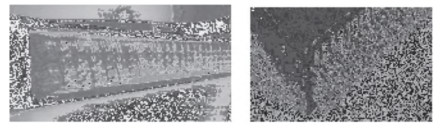


(a)

(b) (c)

Fig. 1 a) Santur b) Sticks (Mezràb) c) Tuning Key

Parallel sides of a 9-bridge[1] Salari[2] Santur, on which we did the measurements, are 90.5cm and 34.8cm. The other sides (left and right) are 38.9cm and 39.0cm. The top and the back plates[3] are 6.4cm apart. Their thicknesses are 5.5mm and 8.0mm respectively. The lengths of the four strings of a note are not exactly the same. They are between 36.8 and 37.8cm for F5-F6 and between 84.8cm and 86.25cm for C3.



(a) (b)

Fig. 2 a) String holders b) tuning pegs

Diameters of strings are between 0.35-0.36mm depending on the string's age and the tension. A bridge has a height of 2.3cm, and there is a metal roll of diameter 2.5mm on top of it (Fig. 3-a). The length of the strings between the right side of top plate and the tuning pegs (Fig. 3-b) is 2-6cm for the first and fourth strings.

(a) (b) (c)

Fig. 3 a) Bridge with a metal roll b) strings between top plate edge and tuning pegs c) Sound hole

Four pieces of flat wood along with some sound posts (rigid wooden bars) support the top plate over the back plate. They bear the force exerted by bridges, string holders and tuning pegs. The resonant body of a Santur is hollow, but the sound posts keep the instrument from collapsing. The two sound holes (Fig. 3-c) are of diameter 5cm. They influence the timbre and serve to enhance the sound quality. The tone range is: E3 (164.8 Hz)-F6 (1396.9 Hz), while the first bass note is usually tuned at C3 (130.8 Hz), instead of E3.

2. PITCH AND HARMONIC DEVIATION

Due to inharmonicities, we expect the overtones of a fundamental frequency (f_0) to move slightly upwards. The positions of overtones for a stiff string is calculated by [2]:

$$f_h = hf_0\sqrt{1+\beta h^2} \qquad (1)$$

Where, f_h is an overtone, h is harmonic index, and β is inharmonicity factor[4]. Therefore, f_0 is slightly shifted to $f_0^1 = f_0\sqrt{1+\beta}$ and Eq. 1 can be rearranged as:

$$f_h = hf_0^1\frac{\sqrt{1+\beta h^2}}{\sqrt{1+\beta}} \qquad (2)$$

Where f_0^1 is the measured fundamental frequency.

Different factors such as the thickness, length and the string tension contribute to the inharmonicities. Increasing the thickness and the tension force or decreasing the string's length result in a higher inharmonicity factor. The following equation describes the inharmonicity factor of a

[1] Modern Iranian Santurs usually have 9 bridges. 11 and 12-bridge Santurs are also prevalent [1].

[2] Master Daryoush Salari is a famous Santur producer in Iran. He has adapted innovative production techniques.

[3] Upper and lower boards

[4] β is around 0.0004 for Piano [3] and 0.00031 for Santur.

string in term of its length, l, diameter, d, and tension, T [2, 3]:

$$\beta = \frac{E\pi^3 d^4}{64 l^2 T} \quad (3)$$

E is the Young's module[5]. Young's module is the constant of elasticity of a substance.

3. RESULTS

Our dataset consists of 10 isolated samples per note. The frame size is 32768 samples and with a sampling rate of 44.1 kHz, frequency resolution becomes 1.35 Hz. The positions of f_0 and its overtones can be used to calculate the inharmonicity factor.

The pitch deviation is measured for the following notes[6]: F3, Aq3, C4, F4, Aq4, C5, F5, Ab5, C6, Eb6. It was observed that the first and second octaves are compressed by 11 and 28 cents respectively, while the third octave is stretched by 20 cents. Except an F4 sample which has a lower pitch deviation and is interpreted as miss-tuned, the bass and middle pitches tend to be less than the tempered values as we move towards higher notes, while the treble pitches tend to be more. Thus, the treble pitches on a Santur are stretched similar to Piano [4], while the bass and middle pitches are compressed in contrast.

Then, the harmonic deviation of the first 8 overtones from multiples of f_0^1 is calculated. Using Eq.2, the inharmonicity factor[7], β can be calculated in terms of the f_0^1 and overtone positions f_h:

$$\beta = \frac{(f_h / h f_0^{\,1})^2 - 1}{(h^2 - (f_h / h f_0^{\,1})^2)} \quad (4)$$

Or in terms of the h^{th} and m^{th} overtones, f_h and f_m:

$$\beta = \frac{\left(\frac{m.f_h}{h.f_m}\right)^2 - 1}{\left(h^2 - m^2.\left(\frac{m.f_h}{h.f_m}\right)^2\right)} \quad (5)$$

The inharmonicity factors of different notes are not the same. As we move towards higher notes, the inharmonicity factor increases regardless of the tone area (bass, middle and treble). The values calculated through different harmonics are also different. We will calculate the average value over different notes and different harmonics.

It should be noted that variations of the inharmonicity factor, using the first few harmonics are high. So, it is generally better to use higher harmonics in the calculations [3]. This improves the accuracy of calculations due to frequency resolution as well.

We have encountered 8 overtones here. So, calculation of the inharmonicity factor based on h8 and h4 or h5, might be a good choice to avoid using close partials. Fig. 4 shows the inharmonicity factor, calculated based on the positions of the 8th overtone and 1st to 7th overtones. The curve with considerable changes at C4, Aq4, Eb5 and Ab5 corresponds to the measurement through neighboring h8 and h7 harmonics and will be excluded from our calculations. The average value of inharmonicity factor for the 9-bridge Santur is 0.00031.

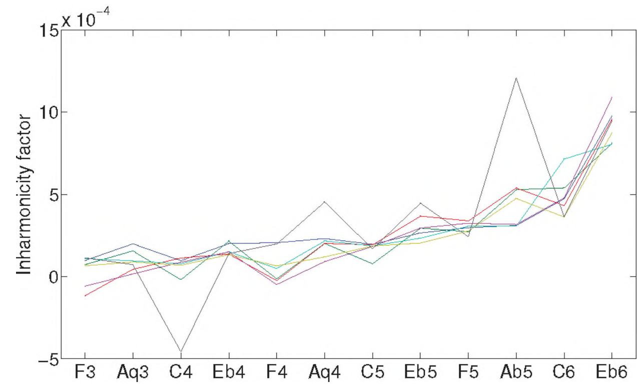


Fig.4 Inharmonicity factor based on the 8th overtone vs. 1st to 7th overtones

4. CONCLUSION

In this paper, the Santur instrument and its parameters were explained. The treble pitches on a Santur are stretched similar to the notes on a Piano, while the bass and middle pitches are compressed in contrast. The inharmonicity factor for a 9-bridge Santur is calculated, based on the average over different notes and overtones. It is 0.00031.

Future work will be on determining a more accurate value for inharmonicity factor by considering other notes, higher harmonics and analysis of the sound of other Santurs.

5. REFERENCES

[1] Heydarian, P. Reiss, J.D., "The Persian Music and the Santur Instrument", *Proceedings of the International Symposium on Music Information Retrieval*, London, UK, 2005.

[2] Fletcher, N.F., Rossing, T. D., *The physics of musical instruments*, 2nd edition, Springer–Verlag, New York, 1998.

[3] Ortiz-Berenguer, L., Casajús-Quirós, F., Torres-Guijarro, M., Beracoechea, J.A., "Piano Transcription using Pattern Recognition: Aspects of Parameter Extraction", *Proceedings of DAFx04 conference*, Napels, Italy, 2004

[4] Martin, D.W., Ward, W.D., "Subjective evaluation of musical scale temperament in pianos", *JAS*, Vol. 94, No. 1, pp. 46-53, 1993

[5] Young's module represents the ratio of stress to strain for a string or a bar of the given substance. It is the force per unit cross section of a material divided by the increase in its length resulting from the force.

[6] "q" shows half-flat and "b" shows flat [1].

[7] Here we ignore the impedance of the bridges and the sound board.

An Overview of Standards and Guidelines Influencing the Implementation of Noise Emission Declarations for Machinery

Stephen H.P. Bly, Ifaz T Haider
Healthy Environments and Consumer Safety Branch, Product Safety Directorate, Consumer and Clinical Radiation Protection Bureau, Acoustics Division, 775 Brookfield Road, Ottawa, Ontario Canada, K1A 1C1
stephen_bly@hc-sc.gc.ca

1. INTRODUCTION

The best known adverse health effect of excessive noise exposure is occupational noise-induced hearing loss (NIHL). The most reliable way of reducing the human and economic cost of excessive occupational noise exposure is to reduce the level of workplace noise. Cost effective methods include purchasing quieter machinery and planning noise control at workplace design stages. Application of these methods can be facilitated by implementing noise emission declarations for machinery (NEDM).

Prolonged use of noisy machinery (e.g., power tools) by consumers has a potential risk of NIHL, which NEDM could help mitigate.

Machinery also creates environmental noise, causing annoyance, sleep disturbance and interference with communication. NEDM can help reduce excessive environmental noise from sources such as factories, construction, or outdoor maintenance projects, thereby mitigating these health effects.

This paper provides an overview of some European, International, U.S., and Canadian standards and guidelines that influence the implementation of NEDM in Canada. Linkages to the development of a planned Health Canada (HC) Guideline: *Noise Emission Declarations for Machinery* will also be discussed.

2. EUROPEAN UNION DIRECTIVES

Noise control in the European Union (EU) is legislated by four Directives on: (i) occupational noise [1], (ii) machinery [2, 3], (iii) outdoor equipment [4, 5], and environmental noise [6]; all containing linkages to NEDM.

The Occupational Noise Directive [1] prescribes employer responsibilities to protect workers from the effects of noise on health and safety. The 2003 revision, effective 2006, increases the stringency of 1986 requirements. The energy averaged, daily exposure limit is now 87 dBA, down from 90 dBA. This Directive also recognizes NEDM's important role in occupational noise control [1].

The Machinery Directive [2, 3] gives requirements for declaration of (emission) sound pressure level and sound power level. The 2006 amendment, in force in 2009, requires a sound power level declaration if the sound pressure level exceeds 80 dBA, down from 85 dBA in the 1998 version. Also, a declaration of measurement uncertainty will be required if a non-harmonized i.e., non-ISO/CEN, measurement standard is used.

The Outdoor Equipment [4, 5] and Environmental Noise [6] Directives regulate environmental noise. The former requires certain outdoor equipment to provide sound power level declarations; a subset is also required to have limits on sound power level. The latter Directive prescribes the development and monitoring of action plans, in EU member states, for the control of environmental noise.

3. INTERNATIONAL

Supporting the EU Directives [1-6] are a noise emission declaration standard and fourteen basic measurement standards from the International Organization for Standardization (ISO) and European Committee for Standardization (CEN); nine for sound power level and five for emission sound pressure level. In addition, there are numerous test codes detailing measurement procedures specific to different classes of machinery [7].

The International Institute for Noise Control Engineering (I-INCE) has published a report: *A Global Approach to Noise Control Policy* [8]. It identifies source controls, such as purchasing quiet machinery, as a cost-effective method of noise reduction and encourages employers to seek noise-limited equipment. NEDM is needed for such actions. The report also identifies NEDM as a useful means of environmental noise control. For consumer machinery, the report suggests that consumer-friendly NEDM ought to be important for noise control.

4. U.S. STANDARDS AND GUIDELINES

The benefits of NEDM are suggested in publications by the Occupational Safety and Health Administration (OSHA) and the Mining Safety and Health Administration

(MSHA). OSHA has identified the purchase of quieter machinery as a way to significantly reduce noise exposure levels [9]. Three recent noise control guides by MSHA [10] recognize the importance of knowing the noise from mining machines, as well as the advantage of purchasing equipment with noise controls already engineered into the device. The importance of NEDM is also recognized by the National Institute for Occupational Safety and Health (NIOSH) via their online Power Tools Database for commonly used power tools in occupational settings [11].

The American National Standards Institute (ANSI) has developed standards to support NEDM, similar to the ISO/CEN standards, described above [7].

5. CANADIAN STANDARDS AND GUIDELINES

Canadian occupational noise exposure limits tend to be falling to 85 dBA with a 3 dB exchange rate. Since 2000, five provinces and one territory have adopted this limit, bringing the total to nine provinces and one territory [12, 13]. NEDM can help meet these limits. For example, an Alberta 2006 guide [14] recommends that employers: (i) target noisy equipment for replacement, (ii) set noise level criteria for new equipment purchases and (iii) request noise level specifications from manufacturers.

All of the above developments indicate the need for a Canadian NEDM standard. The Canadian Standards Association (CSA) has already fulfilled this need, publishing *Z107.58-02 Noise Emission Declarations for Machinery* [15]. This voluntary standard requires NEDM be provided in a form consistent with EU Directives [2-5]. The document also provides guidance on the use of the ISO standards for determining, providing and verifying noise specifications. CSA's *Z107.10-06 Guide for the Use of Acoustical Standards in Canada* [16] also facilitates NEDM by describing international acoustical standards recommended by CSA for use in Canada and providing guidance on their appropriate application.

To reduce potential health hazards from noise, the *Radiation Emitting Devices Act,* administered by HC, controls the sale, lease, importation and advertising of noisy machinery in Canada. Therefore, given the EU Directives [2, 4, 6] and CSA Z107.58-02, HC decided to gauge the feasibility of implementing NEDM here voluntarily, by commissioning two reports. Mixed results were obtained. Baseline data, for the extent and reliability of NEDM in Canada, appeared somewhat encouraging in one report [17]. The conclusions of the second [18], however, suggested that considerable promotion of NEDM and supporting standards would be needed for voluntary compliance to be successful.

Therefore, HC is drafting a three-part guideline, *subject to management approval*, recommending the provision and use of NEDM according to *CSA Z107.58-02* and *Z107.10-06*. Parts one and two will provide guidance to manufacturers and purchasers of machinery for the reduction of occupational and environmental noise, respectively. Part three will provide consumer advice on the use of NEDM and the importance of purchasing quieter machinery. Each of the three guidelines will contain supporting rationales pertinent to their respective foci. A three-stage stakeholder consultation is planned, one for each part, starting with the document focusing on occupational noise reduction.

REFERENCES

[1] Directive 2003/10/EC of 6 February 2003 on the minimum health and safety requirements regarding the exposure of workers to the risks arising from physical agents (noise). OJ L 42, 15.2.2003, p.38-44

[2] Directive 98/37/EC of 22 June 1998 [on the approximation of the laws of the Member States relating to] machinery. OJ L 207, 23.7.98, p. 1-46

[3] Directive 2006/42/EC of 17 May 2006 on machinery. OJ L 157, 9.6.2006, p. 24-86.

[4] Directive 2000/14/EC of 8 May 2000, […see reference 2] the noise emission in the environment by equipment for use outdoors. OJ L 162, 3.7.2000, p.1-78

[5] Directive 2005/88/EC of 14 December 2005 amending Directive 2000/14/EC. OJ L 344, 27.12.2005, p.44-46

[6] Directive 2002/49/EC of 25 June 2002 relating to the assessment and management of environmental noise OJ L 189, 18.7.2002, p. 12–25

[7] Commission communication in relation to machinery OJ C 160, 24.06.2008 p. 1-58.

[8] I-INCE (2006). A global approach to noise control and policy. Noise Control Engineering Journal. Vol 54, No. 5. 2006 Sept-Oct.

[9] OSHA (1983 accessed July 11, 2008). 29 CFR 1910.95(b)(1), Guidelines for Noise Enforcement, Appendix A.

[10] MSHA (1999). Noise Control Resource Guides. http://www.msha.gov/1999noise/noiseresources.htm

[11] NIOSH (accessed July 10, 2008). Power Tools Database. http://www.cdc.gov/niosh/topics/noise/workplacesolutions/toolsdatabase.html

[12] Canadian Center for Occupational Health and Safety (CCOHS) (2007). Noise – Occupational Exposure Limits in Canada. http://www.ccohs.ca/oshanswers/phys_agents/exposure_can.html

[13] CCOHS (2000). Noise Control in Industry: A Basic Guide. Canadian Center for Occupational Health and Safety: Hamilton, Ontario.

[14] Government of Alberta (2006). Occupational Health and Safety Code 2006 Explanation Guide - Part 16: Noise. http://employment.alberta.ca/documents/WHS/WHS-LEG_ohsc_2006_p16.pdf

[15] CSA (2002). Z107.58-02 Noise Emission Declarations for Machinery. Canadian Standards Association: Mississauga, Ontario

[16] CSA (2006). Z107.10-06 Guide for the Use of Acoustical Standards in Canada. Canadian Standards Association: Mississauga, Ontario

[17] CCOHS (2003) A pilot survey on the availability of noise emission data for equipment and machinery sold in Canada. Final Report, June 23, 2003.

[18] HGC (2004) An experience - based investigation of machinery noise declarations in Canada. HGC Engineering. March 31, 2004.

ISO/IEC GUM APPLIED TO ESTIMATION OF SOUND POWER MEASUREMENT UNCERTAINTIES

Stephen E. Keith
Healthy Environments and Consumer Safety Branch, Product Safety Directorate, Consumer and Clinical Radiation Protection Bureau, Acoustics Division, 775 Brookfield Rd. 6301B, Ottawa, Ontario Canada, K1A 1C1 skeith@hc-sc.gc.ca

1. INTRODUCTION

Over the last 40 years there have been significant improvements in ISO machinery sound power measurement standards using sound pressure. However, measurement uncertainties have essentially stayed the same, representing typical conservative situations. This results in users not reporting these uncertainties because they are too large.

In reality, the ISO standards can under or over estimate measurement uncertainty. The same uncertainty could be assumed regardless of whether the measurement is in a quiet abandoned airstrip, or a noisy crowded outdoor construction site. Using information specific to a given measurement allows a better estimate of uncertainty. This paper implements the methods of the *ISO/IEC Guide to the expression of uncertainty in measurement* [1] (GUM) and the ISO engineering grade sound power standard [2] to show uncertainty calculations for a specific measurement.

2. METHOD

A functional relationship for sound power, L_W, is:

$$L_W = \overline{L_p} + 10\lg\frac{S}{S_0} - K_1 - K_2 + \delta_{\text{mic}} + \delta_{\text{angle}} + \delta_{\text{other}}$$

where $\overline{L_p}$ is the mean sound pressure over a measurement surface with area, S (S_0 = 1 m), and K_2 is an environmental correction equal to the difference between the measured sound level versus the level that would be measured in a hemi free field. The δ 's are additional uncertainties, δ_{mic} is due to sampling, δ_{angle} is due to the difference between intensity and pressure and δ_{other} is due to other factors such as instrumentation or method. K_1 is a background noise correction given by:

$K_1 = -10\lg\left(1-10^{-0,1\left(\overline{L_p}-\overline{L_{p(B)}}\right)}\right)$, where $\overline{L_{p(B)}}$ is the mean background noise level averaged over the measurement surface.

Table 1: Sensitivity coefficients and standard uncertainties, where s with subscript is the standard deviation of the quantity in the subscript, and n refers to the number of measurement points.

variable	sensitivity coefficient, c_i	standard uncertainty, u_i
$\overline{L_p}$	$1+1\Big/\left(10^{0,1\left(\overline{L_p}-\overline{L_{p(B)}}\right)}-1\right)$	$s_{\overline{L_p}}$
S	see [2]	
K_1	$-1\Big/\left(10^{0,1\left(\overline{L_p}-\overline{L_{p(B)}}\right)}-1\right)$	$s_{\overline{L_{p(B)}}}$
K_2	1	$K_2/4$
δ_{mic}	1	$s_{L_{p_i}}\Big/\sqrt{n}$
δ_{angle}	$10^{-K_2/10}$	u_{angle} (Fig. 2)
δ_{other}	see [2]	

For each variable, the uncertainty contribution to the sound power, L_W, is $uc_i = u_i \cdot c_i$ where u_i is the standard uncertainty and c_i is a sensitivity coefficient, (Table 1). Sensitivity coefficients are partial derivatives of L_W. The combined standard uncertainty is the summation in quadrature of all uncertainty contributions to L_W.

3. DISCUSSION

Fig. 1 shows the standard uncertainty contributions uc_i for $\overline{L_p}$ and K_1 for an extreme case when the background noise measurement standard deviation, $s_{\overline{L_{p(B)}}}$, is 3 dB. Each uncertainty contribution is calculated from the measurement standard deviation, $s_{\overline{L_p}}$, and sensitivity coefficient, as in Table 1. The bulk of this uncertainty is due to the K_1 correction of the measured levels to account

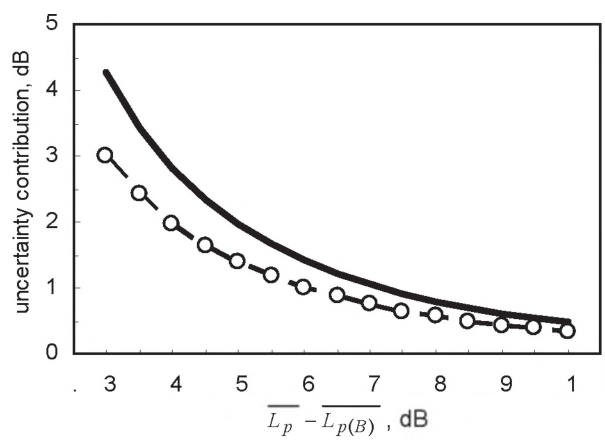


Fig. 1: Standard uncertainty contributions to L_W for an ideal stable source when background noise standard deviation, $s_{\overline{L_{p(B)}}}$, is 3 dB: background noise contribution uc_{K1} (dashed line); source contribution uc_{Lp} (open circles); and their combined standard uncertainty (solid line).

for background noise. For example, if the background noise is not stable, then there is more uncertainty in this correction. Similarly, if the measured noise source levels vary, then the K_1 correction will also introduce uncertainty. In this figure the source was ideally stable (i.e., a reference sound source). However, even measured levels from an ideal source would not be exactly repeatable due to the influence of the background noise on the measurement.

Fig. 1 assumed background noise with a standard deviation, $s_{\overline{L_{p(B)}}}$, of 3 dB to make the combined uncertainty consistent with the standard deviation of reproducibility in the ISO standard [2]. This engineering grade standard has $\overline{L_p} - \overline{L_{p(B)}}$ limited to 6 dB and the associated standard deviation of reproducibility is 1.5 dB, the same as the combined standard uncertainty in Fig 1. Using this standard in an outdoor construction site, background noise standard deviation, $s_{\overline{L_{p(B)}}}$, could easily exceed 3 dB, and the standard uncertainty could exceed the published standard deviation of reproducibility. The situation is similar in other ISO precision and survey grade standards.

ISO also puts a limit on the decibel range of measured levels. For engineering grade, this range is equal to the number of measurement points. Using 10 measurement points a worst case for δ_{mic} occurs when the measured levels are equally split and lie at the upper and lower limits of the allowable range. This situation could occur on a machine with a shielded operator area. This results in $s_{L_{p_i}}$ =5.3 dB and, from Table 1, a standard uncertainty contribution of 1.7 dB. This exceeds the 1.5 dB standard deviation of reproducibility in the standard. A similar situation exists in the precision grade standard.

Measurements close to the source affect δ_{angle} due to an overestimate of sound power, which is proportional to the average component of intensity normal to the measurement surface. Sound pressure obtained on a very large planar measurement surface very close to a very large piston can give a good estimate of sound power, since the sound pressure is equal to the normal component of sound intensity (when both are measured in decibels). However, if the sound was to originate from a point source at the centre of the piston, sound pressure would overestimate the sound power since the direction of the sound intensity at the edges of the measurement surface would be almost parallel to the surface, and the desired normal component would be much smaller. When these two situations are extended to an enclosing measurement surface, the range of uncertainty is given in Fig. 2 as a function of S/d^2, the ratio of measurement surface area S to the distance, d, between the measurement surface and the source. The shaded area represents the range of values obtained with an infinite number of measurement points on a box shaped measurement surface. The worst case for a cubic 9 point measurement surface from the engineering grade standard falls somewhat above this shaded range, i.e., S/d^2 is 45 and the standard uncertainty is 2.3 dB (for a machine with a point source in the middle of the top edge). Table 1 shows when the environmental correction, K_2, is zero (such as in a hemi anechoic space) the worst case sensitivity coefficient is 1, with the resulting worst case uncertainty contribution of 2.3 dB. For relatively small sources a hemisphere can be used, which significantly reduces this error, as shown by the hatched area in Fig 2.

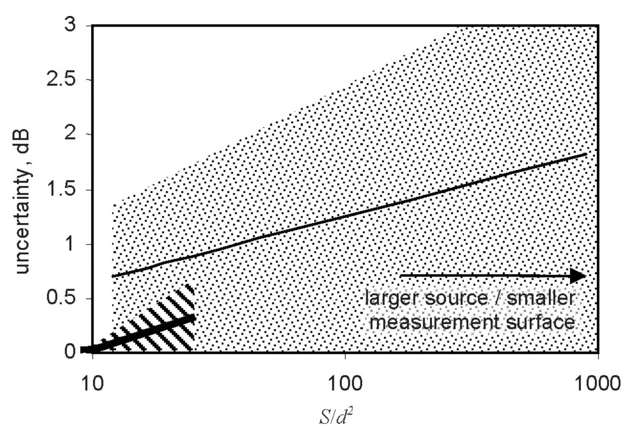


Fig. 2: Standard uncertainty u_{angle} due to approximation of intensity level using sound pressure. Shaded area shows range for infinite number of points over a box measurement surface, where thin line is an average value taken from [3]. The hatched area shows a similar range for a hemisphere measurement surface, with the thick line an average value.

In contrast, the best case scenario for any of the above uncertainty contributions approaches zero dB. Uncertainties in Fig 1 can be reduced by controlling background noise, longer averaging times, or measurement closer to the source. δ_{mic} is reduced by increasing the number of microphone positions, measurement farther from the source or an increase in reverberation time. δ_{angle} is reduced by increasing reverberation time, or a larger preferably hemispherical measurement surface.

4. CONCLUSION

This paper has shown that existing standards can underestimate or overestimate uncertainties. This is entirely appropriate since basic statistics tell us that 5% of measurement situations should differ from the mean by more than 2 standard deviations. However, 20% of measurement situations will be within ¼ of a standard deviation from the mean. Given that it is possible to identify these situations, it seems inappropriate to base uncertainties on a somewhat conservative general case. Using information specific to a given measurement allows a better estimate of uncertainty.

REFERENCES

[1] ISO/IEC Guide 98-3 (1995) Uncertainty in measurement - Part 3: Guide to the expression of uncertainty in measurement

[2] ISO 3744 (1994) Acoustics - Determination of sound power levels of noise sources using sound pressure - Engineering method in an essentially free field over a reflecting plane

[3]. Probst, W. (1999) Checking of sound emission values. BAuA report, Fb 851, ISBN 3-89701-375-4, Dortmund

TESTING OF HEALTH CANADA'S ACOUSTIC CHAMBER AT THE CONSUMER AND CLINICAL RADIATION PROTECTION BUREAU BASED ON ISO STANDARD 3745:2003

Jason T. Tsang[1], Stephen E. Keith[1], William J. Gastmeier[2]

[1] Healthy Environments and Consumer Safety Branch, Product Safety Directorate, Consumer and Clinical Radiation Protection Bureau, Acoustics Division, 775 Brookfield Road, Ottawa, Ontario Canada, K1A 1C1

[2]HGC Engineering, 2000 Argentia Road, Plaza 1, Suite 203, Mississauga, Ontario, Canada, L5N 1P7

1. INTRODUCTION

Third party acoustical certification testing was performed in Health Canada's acoustic chamber at the Consumer and Clinical Radiation Protection Bureau in Ottawa, Ontario. Testing was performed prior to installation of a fire suppression system to document baseline acoustical conditions. The installation required considerable modification to the chamber, including drilling holes through both acoustical absorbing wedges and the ceiling. Also sprinkler heads were installed at the edge of some wedges.

Tests were made using the pure tone method prescribed in Annex A of ISO 3745 (2003) (Standard) [1]. This paper describes the results of the free field performance testing and discusses the characteristics of the sound sources.

1.1 Chamber Description

The chamber is constructed as a room within a room, with concrete inner and outer walls, 25 cm and 15 cm thick, respectively, separated by a 127 cm air gap. The inner room sits on 56 vibration isolating springs mounted on rubber pads. The springs have a resonance frequency for rocking oscillations of 1.5 Hz. The inner chamber weighs approximately 1000 tons. The walls, floor, and ceiling are lined with flat-tipped fibreglass wedges designed for a cut-off frequency of 50 Hz. The interior (wedge tip to wedge tip) is approximately 13 m long, 9 m wide and 8 m high. The chamber is anechoic in design with removable concrete floor tiles to convert to a hemi-anechoic chamber. A B&K 9654 robot is suspended from the chamber roof at a height of approximately 5.3 m above the floor. The robot consists of stepper motors travelling along a frame of 5 cm square tubing.

2. METHOD

Measurements were conducted with a Gras Model 40 BE 1/4 inch microphone and Model 26CB preamplifier connected to a Bruel & Kjaer Type 2260 Precision Integrating Sound Level Meter.

Two sound sources were used to cover the design frequency range of the chamber (50 Hz to 10 kHz): a dynamic loud speaker and a compression driver. The sound sources were positioned in a hole in the reflecting floor near the centre of the chamber. The compression driver was mounted to the underside of the concrete tile floor through a 5/16 inch hole to improve its directional properties. The source outlets were flush with the floor surface to eliminate potential image sources.

The signal was supplied from a Dell Precision M65 laptop through a Tascam GE20B equalizer and a Stewart PA50B power amplifier. The measured signal was at least 20 dB greater than the background noise. Measurements were recorded for 7 to 10 seconds at each position.

Measurements were taken along six radial traverses, extending 0.5 m from the effective acoustic centre of the source to 0.5 m distance to the walls, ceiling, and corner. Traverses were performed using pure-tone signals at 1/3 octave band centre frequencies. The microphone was moved at 10 cm intervals.

For each traverse measurements, source strength [1] was calculated from:

$$a = \frac{Nr_0^{\,2} + \sum_{i=1}^{N} r_i^{\,2} - 2r_0 \sum_{i=1}^{N} r_i}{\sum_{i=1}^{N} r_i q_i - r_0 \sum_{i=1}^{N} q_i} \qquad \textbf{(1)}$$

where, for each traverse, $q_i = 10^{-0.05Lpi}$, Lpi is the sound pressure level at the ith measurement position in decibels, r_i is the distance of the ith measurement position from the centre of the measurement hemisphere.

The sound source must meet the requirements in Table 1 to minimize the uncertainties in chamber characterization which can arise from a directional source. Ideally the test source should behave like a point source with a uniform sound distribution [1].

Table 1. Allowable deviation in directionality of the test source

One-third-octave band frequency (Hz)	Allowable deviations in directionality
≤ 630	± 2.0
800 to 5,000	± 2.5
6,300 to 10,000	± 3.0

Qualification of the chamber requires that the differences between the measured and theoretical levels do not exceed the values in Table 2 [1]. The theoretical free-field decay of sound, from a point source, follows the inverse square law of 6.0 dB per doubling of distance. Reflection and scattering effects from the wedges or any other surface in the room will yield deviations from the inverse square law.

Table 2. Maximum allowable difference in hemi-anechoic rooms between measured and theoretical free-field levels

One-third-octave band frequency (Hz)	Allowable difference (dB)
≤ 630	± 2.5
800 to 5,000	± 2.0
> 6,300	± 3.0

3. RESULTS

3.1 Sound Source Directionality Check

As a check on source directionality, for each traverse, source strength was calculated from Eq. (1), for points between 0.5 m and 1.5 m, assuming r_0=0. Figure 1 shows the maximum deviation in source strength (half the difference between maximum and minimum source strength levels), across all traverses, in each frequency band of interest. These deviations are compared to the deviations limits in Table 1.

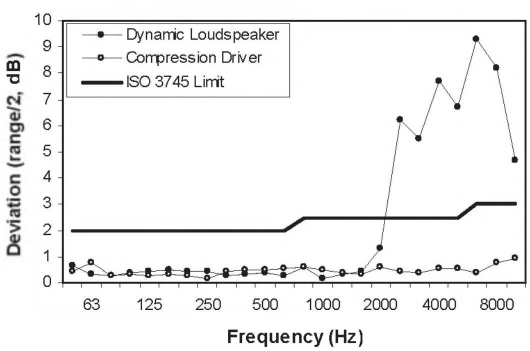


Fig. 1: Measured deviations (range/2) in directivity for the dynamic loudspeaker and compression driver compared to ISO 3745 limits for a hemi-anechoic test room.

Based on the results in Figure 1, it is evident that the dynamic loudspeaker directivity effects are more prevalent at the higher frequencies, above 2 kHz. Deviations are less than 2 dB for frequencies 2 kHz and below. Above 2 kHz deviations continue to climb to 9 dB at 6300 Hz and then decline to 5 dB at 10 kHz.

Directional effects of the compression driver are subtle, with deviations less than 1 dB over most of the frequency range of interest. At 8 kHz and 10 kHz, deviations increase to approximately 1 dB. Below 500 Hz, the compression driver does not yield sound pressure levels high enough to meet the signal to noise requirements of the Standard.

Drawing results from both speakers, the maximum deviation from omnidirectionality is less than 2 dB for all frequencies of interest (50 Hz to 10 kHz). The dynamic loudspeaker was used to cover the frequency range from 50 Hz to 1 kHz, and the compression driver was used for the frequency range from 1.25 kHz to 10 kHz

3.2 Chamber Qualification

The measured and theoretical levels met the allowable limits shown in Table 2, with the exception of the upward traverses near the robot frame, in the higher frequency bands. The deviation from theoretical levels was likely due to reflections off of the robot frame. Limits in the higher frequency bands were exceeded by up to 4 dB, 10 cm from the robot frame. Measurement results are shown in Figure 2 for 50 Hz and 10 kHz.

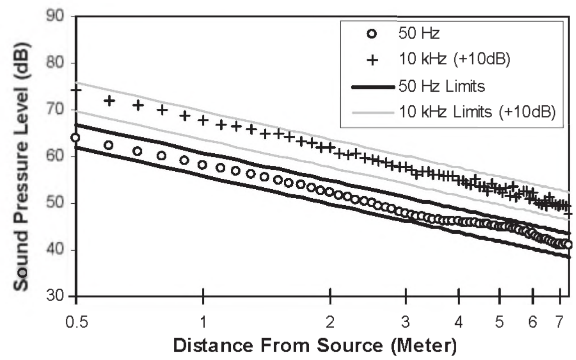


Fig. 2: Measurement results for Up and Over Traverse at 50 Hz and 10 kHz.

4. CONCLUSIONS

Two sound sources were used to cover the design frequency range of the chamber, 50 Hz to 10 kHz. The sources were in accordance with the directionality and signal to noise requirements of ISO 3745, within the respective frequency ranges used.

These tests showed that, over its useable volume, the chamber complied with the ISO 3745 pure tone option for chamber characterization in all 1/3 Octave bands over the design frequency range of 50 Hz to 10 kHz.

These tests were performed pre-installation of a fire suppression system. Testing is planned in the future to determine if the installation has caused any changes in the acoustical conditions of the chamber.

REFERENCE

[1] ISO 3745 (2003). Acoustics - Determination of sound power levels of noise sources using sound pressure - Precision methods for anechoic and hemi-anechoic rooms. International Organization for Standardization.

The ANSI Standard S12.68-2007 Method of Estimating Effective A-Weighted Sound Pressure Levels When Hearing Protectors are Worn: A Canadian Perspective

Jérémie Voix[1], Alberto Behar[2], Willy Wong[2]
[1] Sonomax Hearing Healthcare Inc, Montréal (QC) H4P 2E2, Canada, jvoix@sonomax.com
[2]Department of Electrical and Computer Engineering, University of Toronto, 164 College Street, Toronto, ON M5S 3G4, Canada , alberto.behar@utoronto.ca; willy@eecg.utoronto.ca

1. INTRODUCTION

The most serious and difficult issue with hearing protection devices (HPD) is to estimate the protected noise level at the wearer's ear. Such estimation is difficult for two reasons. First, the effective field attenuation of an HPD on a given user is rarely known. Instead, the Noise Reduction Rating (NRR), specified in the USA by Code of Federal Regulations CFR 40 Part 211, that represents the 98th percentile of the group attenuation of test-subjects tested in a laboratory, is currently used. Such use of the NRR provides a highly unrealistic assessment, not only because the NRR value dramatically overestimates the attenuation with respect to real-life situations, but also because the NRR does not reflect the inter-subject variability in terms of attenuation for a given HPD. Second, is the issue associated with the practical use of the NRR is that it requires the C-weighted exposure level, which is rarely available in practice, and assumes that the noise spectrum is flat per octave-band (pink noise).

Finally, although the Canadian Standards for HPD [1] is referring to classes and grades to assist the end-user with the proper selection of HPD, it should be noted that these classification are based on the average attenuation values achieved by the HPD in laboratory. Hence, they are affected in a similar way by the lack of realistic laboratory attenuation data (that are to be measured using ANSI S3.19 [2] or ANSI S12.6 [3]) and furthermore, it can be easily demonstrated that a very minor change (less than one dB) of the average attenuation at only one frequency band can change the HPD from one class or grade to another. This demonstrates the weakness of such classification scheme and the need for the hearing conservationist to have access to a single number rating of the attenuation of a given HPD..

For the reasons mentioned above, the current NRR is a poor indicator ("not really relevant") of the attenuation that a worker will experience in the field. Such inadequacy of the NRR is not new [4] and various attempt have been made in the past to either come up with more realistic attenuation data (like in the Method B of ANSI S12.6-1997 that describe an experimental protocol on naïve subjects that would fit the HPD themselves, therefore hopefully providing more realistic attenuation values) or like in the past attempt to promote a "derating" rule [5] that would be applied on the laboratory data in order to lower the attenuation values and hopefully obtain values that would be more realistic. None of these approach have been found to be satisfactory and the ANSI standard S12.68 [6] has been recently developed to specifically address these the issues of the inter-subject-fit variability and noise-spectrum content variability previously mentioned. The standards presents 3 methods to estimate from laboratory attenuation data, the sound pressure levels when HPD are worn. The first one is the use of the Noise Reduction Level Statistics, that will be further explained in this paper. The second one is a graphical representation of the Noise Reduction Level as a function of the C-A values of the exposure level. The third one is the use of the exact "octave-band" method. These last two methods won't be detailed here, but have already been vulgarized [7].

2. A new proposal: the Noise Reduction Level Statistics, NRS_A

A substantial divergence in this standard from prior publications and other standards already cited is the recommendation that the simplified ratings be presented as pairs of numbers at the 80th and 20th percentile level. Furthermore, an exhaustive set of 100 actual industrial spectra, denoted "NIOSH 100", will be used for the computation of this single number rating rather than only 8 spectra, as in the previous section.

$$\Delta L_{A_{pn}} = 10 \cdot \log_{10}\left(\sum_{i=1}^{7} 10^{\frac{L_n^i + A^i}{10}}\right) - 10 \cdot \log_{10}\left(\sum_{i=1}^{7} 10^{\frac{L_n^i + A^i - REAT_p^i}{10}}\right) \quad (1)$$

where $L^i n$ is the sound pressure level in decibels for the octave centered on i for the n th noise in an industrial noises database; $REAT^i{}_p$ is the attenuation in decibels measured for the hearing protector on the p th subject at octave-band center frequency i, averaged across several trials (usually 2, as in the ANSI S12.6).

The NRS_{Ax} is defined as :

$$NRS_{Ax} = m - \alpha \cdot \sqrt{s^2_{subject} + s^2_{spectrum}} \quad (2)$$

where m is the average attenuation across subjects across spectrum, obtained as:

$$m = \frac{1}{P \cdot N} \sum_{n=1}^{N} \sum_{p=1}^{P} \Delta L_{A_{pn}} \tag{3}$$

and the standard deviations $s_{subject}$ and $s_{spectrum}$ are respectively -classically- defined by:

$$s_{subject} = \sqrt{\frac{1}{P-1} \sum_{p=1}^{P} (m_p - m)^2} \tag{4}$$

and

$$s_{spectrum} = \sqrt{\frac{1}{N-1} \sum_{n=1}^{N} (m_n - m)^2} \tag{5}$$

The effective A-weighted sound pressure level for protection performance x percent is computed by subtracting the NRS_{Ax} from the A-weighed exposure level, using the following equation:

$$L'_{Ax} = L_A - NRS_{Ax} \tag{6}$$

The attenuation values to be used in the calculation are not specified and both the use of Method B ("subject-fit", already described, where naïve or inexperiences subject are used) attenuation data or the use of Method A ("supervised-fit", where the testing laboratory would supervise the HPD insertion and fit) are possible. The NRS_A, is finally expressed at the 20th and 80th percentile values, in order to reflect both this variability's on the noise spectrum content and on the change in the attenuation from individual to individual. It is also expected that theses 2 percentile values would respectively represent the attenuation that "most individually trained users to achieve or exceed" and that "a few motivated proficient users to achieve or exceed."

3. NRS_A: some concerns for use in Canada?

Although the NRS is much more easy to use and also more realistic, its direct use in the Canadian workplaces may require some further investigations on three aspects. The first aspect is the representativity for the Canadian workplaces of the NIOSH 100 spectrum used: although a comprehensive analysis has been conducted by Berger and Gauger [8], it is not clear how the relative weight of the industrial workplaces has been taken into account in the database used. At least five spectrum are repeated more than once in that dataset, and may be intentionally or not, making some spectrum more preponderant. Furthermore it could be argued that some Canadian industrial areas (like the mining, wood and forestry industries) may be significantly underrepresented in that dataset [9].

The second issue associated with the use of the NRSA is that the attenuation values to be used in that calculation are not specified and both the use of Method B ("subject-fit", already described, where naive or inexperiences subject are used) attenuation data or the use of Method A ("supervised-fit", where the testing laboratory would supervise the HPD insertion and fit) are possible.

The third issue is related to the fact that field validation are still missing to ascertain what percentile should be used to give a realistic assessment of the field attenuation values that a group of user can achieve. Such value was assumed to be 80%, but the foreseen use of "Method A" attenuation data for the calculation of the NRSA urged some authors to ask for consideration of other percentile value (such as the 90% value discussed among the ANSI S12 working group WG11). A field validation in the Canadian workplace is definitely required and would consist in the statistical comparison of field attenuation data to NRS_{Ax} expressed at various percentiles x.

4. CONCLUSION

The recent developments of ANSI S12.68 is certainly very good news for the Canadian hearing conservationist: it provides 3 different practical tools to estimate from laboratory data the attenuation that user may achieve in the field, with a built-in uncertainty that accounts for the inter-subject-fit variability and noise-spectrum content variability. Its use and reference by the Canadian standards could also be considered, after several validation have been successfully conducted.

REFERENCES

[1] *CSA Z94.2-02, Hearing Protection Devices - Performance, Selection, Care, and Use*, 2002.

[2] "ANSI S3.19-1974 (R 1990), American National Standard Method for the Measurement of Real-Ear Protection of Hearing Protectors and Physical Attenuation of Earmuffs," 1974.

[3] ANSI, *ANSI S12.60-2002 American National Standard Acoustical Performance Criteria, Design Requirements, and Guidelines for Schools.*

[4] E. Berger et al., *The Noise Manual*, American Industrial Hygiene Association (AIHA) Press, 2000.

[5] NIOSH, "Method for Calculating and Using the Noise Reduction Rating -NRR"; http://www2a.cdc.gov/hp-devices/pdfs/calculation.pdf.

[6] A. Accredited Standards Committee S1, *ANSI S12.68-2007 AMERICAN NATIONAL STANDARD Methods of Estimating Effective A-Weighted Sound Pressure Levels When Hearing Protectors are Worn* , 2007.

[7] W.J. Murphy, "How to assess hearing protection effectiveness: what is new in ANSI/ASA S12.68," *Acoustics Today*, pp. pp.40 - 42.

[8] D. Gauger and E.H. Berger, *A New Hearing Protector Rating: The Noise Reduction Statistic for Use with A Weighting (NRSA)*, 2004.

[9] J. Voix, F. Laville, and J. Zeidan, "Filter Selection to Adapt Earplug Performances to Sound Exposure," *Journal of the Canadian Acoustical Association*, vol. 30, 2002, p. 122.

CSA APPENDIX ON MEASUREMENT OF NOISE EXPOSURE FROM HEADSETS

Alberto Behar[1], Christian Giguère[2], and Tim Kelsall[3]
[1]Institute of Biomaterials and Biomedical Eng., University of Toronto, Toronto (ON) M5S 3G9
[2]School of Rehabilitation Sciences, University of Ottawa, Ottawa (ON) K1H 8M5
[3]Hatch Consulting Engineers, 2800 Speakman Drive, Mississauga (ON) L5K 2R7

1. INTRODUCTION

Several national and international standards, such as CSA Standard Z107.56 [1], provide procedures for the measurement of occupational noise exposure. They concentrate on situations where sources are far from the worker's ears. In some situations, the sound source is close to or even occludes the ears (e.g., headsets) and measurements have to be performed using different equipment and techniques. A new appendix to CSA Standard Z107.56 is currently being developed to cover measurements of noise exposure from employees wearing headsets for communication. This paper reviews different assessment methods for this application.

Standardized methods to measure sound levels directly under the device include the use of acoustic manikins, artificial ears and real-ear procedures [2-4]. For the final assessment, occluded-ear measurements must be converted to equivalent far-field levels. This conversion is required to compare the noise exposure under the occluded ear to the applicable regulatory limit (e.g. 85 dBA), the latter being referred to the exposure at the position of the worker and not inside the ear.

An alternative indirect calculation method is proposed here that includes the main determinants of exposure as input parameters into the assessment, such as the background noise around the worker, the attenuation of the device and the expected signal-to-noise ratio under the device. This method has the advantage that it facilitates the implementation of solutions or treatments to reduce exposure.

2. EXISTING MEASUREMENT STANDARDS FOR HEADSETS

2.1 ISO 11904

ISO 11904 describes a set of two related standards for the measurement of sound levels from sources located close to the ear.

ISO 11904-1 specifies acoustic measurements in the real ears of human subjects using miniature or probe microphones (MIRE technique) [2]. Acoustic measurements must be performed in 1/3-octave bands and transformed to equivalent A-weighted free or diffuse sound levels. The measurement microphone or probe can be located anywhere from the ear canal entrance to the eardrum, under open or blocked ear canal conditions. Transfer functions are provided to transform measurements to free or diffuse-field equivalent sound levels for predefined microphone measurement locations. A main advantage of this method is that it provides the most direct estimate of sound exposure for the worker. There is no need for a duplicate matched headset/headphone or modifications to the electrical connections to the device. The main disadvantages are that the method is invasive and may restrict head and body movements, and thus is difficult to implement in a real workplace for a sustained period of time. Sound leakages are also possible due to electrical wires or flexible tubing breaking the seal of the device against the ear or head, and flanking background noise through the flexible tubing outside the device may be an issue in some cases.

ISO 11904-2 specifies sound measurements on a manikin modeling the mechanical parameters and acoustical effects of the human ear, head and torso (manikin technique) [3]. The microphone is located in the ear simulator of the manikin. Acoustic measurements must again be performed in 1/3-octave bands and transformed to equivalent A-weighted free or diffuse sound levels. A main advantage of this method is that it is not invasive. The main disadvantage is that it typically requires parallel measurements using separate matched headsets for worker and manikin. Also, the equipment is not widely available and can be cumbersome to use in the workplace. Another limiting factor may be the difficulty to fit or couple the device to the pinna simulator and ear canal extension in a realistic manner owing to the different shape and mechanical properties between the manikin and ears of the workers under study.

2.2 AS/NSZ 1269.1

The Australian/New Zealand Standard AS/NZS 1269.1 contains Appendix C: "*Recommended procedures for measurement of sound pressure levels from headphones or insert earphones*" [4]. It describes several measurement methods. The primary method requires that an identical type of headphone or earphone (with similar response characteristics) be connected in parallel to the signal source (with proper matching impedance network) used for the headphone or earphone of the worker. This additional headphone is applied to a wide-band artificial ear or an acoustic manikin in the case of headphones, or to an occluded-ear simulator in the case of insert earphones. The headphone signal applied to the measuring device is deemed identical to the one applied to the worker.

The advantage of the AS/NZS 1269.1 method is that it allows using an artificial ear or occluded-ear simulator that is inexpensive, easy for transportation and to be used in the workplace. However, there are several issues that largely affect the accuracy of the measurements. Also, the authors are not aware of a study that compares measurement results obtained using the manikin with the wide-band artificial ear or occluded-ear simulator.

3. CALCULATION METHOD

The calculation method provides a simpler approach which can be carried out by an industrial hygienist or safety officer using the same equipment used to measure noise exposure. In most cases, the listener adjusts the sound level under the device to be able to communicate properly. The signal-to-noise ratio (SNR) is usually set at around 10-15 dB. This fact provides another way of assessing the noise level at the ear, by measuring the room background noise, subtracting the attenuation of the headset (if there is any) and then correcting for the expected SNR and speech signal duration [5].

In practice, the measurement procedures are the same as used for employees without headsets. For a regulated limit of 85 dBA, this would mean that the combination of the background noise coming through the headset and the expected noise produced by the headset signal should be no louder than a sound-field equivalent level of 85 dBA. Most headsets provide little or no protection against external noise. Accordingly, the noise reduction of the headset is assumed to be zero unless the manufacturer can provide user fit octave band attenuation data. The calculation must also account for the duration the headset signal is ON.

An example of a simple calculation is given in Table 1, where the headset attenuation is assumed to be zero. Note that unless the use of the headset is extremely intermittent, the L_{ex} from the noise inside the headset is much lower than the signal from the headset. If the headset is used more than 1 hour per day, the background noise has less than 1 dB effect on the result. In such cases, the exposure under the headset can be calculated by simply adding 15 dB to the L_{ex} measured outside the headset (corrected for headset signal duration) and subtracting the NR of the headset (which is zero for most general purpose headsets). Thus, the L_{ex} under the headset equals the L_{ex} outside the headset (corrected for duration) plus 15 dB, provided the headset is used for more than 1 hour per day and no standardized attenuation data is available for the headset, which covers most applications.

Table 1. Example exposure calculation

	SL (dBA)	Duration (hr)
Room background noise level	70	8
Headset attenuation (NR)	0	
Background noise under headset	70	
Headset signal level when ON	85	
Hours headset signal is ON		1
Hours headset signal is OFF		7
Lex from background noise	**70**	**8**
Lex from headset signal	**76**	**8**
Total Lex	**77**	**8**

4. CSA WG

The Canadian Standard Association (CSA) has produced Standard Z107.56 "*Method for the measurement of occupational noise exposure*" [1]. It does not contain provisions for the measurement of noise from sources located close to the ear of the exposed person. In 2007, an ad hoc Working Group (WG) was created to prepare an Annex to the Standard that will deal specifically with the measurement of noise exposure from headsets. Members of the WG, selected from a broad range of practitioners and academia, are as follows:

Alberto Behar – University of Toronto
Art Thansandote – Health Canada
Christian Giguère – University of Ottawa
Christine Harrison – WorkSafeBC
Hans Kunov – University of Toronto
Hilmi Dajani – University of Ottawa
Joe Principato – RCMP/GRC
Marshall Chasin – Musician Clinic
Michael Sharpe – HCCA Ltd.
Stephen Keith – Health Canada
Warwick Williams – National Acoustic Labs (Australia)

It is expected that a first draft of the Annex will be prepared before the end of 2008. The authors of the present paper acknowledge the contributions by all members of the WG.

REFERENCES

[1] CSA Z107.56, *Procedures for the Measurement of Occupational Noise Exposure* (Canadian Standards Association, 2006).

[2] ISO 11904-1, *Acoustics – Determination of sound immission from sound sources placed close to the ear: Part 1: Technique using a microphone in a real ear (MIRE technique)* (International Organization for Standardization, 2002).

[3] ISO 11904-2, *Acoustics – Determination of sound immission from sound sources placed close to the ear: Part 2: Technique using a manikin (manikin technique)* (International Organization for Standardization, 2004).

[4] AS/NZS 1269-1, *Occupational noise management Part 1: Measurement and assessment of noise immission and exposure* (Standards Australia and Standards New Zealand, 2005).

[5] Giguère and C. Laroche, "Hearing loss prevention program in the military environment," *Canadian Acoustics* **33**(4), 21-30 (2005).

NOISE EXPOSURE FROM COMMUNICATIONS HEADSETS: THE EFFECTS OF ENVIRONMENTAL NOISE, ATTENUATION AND SNR UNDER THE DEVICE

Christian Giguère[1] and Hilmi R. Dajani[2]
[1]School of Rehabilitation Sciences, University of Ottawa, Ottawa (ON) K1H 8M5
[2]School of Information Technology and Engineering, University of Ottawa, Ottawa (ON) K1N 6N5

1. INTRODUCTION

A Working Group (WG) was created to prepare an extension to CSA Standard Z107.56 [1] to cover situations when the worker is wearing communication headsets or other listening devices placed close to the ear. Currently, the CSA standard describes noise measurement and assessment procedures that are restricted to situations where the sources are far from the ears. With the increasing use of headsets in the workplace, these general procedures are unsuitable in a growing number of cases such as found in call centers, retail stores and fast food outlets, airport ground and control tower operations, the military and law-enforcement agencies, etc.

The new appendix will likely include a variety of methods to measure sound levels directly under the listening devices, through the use of manikin, artificial ear or real-ear procedures. While these specialized methods provide the most direct assessment of exposure, they require equipment, expertise and field logistics well beyond the usual range for industrial hygienists and safety personnel. To make the appendix as widely used as possible, the WG is exploring simpler survey methods that could be carried out using the same equipment as for general noise surveys, namely a sound level meter or noise dosimeter [2].

Headset users are exposed to both the environment background noise around them and the audio signals from their device. These two sources are not independent. Users will typically adjust their headset to ensure proper reception of speech and audio signals above the noise entering the device. This suggests an alternative way of assessing headset exposure by measuring the environment background noise around the worker, correcting for the device attenuation and then accounting for the expected signal-to-noise (SNR) under the device. This indirect assessment procedure is referred to as the calculation method [2].

This paper reviews earlier Canadian studies on headset exposure, primarily the surveys of Dajani et al. (1996) at several industrial sites and Crabtree (2002) in military aircraft. The purpose of this review is to gain more insight into the main determinants of headset sound exposure and to provide an empirical basis for the new calculation method under development by the CSA WG.

2. REVIEW OF CANADIAN STUDIES

2.1 Inst. of Biomaterials and Biomedical Eng. (IBBME) of the University of Toronto

Under the leadership of Hans Kunov, a major research effort on the development and application of acoustic manikins was undertaken at IBBME circa 1985-1995. The basic test fixture consisted of a modified KEMAR manikin, adapted to increase the sound isolation and to provide artificial skin lining around the circumaural area and ear canal [3]. While initially designed for the objective evaluation of hearing protectors, the manikin was also used for headset exposure assessments through a series of contracts with Labour Canada [4-5].

The field method required two similar communication headsets, one worn by the worker to carry out normal tasks and one placed on the manikin to measure sound levels under the device. This necessitated the design of a special splitter box to duplicate the electric signal to the headsets or the availability of parallel output connectors in the audio console. The manikin was positioned near the worker, and measurements of the environment background noise were taken in addition to manikin recordings. Manikin data were transformed into diffuse-field equivalent levels using a 1/3-octave band calculation procedure [4] or through a filter module connected to the recording equipment [5].

Figure 1 shows the relationship between the headset diffuse-field transformed sound levels and the background noise around the user. The data set includes 2 measurements from [4] and 31 measurements from [5] covering 9 workplaces in a variety of settings (telephone operators, cable maintenance, control towers, ground crew) in 3 provinces (ON, SK, AB). The distribution of headset types was as follows: 9 intra-aural, 14 supra-aural and 10 circumaural.

The correlation coefficient in Figure 1 is 0.77, indicating that about 59% of the variation in headset sound level can be explained by the environment background noise around the user. The regression equation gives the predicted headset sound level (y) given the background noise level (x). The standard deviation of the headset level from the regression line is 5.2 dB. This can be used to establish empirically-based environment background noise levels that should not be exceeded to ensure compliance with the regulatory limit (e.g. 85 dBA). The thin lines above the regression line (50^{th}) depict the 90^{th} and 95^{th} percentile headset exposure given the background noise. At the 95^{th} criterion, headset sound levels do not exceed an equivalent 85 dBA exposure limit if the background noise is below 69.6 dBA. At the 90^{th} criterion, the maximum background noise is 74.1 dBA.

In Figure 1, the slope of the regression line is 0.42, well below 1. Thus, headset exposure rose by only 0.42 dB for each 1 dB increase in background noise over the data set. Analysis of headset selection in the different work sites reveals that low-attenuation intra and supra-aural devices were used in quieter settings (e.g. office) while high-attenuation circumaural devices (Peltor, David Clark) were chosen in the noisier settings (e.g. airport ground crew). As a result, the difference between headset equivalent sound levels and the background noise was about +12 to +15 dB in the quieter settings (50-65 dBA) and around -5 to 0 dB in the noisier settings (75-95 dBA). Figure 2 summarizes this data. Device attenuation appears to be an important determinant of exposure.

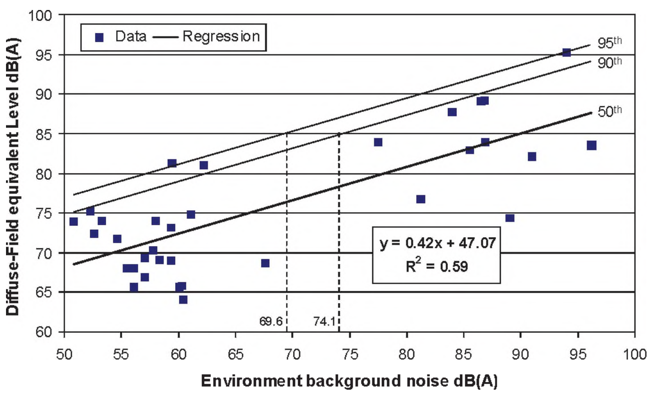


Fig. 1: Correlation between headset diffuse-field equivalent level and the environment background noise. Data from [4-5] (n=33).

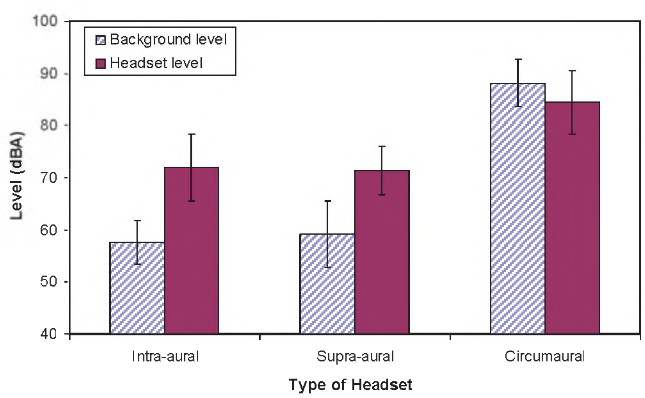


Fig. 2: Background and headset level in relation to headset type.

2.2 Defense Research and Development Canada (DRDC)

The evaluation of communication headsets for military applications has been an ongoing research stream at DRDC Toronto, using manikins and real-ear procedures. In one study [6], at-ear sound levels were measured using miniature microphones under 5 passive and active communication devices worn by 3 crew members inside a Hercules Aircraft. The crew members were asked to adjust the listening volume of the audio channel for adequate speech discrimination and comfort. Measurements were carried out both during communications (ON) and when no communications took place (OFF), allowing estimating the effective SNR under the device. From these data, a mean SNR of 13.8 dB and standard deviation of 6.7 dB across devices and crew members (n=22) can be calculated.

The result above can be used to further analyze the data from Figure 2. If we assume an SNR of +13.8 dB under the devices, the estimated attenuation of intra and supra-aural devices is -2 to +1 dB (as expected from general purpose headsets without rated attenuation) and about 18 dB for circumaural devices (in line with listed NRR).

3. CONCLUSIONS

This paper reviewed two earlier Canadian studies on communication headset exposure, covering both civil and military workplace settings. The headset equivalent sound level appears well correlated with the environment background noise around the user. Typical SNR under the device is in the order of 12-15 dB. Altogether, the studies provide good support for an indirect calculation method for CSA Z107.56 based simply on the measurement of background noise around the user and an estimate of the attenuation of the headset. The total exposure also depends on the proportion of time that audio communications take place during use of the device [2]. A similar approach was recently proposed in a review of the hearing loss prevention program for the Canadian military [7].

REFERENCES

[1] CSA Z107.56 (2006), *Procedures for the Measurement of Occupational Noise Exposure* (Canadian Standards Association).

[2] A.Behar, C.Giguère & T.Kelsall (2008). "CSA appendix on measurement of noise exposure from headsets," *Canadian Acoustics* 37(3).

[3] H.Kunov & C.Giguère (1989) "An acoustic head simulator for hearing protector evaluation. I: Design and construction," *J. Acoust. Soc. Am.* **85**(3), 1191-1196 (1989).

[4] H.Kunov, C.Giguère & R.Simpson (1989). *Method for measuring noise exposure from communication headsets*, Final report to Labour Canada (Contract #1170-4-88-084).

[5] H.Dajani, H.Kunov & B. Seshagiri (1996). "Real-time method for the measurement of noise exposure from communication headsets," *Appl. Acoust.* **49**(3), 209-224.

[6] R.B. Crabtree (2002). *Hercules Audio Enhancement: Project Report* (DRDC Toronto, Letter report 20 September 2002).

[7] C.Giguère & C.Laroche (2005) "Hearing loss prevention program in the military environment," *Canadian Acoustics* **33**(4), 21-30.

ADJUSTING HISTORICAL NOISE ESTIMATES BY ACCOUNTING FOR HEARING PROTECTION USE: A PROBABILISTIC APPROACH AND VALIDATION

Hind Sbihi[1], Kay Teschke[1, 2], Ying MacNab[2], Hugh W. Davies[1]
[1]School of Environmental Health, University of British Columbia, Vancouver Canada
[2]School of Population and Public Health. University of British Columbia, Vancouver Canada

1. INTRODUCTION

Earlier retrospective noise exposure assessments were not adequately characterized for several methodological limitations, including not properly accounting for use of hearing protection devices (HPD), which would result in the over-estimation of noise exposure and, consequently, in potential misclassification.

Exposure misclassification has been shown to attenuate exposure-outcome relations. In the case of already subtle relationships such as noise and cardiovascular diseases, this would potentially obscure any association. Our study was motivated by an earlier investigation of noise and ischaemic heart disease in a cohort of sawmill workers where the exposure-response relation was strengthened when the authors limited their analyses to workers who terminated their work before 1970, a date where HPD were presumably not used [1]. This finding supports the fact that HPD use contributes to a misclassification of exposure.

In this study, we re-examined this cohort of workers and examined an approach to account for HPD and to validate the new exposure measures by testing the predictive ability of the HPD-adjusted noise estimates to predict noise-induced hearing loss.

2. DATA

2.1 Population data

A cohort of 27,500 sawmill workers was enumerated in 14 BC lumber sawmills selected because of the high quality of their work history data [2]. We utilized a subset of this cohort (referred to as the study cohort) linked to an audiometric database, described below, who were employed for at least one year between 1950 and 1998, and who had at least one audiogram test.

2.2 Noise exposure data

Noise exposure data was gathered from research, industry and regulatory sources. 1,900 full-shift dosimetry measurements, from cohort mills, were used in modeling the determinants of noise exposure. Defining an exposure data matrix, and using the determinants of noise exposure model, we obtained for 3700 unique combinations of job/mill/time, the predicted A-weighted dB(A) noise exposure estimates [2]. These retrospective noise exposure estimates did not account for use of HPD.

2.3 Audiometric data

Audiometric data was obtained from WorkSafe BC, the local regulatory agency that coordinates hearing conservation programs in workplaces where noise is deemed to be higher than the regulatory limit, 85 dB(A). As part of the program, WorksafeBC archives routine audiometry surveillance data. Approximately 90,000 hearing tests were linked to the cohort sawmills.

Since each hearing test was also accompanied by a questionnaire administered by an audiometric technician, this data offers three distinct types of information:
(1) Information on HPD use: self-reported by worker at time of hearing test;
(2)Audiograms which comprise binaural hearing threshold levels for frequencies of 500Hz to 8 Khz;
(3) Personal information on subjects' otological health history, occupational and leisure noise exposures.

3. METHODS

We linked the three data sources to obtain the study cohort. We re-estimated the historical noise estimates in three steps: (1) the 'real-world' attenuation of HPD used were derived; (2) determinants of use of hearing protection devices were modeled; (3) we used the results from the two previous steps to predict use of HPD for workers in the study cohort, and to adjust the exposure according to the HPD field performance estimates.

3.1 'Real-World' attenuation

Performance data literature on earplugs and earmuffs was used [3] and calibrated to Canadian Standards for class A and class B ear protectors [4], resulting in a nominal attenuation factor, denoted A. Given that workers seldom use their ear protectors throughout the work shift, we accounted for partial compliance of usage of HPD, and obtained the effective attenuation [5].

3.2 Hearing protection use modeling

A subset of the study cohort was used to model the determinants of use of HPD. Mixed effects models were used to handle the binary response (yes/no) for use of HPD as well as the nested structure of the data (workers within mills). We applied this model to the study cohort and obtained predicted probability of use of HPD, noted π thereafter, for each combination of calendar year/job/exposure level.

3.3 Retrospective noise exposure re-estimation

A correction factor was computed for each job-exposure observation using a time-varying HPD-specific effective attenuation weighted by the predicted probability of use of HPD, π.

We calculated the cumulative exposure for the noise metric, (for both HPD-adjusted and unadjusted). Cumulative exposure is the sum of products of noise intensity and duration of employment, in units of dB(A)×year, in a given job j, for jobs 1 to k, as follow:

$$\text{Cumulative exposure} = 10\log\left[\sum_{j=1}^{k} Tj * 10^{\frac{Leq_j}{10}}\right] \qquad (1)$$

3.4 Validation

Ideally the validity of any exposure measures should be tested. In the absence of a 'gold standard', we proposed to examine in a further study the predictive validity of the re-estimated noise estimates against a well-established noise health effects, namely noise-induced hearing loss.

Using the archived audiometric data, we will define the hearing loss for all cohort workers.

Additional information gathered by audiometric technicians, including potential confounders for the noise-induced hearing loss association can also be used and accounted for in the noise-hearing loss relation.

Using a linear mixed-effects modeling to account for within-subject correlation in repeated hearing loss measurements, we would have the following general formulation of the validation model, where the indices i and j are for hearing tests and workers, respectively, and where only the intercept is allowed to have random effects u0ijj among workers:

$$y_{ij}^{*} = \beta_{0j} + \sum_{p=1}^{P} \beta_{pij} x_{ij} + e_{0ij} \qquad (2)$$

$$\beta_{0j} = \beta_0 + u_{0ij}, u_{0ij} \sim N(0, \sigma_u^2), e_{0ij} \sim N(0, \sigma_e^2) \qquad (3)$$

The βs here are fixed effects for P predictors (age, sex, ethnicity, noise exposure, significant risk factors), and var(u_{0j}) and var(e_{0ij}) are two variance parameters to be estimated. The slopes (β) of the relationship between hearing loss and the adjusted or unadjusted metric can then be compared to determine whether adjustment for HPD use improved the exposure-response relation.

4. RESULTS

4.1 Descriptive

The study cohort comprised 13147 workers representing a total of 183,115 records defined by job, exposure, and self reported use of HPD. It was predominantly male (99%) and composed of three major ethnic groups with 8.8% East Indian (mostly Sikhs), 1.5% Chinese, and the remaining majority being Caucasian, (mostly of European descent 89.7%). Workers were highly exposed to noise; without accounting for HPD the average measured exposure was 90.6 dB(A) and the mean unadjusted cumulative exposure was 101.4dBA)×years.

The validation sub-cohort had a slightly lower exposure than in the study cohort with a mean adjusted cumulative exposure of 99.7 dB(A))× year. This difference was driven by the job length as workers in the validation subgroup had shorter job tenure (716 days on average) than in the study cohort (894 days on average). In this subgroup, workers had between 2 and 16 hearing tests (on average 4.3 hearing tests).

4.2 HPD-adjusted noise estimates

The mean predicted probability of using hearing protection was 0.82 (standard deviation 0.27).

The mean correction factor for all types of self-reported hearing protection devices (earplugs, or earmuffs, class A or class B) after accounting for partial use was 9.7dB. This value was based on yearly prevalence of use of hearing protection devices, the probability of use of HPD, and the HPD filed attenuation values.

The mean adjusted cumulative exposure since first entry in the study cohort was 98.26 dB(A))× year and the mean corrected exposure was reduced to 84.6dB(A).

4.3 Validation

The results of the validation study will show whether an increase in the dose-response relationship does support the hypothesis that accounting of HPD lessens misclassification of exposure.

5. CONCLUSION

We showed in this study, that adjusting for HPD use led to a stronger and more significant noise-hearing loss relationship than exposure estimates with no adjustment, thereby demonstrating that HPD use contributes to non-differential misclassification.

REFERENCES

[1] Davies HW, Teschke K, Kennedy SM, Hodgson MR, Hertzman C, Demers PA (2005). Occupational exposure to noise and mortality from acute myocardial infarction. *Epidemiology*, 16:25-32.

[2] Hertzman C, Teschke K, Ostry A, et al (1997). Mortality and cancer incidence among sawmill workers exposed to chlorophenate wood preservatives. *Am J Public Health*, 187:71-79.

[3] Berger EH, Franks JR, Lindgren F. (1996). International review of field studies of hearing protector attenuation. In: Axlesson A, et al., ed. Scientific Basis of Noise-Induced Hearing Loss. Thieme ed. New York; 361.

[4] Canadian Standards Association (1994). Hearing protectors.

[5] Arezes PM, Miguel AS. Hearing protectors acceptability in noisy environments (2002). Ann Occup Hyg.46:531-536.

ACKNOWLEDGMENTS

Funded by the Canadian Heart and Stroke Foundation.

CROSS-MODAL INTEGRATION IN MUSIC PERCEPTION

Michael Maksimowski and Frank Russo
Dept. of Psychology, Ryerson University, 350 Victoria St. Toronto, Ontario, Canada, M5B 2K3

1. INTRODUCTION

Music is an ideal domain through which to study issues concerning cross-modal integration. In addition to auditory information, the perception of music typically involves visual information and can in some instances involve vibrotactile information. Music is also a complex auditory stimulus; pitch, dynamics, mode and temporal information must all be integrated in order to generate meaning. According to the inverse effectiveness rule (Meridith & Stein, 1986), the greater the object ambiguity from a perceptual standpoint, the greater the opportunity for multiple modalities to enhance perception. Recent research has demonstrated a strong influence of visual information on auditory judgments concerning music (e.g., Thompson, Graham, & Russo, 2005; Thompson & Russo, 2007). Similarly, visual and vibrotactile information has been shown to influence speech recognition (Fowler & Dekle, 1991; Gick, Jóhannsdóttir, Gibraiel, & Mühlbauer, 2008). However, we have very little empirical information regarding integration of vibrotactile information in music. In the following experiment, participants made judgments of interval size for cross-modal presentations of intervals comprised of stimuli presented using audio alone, audio-visual, and audio-vibrotactile signals. In the latter two conditions, participants were instructed to base judgments on the auditory information alone. Results showed that accuracy of interval size was significantly greater in both the auditory-visual and auditory-vibrotactile conditions compared to audio-alone. Audio-visual and audio-vibrotactile conditions were not significantly different from one another. In light of these findings, differences in the extent of visual and vibrotactile influences on auditory judgments and the role of learning in cross-modal integration in music are discussed.

Music perception is fundamentally a multimodal experience. Yet the extent to which cross-modal integration in music is learned is relatively unknown. The primary purpose of this paper is to examine whether cross-modal integration in music requires learning. To test this, we compare the benefits of adding an additional modality of information to participants who are engaged in making judgments about the auditory stimulus: visual or vibrotactile.

Although visual features are frequently paired with auditory information in natural displays of music, the same is not true of vibrotactile information. Moreover, vibrotactile information tends to be dominated by low-frequency components. If we find that vibrotactile information supports music perception to the same extent as visual information, this would suggest that cross-modal integration in music is not entirely dependent on learning. We hypothesize that visual and vibrotactile signals, when added with audio, serve to enhance the information acquired.

2. METHOD

Ten university students (2 male, 8 female) participated in the experiment, ranging in age from 19 to 26 years. Years of formal musical training ranged from 0 to 7 years, with an average of 3.1 years (SD = 2.5). Only one participant reported current musical activity, and no participant reported any problems with hearing.

The music stimuli consisted of audio-visual recordings of 22 ascending intervals sung by two trained female vocalists. Each interval consisted of two tones in sequence. Each tone was approximately 1500 ms in duration and the pitch separation between tones ranged from 1 to 11 semitones. All tones fell within the range of 233.1 Hz (Bb3) and 440 Hz (A4). These intervals were presented in 1 of 3 ways: audio-alone, audio-visual, and audio-vibrotactile. Auditory information consisting of intervals and white noise (included to enhance difficulty) was played over headphones (Sennheiser HD580). Visual information was displayed on a 13-inch MacBook. Vibrotactile information consisted of vibrations applied to the palm of each hand using a pair of skin stimulators (Tactaid VBW32) with a useable output of 100 to 800 Hz and a peak frequency of 250 Hz. The perceived vibrotactile intensity of all stimuli was equalized based on a pilot experiment involving a separate group of participants. A channel box (M-Audio) was used to play both auditory and vibrotactile signals simultaneously.

Participants were seated at a desk in a well-lit double-walled sound-attenuated chamber (Industrial Acoustics Company). Trial presentation and responses were computerized using Experiment Creator X (Thompson & Kosalev, 2000). A within-subjects design was utilized; all participants began with the audio-alone condition and then proceeded to the audio-visual and audio-vibrotactile conditions. The order of presentation of the latter 2 conditions was counter-balanced across participants. At the beginning of each condition, participants were given practice intervals of 0 and 12 semitones. These intervals were not used in the actual experiment. Accuracy was calculated by taking the absolute difference between the participant's response and the actual semitone range (i.e., lower is more accurate). At the end of each trial, participants were prompted to press the key that corresponded with the semitone interval they just heard. All statistical analyses used an alpha-level of .05.

3. RESULTS

We evaluated the hypothesis that accuracy would improve in the bimodal conditions (audio-visual and audio-vibrotactile) relative to audio alone. Figure 1 displays mean accuracy (and standard error) for each condition while Figure 2 displays mean accuracy for each participant based on condition. A one-way within-subjects ANOVA showed that there was a significant effect of presentation condition, F (2, 1317) = 12.701, p < .001. Post-hoc analyses revealed that the significant differences existed between audio (M = 2.58, SE = 0.102) and audio-visual (M = 2.05, SE = 0.083) and audio and audio-vibrotactile (M = 2.01, SE = 0.084), but there was no significant difference between the audio-visual and audio-vibrotactile conditions.

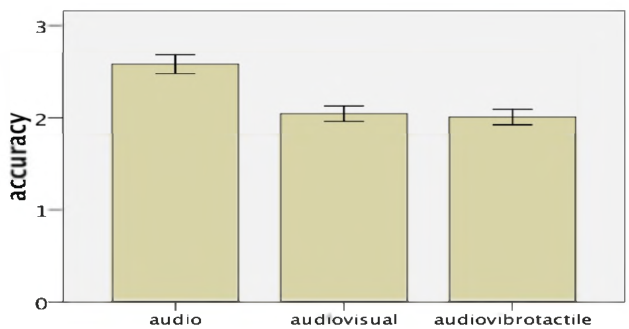


Fig 1. Mean accuracy levels for each condition, with error bars denoting standard error.

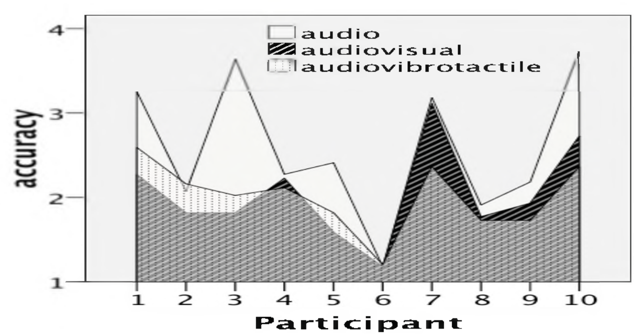


Fig 2. Participant-by-participant improvement in accuracy across conditions.

4. DISCUSSION

The results indicate that visual and vibrotactile inputs enhance accuracy for judgments of interval size compared to audio alone. Most importantly, however, is that participants gained equal enhancements of accuracy from both audio-visual and audio-vibrotactile conditions. Similar to speech (Fowler & Dekle, 1991), congruent vibrotactile information minimizes uncertainty and supports perception of the auditory signal.

Relative gains from vibrotactile exposure are not surprising considering that auditory-vibrotactile associations are both lawful and stem from same environmental signal (i.e., pressure waves). Cross-modal pairings of auditory and vibrotactile information are relatively unfamiliar to participants. Although vibration can be a desired aspect of music listening, the prominent vibrational components tend to be lower in frequency than those presented in the current experiment. Moreover, in ecological settings, lower- and higher-frequency vibrational components are available simultaneously, resulting in masking of higher-frequency components.

Our findings suggest that benefits of cross-modal integration in music are not *dependent* on learning. This view falls in line with direct theories of perception (Gibson, 1966; Liberman *et al.*, 1967). We do not suggest however, that integration of non-auditory information is devoid of learning in the case of music. Many visual influences in music are clearly best understood from a semiotic perspective (Thompson *et al.*, 2005). Future research should clarify mechanisms of integration and investigate the manner and extent to which vibrotactile information may support the musical experience.

REFERENCES

Fowler, C. A., Dekle, D. J. (1991). Listening with eye and hand: Cross-modal contributions to speech perception. *Journal of Experimental Psychology: Human Perception and Performance, 17* (3), 816-828.

Gibson, J. J. (1966). *The senses considered as perceptual systems*. Boston: Hougton Mifflin.

Gick, B., Jóhannsdóttir, K. M., Gibraiel, D., Mühlbauer, J. (2008). Tactile enhancement of auditory and visual speech perception in untrained perceivers. *Journal of the Acoustical Society of America Express Letters, 123* (4), 72-76. doi: 10.1121/1.2884349

Liberman, A. M., Cooper, F., Shankweiler, D., Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review, 74*, 431-461.

Meredith, M. A., Stein, B. E. (1986). Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. *Journal of Neurophysiology, 56*, 640-662.

Thompson, W. F., Graham, P., Russo, F. A. (2005). Seeing music performance: Visual influences on perception and experience. *Semiotica 156*, 177-201.

Thompson, W. F., Russo, F. A. (2007). Facing the music. *Psychological Science*, *18* (9), 756-757.

A CASE OF SUPERIOR AUDITORY SPATIAL ATTENTION

Robert Quelch[1], Gurjit Singh[1,2] , Kathy Pichora-Fuller[1,2]
[1]Dept. of Psychology, University of Toronto, 3359 Mississauga Rd North, Mississauga, Ontario, Canada L5L 1C6
[2]Toronto Rehabilitation Institute, 550 University Ave, Ontario, Canada M5G 2A2

1. INTRODUCTION

A growing body of research investigating the behavioural and neurophysiological responses of professional musicians to simple auditory stimuli has revealed enhanced responding to a number of auditory cues, including spatial location, pitch, and timing [1]. With regard to auditory spatial attention, the research suggests that musical experience involving the simultaneous deployment of attention toward both the entire musical 'scene' and its constituent components (i.e., orchestral conducting) can enhance mechanisms associated with both the acoustic (bottom-up) and cognitive (top-down) processing of sound [2].

Using more complex stimuli, other researchers have been able to show how the allocation of attention to spatial locations can significantly improve speech recognition scores in listening situations involving multiple talkers and uncertainty about the spatial location of the target [3]. Further, it appears that the top-down cognitive processing of sounds is influenced by a listener's ability to take advantage of the binaural auditory cues associated with spatial separation [4]. In the present experiments, we investigate the hypothesis that individuals with prolonged musical experience are better able to report on speech in multi-talker listening situations.

2. EXPERIMENT 1

2.1 Participant

RQ is a native English speaking 40-year old musician with 7 years of formal piano training. For 20 years, he played music professionally in a number of bands. RQ has a moderate bilateral hearing loss above 4 kHz and chronic bilateral tinnitus (Table 1).

2.2 Stimuli

The stimuli were sentences from the Coordinate Response Measure (CRM) corpus spoken by four male talkers [5]. Sentences have the format: "Ready [callsign] go to [colour] [number] now". For each talker, the CRM corpus contains sentences with all possible combinations of eight callsigns, four colours, and eight numbers.

2.3 Equipment and Presentation Conditions

All data were collected in a 3.3m^2 single-walled sound–attenuating booth. The stimuli were presented via Grason-Stadler Inc. loudspeakers: two when there was simulated spatial separation or three when there was real spatial separation. All stimuli were routed from a computer to a Tucker-Davis Technologies (TDT) System III to a Harmon/Kardon (HK3380) amplifier. Sentences were converted from digital files (sampling rate 24.414 kHz) to analogue using two TDT System III RP2.1s. Visual cueing and feedback was displayed on a Compar 17" NEC touch screen monitor positioned on a table located in front of the listener, below shoulder level.

2.4 Design

The dependent variable was accuracy of word identification. Three independent variables were systematically manipulated: callsign cue, location certainty, and presentation method (real or simulated spatial separation). Each block included 30 trials and the starting callsign cue condition for a block of trials was randomly determined. Within each session, for each callsign condition, the probability specifications were selected randomly without replacement. Data was collected on a total of 240 trials for each of the 16 conditions.

2.5 Procedure

On each trial, three sentences were selected randomly from the CRM corpus and presented simultaneously to the listener. All sentences were presented at 60 dBA. Participants were instructed to face forward (0° azimuth). One of the three sentences in each trial was randomly designated as the target sentence by providing the listener with its callsign and the other two sentences were considered maskers. The task was to select the colour and number in the target sentence using the touch screen. Both the correct colour and number were required for a correct response. The callsign cue was provided either 1 s before ('before' condition) or immediately after ('after' condition) each trial in a block. Four probability specifications were provided in advance to indicate the proportion of trials that the target sentence would be presented from the left, center and right spatial locations (*0-100-0, 10-80-10, 20-60-20, 33-33-33*). On any given trial, the location of the target sentence was randomly selected from among the three loudspeakers with the limitation that the probability cue was accurate across a block of 30 trials.

2.6 Results

In the callsign 'after' conditions, RQ performed similarly to the normal hearing younger adults tested in prior studies [3,4] with the exception of the certain location condition (0-100-0) in which he performed better than the 99% confidence interval (CI) of the younger adult group tested previously in identical conditions [4]. In the callsign 'before' conditions, RQ significantly outperformed the other listeners in all but one of the probability conditions (0-80-0). When the stimuli were presented using the precedence effect to simulate spatial separation, the reduced auditory cues associated with real spatial separation had an overall negative impact on RQ's performance, although, he significantly outperformed the young group in all of the location probability conditions (Fig. 1).

Table 1. Left (L) and right (R) hearing thresholds (dB HL) for RQ.

	Frequency (kHz)									
	0.25	0.5	0.75	1	1.5	2	3	4	6	8
L	0	15	20	20	25	10	10	20	35	45
R	0	5	10	5	20	10	10	15	40	50

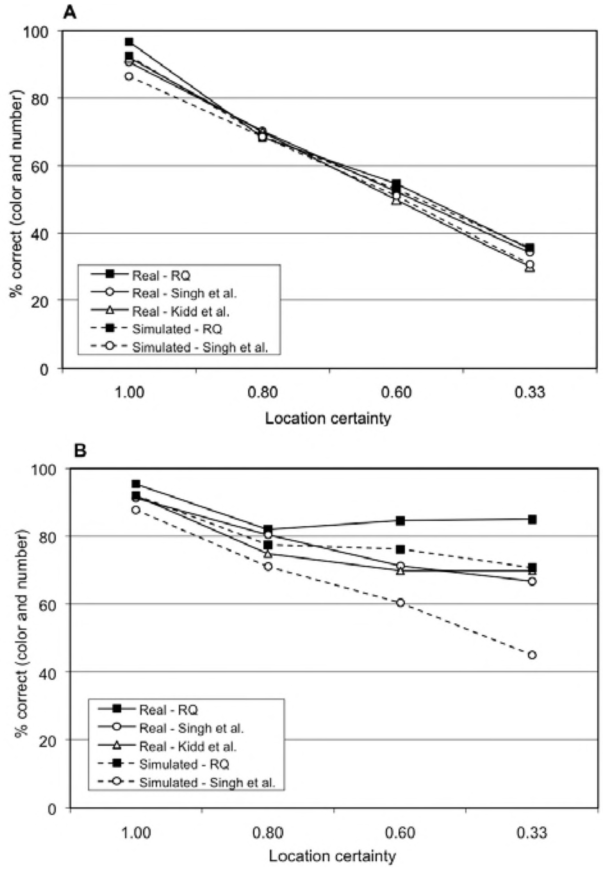


Fig 1. Mean proportion correct identification scores for RQ, Kidd *et al.*, Singh *et al.* plotted as a function of increasing target location uncertainty when the target identity was (A) unknown prior to the presentation of the stimuli and (B) known prior to the presentation of the stimuli.

3. EXPERIMENT 2

3.1 Participants

The listeners were 6 normal hearing musicians aged 17-25 years, each with a minimum of 5 years of music experience (mean = 6.5 yrs, *SD* = 3.2 yrs). All but one of the participants had formal music training (mean = 10 yrs, *SD* = 3.4 yrs) and all reported playing music at least weekly.

3.2 Procedure

The stimuli and procedures were similar to Experiment 1 with the exception that the callsign cue after and simulated spatial separation conditions were not presented. Data were collected on a total of 120 trials for each of the 4 conditions for each participant.

3.3 Results

When the location of the target sentence was certain, 4 of the 6 musicians significantly outperformed the non-musicians (P1, P3, P4, P5). When the spatial location of the target was uncertain, two of the musicians significantly outperformed the nonmusicians (P3, P4) (Fig. 2).

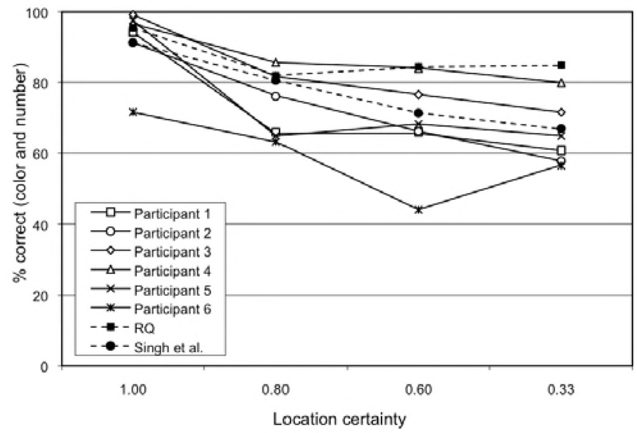


Fig 2. Mean proportion correct identification scores for the musician group, RQ, and the previously tested young group [4] plotted as a function of increasing target location uncertainty when the target identity was known prior to stimulus presentation.

4. DISCUSSION

Although the results of these experiments are not sufficient to define the precise mechanisms supporting the high levels of performance reported for RQ and some of the musicians, overall, they suggest that prolonged musical experience may, in some cases, enhance a listener's ability to comprehend speech in complex listening environments. This may be due to an improved ability to process the various auditory cues due to spatial separation resulting in an enhanced ability to selectively attend to and/or divide attention among the competing talkers. The results obtained for RQ in Experiment 1, in which the precedence effect was used to reduce the auditory cues associated with real spatial separation, underscore the importance of top-down cognitive processing of auditory information. Further, it appears that this ability may help offset the detrimental effects that high-frequency hearing loss and chronic tinnitus might have on hearing, listening and comprehending in complicated listening environments.

REFERENCES

[1] Münte, T.F., Nager, W., Beiss, T., Schroeder, C., & Altenmüller, E. (2003). Specialization of the specialized: Electrophysiological investigations in professional musicians. *Ann. N.Y. Acad. Sci.*, *999*, 131-139.

[2] Nager, W., Kohlmetz, C., Altenmüller, E., Rodriguez-Fornels, A., Münte, T.F. (2003). The fate of sounds in conductor's brains. *Cognitive Brain Research*, *17*, 83-93.

[3] Kidd, G., Arbogast, T.L., Mason, C.R., & Gallun, F.J. (2005). The advantage of knowing where to listen. *J. Acout. Soc. Am.*, *118*, 3804-3815.

[4] Singh, G., Pichora-Fuller, M.K. & Schneider, B. (in press). The effect of age on auditory spatial attention in conditions of real and simulated spatial separation. *J. Acout. Soc. Am.*

[5] Bolia, R.S., Nelson, W.T., Ericson, M.A., & Simpson, B.D. (2000). A speech corpus for multitalker communications research. *JASA, 107*, 1065-1066.

This work was funded by the Canadian Institutes of Health Research and the Natural Sciences and Engineering Research Council of Canada.

THE EFFECT OF INFORMATIONAL MASKING AND WORD POSITION ON RECALL

Payam Ezzatian[1], Liang Li[1,2], Kathy Pichora-Fuller[1], and Bruce Schneider[1]
[1]Dept. of Psychology, University of Toronto, 3359 Mississauga Rd North, Mississauga, Ontario, Canada L5L1C6
[2]Dept. of Psychology, Peking University, Beijing, China, 100871

1. INTRODUCTION

In order to perceive speech in noisy backgrounds, listeners need to perceptually separate the target stream from competing streams. Although competing streams such as construction noise may reduce target audibility (energetic masking), they are easy to distinguish from target speech because they are highly dissimilar to speech. Speech maskers on the other hand, can reduce target audibility as well as interfere with the processing of the target due to linguistic and acoustic similarities to target speech (informational masking, [1]). Because of these similarities, it may initially be difficult to perceptually segregate a speech signal from competing speech. Due to this initial delay in segregation, we might expect speech perception to improve as sentences unfold over time in situations where informational masking is prevalent. Informational masking may especially trouble older adults due to age-related declines in ability to take advantage of the F_0 difference between speakers' voices [2], or a reduced ability to take advantage of modulations in background maskers, thereby making stream segregation more difficult for older adults.

The current study examined age-related differences in the time course of perceptual streaming in an informational masking paradigm. In Experiment 1, younger and dder adults' ability to repeat sentences was measured in the presence of a two-talker speech masker (informational masking) or a speech-spectrum noise masker (energetic masking). In Experiment 2, the speech masker was noise-vocoded using 3 bands to determine the extent to which performance in the presence of speech masker is related to the fluctuations in the amplitude envelope of such maskers.

2. METHOD

2.1 Participants

Two independent groups of 16 younger and 16 older adults participated in each experiment. All participants were native English speakers, and had clinically normal audiograms (thresholds $\leq$ 25 dB HL) from .25 to 3 kHz, with interaural differences not exceeding 15 dB.

2.2 Stimuli

Target sentences were 208 nonsense sentences spoken by a female talker, e.g. "A house should dash to the bowl" [3]. In Experiment 1, a speech-spectrum noise masker was used in one half of the conditions, and a two-talker nonsense speech masker was used in the other half. In Experiment 2, the same speech-spectrum noise masker from Experiment 1 was used, but the two-talker nonsense speech masker from Experiment 1 was noise vocoded using 3 bands. The speech masker was noise-vocoded as follows: First, it was divided into 3 frequency bands using the following boundaries: 300, 814, 1528, and 6000 Hz. Then, the amplitude envelope was extracted in each band and used to modulate bands of noise having the same widths and center frequencies, thus creating a "vocoded" signal [4]. This procedure preserves amplitude envelope cues while removing fine structure cues. By using only 3 bands, the signal is largely unintelligible.

All stimuli were digitized at 20 kHz using a 16-bit Tucker Davis Technologies (TDT, Gainesville, FL) System II. The stimuli were converted to analog using the TDT system. The stimuli were then low-pass filtered at 10 kHz and amplified by a Harmon Kardon amplifier (HK 3370) prior to transmission via a single 40 watts loudspeaker. The loudspeaker was situated in the left far corner of a 9.3 x 8.9 x 6.5 foot Industrial Acoustic Company (Bronx, NY, USA) double-walled sound-attenuated chamber, and participants were seated at the center of the chamber facing the loudspeaker at a distance of 1.03 meters. Target sentences were presented at 60 dBA. Masker levels were adjusted to produce 4 signal-to-noise ratios (SNRs) for the younger adults: -12, -8, -4, and 0 dB, and 4 SNRs for the older adults: -8, -4, 0, and 4 dB. Sentence lists, SNRs, and masker types were counterbalanced across participants such that each list was presented in each masker and SNR condition an equal number of times.

2.3 Procedure

Prior to the start of each experiment, participants were familiarized with the task by being presented with one of the practice sentences ("A house should dash to the bowl") at the easiest SNR. The experimenter initiated the presentation of each sentence with the press of a keyboard button, which was followed immediately by the onset of background noise. Exactly 1 second later, the female target speaker uttered a target sentence, at the end of which the masker terminated as well. Participants were asked to repeat each target sentence, and were scored for 3 keywords (e.g., in "A rose can paint a fish", keyword 1 = Rose, keyword 2 = Paint, keyword 3 = Fish).

2.4 Analysis

The dB SNR required for 50%-correct performance was computed for each individual at each of the three keywords. These thresholds were then entered into an analysis of variance (ANOVA) to examine the effect of Age, Masker, and Word Position on performance.

3. RESULTS AND DISCUSSION

3.1 Experiment 1: Two-talker speech masker vs. speech-spectrum noise masker

As shown in Figure 1, younger adults outperformed older adults, and both groups performed worse with the speech masker than the noise masker in the background. The mean thresholds for both age groups improved in a linear fashion as a function of word position, but for the speech masker only. These observations were confirmed by 2 Masker by 3 Word Position ANOVA, with Age as a between-subjects factor, which revealed significant main effects of Age; $F(1, 30) = 7.12$, ($p = 0.01$), and Masker; $F(1, 30) = 23.43$, ($p < 0.001$), and Word Position on thresholds; $F(2, 60) = 4.23$, ($p = 0.019$). On average, older

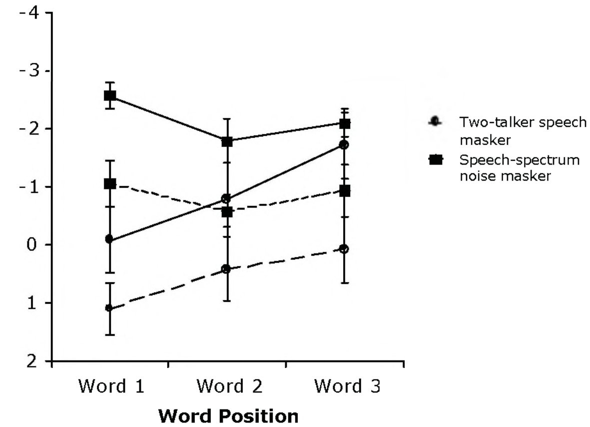


Fig. 1. Average thresholds from Experiment 1 for the younger adults (solid lines), and the older adults (dashed lines), are shown as a function of word position for the two-talker speech masker (squares), and speech-spectrum noise masker (circles) conditions. Error bars represent ± 1 SE.

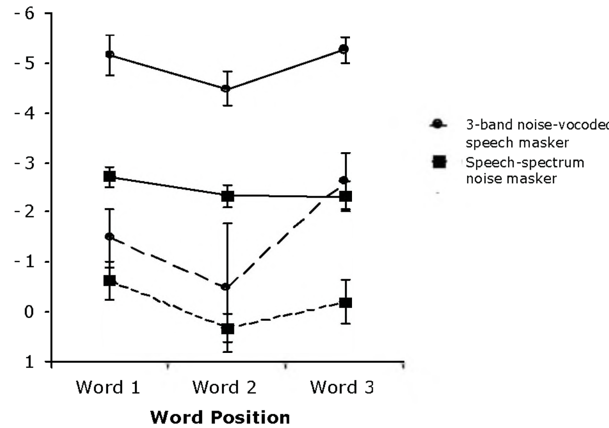


Fig. 2. Average thresholds from Experiment 2 for the younger adults (solid lines), and the older adults (dashed lines), are shown as a function of word position for the 3-band vocoded masker (squares), and speech-spectrum noise masker (circles) conditions. Error bars represent ± 1 SE.

adults required a 2.3 dB higher SNR to have the same performance as younger adults, and performance was an average 1.34 dB SNR better with the noise masker in the background (supporting the notion that speech maskers result in informational masking). Finally, overall performance was better for the third word, but the interaction of Word Position by Masker was significant; F (2, 60) = 8.82, $p < 0.001$, confirming the trend in the figure whereby performance improves over time with the speech masker, but not the noise masker. More importantly, the three-way interaction of Age by Masker by Word Position was not statistically significant, suggesting that the two age groups do not differ in the way in which they perceptually segregate target streams from background maskers.

3.2 Experiment 2: 3-band noise-vocoded speech masker vs. speech-spectrum noise masker

To confirm that the results from Experiment 1 are due to informational masking and not energetic masking, the speech masker from Experiment 1 was noise-vocoded using 3 bands (which removes content and fine structure cues) and tested in Experiment 2. The speech-spectrum noise masker was also tested in Experiment 2. Figure 2 plots the average performance for younger and older adults as a function of masker and word position. As can be seen from this figure, younger adults outperformed older adults, and both groups performed better with the noise-vocoded masker than with the speech-spectrum noise masker in the background. An ANOVA confirmed a main effect of Age; F (1, 30) = 21.11, ($p < 0.001$). On average, older adults required a 2.87 dB higher SNR to match the performance as younger adults. The interaction of Masker by Age was also significant; F (1, 30) = 5.69 (p = 0.024). Overall, younger adults improved by an average 1 dB more than older adults when going from the noise masker to the vocoded masker. Finally, there was a significant main effect of Masker; F (1, 30) = 65.27 ($p <$ 0.001), and a significant interaction of Masker by Word Position; F (2, 60) = 4.53 (p = 0.014). Performance was an average 1.9 dB SNR better with the vocoded masker in the background, and the positive change in performance between Word 2 and 3 was greater for the speech masker.

4. DISCUSSION

Results from both experiments replicate previous findings of negative age-related differences when listening to speech in noisy background. Results from Experiment 1 show that for both younger and older adults, performance improves systematically with word position when the background consists of a speech masker, but not when it consists of a noise masker, indicating that stream segregation takes time to build up in informational masking situations. In Experiment 2, the pattern of improvement for the speech masker was no longer present once it was noise-vocoded. Older adults experienced less benefit from the noise-vocoded masker than did younger adults (possibly due to a decline in the ability to benefit from temporal fluctuations in a masker); however the pattern of performance as a function of Word Position was again equivalent for both groups, indicating that the time course of stream segregation is not affected by aging.

REFERENCES

[1] Schneider, B. A., Li, L., & Daneman, M. (2007). How competing speech interferes with speech comprehension in everyday listening situations. *Journal of the American Academy of Audiology*, 18, 559-572.

[2] Vongpaisal, T., Pichora-Fuller, M.K. (2007). Effects of age on F0 difference limen and concurrent vowel identification. *Journal of Speech Language and Hearing Research*, 50, 1139-1156.

[3] Helfer, K. S. (1997). Auditory and auditory-visual perception of clear and conversational speech. *Journal of Speech Language and Hearing Research*, 40(2), 432-443.

[4] Eisenberg, L. S., Shannon, R. V., Martinez, A. S., Wigonski, J. Boothroyd. (2000). Speech recognition with reduced spectral cues as a function of age. *Journal of the Acoustical Society of America*, 107(5), 2704-2710.

ACKNOWLEDGMENTS

This project was funded by CIHR.

The Effect of Types of Acoustical Distortion on Lexical Access

Marianne Pelletier, Marco Coletta, Renée Giroux, Huiwen Goy and Kathy Pichora-Fuller
Department of Psychology, University of Toronto, 3359 Mississauga Rd N., Mississauga, Ontario, L5L 1C6.

1. INTRODUCTION

Lexical access is crucial to understanding spoken language. According to the Cohort model by Marslen-Wilson & Welsh (1978), the process of lexical access is accomplished according to three steps: the *activation* of specific candidates based on the acoustic input, the *selection* of one candidate from a pool of candidates, and the *integration* of the term with others according to the semantic context of the discourse. The process of lexical access and comprehension enables accurate understanding of language. Therefore, when distortions are introduced, there may be difficulties in comprehension because it is more difficult to decipher the acoustic signal. Distortions induced in the laboratory are analogous to difficulties in understanding spoken language in everyday situations, such as when there is noise in crowded streets, or in busy workplaces, or over a noisy telephone line, or when talkers have unclear speech.

Aydelott and Bates (2004), in their study of the effects of acoustic distortion and semantic context on lexical access, argue that different types of distortions affect lexical access at different stages. In their research, these authors argue that low-pass filtering affects the early stages of lexical access (encoding and activation of lexical-semantic information), whereas time compression inhibits the later stages of selection and integration. In their study, these authors appeared to have arbitrarily selected the amount of low-pass filtering (1Hz cutoff) and time compression (50 %). The goal of the present work was to replicate the findings of Aydelott and Bates (2004) using the same type and amount of distortion, while also extending their work by including a third type of distortion that is encountered in everyday life, namely multi-talker babble noise.

1.1 Preliminary Investigation

A preliminary investigation was conducted to guide the selection of the amounts of distortion used in the present study. Specifically, psychometric functions were determined using the Northwestern University Auditory Test No. 6 (NU6), which consists of four lists of 50 monosyllabic words. There are no differences in word recognition performance among the four lists if they are administered in quiet (Stuart, Green, Phillips, & Stanstrom, 1994). These word lists were distorted to varying degrees by low-pass filtering, time compressing, and adding background babble to create different signal-to-noise ratio (SNR) conditions. Three groups of 16 (young) adults with normal hearing were tested using the NU6 lists distorted by each of the three methods (each group heard only one type of distortion). The psychometric functions enabled us to estimate speech intelligibility in the two conditions used in the study of Aydelott and Bates (2004): 1000Hz low-pass filtering corresponded to 50% correct on NU6 lists, whereas 50% time compression corresponded to 76% correct on NU6 lists. In the present study of lexical access, we replicated the two original conditions tested by Aydelott and Bates (2004). The initial SNR level to be tested was chosen to correspond to 50% correct on NU6 lists so that the effect on intelligibility of the degree of distortion was the same for SNR (-15 dB) condition and the original filtering condition, but the degree of distortion chosen for SNR was greater than the degree of distortion resulting from the original 50% time-compression (see companion paper for comparisons between different degrees of distortion due to filtering). Table 1 summarizes the levels of distortion selected for each distortion type reported here, as well as those reported in our companion study and those being tested in ongoing studies.

Table 1. *Performance for degrees and types of distortion*

Type of Distortion	% Correct	Degree of Distortion
Low-Pass Filtering	50%	1000 Hz
	76%	1750 Hz[c]
Time Compression	50%*	72%*
	76%	50%
Signal-to-Noise Ratio	50%	-15dB
	76%*	-9dB*

c See companion paper in this issue.
** These levels of distortion are being tested in ongoing experiments.*

2. METHOD

2.1 Participants and General Set-Up

Participants were tested individually in a single experimental session lasting 45-60 minutes. They completed a hearing and language history, a vocabulary test and an audiogram. They were randomly assigned to conditions so that the stimulus list and the hand used to press the response button were counterbalanced across participants.

2.2 Stimuli, Task, and Equipment

The stimuli were sentences consisting of an initial context, a 50-ms pause, and a sentence-final target utterance. The target utterance spoken by a male was always presented intact. The task of the listeners was to decide if the target utterance was a word or non-word. The initial portion of each sentence was spoken by a female and was presented either intact or distorted, and it provided a semantically congruent, incongruent, or neutral context for the following intact target (when it was word). All stimuli were presented binaurally at 70dB SPL using Sennheiser HD 250 headphones. A light signal on the response box at the beginning of the target word indicated to the participant that a response could be entered by pressing the correct button as quickly and accurately as possible. Participants were allotted 3000 ms to answer, after which, regardless of whether a response was entered, the next trial was presented. The participant's response time started at the beginning of the target utterance (same time as the light signal on the response box) and ended when they pressed a button. Participants performed a block of nine practice trials prior to the test-trial sequence of 48 items, which contained 8 examples of both the altered and unaltered conditions for each of the congruent, incongruent and neutral sentence contexts, half of which were followed by word targets and the other half by non-word targets.

3. RESULTS

Inclusion of participants into the analysis was based on the criterion that participants correctly answered 90% of the test trials. RT's were eliminated from analysis if an incorrect or no response was entered. Only the reaction times for word targets were analysed. Priming effects in the unaltered and altered conditions were calculated for both facilitation (congruent minus neutral contexts) and inhibition (incongruent minus neutral contexts) effects. A repeated measures analysis of variance (ANOVA) with distortion (altered vs. unaltered) and priming (facilitation vs. inhibition) as within-subject variables was conducted; distortion type (low-pass filtering, time compression, or signal-to-noise ratio) was included as a between-subject variable. Figure 1 depicts the priming effects for each distortion type, where values above zero reflect slower reaction times than in the neutral baseline conditions and values below zero reflect faster reaction times.

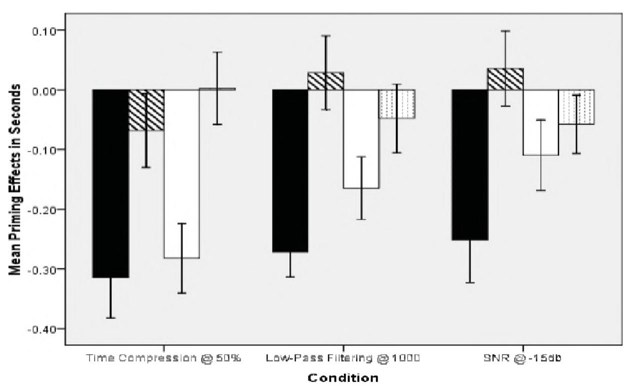


Fig. 1. Mean priming effects (Facilitation and Inhibition) for both unaltered and altered contexts for each type of distortion. Black bars show facilitation for the unaltered context; stripped bars show inhibition for unaltered context; white bars show facilitation for altered contexts; dotted bars show inhibition for altered context. Error bars represent 95% confidence intervals.

There was a significant main effect of type of distortion, (F (2, 105) = 3.945, $p < 0.05$). A significant priming effect also emerged, F (1, 105) = 214.508, $p < 0.01$. The priming effects varied from -0.232 sec (*SE* = 0.013) for the facilitation effect to -0.018 sec (*SE* = 0.012) for the inhibition effect. No main effect of distortion (unaltered vs. altered) was found. A significant distortion type X priming effect interaction emerged, F (2, 105) = 3.629, $p = 0.03$, as well as a significant distortion X priming effect interaction, F (1, 105) = 30.819, $p < 0.001$. Finally, a significant distortion X priming effect X type of distortion interaction emerged, F (2, 105) = 13.521, $p = 0.00$.

4. DISCUSSION

The goal of the present study was to replicate and extend the claims of Aydelott and Bates (2004) regarding the effects of acoustic distortions on lexical access. For intact sentence contexts, it was expected that reaction times to targets in congruent contexts would be faster compared to reaction times to targets in a neutral context, whereas those in an incongruent context would be slower. Furthermore, if reaction times differ depending on the type of distortion then the patterns suggest how priming (facilitation or inhibition) may differ in lexical access for different types of distortion.

The analysis revealed that different types of distortion did affect lexical access to different degrees. Low-pass filtering at 1000Hz and at -15dB SNR both appear to affect the earlier stages of lexical access, which replicated the findings of Aydelott & Bates (2004). However, in the time compression condition, an inhibition effect was expected based on previous findings, but instead, the results of the replication revealed a facilitation priming effect. Importantly, the distortions due to low-pass filtering and SNR produced a release from inhibition and significant facilitation decrease, whereas time compression significantly reduced facilitation and increased inhibition.

4.1 Future Research

These results indicate that different types of distortion have similar effects on lexical access when the degree of distortion is matched (filtering and SNR). The differences observed between time-compression and the other two types of distortion are confounded with differences in degree of distortion referenced to percent correct scores on NU6 words. Extensions of the current study will investigate the effects of different degrees of distortions (e.g., comparing low-pass filtering at 1000Hz and 1750Hz as shown in Table 1 and as described in our companion study in this issue).

REFERENCES

Aydelott, J., & Bates, E. (2004). Effects of acoustic distortion and semantic context on lexical access. *Language and cognitive processes*, 19, 29-56.

Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology*, 10, 29-63.

Stuart. A., Green, W. B., Phillips, D. P., & Stenstrom, R. (1994). List equivalency of the Northwestern University Auditory Test No. 6 in quiet and in continuous broad band noise. *Journal of Speech-Language Pathology and Audiology*, 18, 121-125.

The Effect of the Degree of Acoustical Distortion on Lexical Access By Younger Adults

Marco Coletta[1], Marianne Pelletier[1], Renée Giroux[1], Huiwen Goy[1] and Kathy Pichora-Fuller[1, 2]
[1]Dept. of Psychology, University of Toronto, 3359 Mississauga Rd N., Mississauga, Ontario, Canada L5L1C6
[2]Toronto Rehabilitation Institute, 550 University Ave, Ontario, Canada M5G2A2

1. INTRODUCTION

Comprehension of spoken language is dependent on the accurate transmission of the acoustic speech signal. If the signal becomes distorted, then comprehension becomes more difficult and the specific nature of language processing may be altered. As discussed in our companion paper [1], it has been suggested that some types of acoustic distortions, such as low-pass filtering, may disrupt the early stages of lexical access (encoding and activation of lexical-semantic information), whereas other types of distortion, such as time compression, may disrupt later stages of lexical access (word selection and integration) [2]. However, previous research investigating this hypothesis [1, 2], has confounded the type of distortion with the amount of distortion.

The goal of the present study was to investigate how the speed of processing a sentence-final target word is influenced by varying the degree of acoustic distortion applied to three different types of sentence contexts: semantically congruent, semantically incongruent, and semantically neutral. The effects of acoustic distortion and semantic context on the speed of a lexical decision were measured. The results on the lexical decision task were interpreted as evidence of semantic priming, encompassing both facilitation and inhibition effects [3]. Facilitation of lexical access by the preceding sentence context is evidenced by faster reaction times (RT) in response to targets preceded by a semantically congruent sentence context relative those preceded by a neutral context [4]. Inhibition is reflected in increased RT`s to targets preceded by incongruent contexts, relative to a neutral context [4].

2. METHOD

2.1 Participants

Two experimental groups, each composed of 36 undergraduate students, participated in the experiment: Group 1 (mean age = 19.11 years, *SD* = 1.56) and Group 2 (mean age = 19.06 years, *SD* = 1.47). Each participant had normal audiometric thresholds (= 25 dB HL) at frequencies below 4 kHz. Participants in Group 1 obtained a mean score of 11.78/20 (*SD* = 2.73) on the Mill Hill vocabulary test, while participants in Group 2 obtained a mean score of 11.94/20 (*SD* = 2.47).

2.2 Stimuli

The experimental stimuli consisted of six lists, where each list contained 48 sentence contexts, 24 in which the sentence-final target word was a real word and 24 in which the target was a non-word distractor. The non-word distractors were phonologically permissible strings, had no meaning, and generally resembled the real word targets in length, number of syllables, and phonetic content.

Within each list of 48 sentence contexts, 24 sentences remained acoustically unaltered and 24 sentences were acoustically distorted using low-pass filtering. All sentence-final word targets remained unaltered. Both unaltered and altered sentence contexts were equally divided into 3 semantic categories forming 6 experimental conditions: Congruent-altered, incongruent-altered, neutral-altered, congruent-unaltered, incongruent-unaltered and neutral-unaltered. The acoustically altered sentence contexts were low-pass filtered using Praat [5]. Participants in Group 1 received altered contexts low-pass filtered at 1000 Hz and those in Group 2 received altered contexts filtered at 1750 Hz. It was determined during pilot testing that these degrees of low-pass filtering yielded word recognition scores of 50% and 76%, respectively, for the standard NU6 word lists used in speech audiometry (see companion paper [1]).

2.3 Procedure

Within each group, each participant was randomly selected to be presented one of the six lists, with the lists counterbalanced across participants. The stimuli were presented to participants binaurally over Sennheiser 265 headphones at 70 dB SPL in a double-walled sound-attenuating booth. Participants were instructed to listen to each sentence context and the accompanying sentence-final (non-)word and to make a word/non-word lexical decision by pressing the YES or NO button, respectively, on the response box as soon as the light was illuminated to cue the beginning of the response period. Button order (Y/N vs. N/Y) on the response box, corresponding to the hand participant used to press the response button, was counterbalanced across participants.

3. RESULTS

3.1 Analysis

Inclusion of a participant's data in the final analysis was determined based on the criterion that the participant correctly answered 90% of the test trials. RT's for a word were eliminated if no response was entered or if the response entered was incorrect. An independent samples t-test comparing the mean RTs for words and non-words was significant (Group 1: $t\ (70) = -3.79$, $p = 0.00$; Group 2: $t\ (70) = -7.16$, $p = 0.00$). Only the RT's for real word targets were considered in subsequent analyses.

3.2 Reaction Time

Initially, button order was analyzed as a between-subjects factor; however, since no significant effects of button order emerged, this factor was dropped from further analyses. Semantic context and acoustic distortion (altered or unaltered) served as within-subjects variables for the separate analyses conducted for each group. A repeated measures analysis of variance (ANOVA) revealed significant main effects of semantic context (Group 1: $F\ (2, 68) = 67.60$, $p < 0.005$; Group 2: $F\ (2, 68) = 92.19$, $p < 0.005$). There were no significant main effects of distortion for either experimental group. For Group 1 only, there was a significant interaction of distortion and context (F (2, 68) = 12.20, $p < 0.005$). No other interactions were significant.

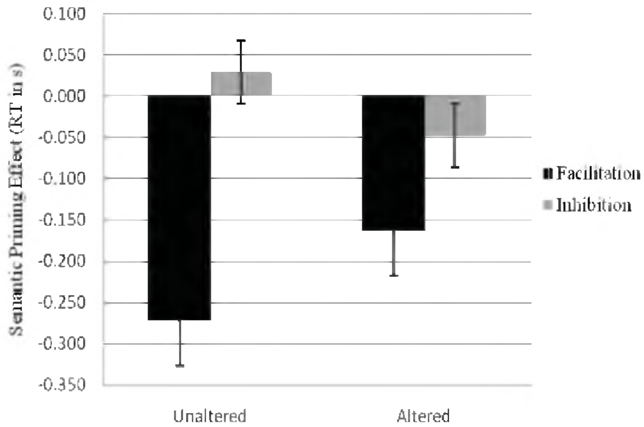


Fig.1. Semantic priming effects for unaltered contexts and contexts altered using low-pass filtering at 1000 Hz. Facilitation is plotted as the difference between mean RT's for congruent and neutral contexts. Inhibition is plotted as the difference between mean RT's for incongruent and neutral contexts. Error bars represent standard errors.

3.3 Facilitation and Inhibition

The finding of a significant main effect of context was explored further by examining facilitation and inhibition of the lexical decision by the semantic context of the sentence. Recall that facilitation is determined by calculating the difference between RT's in the congruent and neutral conditions, whereas inhibition is determined by calculating the difference between RT's in the incongruent and neutral conditions. Figure 1 and Figure 2 show the facilitation and inhibition effects for Group 1 and Group 2, respectively. As expected, the results for both groups are similar in the unaltered condition. Not surprisingly, facilitation was similar in the unaltered and altered conditions when less distortion was applied (Group 2), but facilitation was reduced relative to the unaltered condition when more distortion was applied (Group 1), even though the extent of inhibition was similarly reduced relative to the unaltered condition regardless of the degree of distortion. This illustrates the interaction revealed in the ANOVA above.

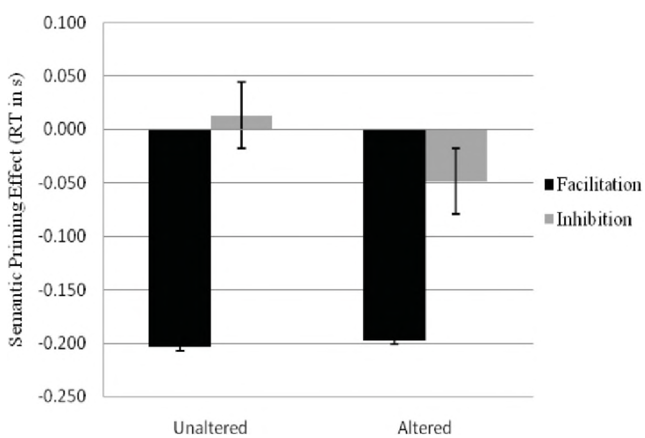


Fig.2. Semantic priming effects for unaltered contexts and contexts altered using low-pass filtering at 1750 Hz.

4. DISCUSSION

Acoustically distorting the semantic context using low-pass filtering disrupts the encoding of linguistic information as evidenced by changes in the speed of lexical decision for a following target word that was spoken without distortion. The extended response times arising from low-pass filtering contributed to a smaller facilitation priming effect that varied with the degree of distortion that was applied. As shown in Figures 1 and 2, this decrease in facilitation was more pronounced for experimental Group 1 as compared to Group 2. Low-pass filtering at 1750 Hz allows more high frequency spectral information, which could account for the smaller decrease in facilitation evidenced by experimental Group 2 as compared to Group 1. Thus, the amount and/or quality of available acoustical information seems to be more important to lexical access and lexical decision operations than is the type of distortion as was initially argued by previous researchers [2]. Further exploration of precisely how different degrees of acoustical distortion alter the speed and nature of lexical processing warrants further study. This research may provide insights into the effort needed to process words heard in non-ideal acoustical conditions that cannot be provided by traditional 'off-line' studies of word recognition because performance remains at ceiling.

REFERENCES

[1] Pelletier, M., Coletta, M., Giroux, R., Goy, H., & Pichora-Fuller, K. (this issue). The effect of types of acoustical distortion on lexical access. *Canadian Acoustics.*

[2] Aydelott, J., & Bates, E. (2004). Effects of acoustic distortion and semantic context on lexical access. *Language and Cognitive Processes, 19,* 29 – 56.

[3] Stanovich, K.E., & West, R.F. (1983). On priming by sentence context. *Journal of Experimental Psychology: General, 112,* 1 – 36.

[4] Neely, J.H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In D. Besner & G. W. Humphreys (Eds.), *Basic Processes in Reading: Visual Word Recognition* (pp. 264 – 336). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

[5] Boersma, P. & Weenik, D. "Praat: Doing phonetics by computer (version 4. 6. 20)" [computer program], 2007 Sep. 2 [cited 2008 July 8], Available HTTP: http://www. Praat.org

ACKNOWLEDGMENTS

This project was funded by CIHR and NSERC.

IS THERE A 'RISE-FALL TEMPORAL ARCHETYPE' OF INTENSITY IN ELECTROACOUSTIC MUSIC?

Roger T. Dean[1] and Freya Bailes,
MARCS Auditory Laboratories, University of Western Sydney, Australia, NSW 1797. roger.dean@uws.edu.au

1. INTRODUCTION

We seek to identify recurrent temporal structures in a range of musics, which may relate to affect. Huron proposed a 'ramp archetype' in intensity patterns, based on his study of Western classical score notations[2, 3]. He observed more notations for 'crescendo' (increases in loudness) than 'diminuendo' (decreases), and estimated that crescendi occupied greater lengths of notational time than diminuendi. Since the word ramp can refer to both rises and falls, we will discuss a possible 'rise-fall temporal archetype'. Huron proposed that rises occupy greater time than falls, and that this may optimise attention, since rises may share features of environmental sonic looming.

We made specific predictions from this theory: first, that within most successive rise-fall pairs, rises last longer than falls. Since there are limits to the maximum and minimum intensities, commonly the intensity range covers the whole possible range, and so rises and falls on average involve similar magnitude intensity changes. Thus we predict, second, the rate of intensity change during a (longer) ramp is smaller than that in the succeeding (shorter) damp. The critical feature for a listener is probably perceptual loudness, but here we investigate the only possible surrogate for it in the case of the complex sonic textures of computer music, physical sound intensity. Huron did not make acoustic measurements. We are undertaking computational analysis of recordings within various music genres, to asess whether these predictions are upheld, and to extend them to other possible analogues of such archetypes beyond dynamics. Here we present such analysis of an anthology of Canadian computer music.

2. METHOD

We studied the 25 short pieces presented in the 1st CD of the 'DISContact! II' anthology of the Canadian Electroacoustic (ea) Community(1995). The pieces (from 80 to 188 seconds in length) are not primarily notated, rather their embodiment is sonic; they depend little on enculturated Western devices such as tonality.

Praat [1]scripts analysed durations and intensity rate changes of successive rises and falls in the pieces. First, the mean intensity (SPL) over time frames of 40msec, 0.5, 5 and 10 seconds was measured and each peak and trough identified. By definition, every intensity peak is followed by a trough (the 'all peaks' measure). There were between 235 and 792 rise-fall pairs in the 40msec analyses of these pieces, but in some cases there were 0 in the 5 and 10 second window analyses. Second, 'significant peaks' and troughs were measured for the 40msec data. Here only peaks/troughs which differ from the immediately preceding peak/trough by >=1/7 (the 'dynamic step') of the range between the 10 and 90% quantiles of intensity means are recorded as significant. The 1/7 criterion parallels the dynamic ranges used in contemporary notated music (ppp-pp-p-mp-mf-f-ff-fff). Rise times and corresponding intensity increase rates are measured by cumulating the parameters of each immediate succession of peaks, and fall parameters for each succession of troughs. There were between 5 and 266 rise-fall pairs in the analyses. Only analyses 1 and 2 with 5 or more pairs were considered in the statistical analysis.

In a third analysis, we tested whether acoustic 'crescendi' are more common than 'diminuendi'. We used the 'significant' peaks and the 0.5s window, since musically significant crescendi and diminuendi operate over at least this time frame, and it permits the detection of any longer patterns. A 'crescendo' is an increase in intensity of > (¼*dynamic step) from the reference value current at any particular time (the most recent peak or trough); and a diminuendo is a comparable decrease. Successive values which oscillate within ±(¼*dynamic step) of the current reference are 'plateaux'. Here, as in musical notation and in contrast to the earlier determinations, if crescendi precede and succeed a plateaux, this is counted as two crescendi.

The main parameters determined were the duration of each pair of rises and falls in the 'all peaks' and 'significant peaks' approaches, using the indicated time windows; and the log-ratio of the times within each such pair. The overall times occupied by ramps and damps were also computed for each piece. Using the 'crescendo' approach, we determined the number of crescendi and of diminuendi (not necessarily paired). The statistical significance of the hypotheses that the ramp-damp time log-ratio and the log(crescendo-rate/diminuendo-rate) differ from 0, was tested for each sound file in each data set, using a two-tailed t-test. The hypothesis that numbers of crescendi and decrescendi differ significantly was tested for each sound file in each data set with the chi-square test. The hypothesis that overall ramps are on average longer than damps (i.e. not considering them as pairs) was tested using a two-tailed paired t-test. The alpha level was set at 0.05.

3. RESULTS

Table 1 shows a summary of the data. The hypothesis that rises are longer than paired falls is not supported: they are generally shorter. Concordantly, determinations of the average lengths of crescendi and decrescendi in individual pieces are in 34 cases indicative of shorter and only 7 of longer crescendi than decrescendi. Crescendi and diminuendi are not significantly different in number, with one exception, in which there are fewer crescendi. Of the 25 pieces, only 7 show a crescendo count greater than their diminuendo count. This is in contrast to Huron's observations of scored dynamics. Given that rises are shorter than falls, it is perhaps not surprising that rates of intensity change during crescendi are generally greater than those during the succeeding decrescendo. The global means and the mean of all significant determinations were as follows, respectively: log(paired time ratios): -0.04, -0.09; log (paired intensity change rate ratios) : 0.03, 0.15; ratio of total rise: fall time : 0.89, 0.91. Taking all the pieces together, there were 989 crescendi, and 1064 diminuendi.

	No. of paired log(time) ratios < 0	No. of paired log(time) ratios > 0	No. of paired log(change-rate) ratios < 0	No. of paired log(change-rate) ratios > 0	No. of pieces with total rise time < fall	No. of pieces with total rise time > fall
0.5secAllpeaks	4	1	1	0	3	0
5secAllpeaks	1	0	1	0	2	2
10secAllpeaks	1	0	1	0	1	0
Allpeaks	11	0	1	1	13	0
0.5secSigpeaks	3	2	2	0	6	2
5secSigpeaks	1	1	4	0	2	1
10secSigpeaks	0	0	0	0	0	0
Sigpeaks	8	1	9	3	7	2
Totals	29	5	19	4	34	7

Table 1. 'Allpeaks' and 'Sigpeaks' are defined in the text, and the seconds value indicates the window duration for the analyses (40msec when not stated). Only determinations which reach statistical significance ($p<0.05$) are shown. The composers represented are: Bouhalassa; Chuprun; Dhomont; Feist; Frigon; Gobeil; Jean; Kennedy; Koustrup; Leduc; Normandeau; Polen; Radford; Routhier/ Migone/ Coté; Roverselli; Schryer; Trudel; Winiarz; Bebris; Ciamaga; Cross; Cruickshank; Degazio; Buono; and Lekkas. The text gives additional data.

4. DISCUSSION

We observe a temporal archetype of intensity, in which rises take longer than their succeeding falls, and intensity changes faster during rises than falls. It will be interesting to see whether such acoustic features characterise longer pieces and Huron's classical canon, or whether they are particular to short works or ea music.

The features of the archetype, although different from prediction, may still relate to attention and the affective response. Intensity rises are arousing in rate- and length-dependent way [5]. So if a continual long increase gradually becomes less effective (habituation), rises might be shorter than falls, and increase rates might intensify with rise duration, or within a ramp, as does the looming of a sounding object approaching at constant velocity [4].

Our data may reflect the force-energy input to the music, judged by listeners as effort-loudness-affect (our 'FEELA' interpretation). Just as a performer uses force and effort to vary the intensity of their sound, an ea composer creates sonic counterparts to this. A listener may be influenced by the observed force and effort (FE) of a live performer, and perhaps gain similar stimulus range from sculpted ea sound. In each case they may perceive effort and loudness (EL) from the signal.

Loudness (closely related to measures of acoustic intensity) is a known correlate of the real-time affective (A) responses to certain classical music[6], and this relationship is probably more general. Thus we have perceptual evidence to link segmentation and affect in ea music; segments may be constituted by rises and falls. As yet it is not known whether loudness is an important determinant of affect in such music, but we suggest that it is: so a piece aiming for affective expression will use the intensity structures we describe (i.e. the FEELA process).

REFERENCES

[1]. Boersma, P. and Weenink, D., *Praat., v 4.6.01*. 1992-2007: www.praat.org.

[2]. Huron, D., *The ramp archetype: A study of musical dynamics in 14 piano composers*. Psychology of Music, 1991. **19**: p. 33-45.

[3]. Huron, D., *The ramp archetype and the maintenance of auditory attention.* Music Perception, 1992. **10**: p. 435-444.

[4]. Neuhoff, J.G., *Perceptual bias for rising tones*. Nature, 1998. **395**: p. 123.

[5]. Olsen, K.N., Stevens, C., and Tardieu, J., *A perceptual bias for increasing loudness: Loudness change and its role in music and mood*, in *International Conference on Music Communication Science*. 2007, http://marcs.uws.edu.au/links/ICoMusic/Full_Paper_PDF/Olsen_Stevens_Tardieu.pdf: Sydney.

[6]. Schubert, E., *Modelling perceived emotion with continuous musical features*. Music Perception, 2004. **21**: p. 561-585.

EFFECTS OF EMOTIONAL CONTENT AND EMOTIONAL VOICE ON SPEECH INTELLIGIBILITY IN YOUNGER AND OLDER ADULTS

Kate Dupuis[1] **and Kathy Pichora-Fuller**[1,2]
[1]Dept. of Psychology, University of Toronto, 3359 Mississauga Rd North, Mississauga, Ontario, Canada
[2]Toronto Rehabilitation Institute, 550 University Ave, Ontario, Canada

1. INTRODUCTION

Everyday speech contains emotional information which is conveyed through emotional inflection. (e.g., a disgusted tone of voice) and through the lexical content of the message (e.g., "I hate your dress."). Recent advances have been made in determining the exact nature of emotional representations in speech (e.g., [1]), and in describing the numerous methodologies which have been used to examine understanding of emotion based on both content and tone (e.g., [2]). Standardized speech tests are used extensively by clinicians and researchers to measure speech perception and spoken language understanding in different listening conditions and in people of all ages. However, the emotion represented in the tone and content of most speech intelligibility test materials seems to be either minimized or poorly controlled. Speech stimuli are typically recorded in an artificially neutral way, devoid of any affective cues. Additionally, the lexical content of many of the stimuli have emotional connotations which are not controlled for or equalized across lists. For example, the sentence "The airplane went into a dive." in the Speech Perception in Noise test [3] and the sentence "The young people are dancing." in the Hearing in Noise Test [4] can have negative and positive connotations, respectively.

Literature in the visual and audio-visual domains suggests that printed words or spoken sentences with an emotional meaning are more efficiently recognized and recalled better than stimuli with neutral meaning [5, 6]. Indeed, databases of words, pictures, and sounds have been developed to provide normative ratings to allow researchers to equate stimuli on the emotional characteristics of arousal and valence [7, 8, 9]. Nevertheless, most commonly-used speech tests were developed before the advent of such databases. The effect of the emotional implications of an utterance's emotional content and/or affective tone on intelligibility has not been controlled for or investigated. In addition, the interpretation of emotional cues in speech differs depending on age [10, 11], and it may be a factor influencing older adults' reported difficulties understanding speech in noise (e.g., [12]). Thus, performance on speech intelligibility tests may be influenced by an interaction of age with emotional content and/or tone, but this possibility has not been examined.

The first purpose of the current study was to determine the emotional valence and emotional arousal properties of 200 words spoken in a neutral tone of voice which were taken from a standardized test (Northwestern University Auditory Test No. 6; NU6 [13]). Each target word is presented in the carrier phrase "Say the word…". The relationship between the emotional ratings and the ability of listeners to identify the target words under varying signal-to-noise conditions was then examined; pilot data from younger adults will be described in the present paper. The secondary purpose of the study was to create an emotionally-spoken version of the NU6 words so that the effect of emotional speech production on intelligibility could be tested. A younger and an older female actor each re-recorded the 200 NU6 words in seven emotional tones of voice (happy, sad, neutral, angry, fearful, pleasantly surprised and disgusted). Numerous tokens of each stimulus were recorded and three raters chose the token which they felt best represented each emotion. The ratings of the three judges were compared to analyze inter-rater reliability. A final set of 2800 stimuli (200 sentences x 7 tones of voice x 2 speakers) will be presented to both younger and older listeners in subsequent studies to determine the accuracy with which each portrayed emotion can be identified and how emotion influences speech intelligibility.

2. EXPERIMENTS

2.1 Experiment 1

2.1.1 Method

To date, twenty-eight younger adults (mean age = 19.9 years, *SD* = 2.5) with good health and clinically normal hearing thresholds in the speech range have been tested. The stimuli used were the 200 sentences from the NU6 lists. These stimuli were presented to one group (N=12) visually as text on paper, and to two groups auditorally, through two loudspeakers in a sound-attenuating booth. In the auditory conditions, one group (N=9) heard the sentences presented by a recorded female voice [14] and the other group (N=7) heard presentations by a recorded male voice [15]. In all cases participants rated emotional valence from 1-9 (1 was "extremely negative"; 9 was "extremely positive") and emotional arousal from 1-9 (1 was "least arousing"; 9 was "most arousing"), either by circling the appropriate number in the visual conditions or by pressing the appropriate box on a touch-screen in the auditory conditions. The participants were tested individually and provided ratings for each of the 200 words.

2.1.2 Results

Means were obtained for both valence and arousal ratings for all three groups (visual, auditory female voice and auditory male voice). These means were then correlated with ratings of frequency, familiarity, and neighbourhood density for each word, as well as the mean level (in dB SNR) at which participants could reliably identify each word 50% of the time. The SNR threshold data were collected on young adults with normal hearing by Richard Wilson and colleagues using the female voice. Analyses revealed a significant positive correlation between valence and arousal for participants in the visual condition, $r = .314$, $p < .001$, and auditory female voice condition, $r = .360$, $p < .001$. However, participants who listened to the stimuli spoken by the male voice exhibited a negative correlation between valence and arousal, $r = -.143$, $p = .044$.

Furthermore, negative correlations were found between arousal rating and mean SNR threshold level only for the visual presentation group, $r = -.153$, $p = .031$. Neither valence nor arousal ratings correlated with ratings of word frequency, familiarity or neighbourhood density in any of the three conditions.

2.1.3 Discussion

This experiment was the first to our knowledge to gather emotional valence and arousal ratings for the target words of a test commonly used in speech audiometry. The results indicate that the emotional arousal of listeners to a particular word can affect intelligibility, depending on the modality of presentation. More arousing words can be reliably identified at lower SNR than less arousing words. However, this finding of a significant correlation between emotional rating and the SNR threshold for words only holds true for the emotional rating of words which were visually presented to participants. This modality-specific finding likely reflects the success of the auditory test developers in achieving recordings (in both male and female voices) which minimized the emotional response of listeners who rated the words. In contrast, those who read the words were free to react with an unrestricted emotional response during rating. The neutrality of tone of voice could, nevertheless, influence peoples' everyday understanding of the content of the word, as very rarely is natural speech devoid of emotional inflection. These data demonstrate the strong interaction between emotional tone and content in spoken language, but highlight that this aspect of natural communication has not been tapped by traditional speech tests which have been more focused on de-contextualized phonemic and lexical aspects of speech perception.

2.2 Experiment 2

2.2.1 Method

A younger (aged 26 years) and an older (aged 64 years) female actor were recruited from the community and consented to create voice recordings on separate occasions. Both actors spoke English as a first language and had clinically normal hearing thresholds in the speech range. Each actor recorded numerous tokens of 1400 sentences (200 stimuli x 7 emotional tones of voice -- happy, sad, neutral, angry, fearful, pleasantly surprised and disgusted) and instructions were given in an attempt to equate production styles across actors. The actors were given the 50 target stimuli from each of the four NU-6 lists on paper and were asked to speak each word in the carrier phrase "Say the word..." and to produce each item multiple times. The experimenter was present and directed the recording session. Tokens of each word spoken in each emotion were produced until the actors and experimenter were satisfied with the production of each word in each emotion.

2.2.2. Results

Using the Praat [16] acoustical analysis software, each token was saved as a separate sound file. Three English-speaking undergraduates listened to these stimuli and chose what they considered to be the best representation of the target emotion for each sentence from the set of tokens. Agreement was over 90% for at least two raters across all sentences.

2.2.3 Discussion

With knowledge that the emotional tone of presentation can have an effect on spoken language understanding and that this effect interacts with a listener's age, a novel battery of stimuli has been developed. In this battery, the 200 stimuli of the NU-6 test have been re-recorded by both a younger and an older female actor in seven different tones of voice. Agreement on the specific tokens which are the best representations of a particular emotion is high. The creation of these materials is the first step towards creating an emotional-tone speech intelligibility task which will enable us to determine whether the way in which a sentence is presented (e.g., in a sad or fearful tone of voice) will affect how well a listener can perceive and understand it.

REFERENCES

[1] Cowie, R. & Cornelius, R.R. (2003). Describing the emotional states that are expressed in speech. *Speech Communication, 40,* 5-32.

[2] Scherer, K.R. (2003). Vocal communication of emotion: A review of research paradigms. *Speech Communication, 30,* 227-256.

[3] Bilger, R.C., Nuetzel, J.M., Rabinowitz, W.M., & Rzeczkowski, C. (1984). Standardization of a test of speech perception in noise. *J Speech and Hearing Research, 27,* 32-48.

[4] Nilsson, M., Soli, S., & Sullivan, J. (1994). Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise. *J Acoustical Society of America, 95,* 1085-1099.

[5] Hamann, S. (2001). Cognitive and neural mechanisms of emotional memory. *Trends in Cognitive Sciences, 5,* 394-400.

[6] Cahill, L. & McGaugh, J.L. (1995). A novel demonstration of enhanced memory associated with emotional arousal. *Consciousness and Cognition, 4,* 410-421.

[7] Bradley, M.M., & Lang, P.J. (1999a). *Affective norms for English words (ANEW): Instruction manual and affective ratings.* Technical Report C-1,

[8] Ito, T.A., Cacioppo, J.T., & Lang, P.J. (1998). Eliciting affect using the International Affective Picture System: Bivariate evaluation and ambivalence. *Personality and Social Psychology Bulletin, 24,* 856–879.

[9] Bradley, M.M., & Lang, P.J. (1999b). *International affective digitized sounds (IADS): Stimuli, instruction manual and affective ratings (Technical Report B-2).* Gainesville, FL: The Center for Research in Psychophysiology, University of Florida.

[10] Kiss, I., & Ennis, T. (2001). Age-related decline in perception of prosodic affect. *Applied Neuropsychology, 8,* 251-254.

[11] Orbelo, D. M., Testa, J. A., & Ross, E. D. (2003). Age-related impairments in comprehending affective prosody with comparison to brain-damaged subjects. *J Geriatric Psychiatry and Neurology, 16,* 44-52.

[12] Committee on Hearing, Bioacoustics and Biomechanics. (1988). Speech understanding and aging. *J Acoustical Society of America, 83,* 859-895.

[13] Tillman, T. W., Carhart, R. (1966). An expanded test for speech discrimination utilizing CNC monosyllabic words: Northwestern University auditory test no. 6. Technical report no. SAM-TR-66-135. San Antonio, TX: USAF School of Aerospace Medicine, Brooks Air Force Base.

[14] Department of Veteran Affairs (1991). *Speech recognition and identification materials, Disc 1.1 (CD).* Auditory Research Laboratory V. A. Medical Center, Long Beach, CA.

[15] Wilson, R. H., Coley, K. E., Haenel, J. L., & Browning, K. M. (1976). Northwestern University Auditory Test No. 6: Normative and comparative intelligibility functions. *J American Audiology Society, 1,* 221-228.

[16] P. Boersma and D. Weenik, "Praat: Doing phonetics by computer (version 4.5.08)" [Computer program], 2006 Apr 4 [cited 2008 Jun 29], Available HTTP: http://www.praat.org/

VARIABILITY OF THE CONSONANT MODULATION SPECTRUM ACROSS INDIVIDUAL TALKERS

Pamela Souza[1] and Frederick Gallun[2]

[1]Dept. of Speech & Hearing Sciences, University of Washington, 1417 NE 42nd Street, Seattle, WA, USA, 98105, psouza@u.washington.edu

[2]National Center for Rehabilitative Auditory Research, Portland VA Medical Center, 3710 SW US Veterans' Hospital Road, Portland, OR, 97239, USA, Frederick.Gallun@va.gov

1. INTRODUCTION

Our previous work (Gallun & Souza, 2008) demonstrated that vowel-consonant-vowel syllables with similar modulation spectra were likely to be confused with one another. This implied that each consonant is identified by its modulation spectrum. However, if the modulation spectrum for a particular phoneme varies substantially across talkers, that would argue against use of those cues for consonant identification. This follow-up study investigated variability of consonant modulation spectra across individual talkers.

2. METHOD

2.1 Materials.

Speech recordings were drawn from a database (Markham & Hazan, 2002) which included adult male and female, and child male and female talkers. Each talker produced the same set of vowel-consonant-vowel tokens. This included 23 consonants in three vowel contexts (/ɑ/, /μ/, and /i/). The tokens were recorded on digital audio tape and transferred to a computer hard drive for analysis.

2.2 Modulation spectrum and spectral correlation index.

Details of the processing are available in Gallun and Souza (2008). Briefly, the modulation spectrum of each signal was calculated by (a) filtering the VCV into six octave-wide bands centered at .25, .5, 1, 2, 4, and 8 kHz (b) half-wave rectifying each of the filtered signals (c) low-pass filtering the rectified signal at 50 Hz (d) downsampling at 1000 Hz and completing a Fast-Fourier Transform (FFT). Prior to the FFT, the signal was zero-padded such that the duration of the signal was extended to five seconds, thus allowing a frequency resolution in the FFT of .2 Hz (e) the energy in each .2 Hz bin between .2 and 64 Hz was summed with the energy in adjacent bins. The choice of which bins to sum was made such that the summed energy was obtained for the equivalent of six, 1-octave wide rectangular filters with center frequencies stretching from 1 Hz to 32 Hz (f) the summed energy value was divided by the energy in the 0 Hz bin to provide a normalized modulation index value, which indicates the relative amount of modulation. Thus, each phoneme was represented by a matrix of thirty-six values (six modulation frequencies x six carrier frequencies). Note that there is no information about the relative phases of the modulation across channels. Figure 1 shows an example of the modulation spectrum for the /aza/ (for figure clarity, only three of the six bands are shown).

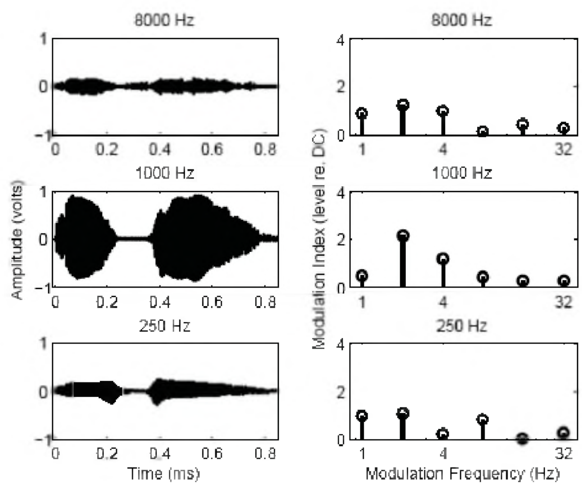


Figure 1. Example of modulation spectra (the .25 kHz, 1 kHz and 8 kHz carrier bands are shown) for the syllable /aza/. Left panels show the output of the first filtering stage; right panels show the amount of modulation at each modulation frequency from 1 to 32 Hz, with larger modulation index values indicating greater modulation.

Next, Spectral Correlation Index (SCI) (Gallun & Souza, 2008) values were calculated across subsets of the stimulus set. The SCI between two signals is obtained by concatenating the 36 modulation spectrum values into a single vector and correlating (Pearson *r*) the vectors. In this way, a single SCI value can be obtained for each pairing of phonemes.

3. RESULTS

3.1 Consonant similarity

Consistent with Gallun and Souza (2008), there was a wide range of SCIs across different consonant tokens produced by the same talker. Figure 2 shows an example for a single adult male talker, with all consonants compared to /aza/. High SCI values indicate the consonants that have the most similar modulation spectrum to /aza/ (e.g., /aʃa/, /adʒa/) and the least similar modulation spectrum to /aza/ (e.g., /apa/, /afa/). The phoneme /aza/, correlated with itself, has SCI = 1.0.

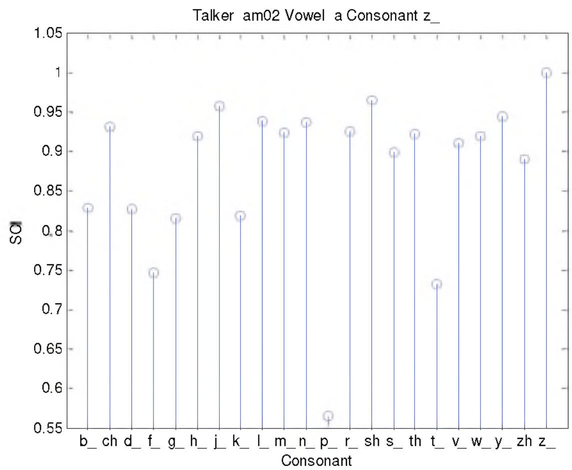


Figure 2. SCI values for a single male talker for all consonants compared to /z/, for the /ɑ/ vowel context. SCI ranges from about 0.5 (for /p/) to 1.0 (for /z/, correlated with itself).

3.2 Talker variability

Results indicated that the modulation spectrum for a single VCV token was very similar across talkers. Typical results are shown for /aza/ in Figure 3. This shows how similar the modulation spectra are (expressed as SCI) for each adult male talker relative to adult male talker #11 (who has a SCI value of 1 with himself). Despite the range in voice pitch, vocal quality, and the fact that all talkers were untrained and produced the VCVs spontaneously, SCI values for all talkers are 0.9 or higher. This supports the idea that consonant identification is based on modulation characteristics which are maintained across individual productions of the consonant.

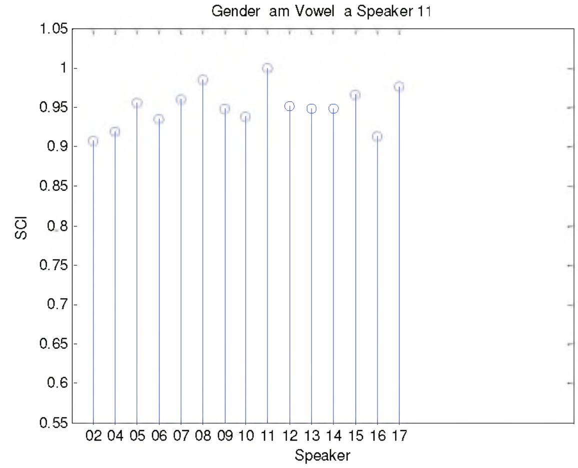


Figure 3. SCI values for all adult male talkers compared to adult male #11, for /aza/. The high (> 0.9) values indicate that all talkers produced /aza/ with a similar modulation spectrum.

4. DISCUSSION

Our previous work showed that similarity of the modulation spectrum, as expressed by the SCI, is significantly related to consonant error patterns. That suggests that modulation spectrum can be used to characterize the temporal (modulation) cues to consonant identity. This study examined how the modulation spectrum varies as the same utterance is produced by a variety of talkers. In general, modulation spectra for the same consonant across speakers are very similar, compared to modulation spectra for different consonants.

REFERENCES

Gallun F., Souza P. (2008). Exploring the role of the modulation spectrum in phoneme recognition. Ear and Hearing, in press.

Markham D., Hazan V. (2002). Speaker intelligibility of adults and children: The UCL Speaker Database. Speech, Hearing and Language: work in progress (pp. 1685-1688).

ACKNOWLEDGMENTS

The authors gratefully acknowledge the assistance of Valerie Hazan and University College London in providing the acoustic recordings and Eric Hoover for his help with stimulus processing. This work was supported by NIDCD (DC 006014 and DC04661), the National Center for Rehabilitative Auditory Research, and the Bloedel Hearing Research Center.

‘*I SAID MADE* (MATE?)’: CATALAN/SPANISH BILINGUALS’ PRODUCTION OF ENGLISH WORD-FINAL OBSTRUENTS

Natalia Fullana[1] and Ian R. A. MacKay[2]
Department of Linguistics, University of Ottawa, 70 Laurier Ave. E., Ottawa, Ontario, Canada, K1N 6N5
[1]fullnat@gmail.com, [2]imackay@uottawa.ca

1. INTRODUCTION

The production of the voicing contrast in English word-final obstruents often poses difficulties for native speakers (NSs) of Romance languages [e.g., 1, 2, and 3], which have largely been attributed to the non-occurrence of voiced obstruents (and/or to the lower frequency of consonants) in word-final position in such languages as Spanish and Catalan. Thus in the implementation of the voicing contrast in English, Romance language speakers have been found to make use of first language (L1) production rules. Specifically, Spanish NSs have been observed to devoice, spirantize, or delete English /b d g/ in word-final position, as indicated by listener transcriptions [1; see also 4]. Besides, acoustic analyses in [5] showed that Spanish speakers further differed from English NSs (E-NSs) in the production of the /t/-/d/ contrast in word-final position in their failure to produce a significantly greater vowel duration difference in /d/ vs /t/ and to maintain voicing in the closure phase of /d/ longer. Spanish speakers were not successful, either, in producing a longer closure phase duration and higher F1 offset frequency for /t/. Similarly, Catalan NSs have been reported to resort to L1 production rules in the realization of the voicing contrast in final obstruents. For instance, [6] noted that English /b d g v ð z ʒ dʒ/ in absolute final position were produced as devoiced 94% – 100% of the time, hence illustrating Catalan NSs’ use of the final obstruent devoicing (FOD) rule. As expected, the voiceless counterparts, /p t k f θ s ʃ tʃ/, were realized as intended in 100% of instances.

In addition to differences in the phonetic inventories between the second language (L2) and the L1, age of onset of L2 learning (AOL) and experience in the L2 might account for Romance language NSs’ (non)native-like production of English final obstruents. The Speech Learning Model (SLM) [7] hypothesizes that an earlier AOL results in better discernment of phonetic differences between L1 and L2 sounds [e.g., 3]. It also suggests that learners might succeed in producing L2 sounds that do not exist in certain allophonic positions of their L1 sound inventory — or at least they approximate NS norms [5]. However, studies conducted in formal learning contexts [8] do not tend to uphold AOL and experience effects as put forward by the SLM.

The present study aimed to further explore the potential effects of age of onset of L2 learning and exposure to English on the production of the voicing contrast in English word-final obstruents /p b t d s z/ by Catalan/Spanish (Cat/Sp) bilingual learners of English in a formal learning context.

2. METHOD

2.1 Participants

Forty-seven Cat/Sp bilinguals (41 female, 6 male) participated. They were undergraduate students in English Teacher Education at the University of Barcelona (mean age = 23.1 years). They differed in age of onset of English learning (before, at, or after age 8), and in exposure to English (school exposure only vs extra exposure through language courses). Four British E-NSs (1 female, 3 male) were included in the study as a control group.

2.2 Materials and procedure

English /p b t d s z/ were elicited in two different words each (*rope*, *tripe*; *robe*, *tribe*; *mate*, *sight*; *made*, *side*; *ice*, *pace*; *eyes*, *pays*) embedded in the carrier phrase “*I said* ___” of a delayed sentence repetition task. The mini-dialogues comprising the task (e.g., *What did you say? I said made*) were modeled by two taped British E-NS voices. Only after hearing the first question-answer exchange were participants asked to repeat the (answer) carrier phrase. Participants were recorded in a quiet room at the phonetics lab by means of a CASIO DA-7 DAT tape recorder and a YU-Brother EM-106 microphone. The resulting recordings were transferred in raw format onto a computer and later saved as 48 kHz, mono, 16-bit WAV files with *CoolEdit2000*.

2.3 Acoustic analysis

Acoustic measurements of vowel, closure phase and fricative duration were made using spectrographic and oscillographic displays in Praat [9]. The low-energy period during the closure phase of stops until the release burst defined the closure duration of stops (94% of stops were released). Fricative duration was measured from the beginning to the end of the typical high-energy profile. Vowel duration was measured from the first peak of periodic energy indicating the vowel onset to the last peak of periodic energy that coincided with a significant drop in energy. Vowel quality was also analyzed in a subset of productions by measuring F1 and F2 at two points: after consonant transitions (where applicable) preceding the vowel and before the start of final obstruent transitions.

3. RESULTS

Mean results for the acoustic measurements of stop closure phase, fricative, and preceding vowel durations are shown in Table 1. As expected, E-NSs produced vowels preceding voiced stops with longer mean duration (153.23 ms) than vowels preceding voiceless stops. Likewise, the average duration of vowels before /z/ was longer (203.95 ms) than before /s/. Accordingly, the closure phase duration of /p/ and /t/ was longer (27.75 ms) than that of /b/ and /d/. So was the mean duration of the voiceless fricative (108 ms) vs its voiced counterpart.

Cat/Sp bilinguals implemented duration differences in their production of vowels before voiced vs voiceless obstruents. However, the extent of those differences was much smaller than E-NSs’ (mean range = 43.17–88.16 ms). As for consonant (closure phase or frication) duration, there was barely a length difference in voiced vs voiceless obstruents (mean range = 1.47–7.24ms). Moreover, contrary to E-NSs, Cat/Sp bilinguals produced a longer duration of closure phase of /b/ than /p/ in *rope-robe* (11.33 ms) and of /z/ vs /s/ in *pace-pays* (3.76 ms). Mann-Whitney U tests revealed that vowel duration differences between E-NSs and Cat/Sp bilinguals before voiced vs voiceless segments were significant (U 0–16, $p < .05$). Differences in consonant duration also turned out to be significant in fricatives and in the closure phase of /p/-/b/ in *rope-robe* (U 0, $p < .05$).

Table 1. Acoustic measurements of vowel, closure phase and fricative durations (in ms).

Obstruent	Preceding vowel duration		Consonant duration	
	Cat/Sp	English	Cat/Sp	English
/p/	153.11	152.87	119.12	87.12
/b/	199.85	289.12	129.95	57.54
/t/	198.04	181.62	113.20	67.87
/d/	262.75	311.75	104.26	42.50
/s/	235.05	181.66	204.81	259.87
/z/	313.37	380.50	203.44	152

Cat/Sp bilinguals' production of English word-final obstruents as a function of age of onset of L2 learning and exposure did not yield any significant differences according to Kruskal-Wallis and Mann-Whitney U tests. Significant differences occurred only between the various age or exposure groups and the English control group. In all cases, the differences had to do with voiced segments, specifically with closure phase and fricative durations. Although Cat/Sp bilinguals produced vowels before voiceless obstruents with a similar duration to that of E-NSs, they tended to lengthen both closure phase and fricative (as well as vowels before voiced obstruents) in the same way for voiceless as for voiced consonants (see Figure 1). Further evidence of Cat/Sp bilinguals' nonnative-like production was found in the preliminary analyses of vowel quality, wherein, unlike E-NSs' typical pattern, the diphthongs were not consistently different between those preceding voiced versus voiceless obstruents.

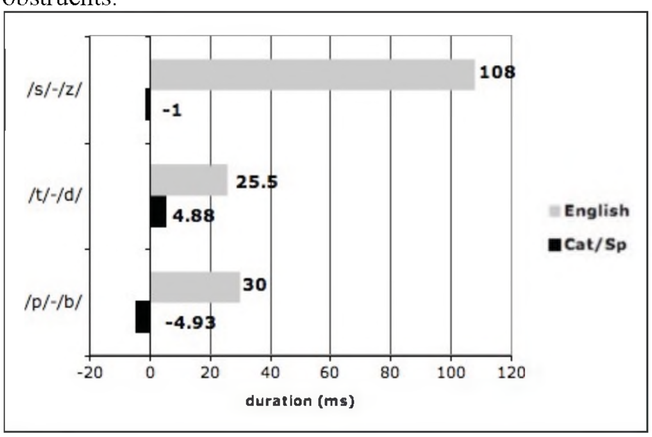


Figure 1. Mean closure phase and fricative duration differences (in ms).

4. DISCUSSION

This study looked at Catalan/Spanish bilinguals' production of the voicing contrast in English obstruents /p b t d s z/ in word-final position. The results obtained further supported previous findings of Romance language speakers' difficulties in producing the voicing contrast in English accurately [1, 3, 6]. In particular, Cat/Sp NSs failed to produce voiced segments at native-like levels. Instead, they appeared to resort to L1 production rules such as FOD, as in [6]. To a certain extent, the finding of Cat/Sp speakers' production of longer vowel duration before voiced obstruents than voiceless consonants is in line with [5] and [10] and might be taken as learners' resembling E-NSs. However, Cat/Sp bilinguals exhibited non-English production patterns in making not only /p t s/ duration longer, but also in realizing the closure phase of /b d/ and the frication of /z/ with considerably longer durations than E-NSs.

The lack of significant differences or a clear pattern in the production of English word-final obstruents as the result of varying ages of onset of L2 learning and degrees of exposure to English corroborated previous findings of formal learning contexts [8], while failing to conform to SLM's predictions.

Further analyses on vowel quality and closure voicing, in addition to native English listeners' perceptions, might provide a greater insight into Catalan/Spanish bilinguals' overgeneralized use of longer and undistinguished duration in their implementation of voiced and voiceless English obstruents in word-final position.

REFERENCES

[1] Flege, J. E., & Davidian, R. D. (1984). Transfer and developmental processes in adult foreign language speech production. *Applied Psycholinguistics, 5,* 323-347.

[2] Flege, J. E., & Hillenbrand, J. (1986). Differential use of temporal cues to the /s/-/z/ contrast by native and non-native speakers of English. *JASA, 79,* 508-517.

[3] Flege, J. E., Munro, M. J., & MacKay, I. R. A. (1995). Effects of age of second language learning on the production of English consonants. *Speech Communication, 16,* 1-26.

[4] Yavaş, M. (1994). Final stop devoicing in interlanguage. In M. Yavaş (Ed.), *First and second language phonology* (pp. 267-282). San Diego, CA: Singular.

[5] Flege, J. E., Munro, M. J., & Skelton, L. (1992). Production of the word-final English /t/-/d/ contrast by native speakers of English, Mandarin, and Spanish. *JASA, 92,* 128-143.

[6] Cebrian, J. (2000). Transferability and productivity of L1 rules in Catalan-English interlanguage. *SSLA, 22,* 1-26.

[7] Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 229-273). Timonium, MD: York Press.

[8] Fullana, N., & Mora, J. C. (2008). Production and perception of voicing contrasts in English word-final obstruents: Assessing the effects of experience and starting age. In A. S. Rauber, M. S. Watkins, & B. O. Baptista (Eds.), *New Sounds 2007: Proc 5th International Symposium on the Acquisition of Second Language Speech* (pp. 207-221). Florianópolis, Brazil: Federal University of Santa Catarina.

[9] Boersma, P., & Weenink, D. (2007). Praat: Doing phonetics by computer (Version 4.6.05) [Computer program]. Retrieved June 2, 2007, from http://www.praat.org

[10] Hogan, J. T., & Rozsypal, A. J. (1980). Evaluation of vowel duration as a cue for the voicing distinction in the following word-final consonant. *JASA, 67,* 1764-1771.

ACKNOWLEDGMENTS

Work supported by a postdoctoral fellowship to the first author from the *Secretaría de Estado de Universidades e Investigación del Ministerio de Educación y Ciencia* and the *Fundación Española para la Ciencia y la Tecnología* (Spain). The authors would like to thank the *Laboratori de Fonètica Aplicada* of the *Universitat de Barcelona* as well as all the participants in the study.

EFFECTS OF TRAINING MODALITY ON AUDIO-VISUAL PERCEPTION OF NONNATIVE SPEECH CONTRASTS

Yue Wang[1], Dawn Behne[2], and Haisheng Jiang[1]
[1]Dept. of Linguistics, Simon Fraser University, 8888 University Drive, Burnaby, Canada, V5A 1S6
[2]Dept. of Psychology, Norwegian University of Science and Technology, Dragvoll, 7491, Trondheim, Norway

1. INTRODUCTION

Speech perception often involves integrated auditory and visual modalities [1,2]. While nonnative perceivers can be facilitated by visual information when perceiving L2 sounds just as the natives, they may also be impeded in correct use of L2 visual cues, as they are not sensitive to the visual cues non-existent in their L1 [3,4,5,6]. Further research indicates that sensitivity to the visual information in L2 sounds can be enhanced through audio-visual (AV) training [3,7,8,9]. Findings also raise the question of the relative effectiveness of AV and V-only training. On the one hand, AV training may be preferred as learners presumably need to extract information from as many input channels as possible to perceive the challenging nonnative sounds [7]. On the other hand, V-only training may force the learners to focus more on the visual speech cues, and therefore reduce the cognitive load required to attend to both auditory and visual input [3]. However, the results from the previous separate AV and V-only training studies cannot be directly compared as they involve different training materials and sessions [9].

On the basis of these previous findings, the current study explored the effects of training on the perception of nonnative speech sounds using A, V, and AV training input modalities. Specifically, we examined how Mandarin learners of English could learn to perceive the A and V cues to the English interdental fricatives non-existent in Mandarin, compared to their familiar labiodentals and alveolars. A study with Mandarin learners is of particular interest. Since Mandarin perceivers demonstrate a lesser degree of attentiveness to visual speech information in their L1 compared to perceivers of other languages [5,10], questions arise as to how this lack of visual attentiveness affect their perception of L2.

2. METHOD

Forty-four young adult Mandarin Chinese natives with less than five years' residency in Canada participated in the study. They were randomly assigned to a control group (n=11) and three training groups (n=11 per group), with each training group receiving training with a different speech input modality: A, V, and AV. All participants took the same pre- and posttest during which they were presented with stimuli from three modalities: A, V, and AV.

Pre/posttest stimuli were recorded of an adult male speaker of Canadian English. The stimuli were based on 18 English CV syllables having a fricative followed by a vowel: [fi, fa, fu, vi, va, vu, θi, θa, θu, δi, δa, δu, si, sa, su, zi, za, zu]. The fricatives differed in place of articulation, (POA: labiodental, interdental, alveolar). The participants′ task was to identify the fricatives while listening to the sounds over a headset, or viewing the speaker mouth movements on the screen, or both.

The perceptual training program followed the high variability procedure demonstrated to be highly effective in auditory perception training [11,12]. Learners were trained to identify the target fricative contrasts appearing in a variety of phonetic contexts, in both natural and nonsense words, and produced by four native speakers of Canadian English (2 males, 2 females). For each training trial, the trainees' task was identification followed by feedback. The stimuli were presented as A-only for the A-train group, V-only for the V-train group, and AV for the AV-train group. The training was completed in two weeks, including six sessions of 45 minutes each.

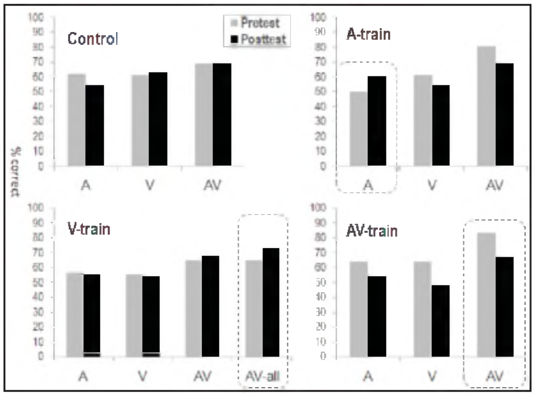
3. RESULTS

Percent correct identification of the fricatives at pretest and posttest was analyzed using a 4-way mixed analysis of variance (ANOVA), with Group (Control, A-train, V-train, AV-train) as a between-subject factor, and Test (pretest, posttest), POA, (labialdental, interdental, alveolar), and Modality as repeated measures. The dependent variable was perceivers' correct identification for POA regardless of voicing since POA was the focus of interest.

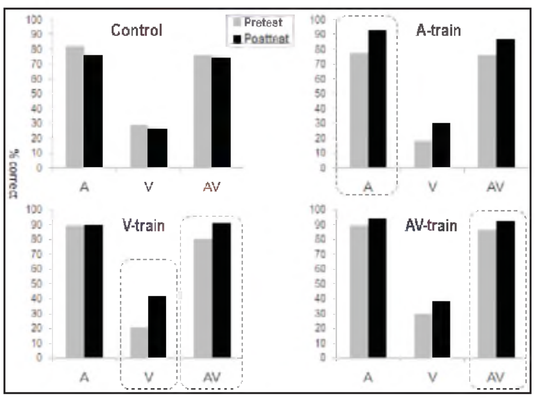
The results showed significant interactions of Group x Test [$F(3, 40)=3.15$, $p<.035$], Group x Test x Modality [$F(6, 80)= 2.30$, $p<.042$], and Group x Test x Modality x POA [$F(12, 160)=22.92$, $p<.001$]. Therefore, data were further analyzed with sets of one-way ANOVAs for each Group to compare the pre- and posttest performance in each Modality and POA. The results of these comparisons are displayed in Figure 1. Since no reliable differences were observed for labiodental perception across groups and tests, the results were excluded.

For the interdentals (Figure 1a), significant improvement from the pretest (50% correct identification) to the posttest (60% correct) was observed for the A-train group in the A modality when only the audio stimuli were presented [$F(1,10)=5.2$, $p<.046$]. For the V-train group, there revealed no reliable test difference across modalities. However, a more detailed analysis (taking voicing into consideration) revealed an increase post-training in perceiving the AV modality for both voiceless (from 65% to 73%) and voiced (from 47% to 59%) interdentals. For the AV-train group, a significant decrease (83% to 67%) was unexpectedly observed in the perception of the AV modality [$F(1,10)=7.6$, $p<.020$].

For the alveolars (Figure 1b), significant improvements were observed for the following groups and modalities: (1) A-train group with A modality [from 77% to 93%; $F(1,10)=7.9$, $p<.019$]; (2) V-train group with V modality [from 21% to 42%; $F(1,10)=31.9$, $p<.001$], and AV modality [from 80%, to 91%; $F(1,10)=7.9$, $p<.019$]; and (3) AV-train group with AV modality [Pretest: 86%, Posttest: 92%; $F(1,10)=5.7$, $p<.038$].



(a) Interdentals



(b) Alveolars

Fig. 1. % correct responses for (a) interdental and (b) alveolar fricative perception at pretest and posttest by perceivers in Control, A-train, V-train, and AV-train groups. (Significant pre- and posttest differences are circled. AV-all: % correct for both POA and voicing.).

4. DISCUSSION AND CONCLUSIONS

The results revealed a noticeable effect of training modality, where the extent of post-training improvement was consistent with the type of training; that is, the A-train group improved most in the perception of the A modality, whereas the V-train group improved most in the perceiving the V or AV modality. An exception to this general pattern of improvement was the decreased correct AV perception of the interdentals after AV training. The result appears to support the view that training with both auditory and visual input may add cognitive load to the learners, resulting in poorer performance [3]. This implicates that at the initial stage (especially with less advanced L2 perceivers learning a difficult L2 contrast), focused training with a single modality may be more effective than with multi-modalities which may be adopted at a later stage.

Regarding POA, the results showed a greater degree of training effect on the perception of the less visually distinct alveolars, as compared to that of the interdentals. Indeed, it has been speculated that learners (such as Mandarin) whose L1 possesses relatively low visual influence may not easily attune to the visual speech cues and therefore need to be trained to attend more to the visual information [3]. The results support this speculation, that despite the alveolars are familiar to the Mandarin perceivers, their perception can still benefit from training.

In sum, the current findings are consistent with the previous research showing the effectiveness of training on the AV perception of L2 speech [3,7,8], with the degree of training effect corresponding to factors such as perceivers' L1 experience, training input modality, and the visual salience of the speech segments.

REFERENCES

[1] Massaro, D. W. (1987). *Speech perception by ear and eye: A paradigm for psychological inquiry*. Erlbaum: Hillsdale, NJ.

[2] Jongman, A., Wang, Y., and Kim, B. (2003). Contributions of semantic and facial information to perception of nonsibilant fricatives. *JSLHR* 46, 1367-1377.

[3] Hazan, A., Sennema, A., Iba, M., & Faulkner, A. (2005). Effect of audiovisual perceptual training on the perception and production of consonants by Japanese learners of English. *Speech Commun.* 47, 360-378.

[4] Hazan, V., Sennema, A., Faulkner, A., & Ortega-Llebaria, M. (2006). The use of visual cues in the perception of non-native consonant contrasts. *JASA* 119, 1740-1751.

[5] Sekiyama, K. (1997). Cultural and linguistic factors in audiovisual speech processing: The McGurk effect in Chinese subjects. *Percept. Psychophys.* 59, 73–80.

[6] Werker, J. F., Frost, P. E., & McGurk, H. (1992). Cross-language influences on bimodal speech perception. *Can. J. Psychol.* 46, 551-568.

[7] Hardison, D. M. (2003). Acquisition of second-language speech: Effects of visual cues, context, and talker variability. *Appl. Psycholing.*, 24, 495-522.

[8] Hardison, D. M. (2005). Second-language spoken word identification: Effects of perceptual training, visual cues, and phonetic environment. *Appl. Psycholing.* 26, 579-596.

[9] Chen, Y., & Hazan, V. (2007). Language effects on the degree of visual influence in audiovisual speech perception. *Proceedings of the 16th ICPhS*, Saarbrueken, 2177-2180.

[10] Burnham, D., Lau, S., Tam, H., & Schoknecht, C. (2001). Visual discrimination of Cantonese tone by tonal but non-Cantonese speakers, and by non-tonal language speakers. *Proceedings of AVSP 2001*, 155-160.

[11] Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *JASA* 89, 874–886.

[12] Wang, Y., Spence, M. M., Jongman, A., & Sereno, J. A. (1999). Training American listeners to perceive Mandarin tones. *JASA* 106, 3649–3658.

ACKNOWLEDGEMENTS

We thank Angela Cooper, Nina Leung, Rebecca Simms, and Jung-yueh Tu for their assistance. This project was funded by SSHRC.

Learning Mandarin Tones at Sentence Level through Training: A Pilot Study

Xinchun Wang
Department of Linguistics, California State University, Fresno, 5245 N Backer Ave. M/S PB 92, Fresno, California 93740 USA

1. INTRODUCTION

Several previous studies have indicated that beginning level Mandarin learners improved significantly in identification and production of individual Mandarin tones in isolation after taking 2-3 weeks of training [1, 2, 3, 4]. However, perception and production of Mandarin tones on larger linguistic units such as phrases and sentences still poses significant challenges to learners who made progress on isolated words. No training studies on lexical tones have explored the effect of training beyond word level. This pilot study investigates the effect of perception with production training for learning Mandarin tones at sentence level. The research question is: will perception with production training using auditory and visual input be effective for learning Mandarin tones at sentence level?

2. METHOD

2.1. Participants

The participants were seven trainees (2 male, 5 female, mean age= 20) and 5 control subjects (3 male, 2 female, mean age = 20). All were beginning level Mandarin learners enrolled in a second semester Chinese course in a US public university. The participants' first languages (L1) include English (4), Hmong (4), and Japanese (4).

2.2. Stimuli

The stimuli used for pre- and post test were three sentences. Two of the sentences were statements and one was a simple question. The mean length of the sentence was 11 characters. The training stimuli consisted of 15 sentences and 48 phrases produced by four native Mandarin speakers (2 male, 2 female).

2.3. Procedure

For pretest, the participants read a list of 40 Chinese phrases and 10 sentences written in characters with pinyin (only 3 sentences were used for the current study). The recordings were made on a PC computer using GoldWave software through a Shure SM 48 microphone connected to an M-Audio MobilePre USB preamplifier. The readings were recorded and saved at a sampling rate of 22050 Hz with 16-bit resolution.

Individual training sessions were performed on a PC computer using Kay Elemetrics Sona Speech II Software as well as the same equipment mentioned in the above. During the training, the trainee opened and played back (through a pair of headphones) each training stimulus with real time display of the pitch contour in the top window of Screen A on the computer screen. The trainee then repeated the target phrase or sentence and recorded his/her own production of the target sentence by speaking into the microphone. The pitch contour of the trainee's production was instantly displayed in the bottom window of Screen B. The trainee could then compare his/her own production with the target phrase/sentence by playing them back repeatedly (auditory input). The trainee could also visually compare the tones by overlaying the pitch contour of the target sentence on that of his/her own production in different colors while alternately playing them back for auditory comparisons.

The training stimuli were blocked by speaker producing four training blocks each of which consisted of 15 sentences and 48 phrases. The trainees recycled the four speakers' phrases and sentences during the six hours of training that spread across 3-4 weeks. Immediately after the training, the trainees took the post test in which they repeated the tasks of the pretest. The control subjects took the pretest and post test at the same time interval but did not take the training.

2.4. Mandarin Listeners' Judgment

Four native Mandarin listeners (2 male, 2 female) residing in the US judged the speakers' productions of three sentences in a goodness rating task. The speakers' pretest and post test sentences were mixed and blocked by sentence. Two native Mandarin speakers' productions of the same sentences were also included in the rating stimuli. In each sentence block, 28 sentence stimuli were presented (1 sentence × 2 repetitions × 14 speakers). Individual rating tasks were performed on a Mac computer using custom designed software. The listeners rated each sentence along a continuum of 1 - 9 where 1 is labeled as "native like tones" and 9 as "very poor tones". They were told to pay attention to tones only while rating. Each listener had a trial session to learn the test procedure before the real rating tasks began. The order of the three sentences being rated was counter-balanced.

3. RESULTS

The mean rating for each of the Mandarin speakers' three sentences was 1.1. Not a single native production was rated above 2, indicating the raters were able to distinguish native from nonnative speech. As the research goal was not to compare native with nonnative production of tones, these Mandarin speakers' sentence data were not included in the analysis. Inter-rater reliability for the Trained and Control groups' productions were computed by sentence for the four raters by using Cronbach's α. The α values for the three sentences were .754, .702, and .831 respectively, which were all acceptable. (In general, the cut-off point of Cronbach's α is .70, above which inter-rater reliability is acceptable.) Additionally, Pearson correlation tests revealed that inter- sentence correlations r ranged from .411 to .686, $p < .01$ for the three sentences. These results suggest that the four listeners had reasonable consensus on their ratings of the three sentences. Therefore, a single mean rating score for each speaker was obtained by averaging the four listeners' mean ratings of the 3 sentences. The Trained and Control groups' mean rating scores along with each individual speaker's data at pretest and post test are summarized in Table 1.

The mean score for the Trained group was 6 at pretest and 5.1 at post test. For the Control group, it was 6.8 and 6.7 at pre- and post test respectively. Paired t-tests revealed a significant difference between the mean rating scores at pretest and post test for the Trained group [t (13) = 5.987, p = .000] but not for the Control group [t (9) = .527, p = .611]. These results suggest that the trainees made significant improvement in their tone productions at sentence level that was not matched by the Control group.

As seen in Table 1, the overall mean ratings for the Trained group were lower (indicating more native-like tone) than the Control group at both pretest and post test. To investigate whether such differences were significant, two Independent-Samples t-tests (2-tailed) on the mean rating scores were performed. The results showed the difference between the groups at pretest [t (22) = 2.798, p =. 01] was significant. The difference was much more significant at post test [t (22) = 4.598, p = .000].

Table 1. Individual speakers' mean rating scores as judged by four native Mandarin listeners

ID	Trained Pre	Trained Post	ID	Control Pre	Control Post
T01	6.6	5.8	C01	6.6	6.7
T02	5.9	5.8	C02	7.3	7.1
T03	6.5	5.8	C03	6.0	6.4
T04	4.8	4.0	C04	7.4	6.8
T05	5.8	4.2	C05	6.5	6.6
T06	5.4	3.9			
T07	7.0	6.3			
Mean	**6.0**	**5.1**		**6.8**	**6.7**

The examination of individual speakers' scores showed that although each trainee's post test score was lower than his/her pretest score, only two of the seven trainees had a size of improvement that was beyond 1.5 along a scale of 9. None of their mean rating score fell within 2 standard deviations of the native Mandarin speakers' mean score, a standard that is often applied to measure whether the learners' performance reaches the level of native-like production in the literature.

4. DISCUSSION

This pilot study investigated the effect of training for learning Mandarin tones on a larger linguistic unit beyond the isolated tones. After taking 6 hours of training during a period of 3-4 weeks, the trainees' productions of Mandarin tones at sentence level were judged by native Mandarin listeners to be significantly better at post test than at pretest. The Trained group's improvement was not matched by the Control group. These results suggest that beginner level learners can improve their production of nonnative tones at larger linguistics units through perception with production training with both auditory and visual input. While the effect of training for learning nonnative tones is well documented in the literature, no previous studies, to my knowledge, have involved training on tones beyond isolated tones. Accurate productions of nonnative tones at phrase and sentence level is noticeably difficult for nonnative speakers due to different reasons. Therefore, it is encouraging to see that laboratory training is effective for learning tones at sentence level.

It is important to point out that although the trainees' improvement was significant, none of them reached the level of native-like production as judged by native Mandarin listeners. This may be due to the fact that six hours of training may not be sufficient for a more significant size of gain in production of nonnative tones at sentence level. More intensive training with more trainees is needed to test the ultimate effect of such training in future studies.

One limitation of the current study is the use of the rating score as the sole judgment of the production of Mandarin tones at sentence level. Although the listeners were instructed to pay attention to tones only, it is possible that their judgment might also be influenced by other factors such as segmental errors and speech rate. Future studies may include other assessment measures such as instrumental analyses of speech properties. Another limitation was the lack of strict control of the speakers' performance at pretest (not by design but due to lack of participants). It would be ideal to have same number of participants in both groups whose production scores were comparable at pretest.

REFERENCES

[1] Leather, J. (1997). Interrelation of perceptual and productive learning in the initial acquisition of second-language tone. In A. James & J. Leather (Eds.*), Second-language speech: Structure and process* (pp. 75-101). Berlin: Mouton de Gruyter.

[2] Wang, X. (2008). Training for learning Mandarin tones. In F. Zhang & B. Barber (Eds.), *Handbook of Research on Computer-enhanced Language Acquisition and Learning.* Hershey, PA: IGI Global.

[3] Wang, Y., Spence, M. M., Jongman, A., & Sereno, J. A. (1999). Training American listeners to perceive Mandarin tones. *Journal of the Acoustical Society of America, 106,* 3649-3658.

[4] Wang, Y., Jongman, A., Sereno, J. A. (2003). Acoustic and perceptual evaluation of Mandarin tone productions before and after training. *Journal of the Acoustical Society of America 113,* 1033-104

ACOUSTIC REALIZATION AND PERCEPTION OF ENGLISH LEXICAL STRESS BY MANDARIN LEARNERS

Yuwen Lai, Joan Sereno, Allard Jongman
Department of Linguistics, University of Kansas, Lawrence KS 66044
yuwen.lai@gmail.com, sereno@ku.edu, jongman@ku.edu

1. INTRODUCTION

The acquisition of English lexical stress by beginning and advanced Mandarin L2 learners was examined through an acoustic study and perception study. Mandarin Chinese is a tone language which adopts F0 height and contour to signal lexical meaning [1]. English is a stress language which utilizes duration, F0, intensity, and spectral composition to signal stress [2]. Given the suprasegmental differences between the two target languages, Mandarin L2 learners are predicted to differ from native English speakers in the acoustic realization as well as the perception of English lexical stress.

First language transfer occurs when speakers or writers carry over the knowledge from their first language (L1) to a second language (L2) [3]. Most research on L2 acquisition focuses on acquisition at the segmental level, i.e. production and perception of individual consonants or vowels [4]. The current study is one of the few studies which focuses on the phonetic realization and perceptual cue weighting of L2 prosodic acquisition.

2. ACOUSTIC STUDY

2.1 Methods

Fourteen disyllabic word pairs with similar segmental composition but contrasting in stress location (e.g. *con*tract and con*tract*) were recorded by 18 Mandarin learners of English (9 advanced, 9 beginning). Mean F0, max F0, intensity, duration, and F2 of stressed and unstressed vowels were measured and compared to the production of native English speakers.

2.2 Results

An asymmetry of acoustic correlate realization in nouns and verbs was found in native speakers' production: native speakers utilize mean F0, max F0, intensity, and duration to signal stressed syllables in noun readings, but use only duration in stressed syllables in verb readings. Advanced and beginning Mandarin L2 learners demonstrate a more consistent use of correlates in nouns and verbs: the four correlates are all adopted in both conditions, and the magnitude (difference between stressed and unstressed vowels) was more consistent across nouns and verbs (see Figure 1).

3. PERCEPTION STUDY

3.1 Methods

Fifty Mandarin L2 learners of English (25 beginning, 25 advanced) and 25 native English speakers participated in a stress localization experiment. The stimuli were resynthesized versions of a naturally produced non-word disyllable [dada]. The target perceptual cues tested were max F0 and duration. Spectral composition of the first

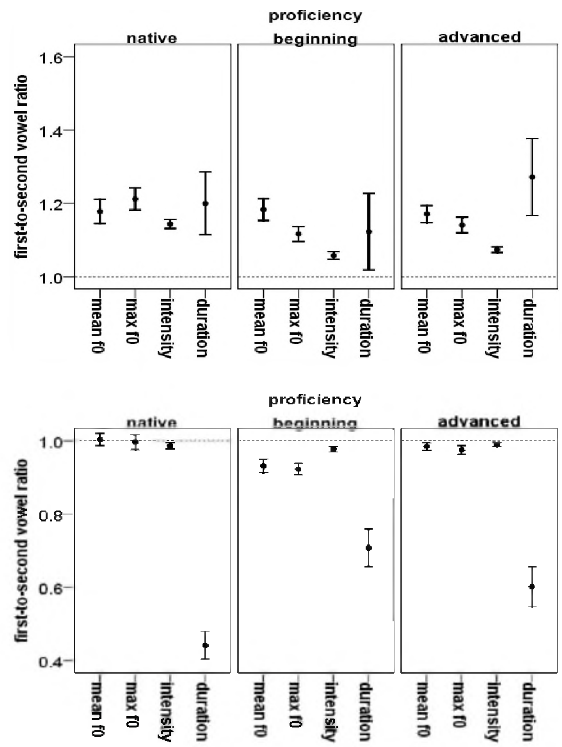


Figure 1. 95% confidence interval of first-to-second vowel ratios for mean F0, max F0, intensity, and duration for nouns (top) and verbs (bottom). The reference line (1.0) is the assumed baseline when a given correlate is identical in both syllables.

vowel of the disyllable was resynthesized to be either full or reduced. Based on the results from the acoustic study, five first-to-second vowel ratios for max F0 and duration were resynthesized for the full and reduced vowel disyllables. The ratios for max F0 are 1.23(F5), 1.115(F4), 1.0(F3), .897(F2), and .813(F1). The ratios for duration are 2.22(D5), 1.11(D4), 1.0(D3), .901(D2), and .45(D1). Participants were instructed to listen to a token and then indicate which syllable was stressed (syllable 1 or syllable 2) by clicking the corresponding text, DAda (first syllable stressed, corresponding to the response '1') or daDA (second syllable stressed, corresponding to the response '2').

3.2 Results

The results indicate that listeners at all proficiencies are sensitive to the spectral composition of the first vowel, i.e., full vowels in the first syllable attracted more noun responses, while reduced vowels in the same location received significantly more verb responses. The interaction of proficiency and max F0 is shown in Figure 2 while the interaction of proficiency and duration is shown in Figure 3. According to the data, native speakers are sensitive to both duration and max F0. Beginning learners are highly sensitive to duration, but not max F0. Advanced learners' responses are more affected by max F0, and to a much lesser extent by duration.

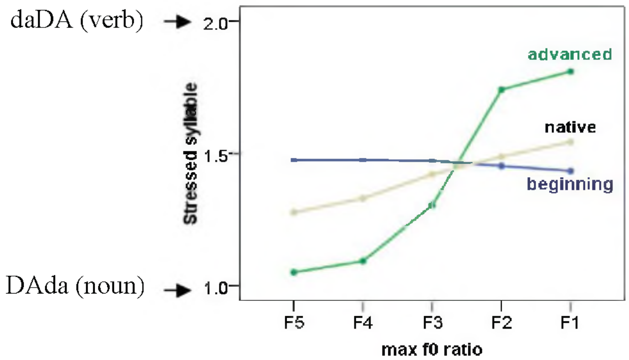


Figure 2. Responses for five max F0 ratios from three participant groups.

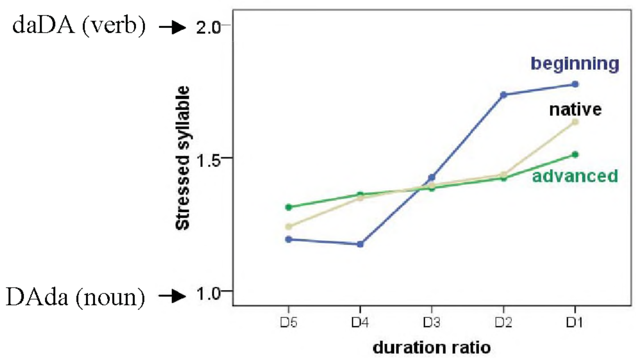


Figure 3. Responses for five duration ratios from three participant groups.

Production and perception of native speakers

Native listeners' responses for each max F0 and duration ratio combination are plotted in Figure 4. The bars represent responses- noun when below 1.5 chance level; verb when above 1.5.

When further compared with the production data, a striking pattern emerges: the role of max F0 is parallel in the production and perception of native speakers. That is, in production, native speakers utilize max F0 to encode stress in nouns but not in verbs. In their perception, the max F0 cue is used in the noun context but not in the verb context. That is, in max F0 cue-only conditions (row D3), native speakers have more noun responses for noun contexts (F5 and F4), but performed at chance level for neutral and verb contexts (F3, F2, and F1).

There is also a correspondence between the production and perception of duration by the native speakers. Native speakers utilized duration in both noun and verb contexts but the magnitude used in verbs was much greater than in nouns. When perceiving a stressed syllable, a greater duration difference is required to trigger verb responses compared to noun responses. In the duration cue-only conditions (column F3), native speakers have more noun responses for noun contexts (D5 and D4), and close to chance levels for neutral and verb contexts (D3, D2, and D1).

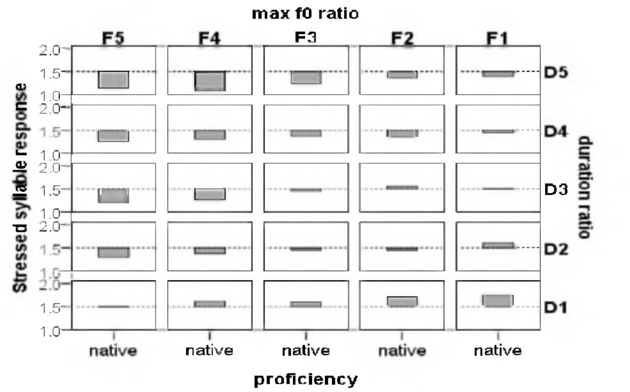


Figure 4. Responses from native English speakers for all max F0 and duration cue combinations.

4. GENERAL DISCUSSION

The current study discovered an asymmetry in correlate realization across noun and verb contexts (*con*tract and con*tract*) for native speakers. Results from a production and perception study revealed a congruence pattern between the utilization of acoustic correlates and the sensitivity of the perceptual system to a given cue in a specific context. More specifically, the magnitude of a correlate required to obtain stressed syllable perception is parallel to the magnitude of its realization in production.

L2 learners did not show such an asymmetry in their productions in correlate realization across noun and verb contexts. The phonemic feature of F0 in the learners' L1 may constrain them to use F0 in a highly consistent manner. Consequently, beginning learners use a consistent magnitude of F0 across noun and verb contexts. In addition, beginning learners carry over their insensitivity to F0 height [5] due to the phonemic role of F0 in Mandarin. As a result, beginning learners are not sensitive to max F0 changes but rely on duration in identifying stress.

The acoustic realization patterns of advanced learners are between beginning learners and native speakers. It is argued that as the learners' proficiency improves, they are able to approach a more native-like acoustic realization pattern in production. In terms of perception, as their awareness of English stress increases, they are able to utilize F0 (a cue familiar to them in their L1) as the major cue when perceiving English stress.

In sum, the present study demonstrates a context-specific congruence pattern in native production and perception. In addition, L2 learners with a phonemic F0 feature in their L1 are found to encode F0 in stress realization in a more consistent manner across contexts. It seems that L2 learners' perception is affected by the phonemic role of F0 in their L1 but that this effect may change as their proficiency improves.

REFERENCES

[1] J. M. Howie. (1976). *Acoustical Studies of Mandarin Vowels and Tones*. Cambridge: Cambridge University Press.
[2] I. Lehiste. (1970). *Suprasegmentals*. Cambridge, MA: MIT.
[3] T. Odlin. (1989). *Language Transfer: Crosslinguistic Influence in Language Learning*. Cambridge: Cambridge University Press.
[4] J. E. Flege. (1995). Second-language speech learning: Theory, findings, and problems. In *Speech perception and linguistic experience: Theoretical and methodological issues*, ed. Winifred Strange, 229-273. Baltimore: York Press.
[5] J. Gandour. (1983). Tone perception in far eastern languages. *J. Phon.* 11:149-175.

THE TRANSFER OF L1 ACOUSTIC CUES IN THE PERCEPTION OF L2 LEXICAL STRESS

Qian Wang, Tae-Jin Yoon
Department of Linguistics, University of Victoria, Canada; wangqian@uvic.ca; tyoon@uvic.ca

1. INTRODUCTION

In second language (L2) acquisition, first language (L1) transfer is observed at different levels, e.g. syntax, morphology, and also phonology. For example, L2 learners' problems with English lexical stress were attributed to the difference between L1 and L2 phonological systems [1]. The present research aimed to explore the nature of such problems from a different perspective, the nonnative use of acoustic cues in stress acquisition. It is proposed that L1 transfer also operates at the level of acoustics, i.e. the transfer of acoustic cue reliance.

It is known that at least three acoustic correlates contribute to English lexical stress perception, i.e. F0, duration and intensity. The acquisition of stress involves the correct use of *all* three cues. Nonnative speakers, even with the correct phonological representation of the L2 stress system, may not be able to employ the correlated cues in a way similar to L1 speakers. The present study focused on the use of acoustic cues by Chinese learners of English in stress perception. Chinese is a tone language, and F0 plays a key role in shaping the L1 phonological system, while the contribution by duration and intensity is marginal. As the acoustic cues of F0, intensity, and duration are used differently in L1 and L2, we expect to find the heavy reliance on F0 by CE will be transferred to the perception of English lexical stress. Two perceptual experiments were conducted to explore the hypothesis. The first experiment involved the stress judgment of synthesized disyllabic nonsense tokens and the second one was an oddity test using real English words extracted from carrier sentences.

2. EXPERIMENT ONE

In the first experiment, F0 difference on disyllabic nonsense words was manipulated in 5 levels, while duration and intensity cues were kept consistent on the both syllables. If there is a positive transfer of F0 from the L1, then Chinese learners would have no problems in perceiving stress on words with only F0 cue.

2.1 Research design

Three disyllabic nonsense words were created, *latmab, nizdit,* and *tetsep*. The nonsense words were produced by a trained phonetician as the basis for manipulation. The two syllables of each nonsense word were first manipulated to have the same duration and intensity. Then, F0 difference between the two syllables was changed in five levels (see Table 1).

Table 1: 5 levels of F0 manipulation

5 levels	1	2	3	4	5
F0 difference: s1 –s2	-50	-25	0	25	50

For example, in the first level of manipulation, the first syllable is 50Hz lower than the second syllable and in the third level, the two syllables have the same F0. This resulted in a total of 15 tokens were created (5 levels * 3 word forms).

Two groups of participants listened to the tokens, an experimental group of 68 CE and a controlled group of 38 NE. They indicated the stress position of each token by clicking **1st Syllable** or **2nd Syllable** on a computer screen.

2.4 Data analysis and results

The number of Initial Stress (IS) judgments (i.e. stress on the first syllable) was collected for each token. The Initial Stress Percentage (ISP) was defined as the percentage of IS judgments over the total number of participants for each token. ISP was calculated for the two groups separately. For example, an ISP of 50% by NE means that half of the NE participants agreed on the IS judgment.

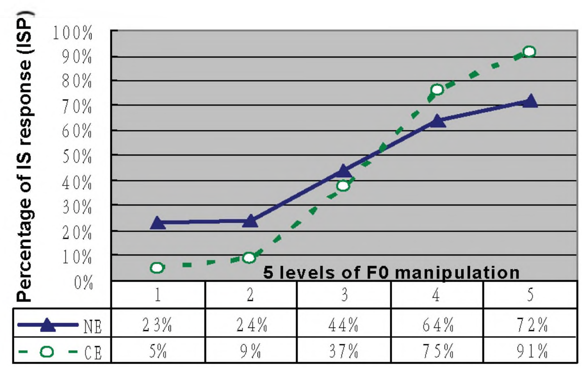


Figure 1: ISP of CE and NE as a function of F0 manipulation

The ISP values of NE and CE groups, respectively, are summarized for each level of F0 manipulation (see Figure 1). The dotted line in Figure 1 shows that CE demonstrated a systematic ISP change as a function of F0 manipulation. This change is in consistence with the change induced by NE (the solid line). A 2 (language group) × 5 (F0 manipulation level) mixed-model ANOVA on ISP showed that there is a main effect for F0 manipulation ($F(4,16) = 39.99$, $p < .001$, $\eta^2 = .909$). No significant effect of group was found ($F(1,4)=0.010$, $p>0.9$). The interaction between F0 and language group was also insignificant ($F(4,16)=2.858$, $p>0.1$). It can be concluded that F0 cue is sufficient for CE to perceive stress. However, we can't conclude that CE behaved in the same way as NE did. If we compare the slope of two lines above, we can see that for NE, F0 change from level one to five led to an increase of 49% in IS responses (23% to 72%). For the CE group, the same F0 change led to a much greater increase of 86% in IS responses in CE group (5% to 91%). This steeper slope of CE may suggest that their's reliance on F0 exceeds the native usage of F0.

3. EXPERIMENT TWO

If CE relies excessively on F0, then we can expect the negative effect when F0 information is absent. That is, CE would perform significantly worse than NE. The second experiment is designed to test this hypothesis.

3.1 Research design

In the second experiment, we chose one word form with two possible stress patterns: PERmit as a noun and perMIT as a verb (where capital letters indicate the stressed syllable). Two versions of either forms were constructed, one with F0 difference between the stressed and unstressed syllable cue and one without such difference. This was realized by extracting target words from two focus contexts: one where the target word is in the focus of the sentence, hence accented, and the other where the target word is not focused and thus unaccented, i.e., where another content word in the sentence is under focus. Previous literature [2] have shown that, the stressed syllable on an accented word in a sentence is reinforced by F0 cue, but stress on an unaccented word is usually cued by duration and intensity, but not F0. As a result, four types of stimuli (2 stress patterns * 2 accent conditions) were constructed and used in an oddity test. Participant listened to 3 tokens in a triad and decided whether the tokens are all the same, i.e. all nouns or all verbs, and if not, which token was the odd one. A total of 64 triad were constructed with two different accent conditions, half of the triads were realized with all three **A**ccented tokens (AAA), and half with all **U**naccented tokens (UUU). Two groups of participants (11 CE and 13 NE) were involved in the second experiment,.

3.2 Data analysis and results

Error rates were collected from each participant and used in a 2 (language group) × 2 (accent condition) mixed-model ANOVA. The results revealed the main effect for language group (F(1, 21[a]) = 10.747, p < .01, η^2 = .339). There was an overall difference in the error rate of CE (M=19.89%) compared to NE (M=6.51%). A significant main effect for the accent condition (F(1, 21)=27.61, p<.001, η^2 = .57), and a significant interaction of language group by accent were also observed (F(1,21)=14.32, p<0.01, η^2 = .405). Two-tailed independent t-tests between NE and CE for the two accent conditions, respectively, indicated that the significant effect of group actually lies in UUU condition (t(21)=3.54, p<.01) but not in AAA (t(21)=.61, p>.5) (see Figure 2 below). In other words, when F0 information is present, CE resembles NE in stress perception, which is consistent with the results of the first experiment. However, when F0 information is absent, CE performed significantly worse.

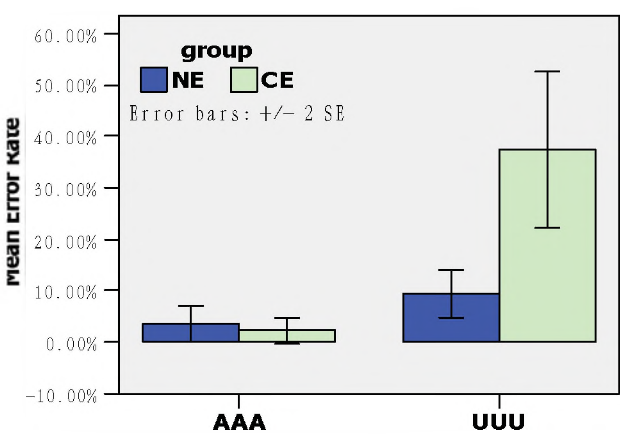


Figure 2. Mean error rates as function of accent effects by two language groups. Two types of accent combination are presented on the x-axis and error rates are represented on the y-axis. The performance of NE is indicated by blue bar, and that of CE is represented with green bar. Error bars are shown.

[a] One NE subject was identified as an outlier.

4. DISCUSSION

The results of the two experiments suggest that the difference between CE and NE in stress perception lies in their reliance on different of acoustic cues. CE rely heavily on F0 information and they don't have problems in stress perception if F0 information is provided (Experiment One). Unlike NE, they could make little use of other acoustic correlates that contribute to stress perception (Experiment Two). Further analysis through paired-sample t-test in the second experiment showed that while NE's performance in UUU condition is worse than in the AAA condition, the difference is not significant (t(11)=2.421, p> 0.01). CE, on the contrary, performed significantly worse in UUU condition then in AAA (t(10) = 4.553, p < 0.01). In UUU condition, NE can rely on the acoustic correlates of duration, intensity and possibly other correlates in stress perception when they are deprived of the F0 information. This is not the case for CE, who have a tonal language background. For CE, F0 is decisive and other cues may not be incorporated into the L2 phonological system and actively exploited in the construction of prosodic elements. Studies with Vietnamese learners and Japanese learners of English have found similar results [3,4]. As Adams and Munro [5] commented, although non-native speaker may not lack the acoustic cues for the appropriate prosodic elements, they nevertheless often fail to use these cues in a native-like way to achieve satisfactory prosody in a second language.

REFERENCES

[1] Archibald, J. (1997b). The acquisition of english stress by speakers of nonaccentual languages: Lexical storage versus computation of stress. *Linguistics, 35*, 167-181.

[2] Sluijter, A.M.C. & van Heuven, V.J. (1996). Spectral balance as an acoustic correlate of linguistic stress. *Journal of the Acoustical Society of America* 100(4), 2471-2485.

[3] Nguyen, T.A.T. (2003). *Prosodic Transfer: The Tonal Constraints on Vietnamese Acquisition of English Stress and Rhythm.* Doctoral dissertation, University of Queensland, Queensland, Australia.

[4] Ueyama, M., *Prosodic transfer: An acoustic study of L2 English vs. L2 Japanese.* 2000, University of California, Los Angeles.

[5] Adams, C. & Munro, R.R. (1978). In search of the acoustic correlates of stress: Fundamental frequency, amplitude, and duration in the connected utterances of some native and non-native speakers of English.

PRODUCTION OF ENGLISH LEXICAL STRESS BY INEXPERIENCED AND EXPERIENCED LEARNERS OF ENGLISH

Yunjuan He[1], Qian Wang[2], Caroline Wiltshire[1]
[1]Program in Linguistics, University of Florida, Gainesville, FL, USA
[2]Department of Linguistics, University of Victoria, Victoria, BC, Canada

1. INTRODUCTION

Misplacing stress might change the lexical categories of English words. For example, the noun 'record' is stressed on the first syllable and the verb 'record' is stressed on the second syllable. Misproduced stress patterns may cause a delay of understanding by native speakers of English (Wiltshire & Moon, 2003). Therefore, in order to produce an English word in a native way, placing stress accurately is a key step for second language speakers to master.

Furthermore, Mandarin Chinese is a syllable-timed tonal language, whereas American English is a stress-timed nontonal language. Mandarin speakers might use different phonetic cues to place English stress. This study was conducted to see whether the placement of stress and its phonetic cues show significant differences between American speakers and Mandarin speakers. In addition, the study tested whether learning experience improves English learners' ability to signal English lexical stress.

2. METHOD

2.1 Participants

Sixteen native Mandarin speakers from Northern China participated in a production experiment which involved two tasks: English real-word reading and English-like non real-word reading. Among the sixteen participants, eight were inexperienced learners of English who were college students in China, and the other eight were experienced learners of English who had studied at a University in the USA for at least three years. Six native English speakers in USA were chosen as a control group.

2.2 Stimuli

Two-syllable words were chosen for the present study: 22 real English words differing only in stress pattern, such as *subject*, and 22 non real English words with *N/i/N/i/+obstruent* syllable structure, such as *mimit*. The stressed syllable is underlined to attract the speaker's attention. 4 Mandarin Chinese words were designed with a tonal combination: either high flat tone + neutral tone, such as *mīmi* or dipping tone + falling tone, such as *mǐmì*.

2.3 Procedure

Each participant read two sets of English words. Each set was divided into two blocks. Between each block, the participant was required to rest for one minute. In addition, the 22 stimuli were preceded by a practice section. The speaker was required to read each word loudly. The words were randomly presented. All Mandarin speakers read the Chinese words as well.

2.4 Acoustic Analysis

With the aid of a script, PRAAT took three kinds of measurement of each vowel: loudness, duration and pitch:
Loudness: mean amplitude
Duration: onset to offset of vowel.
Pitch: initial: 5% of vowel duration
medial: 50% of vowel duration
final: 95% of vowel duration

3. RESULTS

3.1 Error rate

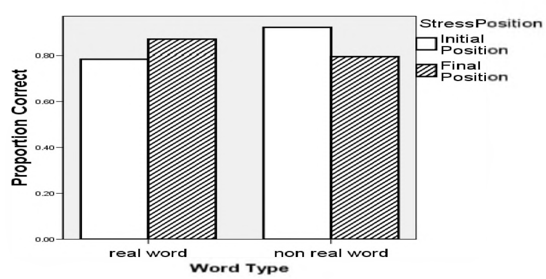


Fig. 1. Proportion correct production of stress

Shown in Fig 1, Mandarin speakers misplaced stress nearly significantly more on final-stressed non real words (p=0.056). However, Mandarin speakers nearly significantly misplace stress more on initial-stressed real words (p= 0.065).

3.2 Acoustic analysis

In the following sections, all data were from the production of non real English words by three groups and Chinese words by two Chinese groups. IE represents the acoustic cues produced by inexperienced learners, EE by experienced learners, AE represents by American speakers, MC_1 represents high flat tone + neutral tone and MC_2 represents dipping tone + falling tone.

Amplitude

Table 1. Ratio of Stressed Vowel/Unstressed Vowel Amplitude Measures on the initial stressed words

	IE	EE	AE	MC_1
Avg	1.1061	1.1027	1.1032	1.0998
StDev	0.04221	0.07444	0.05194	0.03881

A one-way ANOVA showed that there is no significant main effect of group [F(2,276) =0.943, *p*>0.05].

Table 2. Ratio of Stressed Vowel/Unstressed Vowel Amplitude Measures on the final stressed words

	IE	EE	AE	MC_2
Avg	1.0176	1.0155	1.5186	1.0216
StDev	0.03769	0.04540	0.06816	0.05195

A one-way ANOVA showed that there is a significant main effect of group [F(3,234) =0.000, *p*<0.001]. *Post hoc* analyses (Tukey HSD) showed that there is no significant different among IE between EE. However, there is a significant different between Mandarin speakers and English speakers. The ratio of amplitude made by American speaker is significant larger than ratios made Mandarin speakers.

Duration

Table 3. Ratio of Stressed Vowel/Unstressed Vowel Duration on the initial stressed words

	IE	EE	AE	MC_1
Avg	0.8077	0.9749	1.0950	1.2237
StDev	0.19597	0.23424	0.43683	0.40031

A one-way ANOVA showed that there is a significant main effect of group [F(3,276) =0.000, *p*<0.001]. *Post hoc* analyses (Tukey HSD) showed that the ratio of AE is not significantly larger than EE, but is significantly larger than IE.

Table 4. Ratio of S Table stressed Vowel/Unstressed Vowel Duration on the final stressed words

	IE	EE	AE	MC_2
Avg	1.9370	2.1503	2.4081	1.3782
StDev	0.50766	069986	1.10357	0.40696

A one-way ANOVA showed that there is a significant main effect of group [F(3,234) =0.000, *p*<0.001]. *Post hoc* analyses (Tukey HSD) showed that there is no significant difference between EE and AE, However, there is significant difference between IE and AE.

Pitch

Table 5. Ratio of Stressed Vowel/Unstressed Vowel F_0 on the initial stressed words.

p-initial	IE	EE	AE	MC_1
Avg	1.1302	1.1248	1.1707	1.0745
StDev	0.16240	0.27526	0.26774	0.19183
p-medial	IE	EE	AE	MC_1
Avg	1.1917	1.3244	1.3054	1.3980
StDev	0.19104	0.37418	0.38964	0.29895
p-final	IE	EE	AE	MC_1
Avg	1.4975	1.4797	1.2824	1.5483
StDev	0.84079	0.47419	0.46150	0.55694

Table 6. Ratio of Stressed Vowel/Unstressed Vowel F_0 on the final stressed words.

p-initial	IE	EE	AE	MC_2
Avg	1.2995	1.1307	1.1454	1.3620
StDev	0.63843	0.20072	0.27173	0.76671
p-medial	IE	EE	AE	MC_2
Avg	1.2625	1.0950	1.1627	1.2839
StDev	0.60969	0.19725	0.28771	0.45788
p-final	IE	EE	AE	MC_2
Avg	0.9375	0.9119	1.0590	1.0115
StDev	0.30417	0.12501	0.23910	0.34561

A one-way ANOVA showed that there is no significant main effect of group on the ratio of F_0 on each taken points.

4. DISCUSSION

Fig 1 indicates that Mandarin speakers have a phonetic preference for placing stress on the initial position of English disyllabic words. However, Mandarin speakers' learning experience may change their phonetic preference since most of real words in the study were taught as a verb with the stress in the second syllable during their English education in China,

Table 1 and Table 2 show that regardless of language experience, Mandarin speakers performed better at using amplitude cues to place stress on initial stressed words than on final stressed words. It is probably due to transferring the ratio of amplitude of the two tonal combinations in their native language, which show similar phonological pitch patterns.

Table 3 and Table 4 reveal that language experience did help Mandarin speakers to make a more native-like acoustic cue, larger ratios of duration, to stress English words. However, the ratios are still smaller than English native norms. To produce the duration of initial stressed English words, we assume that the final obstruent in the tested syllable structure might trigger Chinese participants to lengthen the second vowel even when they did not tend to stress the syllable.

The tone combination of high flat tone + neutral tone presents the high+low pitch pattern which is similar to the pitch pattern of initial stressed English disyllabic words and the tone combination of low dipping tone+ falling tone presents the low + high pitch pattern which is similar as the pitch pattern of final stressed English disyllabic words. Therefore, Mandarin speakers may easily detect the pitch difference between stressed and unstressed syllables and have no difficulty in producing pitch patterns of English disyllabic words.

REFERENCE:

Wiltshire, C & Moon, R.(2003). Phonetic Stress in Indian English vs. American English, *World Englishes* 22.3: 291-303.

Durational Properties of Stressed Syllables as a Cue for English-Accented French?

Christian Guilbault[1]

[1]Department of French, Simon Fraser University, Burnaby, 8888 University Drive, British Columbia, Canada V5A 1S6

1. INTRODUCTION

This paper reports on some preliminary findings of a corpus of recorded speech by Anglophones learners of French as a second language (FL2 hereafter). Excepts from this corpus were examined in an effort to find the cause of greater temporal syllabic variability, which is associated to the perception of a greater prosodic foreign accent. In order to do so, proper production of stressed syllables is essential. This proves to be a challenge for learners, as past studies [5, 2] have shown notable differences in the temporal rhythmic patterns of learners and those of native speakers. The data presented in [5], which are specific to Anglophones learners of French, did not however isolate the most important factors responsible for these temporal differences.

English rhythm is the result of a succession of accented (longer) and unaccented (shorter) syllables. French rhythm is characterized as having syllables of relatively equal duration, except for the final rhythmic-group syllables which is stressed and noticeably longer. This distinction between the two languages means that Anglophones learning FL2 transfer phonological properties from English and display greater syllabic variability. The analysis presented here examines the duration of accented words in an effort to narrow down the greater temporal variations displayed by Anglophones learners.

2. METHOD

2.1 Participants

A total of nine speakers were used for this first analysis divided into three groups: 3 intermediate English learners of French (EL1), 3 advanced English learners of French (EL2), and 3 native speakers (NS). The assessment of Anglophone learners' proficiency in French was done through a language-background questionnaire that evaluated their level of formal instruction and their overall experience in French. All learners reported having English as their first language. Low proficient learners were enrolled in 200-level language classes at university while advanced learners were enrolled in advanced literature and linguistics courses required for a major in French. Native speakers of French Learners were university professors at Simon Fraser University who had immigrated to an English-speaking environment less than 3 years ago. They speak without excessive traces of regional accent. All subjects were paid a small honorarium upon completion of the experiment.

2.2 Stimuli

The speech samples analyzed for this study are recalled single-sentence utterances which consisted in carrier sentences with a target word embedded word (underlined in (1)) in final rhythmic-group position:

(1) "Le mot <u>pas</u> est très joli."

In order to examine if the presence of the English stress pattern is transferred on polysyllabic words, 8 monosyllabic and 8 trisyllabic words were embedded in the carrier sentences. In total, 16 sentences were analyzed for each of the 9 speakers. The stimuli used for this part of the experiment are: *pas*, *pou*, *bas*, *bout*, *fa*, *fou*, *va*, *vous*, *décampa*, *déroba*, *habita*, *balada*, *étouffa*, *échafaud*, *cérébraux*, *vertébraux*.

2.3 Experiment procedure

Participants were required to say out loud the carrier sentences displayed on computer monitor, wait two seconds, and to repeat it once it had disappeared from the screen. This procedure provides more natural speech than a reading task while the speech material remains identical across subjects.

2.4 Acoustical analysis procedure

Recordings were digitized and analyzed with Computer Speech Lab version 4300. Syllabification of French utterances was done according to the rules generally accepted [3] and verified experimentally [1].

Variability Index: a variability index [3, 5, 6] was used to compare the temporal properties of the entire utterances. The formula to obtain the index, shown below, assesses the average variation of syllable duration for a given sentence, based on a normalized duration to account for variations in tempo.

$$VarIndex = \left(\sum_{k=1}^{n-2} \left| d_{k+1} - d_k \right| \right) / (n-1)$$

d_k=normalized duration (duration of a syllable divided by average duration of all syllables of a phrase/sentence) of the k^{th} syllable, and

n = number of syllables

Perfect syllable isochrony would allow for no durational variability, hence leading to an index of 0 while greater variability in syllable duration would lead to a greater index.

Group-final lengthening: durational variations of stressed syllables were computed by dividing the duration of the stressed syllable by the average duration of all syllables within that sentence, exclusive of the final sentence.

3. RESULTS

3.1 Variability Index

The average indexes presented in Table 1 show a slight tendency for EL1 and EL2 to produce greater variability than NS, which was expected. The figure also shows a greater tendency for EL1 to produce more variability in carrier sentences with monosyllables than EL2 and EL2 greater than NS. On carrier sentences with trisyllables, EL2 displays slightly greater variability than EL1, while both EL1 and EL2 display greater variability than NS. The small sample used for this analysis did not however prove significant when an ANOVA was computed.

Table 1: Variability index

	Monosyllables	Trisyllables	Average
EL1	0.756	0.652	0.704
EL2	0.732	0.707	0.720
NS	0.71	0.612	0.661

3.2 Group-Final Lengthening

Contrary to what was expected, EL2 displayed the greatest average increase in duration, followed by NS and EL1. As can be seen in Table 2 and Figure 1, EL2 speakers produced increases in monosyllabic words of 164.3% compared to 159.3% and 159.8% for EL1 and NS speakers respectively. On trisyllable words, however, EL2 speakers increased the stressed syllable in an almost identical ratio as the NS with values of 161.1% and 161.4% respectively, where EL1 speakers increased their duration by only 154.3%. Again, the small sample did not allow for statistical validation at this point.

Table 2: Group-Final Lengthening

	Monosyllables	Trisyllables	Average
EL1	159.3	154.3	156.8
EL2	164.3	161.1	162.7
NS	159.8	161.4	160.6

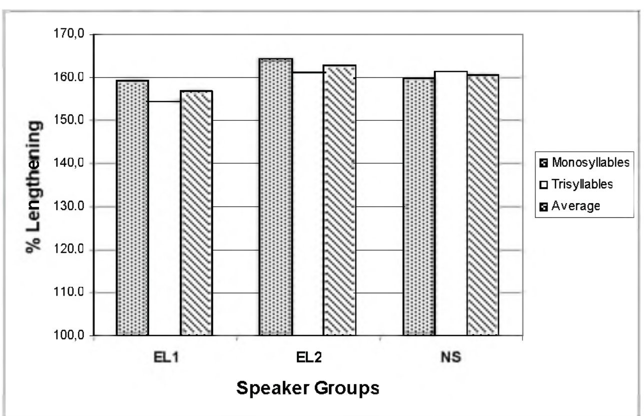


Figure 1: Group-Final Lengthening

3.3 Monosyllabic Words Lengthening

In an effort to determine if the transfer of the aspirated voiceless stops from English to French could account for longer syllables, the data for monosyllabic words containing *pas, pou, bas, bout* were examined. The results, presented in Table 3 and Figure 2, show that EL1 produced longer syllables with a voiced stop was in onset position, with an increase of 170.3% compared to 157.5% for syllables with a voiceless stop in onset. The transfer of aspiration may have had an effect for more advanced speakers who produced increases of 167.5% for stressed syllables with a voiceless stop in initial position and increases of 149.0% for syllables with a voiced stop.

Table 3: Lengthening for Voiceless and Voiced Stops in Monosyllabic Words

	Voiceless	Voiced
EL1	157.5	170.3
EL2	167.5	149.0
NS	163.5	154.3

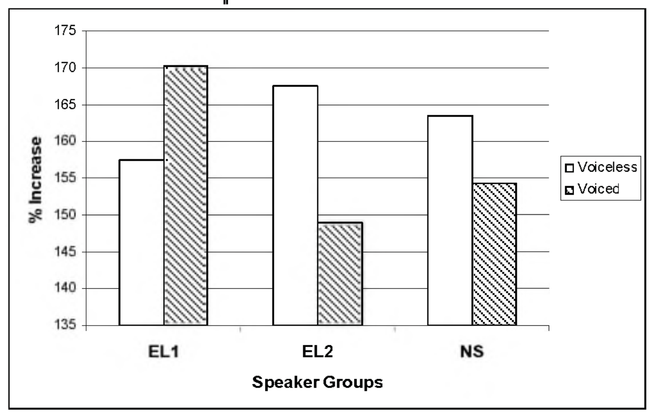


Figure 2: Lengthening for Voiceless and Voiced Stops in Monosyllabic Words

4. DISCUSSION

These preliminary results show that Anglophones do display greater temporal variability than NS. EL1 speakers increase monosyllables more than the last syllables of trisyllabic words. Segmental properties may also have caused increases in syllabic duration, as voiced stops seem to be associated with longer monosyllables produced by EL1.

It is not clear at this point if the main cause for temporal syllabic variability related to the rhythmic properties of a language is related to the way learners produce stressed syllables. Although the results do exhibit tendencies, greater corpora are required to further explore this hypothesis.

REFERENCES

[1] Beaudoin, M. (1996). *The Syllable Structure of French as Perceived by First and Second-Language Speakers,* Unpublished doctoral dissertation, University of Alberta, Canada.

[2] Bila, M. and J. Zimmermann (1999). English rhythm and the Slovak speaker, *International Congress of Phonetic Sciences*, 547-550.

[3] Delattre, P. (1940). Tendances de coupe syllabique en français. *Publications of the Modern Language Association*, *55*(2), 579-595.

[4] Deterding, D. (2001). The measurement of rhythm: a comparison of Singapore and British English. *Journal of Phonetics*, *29*(2), 217-230.

[5] Guilbault, C. (2002). *The Acquisition of French Rhythm by English Second Language Learners*, Unpublished Ph.D. Dissertation, University of Alberta.

[6] Low, E. L. (1998). *Prosodic Prominence in Singapore English*, Unpublished doctoral dissertation, Cambridge University.

ACKNOWLEDGMENTS

This project was funded by the President Research Grant at Simon Fraser University.

Do Mandarin-English bilinguals have an accent in their L1 vowel production?

Haisheng Jiang
Dept. of Linguistics, Simon Fraser University, 8888 University Drive, Burnaby, British Columbia, Canada V5A 1S6

1. INTRODUCTION

It is claimed that the L1 phonetic categories established in childhood do no remain static; instead, they may undergo modification when similar L1 and L2 sounds interact in the process of L2 learning [2]. If the L1 sounds are influenced by the L2 sounds and deviate from the L1 norms, the L1 monolingual listeners should be able to detect it. With regard to L2 influence on L1, previous studies of the effect of L1 use [4, 5] have revealed mixed findings. This gives rise to the need to examine further whether the amount of L1 use is an important factor when examining L2 influence on an L1. The present study examines the Mandarin vowel production by a group of Mandarin-English bilinguals differing in the amount of L1 use and aims at answering the following questions: (a) Do the Mandarin-English bilinguals have an accent in their L1 vowel production? (b) If so, what acoustic properties contribute to this accent? (c) Are the Mandarin-English bilinguals of high L1 use and those of low L1 use equally judged as accented?

2. METHOD

2.1 Participants

Thirteen Mandarin monolinguals (MonoM) (mean age =24 years, *SD*=4), 33 Mandarin-English bilinguals and 12 English monolinguals (MonoE) (mean age =27 years, *SD*=5) were recruited to produce vowel tokens. Based on their self-reported amount of L1 use, the Mandarin-English bilinguals were further divided into a group of high L1 use (BiMH) (mean percentage of L1 use =65, *SD*=8; mean age =22 years, *SD*=2; mean age of arrival =11.6year, *SD*=1.2) and a group of low L1 use (BiML) (mean percentage of L1 use =30, *SD*=9; mean age =21 years, *SD*=2; mean age of arrival =10.9 years, *SD*=1.6). Mandarin monolinguals and Mandarin-English bilinguals produced Mandarin vowel tokens; English monolinguals produced English vowel tokens. Seventeen monolingual Mandarin listeners, all of whom reported normal hearing, participated in a perception test.

2.2 Stimuli

The target Mandarin vowels were /a/, /aj/, /aʊ/, /e/, /i/, /oʊ/, /u/, and /y/. They were in Tone 4 open syllable words with initial /p/ whenever possible and were inserted in the sentence frame "*Zhe ge zi shi* ___. (This word is ___) to elicit production data. The target English vowels were /ɒ/, /aj/, /aʊ/, /e/, /i/, /oʊ/ and /u/. To match the Mandarin words as closely as possible, they were also in open syllable words with an initial /p/ and were inserted in the sentence frame "Now I say ___".

In total, there were 104 Mandarin vowel tokens by Mandarin monolinguals (8 vowels ×13 subjects ×1 repetition), 264 Mandarin vowel tokens by Mandarin-English bilinguals (8 vowels ×33 subjects ×1 repetition) and 84 English vowel tokens by English monolinguals (7 vowels ×12 subjects ×1 repetition). To eliminate the effect of the initial consonants on perception of vowels, the initial consonants of the target words were manipulated to be homogeneous. The motivation to include English stimuli in the perception test was to examine the extent to which the Mandarin-English bilinguals' Mandarin vowel production resembled the English monolinguals' English vowel production.

2.3 Experiment procedure

Stimuli were divided into two blocks, counterbalanced and presented to Mandarin listeners for goodness rating via E-Prime 1.0 on a laptop computer. Mandarin listeners were instructed to rate the goodness of the word they heard on a 7-point scale, with "1" being the worst and "7" the best exemplar of the Mandarin target word.

2.4 Acoustical analysis procedure

The duration, F1, F1 movement (ΔF1), F2, F2 movement (ΔF2), F3 and F0 of each target vowel were measured using Praat script [8].

3. RESULTS

3.1 Inter-rater reliability

The Cronbach's α values for Mandarin listeners in the rating of the 8 Mandarin vowels were all above .70, a cut-off point of an acceptable reliability [6].

3.2 Group differences

The mean ratings for each vowel assigned by Mandarin listeners to each speaker group (pooled across listeners) is given in Figure 1. The general tendency is that MonoM received the highest ratings and MonoE the lowest ratings, with the two groups of Mandarin-English bilinguals receiving intermediate ratings. It is also observed that, in most cases, BiML's ratings were lower than those of BiMH. Despite the trend that the rating scores assigned to the Mandarin-English bilinguals were intermediate between MonoM and MonoE, a significant difference was observed between BiMH and MonoM [$Z = -2.42$, $p < .05$], BiML and MonoM [$t(28) = 2.46$, $p < .05$] only in the rating scores of /y/.

3.3 Individual differences

Based on the accentedness criterion that speakers who obtained a mean rating falling two standard deviations below the mean rating assigned to native speakers were considered to have accented pronunciation [3], some individual Mandarin-English bilinguals had an accent in their L1 vowel production (see Table 1). As can be seen in the numbers in the brackets, there is no evidence showing that BiML outnumbered BiMH in being judged as accented.

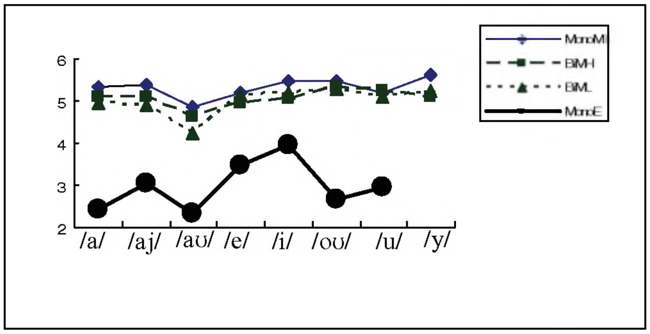


Fig. 1. Ratings assigned to MonoM, BiMH, BiML and MonoE by Mandarin listeners.

Table 1. Number of Mandarin-English bilinguals being judged as accented by Mandarin listeners.

Mandarin Vowel	Number of speakers judged as accented by Mandarin Listeners
/a/	4 (2 BiMH, 2 BiML)
/aj/	1
/aʊ/	1
/e/	2 (2 BiMH)
/i/	4 (2 BiMH, 2 BiML)
/oʊ/	5 (2 BiMH, 3 BiML)
/u/	1
/y/	13 (6 BiMH, 7 BiML)

3.3 Individual differences and acoustic data

As can be seen in Table 2, the acoustic dimensions that possibly contributed to the Mandarin-English bilinguals' accentedness in their L1 vowel production include lower F1(/y/), larger downward ΔF1(/u/), larger upward ΔF2(/aj/), duration(/e/, /oʊ/) and tone deviation(/a/, /aʊ/, /e/, /i/, /oʊ).

Table 2. Possible acoustic properties contributing to Mandarin-English bilinguals' accentedness.

Mandarin Vowel	Possible acoustic properties
/a/	tone deviation
/aj/	Larger upward ΔF2
/aʊ/	tone deviation
/e/	tone deviation, short duration
/i/	tone deviation
/oʊ/	tone deviation, exaggerated duration
/u/	Larger downward ΔF1
/y/	Lower F1

4. DISCUSSION

A clear pattern of accentedness was observed for the Mandarin-English bilinguals (n=13) in the production of /y/. This result is surprising, given the fact that Mandarin /y/ does not have an obvious English counterpart. It is hypothesized that the addition of the English /ɪ/ makes the high front portion of the vowel space more crowded and, as a result, the Mandarin-English bilinguals raised their /y/ to keep it perceptually distinct from its surrounding vowels. Another vowel showing the effect of L2 learning is Mandarin /aj/, in which the Mandarin-English bilinguals judged as accented had larger upward ΔF2 that is a characteristic of English /aj/. Although some individual Mandarin-English bilinguals were judged as accented in /a/, /aʊ/, /e/, /i/, /oʊ/, and /u/, there is no evidence indicating that they have modified the categories of these vowels because the acoustic properties possibly contributing to their accent (e.g. tone deviation and duration) do not necessarily mean the reorganization of L1 vowel category. The Mandarin-English bilinguals' reorganization of /y/ suggests that an L1 sound that does not have an L2 counterpart and is therefore not "similar" to an L2 sound may also be adjusted to maintain perceptual contrast in the shared L1 and L2 vowel space. Therefore, such L1 segments should also be included in the predictions of speech learning theories.

With regard to the effect of L1 use, the Mandarin-English bilinguals sounded accented to Mandarin listeners, whether their amount of L1 use was high or low. It seems that a bilingual's L1 will be affected in one way or another as long as he or she uses and is exposed to L2 on a regular basis. Previous studies of L2 influence on L1 [e.g. 1, 7] provide supporting evidence for this claim.

The present study contributes to the less well-studied field of L2 influence on L1. In particular, it suggests the necessity to include dissimilar L1 segments in speech learning theories.

5. REFERENCES

[1] J. E. Flege. & J. Hillenbrand. (1984). Limits on phonetic accuracy in adult foreign language speech production. *JASA*, 76, 708-721.

[2] J. E. Flege. (1995). Second language speech learning: Theory, findings and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research*. 233-277. Timonium, MD: York Press.

[3] J. E. Flege, Munro, M. J. Munro. & I. R. A. MacKay. (1995). Factors affecting the strength of perceived foreign accent in a second language. *JASA*, *97*, 3125-3134.

[4] S. G. Guion., J. E. Flege & J. D. Loftin. (2000). The effect of L1 use on pronunciation in Quichua-Spanish bilinguals. *J. Phonetics*. *28*: 27-42

[5] Z. McRobbie. (2003). An acoustic study of an uncommon vowel produced by Hungarian-English bilinguals. *Canadian Acoustics*. *31*(3), 34-35.

[6] J. Nunnaly. (1978). *Psychometric theory*. New York: McGraw-Hill.

[7] M. L. Sancier & C. A. Fowler. 1997. Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *J., Phonetics*. *35*, 421–436.

[8] S. Shira. (2003). DurationFormants_F.praat. University of Washington. Retrieved from the Web January 10, 2007. http:/home.myuw.net/ssetsuko/durationFormants_F.praat.

ACKNOWLEDGMENTS

I am grateful to Dr. Zita McRobbie, Dr. Murray Munro and Dr. Yue Wang for their advice and support. Of course, any errors are mine. This project was partly supported by a SHRRC grant held by Dr. Alexei Kochetov.

L2 ENGLISH VOWEL LEARNING BY MANDARIN SPEAKERS: DOES PERCEPTION PRECEDE PRODUCTION?

Ron I. Thomson
Dept. of Applied Linguistics, Brock University, St. Catharines, ON, Canada, L2S 3A1, rthomson@brocku.ca

1. INTRODUCTION

Many researchers argue that inaccuracy in L2 speech perception is the source of L2 accent (Flege, 1995), while others contend that L2 speech perception and production develop simultaneously (Best, 1995). There is little evidence that L2 production can precede L2 perception. Sheldon & Strange (1982) found that Japanese speakers were able to accurately produce the English /l/-/r/ contrast without being able to perceive it. However, that ability might have resulted from prior pronunciation instruction, coupled with a reading task where speakers could more easily attend to articulation. Munro and Derwing's (2008) study of Mandarin and Slavic speakers' acquisition of English vowels found learners' production ability improved in the absence of explicit pronunciation instruction. Because their research elicited L2 English production using auditory prompts, we can assume that the learners' ability to perceive English vowels must also have improved to at least the same extent.

The purpose of this study is to explicitly investigate the relationship between L2 English vowel identification and production by adult L1 Mandarin immigrants to Canada, using the same auditory stimuli to test both identification and production.

2. METHOD

The data analyzed in this experiment were part of a larger study that examined the effect of computer-mediated instruction on the development of L2 English vowels.

2.1 Participants

The participants were fourteen female and eight male Standard Mandarin speakers (*M* age = 36.4; range = 27-50) who were new to Canada (*M* LOR = 11.3 months; range = 4-48 months). They had been attending a local ESL program for an average of 4.8 months (range = 1-14 months). Those who had resided in Canada longest reported having had little interaction in English. All reported normal hearing.

2.2 Identification and production tasks

Participants identified English /i/, /ɪ/, /e/, /ɛ/, /æ/, /ɒ/, /ʌ/, /o/, /ʊ/ and /u/ embedded in /bV/ and /pV/ syllables that were produced by two different native speakers, one female and one male. Open syllables were used in order to adhere to Mandarin phonotactic constraints. Participants heard the stimuli via headphones and were asked to click on one of ten nautical flags presented on a computer screen. They had learned to associate these distinctive flags with the English vowel categories during a larger training study. Stimuli used in this study, however, were not used in training.

Productions were elicited using the same /bV/ and /pV/ stimuli described above, this time embedded in the carrier phrase, "The next word is ____." Participants responded by saying, "Now I say ____." Productions were recorded in a quiet room using a Marantz digital recorder with a sampling rate of 44,100 Hz.

2.3 Analysis

L2 production responses were acoustically analyzed using F1, F2 and F3 measures extracted from 20% and 70% of each vowel token's duration. Mean F0 and vowel duration were also extracted. These measures were tested against a native speaker English discriminant analysis model (Thomson, 2007) that determined to what English vowel categories the L2 productions were most phonetically similar. Results of this approach are highly correlated with human listener responses (Nearey & Assmann, 1986).

3. RESULTS

Mean % correct scores for each vowel on the identification and production tests are shown in Figures 1 and 2, for responses to Voice 1 and Voice 2 stimuli respectively.

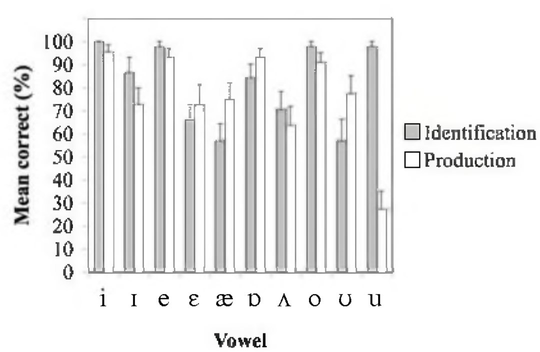


Figure 1. Comparison of average vowel identification and production recognition scores in response to Voice 1 stimuli. Error bars represent standard errors.

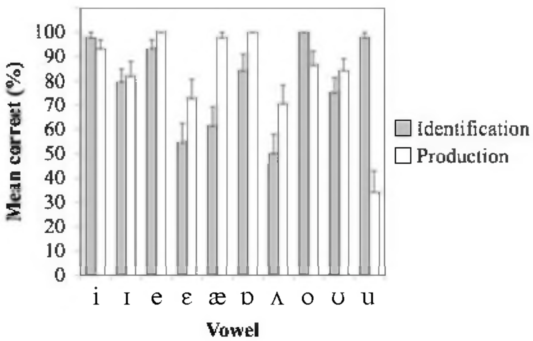


Figure 2. Comparison of average vowel identification and production recognition scores in response to Voice 2 stimuli. Error bars represent standard errors.

A three-way repeated measures analysis of variance was conducted, with Test (2 levels), Stimulus Voice (2 levels) and Vowel (10 levels) as within-speaker factors. Corrected Huynh-Feldt measures are reported because equality of variance could not be assumed for the Vowel variable. The mean scores on the identification and production tests were not significantly different, $F(1, 344) = .300$, $p = .590$, nor were mean scores in response to each stimulus voice, $F(1, 821) = 2.423$, $p = .135$. However, a significant effect for Vowel was found, $F(6.122, 128.561) = 14.966$, $p < .001$. Additionally, the Test x Vowel, $F(6.526, 137.039) = 16.187$, $p < .001$; Test x Voice, $F(1, 21) = 4.886$, $p = .038$; and Voice x Vowel, $F(6.314, 132.602) = 2.30$, $p = .035$ interactions were all significant.

To further investigate the significant Test x Vowel interaction, a series of post-hoc Bonferroni-adjusted *t*-tests were conducted. The mean identification score for /æ/ (59%) was significantly lower than the mean production score (86%), $t(21) = -4.157$, $p < .001$, while the mean identification score for /u/ (98%) was significantly higher than its mean production score (31%), $t(21) = 9.027$, $p < .001$. An examination of error patterns indicated that on the identification test, /æ/ was most often confused with /ɛ/ (20%) followed by /ɒ/ (8%). Conversely, on the production test, /æ/ was most often confused with /ɒ/ (9%), but rarely with /ɛ/ (3%). On the identification test, /u/ was seldom misperceived. In contrast, on the production test /u/ was frequently confused with both /ʊ/ (41%) and /o/ (27%). No other Time x Vowel contrasts were found to be significant.

For the Test x Voice interaction, Bonferroni-adjusted *t*-tests indicated that learner productions in response to Voice 2 stimuli were significantly more accurate (82%) than were productions in response to Voice 1 stimuli (76%), $t(21) = -2{,}968$, $p = .007$. Differences in response to each stimulus voice on the identification test were not significant.

Post-hoc tests examining the Voice x Vowel interaction found no significant differences, likely due to the conservative nature of multiple comparison tests.

4. DISCUSSION

The results of this study indicate that the perception and production of most L2 English vowels develop simultaneously. However, the interaction of test type with vowel suggests that the cues learners use for some contrasts lead to different results on each test type. For example, while /u/ and /ʊ/ can be easily identified by their relative duration, lengthening /ʊ/ in production does not result in a recognizably /u/-like pronunciation. Thomson (2007) found that L2 /ʊ/ productions in response to /u/ were twice as long as /ʊ/ productions in response to /ʊ/. The case of /æ/ is less clear. It is surprising that L2 productions of /æ/ were more accurate than were identification responses, despite the fact that the production task also required perception. Some learners might have had difficulty distinguishing between /ɛ/ and /æ/ on the identification test, resulting in a merged category, the distribution of which was more /æ/-like. If this is true, their production responses would be more similar to /æ/, resulting in higher accuracy in production. Further analysis of the data is needed to test this hypothesis.

Finally, the difference found in learner responses to different stimulus voices has important implications. It suggests that the results of such experiments need to be interpreted with caution when only a single stimulus voice is used. It may also suggest that some voices are easier to perceive, and therefore learn from, than are others.

REFERENCES

Best, C. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues* (pp. 171-204). Timonium, MD: York Press.

Flege, J.E. (1995). Second-language speech learning: Theory, findings, and problems. In W. Strange (Ed.) *Speech perception in linguistic experience: Theoretical and methodological issues*. (pp. 229-273). Timonium, MD: York Press.

Munro, M. J., & Derwing, T. M. (2007). A developmental study of Mandarin and Russian speakers' English vowels. *Language Learning, 58*, 479-502.

Sheldon, A., & Strange, W. (1982). The acquisition of /r/ and /l/ by Japanese learners of English: Evidence that speech production can precede speech perception. *Applied Psycholinguistics, 3,* 243-261.

Thomson, R. I. (2007). *Modeling L1/L2 interactions in the perception and production of English vowels by Mandarin L1 speakers: A training study.* Unpublished Doctoral Dissertation, University of Alberta, Edmonton, AB.

ACKNOWLEDGEMENTS

The author thanks Terrance Nearey, Murray Munro and Tracey Derwing for their input. The Social Science and Humanities Research Council supported this research.

An Acoustic Study of L2 VGn Rime Production

Ya Li
Dept. of Linguistics, University of Victoria, 3800 Ring Rd. Victoria BC, Canada, V8P 5C2 yali@uvic.ca

1. INTRODUCTION

Mandarin speakers are often heard to produce English *down* (/dawn/) as Mandarin *dàng* (/dɑŋ/). Previous studies (e.g., Hansen 2001) revealed that the preceding vowel context has a significant effect on the production of an L2 consonant and may interact with L1 transfer processes. Also, Chen (2000) found that in Mandarin VN (a monophthong vowel followed by a nasal coda /n/ or /ŋ/) rimes, nasal place tends to co-vary with vowel backness. In order to explore vowel context and possible L1 (First Language) influence on L2 (Second Language) nasal coda production, this study adopts an acoustic approach to examine English VGn (a diphthong vowel followed by /n/) rimes produced by Mandarin speakers.

2. METHOD

2.1 Participants & Test Materials

Twenty (10 male and 10 female) native Mandarin Chinese students at the University of Victoria participated in this study. Test words include 5 English words with VGn rimes, *pine/coin/gown/pain/cone,* 5 corresponding words with VG rimes, *pie/coy/cow/pay/go*, and 4 Mandarin words with Vŋ (a monophthong vowel followed by /ŋ/)/VG rimes, *gàng/kào/gòng/gòu* (presented as Chinese characters in the actual test but here in *PinYin* orthography). The 4 Mandarin words were chosen to contrast with the 4 similar sounding English words, *gown/cow/cone/go*, respectively.

2.2 Procedure

The word-reading task was performed in a sound-attenuated room in the Phonetics laboratory of the University of Victoria. Participants were instructed to read the test words presented on-screen through the Microsoft PowerPoint program. Each test word successively appeared 4 times (hence 4 tokens for each word) in a PowerPoint slide. The successive appearance of the 4 tokens was to improve the chance for a word to be produced consistently. There was a 2-second interval following each appearance of a token and the participants were instructed to read each word by the rhythm of its appearance. A total of 1120 tokens (14 words x 4 repetitions x 20 speakers) were collected and analyzed using Praat 4.4.22.

2.3 Acoustic parameters

For each token, mean F1-F0 and F3-F2 (differences between the first and fundamental formant frequencies and between the third and second formant frequencies) over the first half (_fh) and second half (_sh) of vowel duration were respectively used to estimate vowel height and backness changes over the duration. Also, the vowel and nasal duration (V_D & N_D) of each token were measured to infer the degree of vowel-nasal coupling. Last, the band energy difference (ΔdB) between 0~525 Hz and 770~1265Hz bands over nasal duration were calculated to predict nasal place (alveolar /n/, velar /ŋ/, or uvular /ɴ/). Based on previous findings (e.g., Kurowski & Blumstein, 1987), this study assumes that the less the ΔdB, the more backed the nasal place.

3. RESULTS

3.1 Vowel measurements

Figure 1 illustrates vowel height/backness changes over the duration in English *gown/cow* and the contrasting Mandarin *gàng/kào*. The x-axis represents mean F3-F2 in Hz and the y-axis represents mean F1-F0 in Hz. The greater the F3-F2, the more backed the vowel, and the smaller the F1-F0, the higher the vowel.

Mean value	*gown*	*cow*	*gàng*	*kào*
F3-F2_fh	1422	1492	1415	1691
F1-F0_fh	631	727	623	665
F3-F2_sh	1767	1803	1753	1893
F1-F0_sh	587	509	606	526

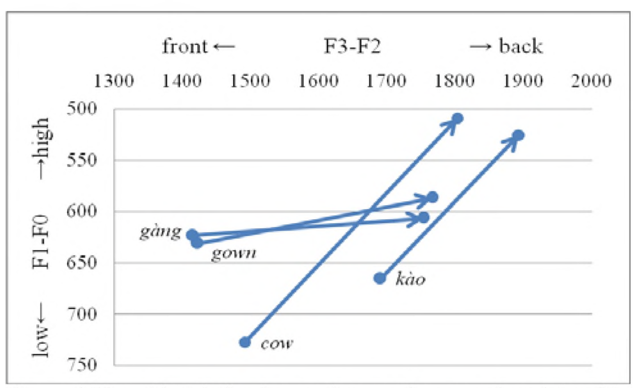


Figure 1 Vowel height/backness changes for *gown/cow/gàng/kào*

In Figure 1, /a/ in *gàng* moving to the further back happens to resemble /aw/ in *gown* because the height of /w/ in /aw/ is greatly lowered. A 2-tailed paired samples t-test revealed that there was a non-significant vowel height change over the duration for /aw/ in *gown* ($t = 1.388$, $p = .181$). In contrast, a significant vowel backness change over the duration was found in *gàng* ($t = -6.127$, $p < .001$). Since significant vowel height/backness changes over the duration are expected for a diphthong but not for a monophthong vowel, both the non-significant vowel height change in *gown* and the significant vowel backness change in *gàng* suggest a strong nasal influence on the preceding vowel.

3.2 Durational measurements

Figure 2 illustrates mean vowel and nasal duration (V_D & N_D) for each of the 7 words, *pine/coin /gown/ cone/pain/gàng/gòng*, across tokens and speakers. The x-axis represents duration in second (s).

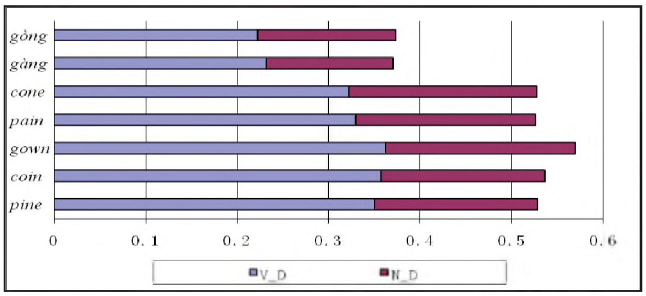


Figure 2 V_D and N_D for *pine/coin/gown/cone/pain/gàng/gòng*

In Figure 2, the 2 Mandarin words *gàng/gòng* have similar duration (V_D + N_D < 0.4s) and are shorter than the 5 English words, *pine/coin/ gown/pain/cone* (V_D + N_D > 0.5s). The short and almost identical duration for *gàng/gòng* is expected due to the fact that Mandarin 4th tone words are normally produced with a very short duration and that Mandarin as a syllable-timed language favors syllables with similar length. Note that short duration (especially N_D) implies a strong degree of vowel-nasal coupling.

3.3 Nasal measurements

Figure 3 illustrates mean ΔdB for *pine/coin /gown/ cone/pain/gàng/gòng*, across tokens and speakers. The y-axis represents the band energy difference in Decibels (dB).

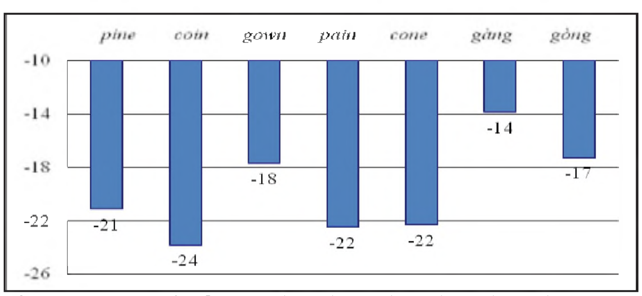


Figure 3 Mean ΔdB for *pine/coin/gown/pain/cone/gàng/gòng*

In Figure 3, *gàng* has the least mean ΔdB (-14dB) among all the 7 words, and *gown* has the least mean ΔdB (-18dB) among the 5 English words, comparable to that in *gòng* (-17dB). A repeated measures one-way ANOVA test revealed that mean ΔdB for *gown* is significantly different from that for *pine/coin/pain/cone/ gàng* (the results from the pairwise comparisons, $p < .001$, $=.003$, $<.001$, $<.001$, $=.016$, respectively), but not for *gòng* ($p = .763$), indicating that /n/ in *gown* can be identified with the velar /ŋ/ in *gòng*. Also, mean ΔdB for *gàng* was significantly lower than for *pine/coin/gown/pain/cone/gòng* (the results from the pairewise comparisons, $p < .001$, $<.001$, $=.016$, $<.001$, $<.001$, $=.039$, respectively), indicating that the nasal place in *gàng* is further back than in the rest of the words and thus can be considered as uvular /ɴ/ rather than velar /ŋ/.

4. DISCUSSION

The Mandarin speakers' production of English VGn and Mandarin Vŋ rimes is subject to a strong vowel-nasal coarticulation effect. For instance, the high glide /w/ in *gown* is greatly lowered and the central vowel /a/ in *gàng* becomes further backed. Conversely, /n/ in *gown* is at the further back than at the presumed alveolar region, and /ŋ/ in *gàng* is at the further back than at the presumed velar region. The nasal place in *gown/gàng* co-varies with the vowel backness change over the duration.

Furthermore, the strong vowel-nasal coarticulation effect in Mandarin speakers' production of English VGn rimes may be related to rhythmic factors rather than L1 transfer; that is, given the controlled 2-second- interval between every 2 tokens presented on-screen in the reading test, the English VGn rimes had to be read within the time frame, resulting in a vowel-nasal backness assimilation effect similar to that in the Mandarin Vŋ rimes.

REFERENCES

Chen, M.Y. (2000). Acoustic analysis of simple vowels preceding a nasal in standard Chinese. *Journal of Phonetics, 28,* 43-67.

Hansen, J.G., 2001. Linguistic constraints on the acquisition of English syllable codas by native speakers of Mandarin Chinese. *Applied Linguistics* 22 (3), 338-365.

Kurowski, K., & Blumstein, S. (1987). Acoustic properties for place of articulation in nasal consonants. *Journal of the Acoustic Society of America, 81(6)*, 1917-1927.

ACKNOWLEDGEMENTS

This study is based on my MA thesis project on L2 nasal coda production, so I would like to thank Dr. Hua Lin (my supervisor) and Dr. John Esling, for having reviewed my MA thesis and offered valuable insights. Any remaining errors are entirely mine.

MANDARIN SPEAKERS' PRODUCTION OF ENGLISH VOWELS IN REAL AND PSEUDO WORDS

Xianghua Wu and Murray Munro
Dept. of Linguistics, Simon Fraser University, 8888 University Dr., Burnaby, British Columbia, Canada V5A1S6

1. INTRODUCTION

Some theoretical perspectives, such as Flege's (1995) Speech Learning Model (SLM) see second language (L2) phonetic learning as a gradual, approximative process whereby learners' productions of at least some L2 segments improve as a function of L2 experience. Empirical support for this view comes from longitudinal investigations, such as Munro & Derwing (2008), who observed that Mandarin speakers produced more intelligible exemplars of some, but not all, English vowels after a year of residence in Canada. An additional outcome of that study was that bVC words had more intelligible vowels overall than did pVC words. However, because the voicing of the initial consonant was confounded with word frequency in their stimulus set and only a limited range of contexts were used, it was not possible to determine whether the intelligibility difference was due to the voicing status of the initial consonant, to word frequency, or to some other influence.

The current study extends Munro and Derwing's (2008) work by unconfounding some of these factors. It examines Mandarin speakers' productions of three English vowels in real and pseudo words. Vowels in pseudo words differing only in the voicing of the initial consonant will be assessed to address the first research question: Does the voicing status of the initial consonant in CVC(C) words influence the intelligibility of the following vowel? Because only pseudo words will be used, the potential effects of word frequency are eliminated. Second, vowel intelligibility in English words varying in initial and final consonants will be compared for words varying in frequency to answer the second research question: Does word frequency influence vowel intelligibility in CVC(C) words? A third question to be addressed is as follows: Is greater L2 experience, assessed in terms of length of residence (LOR) in Canada, associated with greater vowel intelligibility?

2. METHOD

2.1 Participants

Eight native Mandarin speakers, 3 males and 5 females, ranging in age from 22 to 32 (M= 25), participated in this study. The subjects started to learn English at age 7 to 14 years in school and had been in Canada for a mean of 2.4 years (range = 4 months to 6 years). Among them, four had one year or less of residence in Canada. None of them reported any previous pronunciation training or hearing/speech impairment.

2.2 Stimuli

The stimuli were produced in CVC(C) structures containing one of the three target vowels /ɪ/, /ʊ/ and /ʌ/ which are non-native to Mandarin speakers. In total 56 words (see Appendix 1) were paired such that they differed from each other by onset voicing, rhyme and word frequency. These words were arranged into three stimulus sets. Set 1 ('all pseudo') contains paired pseudo words differing in onset voicing. Set 2 ('pseudo/low frequency') consists of pseudo words or low frequency words which differed from the low frequency or high frequency words in set 3 ('low/high frequency'). We utilized the Michigan Corpus of Academic Spoken English (MICASE) to determine the frequencies of the real words.

2.3 Speaking task

Participants produced the target words in a delayed repetition task. They were required to repeat the target word three times in the frame "The next word is _____." The model stimuli were spoken by a male native speaker of Canadian English in a different form "Now I say _____." The best tokens were digitally extracted and normalized, using Sound Forge 6.0. Those tokens with recording errors, background noise and speakers' errors were avoided.

2.4 Classification task

All 56 words were replicated twice, resulting in 960 tokens (60 words × 8 participants × 2 repetitions). All tokens were randomized and grouped into four blocks using E-prime and then presented to a phonetically-trained native speaker of Canadian English, who identified the English vowel category closest to each pronunciation.

3. RESULTS

3.1 Onset voicing

The effect of onset voicing was examined for all pseudo words. A one-way repeated measures analysis of variance (ANOVA) revealed no significant effect of initial consonant voicing on vowel target accuracy [F (1, 7) = .077, p = .789] (see Fig. 1).

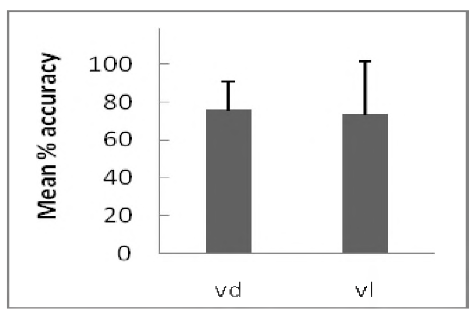


Fig.1. Mean %-ID of vowel targets for 'all pseudo' words with voiced (vd) onsets and those with voiceless (vl) onsets.

3.2 Vowel and word frequency

Effects of vowel and word frequency were measured for 'pseudo/low frequency' and 'low/high frequency' words. A two-way repeated measures ANOVA with vowel and word frequency as within-subject factors indicated a significant effect of vowel [F (2, 14) = 13.449, p = .001] but no significant effect of word frequency [F (1, 7) = 2.294, p

= .174]. Nevertheless, there was a tendency for mean correct identification to be higher for the high frequency items than for the pseudo or low frequency words. The interaction of vowel x frequency was also found to be non-significant [F (2, 14) = .920, p = .421]. Post hoc Bonferroni tests also revealed a significant difference between the identification of /ʌ/ and that of /ʊ/and /ɪ/ (p = .005 and .039, respectively) (see Fig.2).

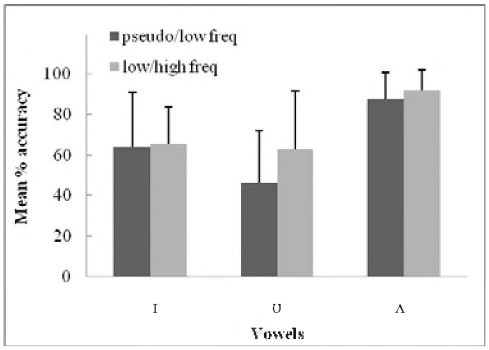


Fig.2. Mean %-ID of /ɪ/, /ʊ/ and /ʌ/ in 'pseudo/low frequency' and 'low/high frequency' words.

3.3 LOR

To answer the third research question, a two-way repeated measures ANOVA was carried out to compare the mean correct identification of /ɪ/, /ʊ/ and /ʌ/ produced by speakers with longer length of residence (LOR-L, LOR = 1.5 to 6 yrs) and those with shorter length of residence (LOR-S, LOR = 0 to 1yr). A significant effect of LOR was found across all three stimulus sets [F (1, 6) = 9.98, p = .02], indicating higher scores for the longer residents (see Table 1). A significant effect of stimulus set was not observed [F (2, 120) = .38, p = .692]. There was also no significant interaction between stimulus set and LOR, indicating that the effect of LOR was comparable for all three stimulus sets. Nevertheless, as shown in Table 1, there was a non-significant trend for the shorter-term residents to score better on the higher frequency words (62% vs. 70%).

Table 1. Mean %-ID (M) and standard deviation (SD) by Mandarin speakers with longer LOR (LOR-L) and shorter LOR (LOR-S).

	All pseudo		Pseudo/low freq		Low/high freq	
	M	*SD*	M	*(SD)*	M	*(SD)*
LOR-S	62	(19)	62	(20)	70	(11)
LOR-L	88	(7)	83	(6)	84	(6)

3.4 Rhyme

A one-way ANOVA was also conducted to examine whether rhyme had any effect on vowel target accuracy. A statistically significant effect of rhyme was observed across stimulus sets 2 ('pseudo/low frequency') and 3 ('low/high frequency') [F (17, 119) = 11.666, p < .0001]. Among all 18 rhymes, /ʊ/ preceding consonant /l/ was identified with the lowest accuracy (9%, compared to the highest score of 97% for vowel /ʌ/ followed by /z/).

4. DISCUSSION

The results of this study provide some insights into the issues raised in the three research questions. First, the voicing status of the initial consonant had no effect on the vowel target accuracy in CVC(C) words, at least in pseudo words. Secondly, word frequency did not appear to affect vowel intelligibility. Nevertheless, a non-significant tendency was observed in this study from two perspectives. The first trend was that the identification of vowels in common words with higher frequency tended to be higher than in those with lower frequency. The second is that the shorter-term L2 learners tended to have better performance when producing the vowels in the words with higher frequency. Because of the small number of participants in this investigation and the limited number of productions considered, it is not possible to draw firm conclusions about the role of word frequency in vowel production. However, the latter tendency is at least consistent with the view that longer-term residence in the target language country improves vowel intelligibility in a significant way. In addition, the results of this study also indicate that final consonants play a role in the identification of the preceding vowel in L2 speakers' productions. Given the wide variety of factors affecting vowel intelligibility, further research is needed to clarify the interplay of the many possible influences.

APPENDIX 1.

Stimulus Set	Phonetic Representation of Stimulus Words
Set 1	bɪm-pɪm, bɪsh-pɪsh, bʊd-pʊd, bʊʤ-pʊʤ, ʧɪsh-ʤɪsh, ʧʌd-ʤʌd, ʧʌf-ʤʌf, ʧʌz-ʤʌz
Set 2	pɪd, bɪk, bɪn, pɪz, pʊk, bʊl, bʊt, bʌb, bʌp, pʌst, pʌz, ʤɪk, ʧɪm, ʤɪp, ʧɪst, ʧʌʤ, ʤʌk, ʤʌkl, ʤʌm, ʧʌst
Set 3	bɪd, pɪk, pɪn, bɪz, bʊk, pʊl, pʊt, pʌb, pʌp, bʌst, bʌz, ʧɪk, ʤɪm, ʧɪp, ʤɪst, ʤʌʤ, ʧʌk, ʧʌkl, ʧʌm, ʤʌst

REFERENCES

Flege, J. E. (1995). Second-language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues* (pp. 233-277). Timonium, MD: York Press.

Munro, M. J. and Derwing, T. M. (2008). Segmental acquisition in adult ESL learners: A longitudinal study of vowel production. *Language Learning*, to appear.

ACKNOWLEDGMENTS

The authors thank the participants for their contribution to this investigation and acknowledge Dr. Y. Wang for the use of her research facility.

VARIABILITY IN CANTONESE SPEAKERS' PRODUCTIONS OF ENGLISH VOWELS

Murray J. Munro
Department of Linguistics, Simon Fraser University, Burnaby, BC V5A 1S6

1. INTRODUCTION

Research on second language (L2) phonetic learning has often emphasized the effects of the first language (L1) sound system on the acquisition of the second. Some work on consonant and vowel production, for instance, focuses on generalizations of the type "Speakers of language *x* have difficulty producing segment *y* when acquiring language *z*." While many of the difficulties that occur in L2 acquisition can indeed be attributed to L1 influences, static accounts of L2 speech do not offer much insight into the *process* of phonetic learning. In the first place, L2 production patterns are not static: the accuracy of some L2 segments, including vowels has been observed to improve in longitudinal research, even without focused instruction [1]. Moreover, generalizations like the one above belie considerable variability both between speakers from the same L1 background and within individual L2 learners when different productions of the same item are compared [2]. Flege's Speech Learning Model (SLM) sees L2 segmental learning as a context-dependent, approximative process whereby learners acquire increasingly better representations of at least some L2 categories over time as a result of experience with language input [3]. This investigation considers the ways in which an evaluation of variability in L2 speaker performance can shed light on the SLM and other models that address the L2 acquisition process. It focuses on English high vowel productions of speakers of Hong Kong Cantonese who are relatively homogeneous with respect to linguistic and social background, but who differ in their length of Canadian residence (LOR). While [ɪ] and [ʊ] occur in closed syllables in Cantonese, they are regarded as allophonic variants of [i] and [u], respectively [4]. Therefore, Cantonese learners of English must learn to produce two vowel distinctions in a phonetic environment in which they do not contrast in L1. In addition, the absence of coda /d/ in Cantonese means that they must also learn to produce all four vowels in a completely new environment. The issues to be considered here are as follows: (1) To what extent is English high vowel acquisition context-dependent for Cantonese speakers, and, in particular, are high vowels produced less intelligibly before /d/ than before other consonants? (2) How much interspeaker variability is evident in high vowel productions, and is that variability related to LOR? (3) What is the relationship between intraspeaker variability in high vowel productions and LOR? and (4) Is there variability in vowel productions across words with the same rhyme?

2. METHOD

2.1 Participants

The participants were 18 native speakers of Cantonese who had been born and raised in Hong Kong. Their mean age of arrival in Canada was 18 yr (range 15-25 yr), and their mean LOR was 4.9 yr (range: 9 months to 6.9 yr). All were attending or had recently attended English-speaking post-secondary institutions in Canada, and all passed a pure-tone hearing screen.

2.2 Test Items

The test items were 30 common English CVC words likely to be very familiar to the speakers. The words contained rhymes consisting of the vowels /i/, /ɪ/, /u/, and /ʊ/ in open syllables or before /t/, /k/, and /d/. Because of the types of rhymes represented and the need for relatively high frequency words, it was not possible to generate a stimulus set such that equal numbers of each rhyme were represented; nor was it possible to match words across rhymes in terms of initial consonants. While most of the words were common nouns, verbs, or adjectives, two were proper names ("Sue" and "Luke").

2.3 Speaking Task

Individual recording sessions were conducted in an audiometric booth using studio-quality digital recording equipment. Words were elicited via a picture-naming task without modeling, so that the productions could be assumed to be based on long-term representations developed through experience with English. Each participant named a randomly-ordered set of drawings, each of which depicted one of the target words. The first letter of each word was provided as an additional cue. Items were first elicited in a practice session, during which the participants guessed each item and produced it in a carrier sentence ("The next word is ___.") If the response was not the target word, the speaker guessed again until the correct word was produced. After all target words had been elicited once, the pictures were shuffled for recording. Each participant then recorded the entire deck three times, with a shuffle after each run-through.

2.4 Vowel Intelligibility Assessment

Four phonetically-trained assistants carried out a vowel intelligibility assessment in a sound-treated room. During multiple individual listening sessions, they focused on the vowel in each word and determined which Canadian English vowel was closest to the one produced. As in previous work, narrow phonetic transcriptions were not used. Responses were registered by clicking computer buttons marked with symbols for the English vowels / i ɪ e u ʊ o / and "other."

3. RESULTS

3.1 Effects of Context

Table 1 shows mean correct identifications, pooled over judges, for the vowels in each rhyme. Because scores on the two open-syllable vowels, /i/ and /u/, were at or near ceiling, these items were excluded from statistical analyses. The lowest scores, 33% and 52%, were observed on /ɪ/ and /ʊ/ before /k/. Separate one-way repeated measures ANOVAs for front and back vowels yielded significant effects of rhyme on vowel intelligibility in both cases, *Fs*(5, 85) = 21.07 and 12.67, respectively, *ps* < .001. Post hoc Bonferroni tests were used to compare intelligibility across rhymes. Overall, /i/ was more accurately identified in all three CV rhymes than was /ɪ/. Moreover, both front vowels were produced just as intelligibly before /d/ as before /t/ and /k/, even though Vd rhymes do not occur in Cantonese. For the back vowels /u/ was more intelligible before /k/ than before /t/, while the opposite was true for /ʊ/. In addition, /u/ was more intelligible than /ʊ/ before /k/, but not in the other contexts. Finally, /ʊ/ was more intelligible before /d/ than before /k/.

Table 1. Mean %-ID by rhyme and numbers of speakers (of 18) who reached criterion on each rhyme.

Rhyme	% ID	Reached Crit.	Rhyme	%ID	Reached Crit.
i	100	18	u	99	17
it	82	11	ut	70	4
ik	93	15	uk	88	14
id	97	17	ud	62	6
ɪt	44	2	ʊt	69	5
ɪk	52	0	ʊk	33	1
ɪd	55	4	ʊd	86	14

3.2 Between-speaker Variability

To facilitate between-speaker comparisons, a total of 85% correct judgments was adopted as the criterion for "acquisition" of a particular rhyme. Table 1 gives the number of speakers who reached criterion on each rhyme. This total varied from a high of 18/18 speakers for open-syllable /i/ to a low of 0 speakers for /ɪk/. Though the number of rhymes reaching criterion did not correlate significantly with LOR in Canada, the intelligibility of rhymes with /ɪ/ and /ʊ/ did show significant positive correlations with LOR, *rs* = .61 and .74, *ps* < .01. For most of the individual /ɪ/ words ('chick,' 'kid,' 'lid,' 'hit,' this statistical relationship held (*p* < .05). For the /ʊ/ words, positive correlations were also observed, but statistical significance was reached in only one of seven possible words ('cook').

3.3 Within-speaker variability

Variability within speakers was assessed by computing two consistency indices based on the three productions of each word. The ALL+ index was the number of times each judge evaluated all three productions of a word correctly, summed over all words. The ALL- index was a parallel total for consistently incorrect judgments. Pearson correlations (*r*) between the ALL+ and ALL- indices and LOR were .379 (ns) and -.589 (*p* < .05), respectively. Thus, there was a non-significant tendency for speakers to produce more words with correct vowels all three times as LOR increased. At the same time, however, there was a significant tendency for them to produce fewer words with incorrect vowels all three times as a function of LOR.

3.4 Differences Across Words

Despite the general patterns noted in sections 3.1 and 3.2, vowel accuracy sometimes varied according to word, even when the rhyme was the same. For instance, mean scores on 'kid' and 'lid' were 75% and 35% respectively, and on 'put' and 'foot' they were 77% and 60%. While the initial consonants in these words might have influenced vowel accuracy, other accounts are possible. To explore this phenomenon further, nine pairs of words with identical rhymes were selected on the basis of word frequencies from the Michigan Corpus of Academic Spoken English (MICASE) database. For each speaker, mean scores were computed for higher and lower frequency words. A paired comparison revealed that the higher frequency words exhibited significantly more intelligible vowels (80%) than the lower-frequency words (66%), t (17) = 6.76, p < .001).

4. DISCUSSION

The results of this study are consistent with a view of L2 segmental acquisition as a context-dependent, approximative, frequency-based process. The following outcomes were observed: (1) As posited by the SLM, vowel intelligibility was context-dependent. However, vowels were not produced less intelligibly before final /d/ than before other consonants, despite the absence of final /d/ in Cantonese. (2) Greater L2 experience (assessed in terms of LOR) was statistically associated with better intelligibility of /ɪ/ and /ʊ/ productions. (3) Intraspeaker variability was also tied to LOR such that speakers with longer residence tended to produce fewer words with consistently wrong vowels. This finding suggests a destabilization of previously non-native-like vowel representations as a function of L2 experience. (4) Vowels were not produced equally well in different words with the same rhyme. Although the reason for this finding cannot be firmly established, higher-frequency words tended to have more intelligible vowels. This outcome provides further support for the view that experience with L2 input played a role in the learners' L2 vowel acquisition.

REFERENCES

[1] Munro, M. J. & Derwing, T. M. (2008). Segmental acquisition in adult ESL learners: A longitudinal study of vowel production. *Language Learning*, to appear.
[2] Munro, M. J., Flege, J. E., & MacKay, I. R. A. (1996). The effects of age of second language learning on the production of English vowels. *Applied Psycholinguistics, 17*, 313-334.
[3] Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233-277). York Press, Timonium, MD.
[4] Barrie, M. (2003). Contrast in Cantonese vowels. *The Toronto Working Papers in Linguistics*, 20, 1-19.

ACKNOWLEDGMENTS

This project was funded by SSHRC. The author acknowledges the important contributions of Herman Li, Susan Morton, and Renee McCallum to data collection and analysis.

PRESSURE/MASS METHOD TO MEASURE OPEN POROSITY OF POROUS SOLIDS

Yacoubou SALISSOU and Raymond PANNETON
GAUS, Department of mechanical engineering, Université de Sherbrooke, Québec, Canada, J1K 2R1
Yacoubou.Salissou@USherbrooke.ca, Raymond.Panneton@USherbrooke.ca

1. INTRODUCTION

The open porosity of a porous solid is a key parameter in the physical and acoustical modeling of porous media, especially to relate the effective properties of the fluid saturating the interconnected pores to the effective properties of the porous solid.[1,2] In the past, several acoustical indirect methods have been proposed to measure this parameter[3,4] However, direct methods are generally preferred. Examples of direct measurement methods include the pioneer work of Beranek,[5] and its improved version standardized as ASTM D 2856-94. The method relies on the equation of state for ideal gases at constant temperature. However it is time consuming and requires several calibration of the device. Other direct measurement methods include the improvement of Beranek method proposed by Leclaire *et al.*[6] and by Champoux *et al.* [7] and the missing mass method proposed by Panneton and Gros[8].

In this paper, a method is presented to measure the open porosity of porous solids. The method needs only a simple apparatus and its accuracy is predictable from the knowledge of the total bulk volume of the tested porous sample.

2. THEORY

The open porosity of porous solid is expressed as

$$\phi = 1 - V_s / V_t \,, \tag{1}$$

where V_s is the volume of the solid phase, and V_t the total bulk volume of the porous aggregate. In general, V_s is unknown and needs to be determined. To determine its value, the four tests presented in fig.1 are used together with the perfect gas law relation. If the process between and during tests is very slow and room conditions are constant, isothermal condition can be assumed. The volume of the solid phase V_s can then be expressed as [see ref [9]]

$$V_s = \left(\left(M_2 - M_1 \right) / \left(P_2 - P_1 \right) - \left(\left(M_4 - M_3 \right) / \left(P_4 - P_3 \right) \right) \right) RT \tag{2}$$

where R is the specific gas constant, T is the temperature in Kelvin and M_i the different masses measured on the balance.

3. ERROR ANALYSIS

3.1 Applicability of the method

As shown, to apply the proposed method, volume V_s needs first to be evaluated. Usually, the target pressures to use are low pressure P_{lo} and high pressure P_{hi} so that $P_1 \cong P_3 \cong P_{lo}$, and $P_2 \cong P_4 \cong P_{hi}$. Also, it can be easily shown that the difference $M_i - M_{i-1} = m_i - m_{i-1}\big|_{i=2,4}$. Furthermore, $m_2 - m_4$ yields the mass of gas occupied by the solid phase of the porous sample at high pressure (i.e., $\rho_{hi} V_s$, where ρ_{hi} is the mass density of the gas at high pressure), and $m_1 - m_3$ yields the mass of gas occupied by the solid phase at low pressure (i.e., $\rho_{lo} V_s$, where ρ_{lo} is the mass density of the gas at low pressure). Consequently, $M_2 - M_4 - M_1 + M_3 = \left(\rho_{hi} - \rho_{lo} \right) V_s$ and ρ_{lo} corrects for the fact that the low pressure condition is not the perfect vacuum condition. Not accounting for $m_1 - m_3$ would only add a bias error proportional to P_{lo} in the evaluation of V_s and ϕ; however it does not prevent the applicability of the method.

From the previous analysis, the most severe condition that remains for applying the method is the readability of the high pressure mass difference (i.e., $m_2 - m_4 \geq \varepsilon$, where ε is the balance readability). This condition gives the maximum porosity the method can determine for a given set of operation conditions

$$\phi_{max} = 1 - RT\varepsilon / P_{hi} V_t \,. \tag{3}$$

Figure 2 gives the minimum bulk volume per balance readability as a function of open porosity for different operating conditions. It shows that larger bulk volume, larger high pressure, and gas with larger density are desirable to increase the applicability of the method, and to detect open porosity closer to unity (typical for sound absorbing porous materials).

3.2 Precision of the method

Using the total differential method and assuming that the errors are random and follow a normal distribution, the expected error committed on the open porosity is

$$\delta\phi = \pm \left(1 - \phi \right) \sqrt{ \left(\delta T / T \right)^2 + \left(\delta V_t / V_t \right)^2 + \left(\delta Z / Z \right)^2 } \,, \tag{4}$$

For a cylindrical test sample, the error on its bulk volume is given by $\delta V_t = \pm V_t \delta_L \sqrt{ \left(2/D \right)^2 + \left(1/H \right)^2 }$, where D and H are its diameter and height, and δ_L is the uncertainty on the dimension measurement. δZ is the error related to variable Z defined, following Eq.(2), as $Z = V_s / RT$. Considering that the uncertainty relative to the masses and pressures are given, respectively, by the balance and the manometer readability ε and p, and Assuming P_1 and P_3 are approximately equivalent and equal to low pressure P_{lo}, P_2 and P_4 are approximately equivalent and equal to high pressure P_{hi}, and test chamber volume $V >> V_s$, one obtains

$$\delta Z \cong 2\sqrt{\varepsilon^2 + \left(pV/RT\right)^2}\Big/\left(P_{hi} - P_{lo}\right). \tag{5}$$

The error predicted by the Eq. (4) is valid only if $V_t \leq V$. Figure 3 shows the expected absolute error on open porosity as a function of bulk volume per test chamber volume for three different open porosity values (0%, 90%, and 99%). The error generated by the individual uncertainties on the mass and pressure, the bulk volume, and the temperature are also plotted. It is noted that the precision of measurement is better when using larger bulk volume. It is also noted that for the three cases, the error is mostly controlled by the uncertainty on the mass and pressure reading. Since the error is mostly controlled by the uncertainty on the mass and pressure reading, the error on the open porosity can be estimated by the approximated expression given by

$$\delta\phi \cong 2\sqrt{\left(\left(RT\varepsilon\right)^2 + \left(Vp\right)^2\right)}\Big/V_t\left(P_{hi} - P_{lo}\right). \tag{6}$$

One can observe that it fits well with the global error in the valid range of bulk volume ($V_t/V \leq 1$).

4. EXPERIMENTAL TESTS AND RESULTS

To validate the gas porosimeter and its precision, two tests have been performed. The first test consists in applying the method to measure the porosity of high porosity samples (95%) of known solid phase volume V_s. The second test consists to applying the method to measure the porosity of low porosity samples (45%) of known solid phase volume V_s. Details on test setup and procedure can be found in ref[9]. Figure 4a compares the measured standard deviation (i.e., measured error) for each of the ten samples to the theoretical prediction. It is observed that the measured error fits well with the one predicted by Eq.(6). Figure 4b shows the measured mean porosity for the 30 individual tests as a function of the bulk volume to test chamber volume ratio to better visualize the scattering of the measurements around the theoretical value. One can clearly observed that the precision on the measurements improves with the bulk volume to test chamber volume ratio.

ACKNOWLEDGEMENTS

N.S.E.R.C., REGAL, and ALCAN supported this work.

REFERENCES

[1]Biot, J. Acoust. Soc. Am., 28, 168 (1956).
[2]Allard, Propagation of sound in porous media, Elsevier, NY (1993).
[3]Umnova et al., Appl. Acous., 66, 607 (2005).
[4]Fellah et al., J. Appl. Phys., 93, 296 (2003).
[5]L.L. Beranek, J. Acoust. Soc. Am., 13, 248 (1942).
[6]Leclaire et al., Rev. Sci. Inst., 74, 1366 (2003).
[7]Champoux et al., J. Acoust. Soc. Am., 89, 910 (1990).
[8]R. Panneton and E. Gros, Acta Acustica, 91, 342 (2005).
[9]Salissou and Panneton, J. Appl. Phys. 101, 124913 (2007)

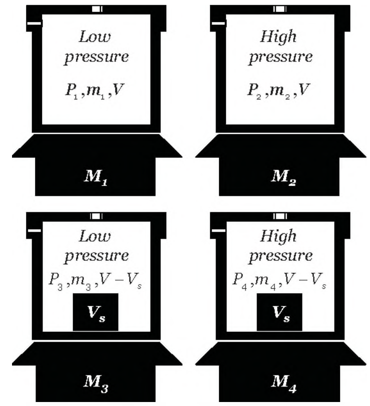


Figure 1: Measurement procedure. For the four conditions shown in the figure, the mass is measured on the balance.

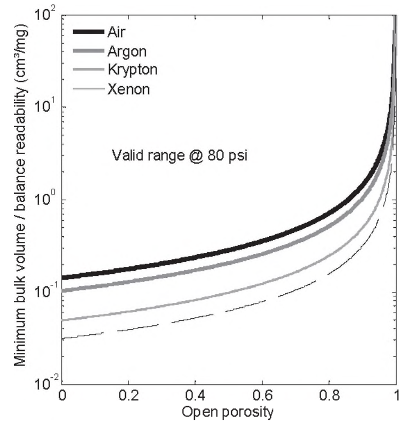


Figure 2: Minimum bulk volume per balance readability as a function of open porosity.

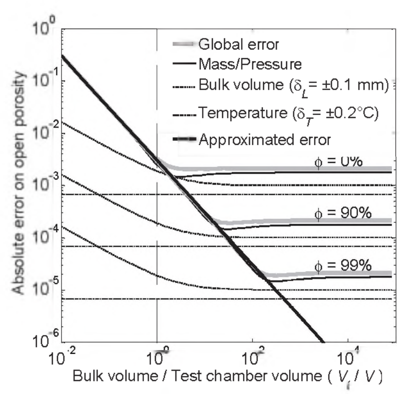
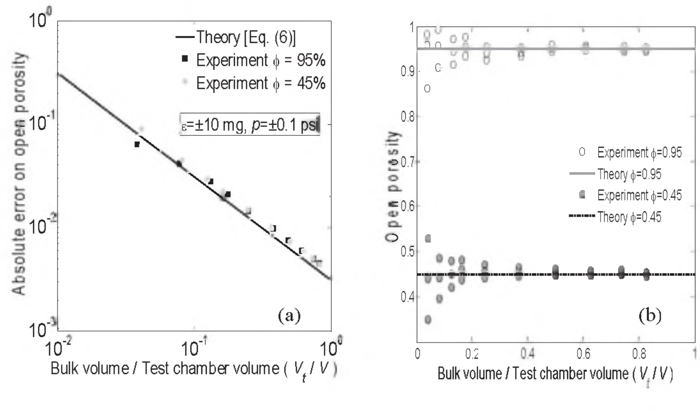


Figure 3: Theoretical absolute error on open porosity as a function of bulk volume per test chamber volume.

Figure 4 : Experimental tests results



FABRICATION OF ACOUSTIC ABSORBING TOPOLOGIES USING RAPID PROTOTYPING

Oliver Godbold[1], Jian Kang[2], Richard Buswell[3], Rupert Soar[1]
[1]Wolfson School Of Mechanical and Manufacturing Engineering, Loughborough University, Loughborough, Leicestershire, LE11 3TU, United Kingdom; oliver.godbold@googlemail.com, r.c.soar@lboro.ac.uk
[2]School of Architecture, Arts Tower, University of Sheffield, Western Bank, Sheffield, S10 2TN, United Kingdom; j.kang@sheffield.ac.uk
[3]Department of Civil and Building Engineering, Loughborough University, Loughborough, Leicestershire, LE11 3TU, United Kingdom; r.a.buswell@lboro.ac.uk

1. INTRODUCTION

Porous absorbing materials traditionally constitute fibrous, granular or open celled foam material types. The manufacturing methods associated with these materials, impose limitations on the topology of the porous microstructure. Pore size and shape are determined by pseudo-random processes such as the layering of bonded fibrous filaments, packing of granulated materials, or the nucleation and growth of gas bubbles within a polymer matrix. These processes only allow a limited level of control over the topology of the porous microstructure, restricting the structure types that can be realised, and introducing uncertainty into the prediction of their acoustic properties. Theories based on extensive empirical measurements such as those described by Delany & Bazely [1], disregard the geometric differences between individual pores by deriving correction factors from measurements of bulk material samples; however, this limits their applicability. Theoretical models which aim to be more generic, describe the visco-thermal effects within individual pores (e.g. Allard & Champoux [2]), but require the use of generalized pore shape factors or characteristic lengths, and an estimation of the distribution of pore sizes, which are often derived from bulk material properties.

Computer controlled additive manufacturing processes are commonly referred to as 'Rapid Prototyping' (RP). These processes precisely control the deposition of material to build any virtually given component geometry. Applying these techniques to the manufacture of acoustic absorbers has the potential to improve pore shape definition, optimize porous structures and deliver a more uniform pore size distribution. RP encompasses a broad family of different processes. The processes are similar in that the additive fabrication machines are driven directly from a computer model. They produce three dimensional shapes from two dimensional cross sectional slices sequentially bonded together from the bottom to the top of the part. The processes differ in their method of producing each cross sectional slice. Some processes selectively solidify, sinter or cut a bulk build material, while others selectively deposit material where it is required. One such process known as Fused Deposition Modelling (FDM) selectively deposits a fine thermoplastic filament and can vary the thickness and spacing between filaments, as well as the angle between filaments on adjacent layers. This process, illustrated in Fig. 1, allows the fabrication of a uniform porous material with a definable structure. The filament diameter can be fabricated down to 0.305mm, 1-2 orders of magnitude higher than that of traditional fibrous materials.

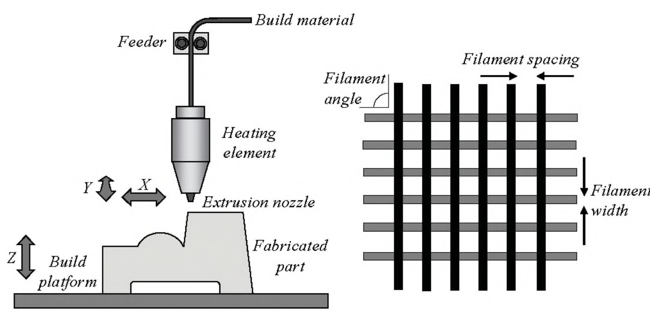


Fig. 1. FDM process schematic and porous structure variables.

The materials and characteristics of each RP process vary, and some are more suited than others for the direct fabrication of acoustically absorbing topologies. All RP processes, however, introduce benefits for design and fabrication that go beyond many traditional methods of manufacture. In particular, the production of complex re-entrant shapes is possible without the requirement for tool clearance which is critical for the realization of intricate or customized acoustic absorbers.

2. FDM POROUS STRUCTURES

To investigate the potential use of RP in the fabrication of acoustically absorbing topologies, porous samples were produced using the FDM process. Samples with different porosities were produced by varying the filament spacing. Samples 25mm thick were produced with 5%, 10%, 25% and 50% porosity, from filaments 0.508mm thick, with a 90 degree filament angle. The absorption properties of each sample were measured using an impedance tube in accordance with BS EN ISO 10534-2 [3]. Two samples of each configuration were produced: one 100mm in diameter and one 28mm in diameter to allow testing using low and high frequency impedance tubes. The measured absorption coefficient, α for each configuration is depicted in Fig. 2.

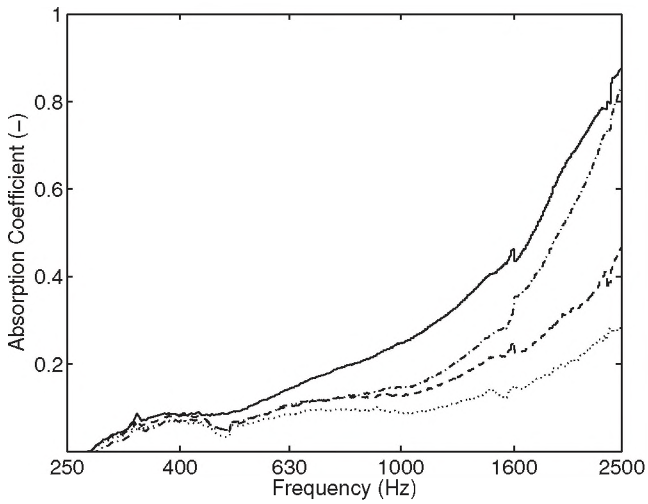


Fig. 2. Measured impedance tube absorption for porous FDM samples (porosity: —— 5%, –·–·· 10%, – – – 25%, ······· 50%).

The absorption results show that a porous structure produced using FDM is capable of achieving significant absorption over 1600Hz. The unconventional pore structure is unlike traditional porous materials in terms of regularity, porosity and fiber size and consequently the comparison of these results against existing theoretical models is difficult; however, a trend of increasing absorption can be noticed as the porosity decreases. Fig. 2 indicates that the acoustic properties of the test samples can be altered through changes to the manufacturing process parameters.

3. FDM POROUS STRUCTURES FOR DAMPED RESONANT ABSORPTION

The relatively large fiber diameter limits the production of small pore sizes to samples with low porosity. A level of flow resistivity high enough to promote porous absorption, therefore, is only possible with the lower porosity samples. The modest flow resistivity characteristics associated with the higher porosity samples, while not immediately suited to porous absorption, may be used to add resistance to Helmholtz type resonant absorbers. The addition of resistance dampens the resonant effect, broadening the bandwidth of effective absorption. This additional application was investigated through the fabrication of a resonant test sample consisting of perforations 4mm in diameter, 5mm in length, arranged with a 12% open area and a 45mm backing air space. The perforations were covered with an 8mm thick porous layer, with 25% porosity. Both the resonant and porous elements of the test sample were fabricated autonomously in a single FDM build, from the same material. The sample was 100mm in diameter for compatibility with the large impedance tube; consequently the measurable frequency range was restricted to below 1600Hz. The measured absorption coefficient for the damped resonant absorber is given in Fig. 3.

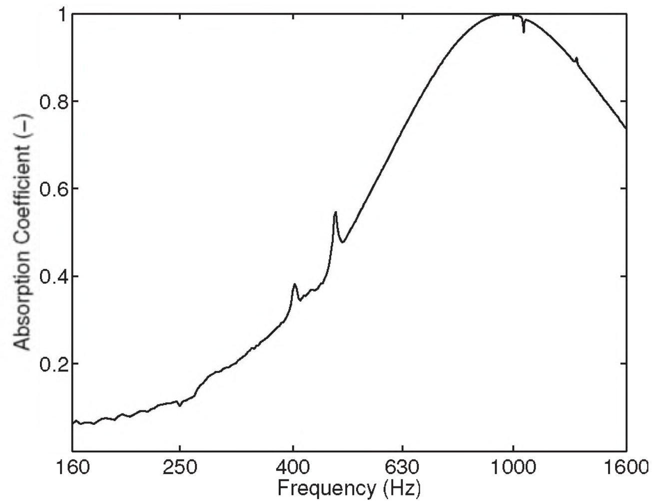


Fig. 3. Measured impedance tube absorption for FDM damped resonant absorber.

The results show that the combined resistive and resonant elements provide significant low frequency resonant absorption over 500Hz. Comparing the measured results to theoretical resonator models indicates that the porous covering offers a flow resistance of 5435 Pa s m^{-2}.

4. CONCLUSIONS AND DISCUSSION

The FDM process has been used to demonstrate the potential of a porous structure, whose characteristics can be established before manufacture. Various acoustic properties can be obtained by changing the manufacturing process parameters. Different configurations have been shown to exhibit significant high frequency absorption, or provide damping to low frequency resonant absorbers.

The ability to accurately produce a uniform, definable pore structure removes much of the uncertainty associated with the prediction of a porous material's acoustic properties, reducing variability between the prediction and performance of fabricated products or parts. It also presents a potential method of validating porous material theoretical models. There is potential to produce topologies with different pore shapes or with variable porosity throughout a single sample. Further investigation into these structure types, the effect of other FDM process parameters, and the application of theoretical models is needed to fully ascertain the potential benefits offered by this technology.

REFERENCES

[1] M. E. Delany & E. N. Bazley. (1970). Acoustical properties of fibrous absorbent materials. *Applied Acoustics,* 3, 105-116.

[2] J. F. Allard & Y. Champoux. (1992). New empirical equations for sound propagation in rigid frame fibrous materials. *Journal of the Acoustical Society of America*, 91, 3346-3353.

[3] British Standards Institute (2001). *BS EN ISO 10534-2:2001 Acoustics -Determination of sound absorption coefficient and impedance in impedance tubes - Part 2: Transfer function method*, BSI: London.

THE TEMPORAL WINDOW OF AUDIO-TACTILE INTEGRATION IN SPEECH PERCEPTION

Bryan Gick[1, 2] and Yoko Ikegami[1]
[1]Dept. of Linguistics, University of British Columbia, Vancouver, 2613 West Mall, British Columbia, Canada V6T1Z4
[2]Haskins Laboratories, 300 George St., 9th Fl., New Haven, CT 16511, USA

1. INTRODUCTION

Asynchronously presented auditory and visual information is integrated asymmetrically in speech perception [1, 2, 3]. For example, Munhall & al. [4] found that audio-visual integration of speech occurred even when the audio signal lagged the video signal by 240ms; however, when the audio signal preceded the video signal, perceivers only integrated 60ms of asynchrony. Munhall & al. suggest that this asymmetrical effect window may be attributable to perceivers' learned awareness of physical properties of the natural world (in this case, of the differing atmospheric speeds of sound and light): "This trend is not surprising since the relative speeds of sound and light would produce many natural occurrences of auditory events lagging their visual counterparts in the natural world" [4, p. 354]. However, this explanation has not been substantiated via comparison with other perceptual modalities.

Replication using the tactile modality should provide a test case for this question: Fowler & Dekle [5] and Gick *et al.* [6] found that untrained perceivers integrate tactile and auditory modalities through direct manual contact with speakers' faces. However, even if realistic and precisely timed synthetic facial (presumably robotic) stimuli could be constructed, this methodology would still fail to provide a natural signal transmission delay comparable to that of light or sound. The present experiment responds to this by coupling an acoustic speech signal with speech-like synthetic tactile stimuli in the form of small bursts of air following aspirated consonants.

The air speed of speech-like turbulent flow is considerably slower than that of sound in air, with flow velocity dropping off log-linearly after expulsion from the mouth [7]. If the physics-based hypothesis (i.e., the explanation based on perceivers' awareness of the relative physical transmission times of different signals) is correct, then the direction of asymmetry in the perceptual integration window should parallel the temporal difference between the relative speeds of sound and air flow. Any other result will fail to support the physics-based hypothesis.

2. METHOD

2.1 Participants

13 adult perceivers participated in the study. All were native speakers of English with no history of speech or hearing problems.

2.2 Stimuli

Acoustic stimuli consisted of recordings of 440 tokens of *pa* and *ba* produced in random order by a single female English speaker. Acoustic stimuli were output through the right channel of a Mac G4 sound card, mixed through a PreSonus mixing board with white noise (at a level such that subjects' baseline correct identification of *pa/ba* was at approximately 75%) and played to participants in stereo through Direct Sound Extreme Isolation headphones.

Tactile stimuli consisted of gentle bursts of air imparted via a vinyl tube at 7cm from the skin. Bursts were released from an air compressor at ~5psi using a Teknocraft 12-volt DC 2-way solenoid valve with a .032-inch orifice. The switch operating the solenoid valve was activated by a voltage initiated by an acoustic square wave output through the left channel a Mac G4 sound card amplified to 5 volts using a Frequency Devices voltage amplifier. Square waves were 60ms long (the average duration of aspiration for "pa" tokens used in the experiment), and offset leftward by 30ms to correct for a 30ms total system latency (see Figure 1).

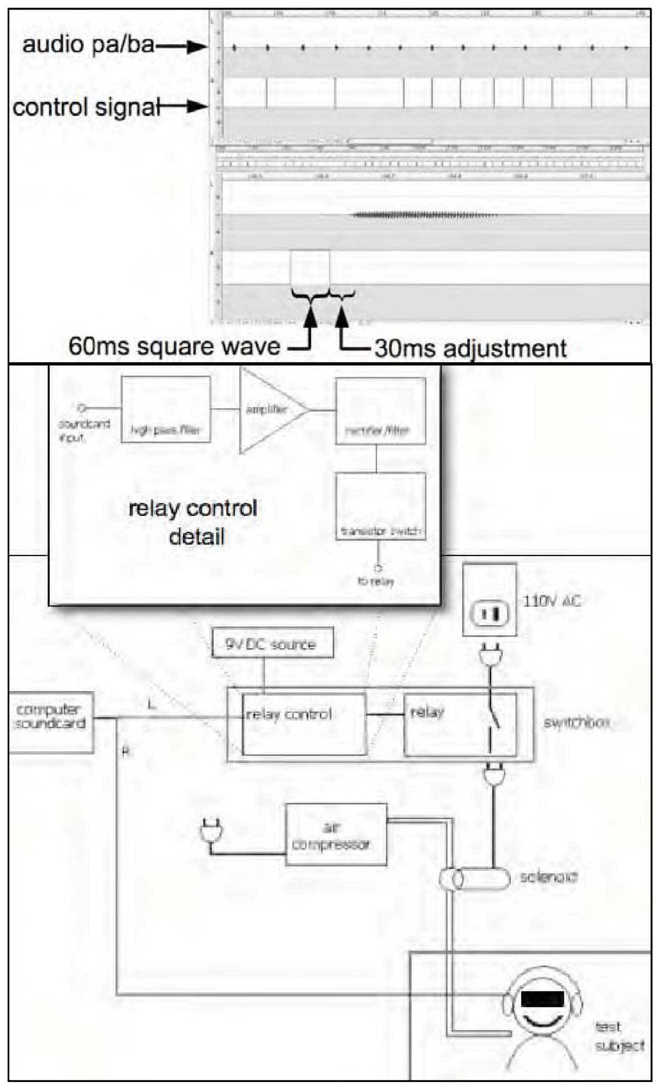


Fig. 1. Example of acoustic control and *pa/ba* signals (top) and flowchart of stimulus presentation system (bottom).

24 experimental conditions were tested, with temporal offsets between air bursts and spoken tokens varying by condition as follows: No Burst, 0ms (Simultaneous), ±50ms, ±100ms, ±200ms, ±300ms, and ±500ms (Distractor). Each

offsets occured in two different conditions (once with *pa* and once with *ba*). Participants heard 20 examples for each condition (except for the ±500ms distractors, for which there were only 10 items each), randomly distributed throughout the experiment, with one example per 3 seconds.

2.3 Procedure

Participants were seated in a soundbooth and read a script describing this experiment as testing their ability to identify different spoken syllables under conditions similar to those experienced by an airplane pilot. No specific mention was made of the air tube (indeed, some subjects reported not being aware of the air burst at all during the experiment). Participants were briefly trained to give forced-choice responses using a button box (with L/R responses balanced across participants), then blindfolded. Headphones were then placed on the participant, and the air tube put in place aiming at the right side of the neck.

3. RESULTS

Figure 2 shows the mean percent of correctly identified *pa* and *ba* syllables across subjects, plotted by condition. Paired t-tests (by subject) indicate significant enhancement to identification of *pa* responses with burst, and significant interference with identification of *ba* responses, in Simultanous conditions (compared with No Burst baseline conditions; $p > .05$). For both *pa* and *ba*, the effect at -50ms was not significantly different from Simultaneous; however, while the effect continued only to +50ms for *ba*, it persisted to a delay of +200ms for *pa* (as indicated by the circles in Figure 2).

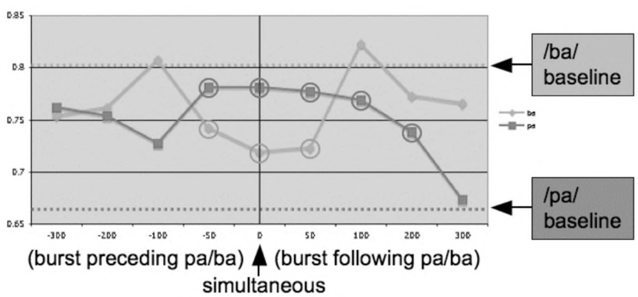


Fig. 2. Mean percent of correctly identified "pa" (dark grey line) and "ba" (light grey line) syllables. Circles indicate contiguous temporal offset conditions where percent identification did not differ significantly from Simultaneous.

4. DISCUSSION

In this experiment, tactile stimuli in the form of small bursts of air were directed at perceivers' necks while they heard productions of *pa* and *ba*. In baseline conditions, a burst occurring immediately prior to vowel onset (i.e., simultaneous with aspiration for *pa*) significantly enhanced perception of *pa* and significantly interfered with perception of *ba*. Asynchronous results showed a similar effect window to previous audio-visual studies: For asynchronously presented bursts, the temporal window of the enhancement effect of air bursts on perception of *pa* (but not the interference effect on *ba*) was asymmetrical, with integration occurring when the air burst followed the audio signal by 200ms, but only by 50ms when the air burst preceded the audio signal. The direction of this perceptual asymmetry parallels the temporal difference between the speeds of sound and air flow, supporting the physics-based hypothesis. Future work will attempt to address the question of whether perceivers' apparent understanding of physical properties of the world is learned or innate.

REFERENCES

[1] N. Dixon & L. Spitz. (1980). The detection of audiovisual desynchrony. *Perception*, 9, 719-721.

[2] P. M. T. Smeele, A. C. Sittig & V. J. Van Heuven. (1992). Intelligibility of audio-visually desynchronized speech: Asymmetrical effect of phoneme position. In *Proceedings of the International Conference on Spoken Language Processing*, pp. 65-68.

[3] Q. Summerfield. (1992). Lipreading and audio-visual speech perception. *Philosophical Transactions of the Royal Society of London: Series B*, 335, 71-78.

[4] K. G. Munhall, P. Gribble, L. Sacco & M. Ward. (1996). Temporal constraints on the McGurk effect. *Perception & Psychophysics,* 58(3), 351-362.

[5] C. A. Fowler, & D. J. Dekle. (1991). Listening with eye and hand: Cross-modal contributions to speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, 17, 816-828.

[6] B. Gick, K. Jóhannsdóttir, D. Gibraiel and J. Muehlbauer. (2008). Tactile enhancement of auditory and visual speech perception in untrained perceivers. *Journal of the Acoustical Society of America,* 123(4), EL72-76.

[7] P. Anderson, D. Derrick, B. Gick & S. Green. (Under review). Characteristics of air puffs produced in English 'pa': Experiments and simulations. *Journal of the Acoustical Society of America*.

ACKNOWLEDGMENTS

This project benefited greatly from technical contributions of Gordon Ramsay and discussions with Douglas Whalen, and was funded by an NSERC Discovery Grant to Bryan Gick and NIH Grant DC-02717 to Haskins Laboratories.

ANALYSIS OF A VOWEL DATABASE

Peter F. Assmann[1], Terrance M. Nearey[2] and Sneha Bharadwaj[1]
[1]School of Behavioral and Brain Sciences, The University of Texas at Dallas, Richardson TX 75083-0688 USA
[2]Department of Linguistics, University of Alberta, Edmonton, Alberta, Canada T6G 2E7

1. INTRODUCTION

To study developmental changes in the acoustic properties of vowels, we have collected a database of recordings from adults and children from the Dallas, Texas region. Across age and gender classes, we find a systematic relationship between the geometric mean of the formant frequencies (F1-F3) across all of the vowels for a given speaker (a measure related to vocal tract length) and the geometric mean fundamental frequency (f_0, reflecting the rate of vocal fold vibration). The aim of this study is to provide a preliminary analysis of the developmental changes in these acoustic properties.

2. VOWEL DATABASE

2.1 Speakers

To date we have recorded 163 children and 39 adults. The children ranged in age from 5 to 18 years of age, with at least 5 boys and 5 girls at each age. The adults were undergraduate students at the University of Texas at Dallas ranging in age from 19-45 years. All spoke English as their first language and were long-term residents of the North Texas region.

2.2 Vowels

Vowels were recorded in hVd words, both in isolation and in a carrier sentence, "Please say the word __ again." The recordings included the 12 monophthongal vowels of American English, /ι/, *heed*; /I/, *hid*; /ε/, *hayed*; /E/, *head*; /Θ/, *had*; /℘/, *hud*; /™/, *herd*; /A/, *hod*; /□/, *hawed*; /o/, *hoed*; /Y/, *hood*; /υ/, *who'd;* and three diphthongs, /Aι/, *hide*; /Aυ/, *how'd*; and /□ι/, *hoyed.*

2.3 Recording procedure

Recordings were carried out over a 2.5-year period at the University of Texas at Dallas. Each recorded token was produced following a screen prompt that displayed the orthographic representation of the hVd word or sentence, along with an audio example spoken by an adult female from the Dallas area. Five repetitions of each of the 12 vowels were elicited from each of the children, and at least 10 repetitions were collected from each adult. Recordings were made in a sound-treated room using a Shure SM-94 microphone, Symetrix SX202 dual-microphone pre-amplifier and Tucker-Davis Technologies data acquisition hardware (MA1, RP2.1). The digital waveforms were stored on computer disk at a rate of 48,828 Hz and 16-bit resolution. Following data collection, each recording was screened by listening and tokens judged to be misarticulated or of low recording quality were omitted from subsequent analysis.

2.4 Acoustical analysis

A semi-automated procedure was used to mark the onset and offset of the vowel in the isolated hVd words. Vowel onset was defined as the beginning of the first pitch period of the voiced segment of the syllable. Vowel offset was defined as the end of the last pitch period before the silent interval created by the final /d/ or by a substantial drop (>15 dB) in the levels of the higher formants (F2-F5) in cases where pre-voicing filled the silent interval. Formant frequencies were measured every 2 ms using an automatic formant tracking program[1]. Fundamental frequency was estimated with 1-ms resolution using a procedure developed by Kawahara.[2]

3. RESULTS

Figure 1 plots the geometric mean f_0 across all of the frames, tokens, vowels, and talkers as a function of age and sex. Consistent with earlier findings[3], there is a progressive decline in f_0 as a function of age. For males there is a sharp break around age 13 and an increase in variability at ages 13 and 14, while females show a more gradual decline up to age 18.

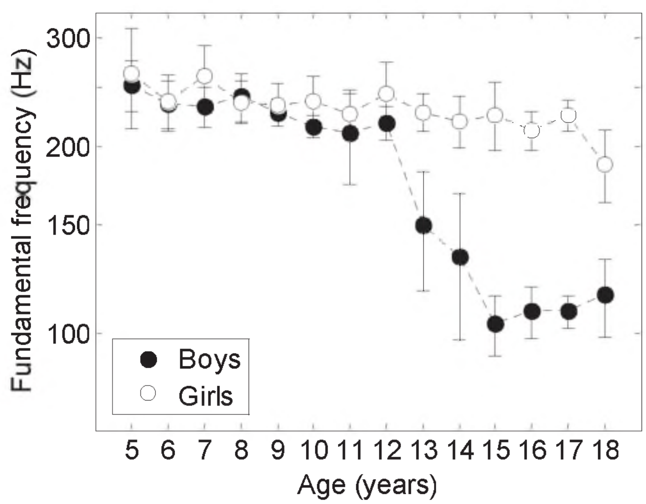


Fig. 1. Geometric mean f_0 averaged across frames, tokens, vowels and talkers. Error bars show standard errors across talkers.

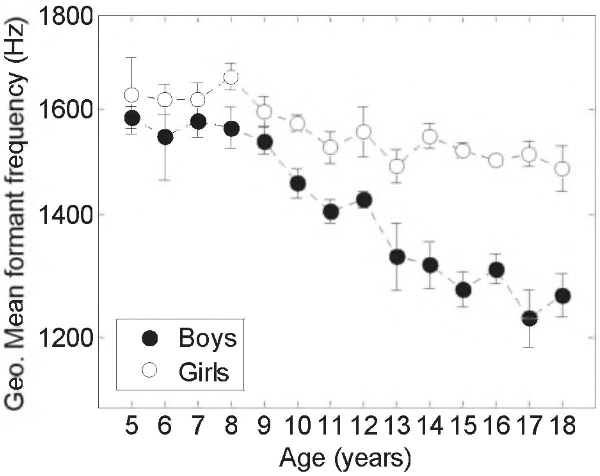


Fig. 2. Geometric mean of the formant frequencies F1, F2, and F3 averaged across frames, tokens, vowels and talkers. Error bars show standard errors across talkers.

Figure 2 plots the geometric mean of the formant frequencies (F1, F2, and F3) across all of the frames, tokens, vowels, and talkers as a function of age and sex. This measure is related to vocal tract length and also shows the predicted decline as a function of age. It is noteworthy that differences between boys and girls in mean formant frequencies appear as early as age 8, consistent with findings of other recent studies[4].

The scatterplot in Figure 3 shows that the geometric mean f_0 increases linearly with the geometric mean of the formant frequencies when the data for males and females are combined. The relationship is weaker for females alone (r=.46) compared to males alone (r=.87). It also appears to be weaker in younger children.

4. DISCUSSION

In a recent study[5] we have investigated the perceptual effects of upward and downward scaling of the formants in combination with changes in f_0. The results indicate that coordinated shifts (formant frequencies and f_0 scaled in the same direction) resulted in higher identification accuracy than uncoordinated shifts (formants and f_0 scaled in opposite directions). One explanation is that listeners are sensitive to the statistical covariation illustrated in Figure 3.

REFERENCES

[1] Nearey, T.M., Assmann, P.F. and Hillenbrand, J.M. (2002). Evaluation of a strategy for automatic formant tracking. *J. Acoust. Soc. Am.* 112: 2323.

[2] Kawahara H., de Cheveigné A., Banno H., Takahashi T. and Irino T. (2005). Nearly Defect-free F0 Trajectory Extraction for Expressive Speech Modifications based on STRAIGHT, Proc. Interspeech2005, Lisboa, pp. 537-540, Sept. 2005.

[3] Lee S., Potamianos A., Narayanan S. (1999). Acoustics of children's speech: Developmental changes of temporal and spectral parameters. J. Acoust. Soc. Am. 105: 1455-1468.

[4] Perry T.L., Ohde R.N. and Ashmead, D.H. (2001). The acoustic bases for gender identification from children's voices. J. Acoust. Soc. Am. 109: 2988- 2998.

[5] Assmann, P.F. and Nearey, T.M. (2008). Identification of frequency-shifted vowels. Submitted to *J. Acoust. Soc. Am.*

ACKNOWLEDGMENTS

This research was funded by NSF and SSHRC.

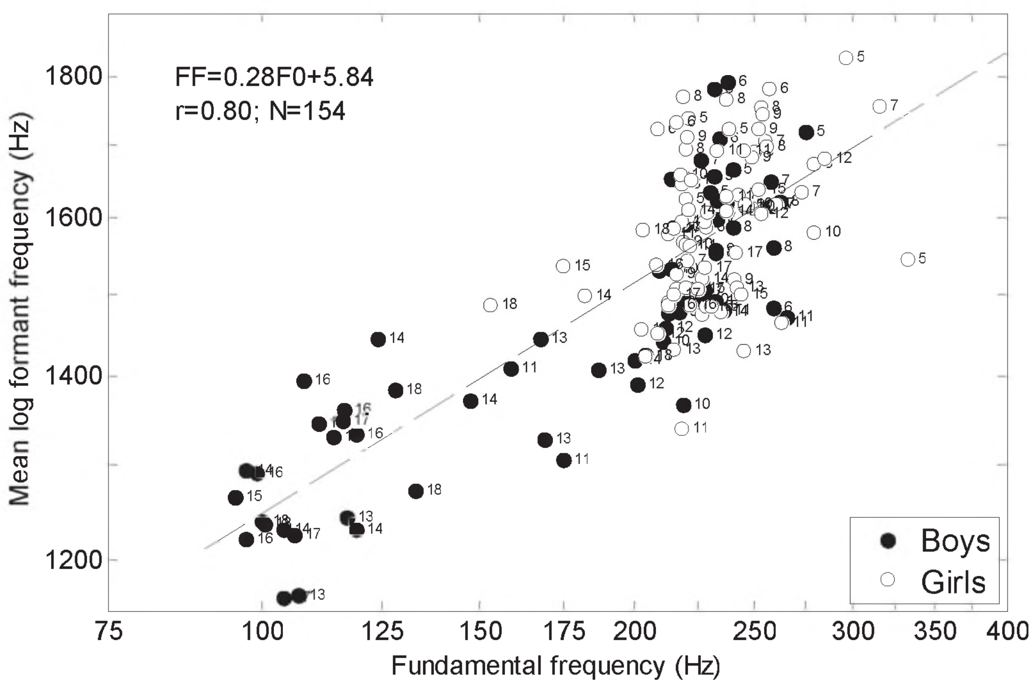


Fig. 3. Geometric mean of the formant frequencies F1, F2, F3 against the geometric mean f_0, both averaged across frames, tokens, and vowels separately for each of the children in the database. The numbers next to each circle indicate the talker's age.

PATTERN RECOGNITION IN SPEECH PERCEPTION RESEARCH

Terrance M. Nearey
Dept. of Linguistics, University of Alberta, Edmonton AB, Canada T6G 2E7
email t.nearey@ualberta.ca

1. INTRODUCTION

We have successfully applied simple pattern recognition techniques to several problems of speech perception. Nearey and Hogan (1986) describe the NAPP (normal *a posteriori* probability) method, based on linear discriminant analysis. Nearey and Assmann (1986) apply NAPP to accurately predict listeners' behavior in the perception of modified natural English vowels. Recently, variations of NAPP have also been applied successfully to cross-linguistic and L2 vowel perception (Morrison 2006, Thompson 2007). Direct application of NAPP involves training a pattern recognizer on natural production measurements and using the 'frozen' model to predict listeners' behavior on new stimuli, without any further tuning. This paper sketches the use of more flexible pattern recognition methods in speech perception research. These include logistic regression and methods imported from automatic speech recognition (ASR) technology.

2. NAPP AND MNLR

The APP (*a posteriori* probability) scores generated by a NAPP model can be expressed in the form of a multinomial logistic regression (MNLR). MLNR has a long history in econometrics (Train 2003) to model discrete choice (by, e.g., consumers). MNLR can be tuned to approximate listeners' response data directly. The question of phonetic compositionality of perceptual choices among (e.g.) CV or VC syllables has been investigated extensively by Nearey (e.g., 1997) using MNLR techniques. For the cases studied, listeners' sensitivity to stimulus properties seems to be linked phoneme- or subphoneme-sized units. Larger units such as syllables, do not appear to associate with specific stimulus patterns in the ways that lower level units do. A related application is discussed below.

3. VC(C)V SYLLABLES

Nearey and Smits (2002) describe a variation of an experiment by Repp (1983) which involves variable (phoneme) length utterances of the form of VC(C)V. We were not at all clear that MNLR models would reveal the kind of compositionality we found with simpler response sets. Our experiment spanned the following responses {aba, ada, ab#ba, ad#da, ab.da and ab.da}, where '#' indicates a phonotactically necessary (in English) word boundary and '.' Indicates a syllable (and possibly word) boundary. The vowel denoted 'a' is low back and slightly rounded in the dialect under study.

3.1 Method

A total of 144 (6 x 6 x 4) stimuli were created by a standard Klatt80 synthesizer. The stimuli were arrayed in *fully crossed* 3-factor design with the following values:
1) *Closing F2* (and correlated F3) associated with VC (ab- and ad-) offset [1060 (2180) to 1450 (2539) Hz in 6 steps]. 2) *Opening F2* (and correlated F3) with CV (-ba and –da) onset [1099 (2262) to 1635 (2500) Hz in 6 steps]. 3) *Gap Duration* at 4 levels 80, 120, 190 and 300 ms. The initial vocoid, V1 had [F1 F2 F3] targets of [777 1147 2466] Hz and a fixed duration of 190 ms including 50 ms V1C transitions; The final vocoid, V2, had the same target frequencies and a duration 300 ms including the CV2 transitions, F0 was fixed at 125 Hz for V1. And a linear downward trend from 125 to 100 Hz for V2. The amplitude of voicing (av) was set to 60 dB at the beginning of V1, it fell abruptly to 0 dB at V1 offset and rose abruptly to 60 dB at V2 onset. Participants were 13 native speakers of Canadian English. Each responded to 10 repetitions of each of the 144 stimuli. Response button layout on a PC screen was as follows:

[b] [bb] [bd]
[d] [dd] [db]

3.2 Results

An initial MNLR analysis was conducted (Nearey and Smits 2002). The response factor comprised the 6 response categories above. The independent variables were the *Closing F2*, *Opening F2* and *Gap Duration* (expressed as square root of ms). This model provided a relatively good fit, with a residual RMS error of about 6 percentage points. The model predicted the modal (most popular) response of listeners for almost 94% (135/144) of the stimuli.

The clustering patterns of consonant responses in Figure 1 (and associated t-tests) suggested a factored (compositional) solution, whereby judgment log-odds were tuned continuously by only the following three factors: (1) Closing place (ab-/ad-) was tuned only by *Closing F2;* (2) opening place (-ba-/da-) only by *Opening F2*. Finally, define a third phonetic factor, cluster type, comprising singletons (-b-, -d-), geminates (-b#b-, -d#d-) and true clusters (-b.d-, -d.b-). Then cluster type is tuned by (3) *Gap Duration* only. A reduced logistic model enforcing the decomposition above shows RMS error and modal agreement that are nearly indistinguishable from the full CC model. Note that this analysis involves splitting even singleton stops (e.g., -b-) into *closing*, *gap*, and *opening* subparts that can be shared with other C(C) patterns.

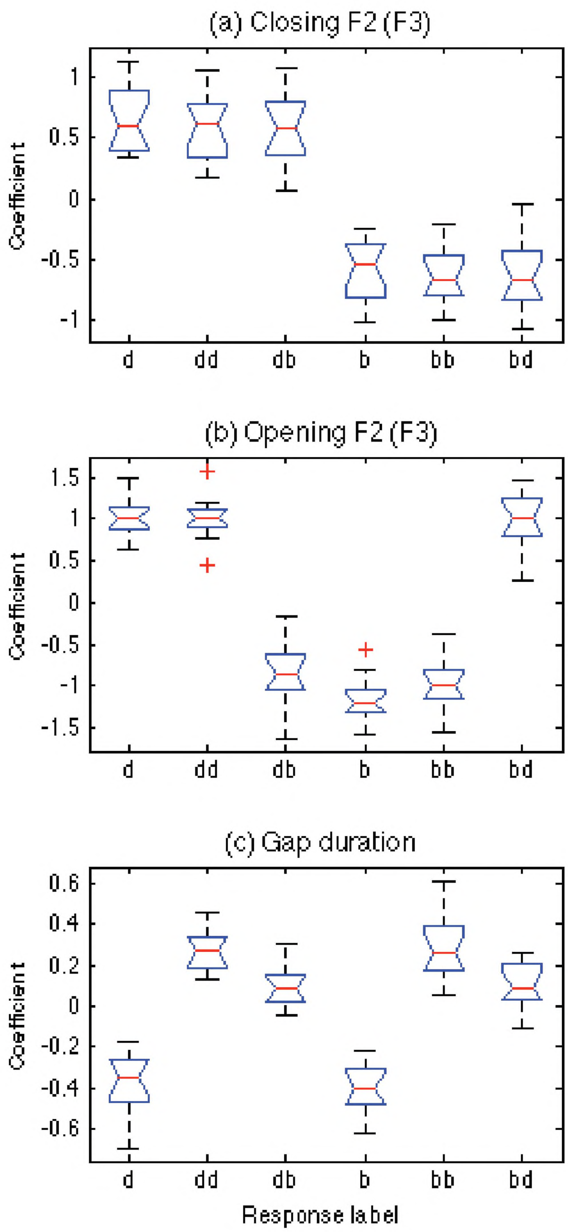


Fig. 1. Coefficients of logistic regression for stimulus properties in experiment 1.

4. BRIDGING TO ASR METHODS

The methods describes so far involve *static* pattern recognition, requiring exactly the same number of signal properties for each stimulus. Outside the laboratory, speech chunks come in many sizes. Longer ones have more properties than shorter ones (e.g., *Albuquerque* vs. *Al*). ASR technology has developed several *dynamic* pattern recognition methods that handle such variable inputs. Preliminary experiments (Nearey 2004) match the results of section 3 using a (dynamic) hidden semi-Markov model (HSMM, Guédon, 1992). The model uses sub-phone states (e.g, *b-closing, cluster-gap, d-opening*) directly related to the subphone elements of section 3. Using a maximum mutual information criterion, the HSMM can be tuned to listeners' responses to performs as well as the MNLR models above. The HSMM uses a conventional frame-based, mel frequency cepstrum representation The resulting system provides a complete framework for modeling speech perception that starts with raw waveforms and culminates in accurate predictions of listeners' responses. This first step bodes well for the future of incorporation of modeling technologies from ASR directly into speech perception research. With time, it may facilitate feedback in the other direction.

REFERENCES

Guédon, Y. (1992). Review of several stochastic speech unit models. Comput. Speech and Lang., 6(4), 377-402.
Nearey, T. M. (1997). Speech perception as pattern recognition. J. Acoust. Soc. Amer., 101(6), 3241--3256.
Nearey, T. M., & Hogan, J. (1986). Phonological contrast in experimental phonetics In J. Ohala, & J. Jaeger (Eds.), Experimental Phonology (pp. 141-161.). New York: Academic Press.
Nearey, T. M., & Smits, R. (2002). Patterns in the Perception of VC(C)V strings. J. Acoust. Soc. Amer.,, 112, 2323 (abstract).
Nearey, T. M. (2004). Adapting automatic speech recognition methods to speech perception: A hidden semi-Markov model of listener's categorization of a VC (C) V continuum. J. Acoust. Soc. Amer., 116, 2570 (abstract).
Repp, B. H. (1983). Bidirectional contrast effects in the perception of VC-CV sequences. Perception and Psychophysics, 33(2), 147-155.
Train, K. (2003). Discrete Choice Methods with Simulation. Cambridge: Cambridge University Press.

ACKNOWLEDGMENTS

This research was funded by SSHRC.

MISMATCH BETWEEN THE PRODUCTION AND PERCEPTION OF F_0 IN NEW ZEALAND ENGLISH ETHNOLECTS

Anita Szakay

Dept. of Linguistics, University of British Columbia, Totem Field Studios, BC, Canada, V6T 1Z4

szakay@interchange.ubc.ca

1. INTRODUCTION

Previous production experiments demonstrated that the two main ethnic dialects of New Zealand English significantly differ in prosodic features, such as rhythm and intonation (e.g. Britain 1992, Warren 1998, Szakay 2006). Szakay (2006) also showed that Maori English mean pitch is significantly higher than that of Pakeha English, the dialect spoken mainly by speakers of European descent. The study also indicated that the increasing F_0 values of Maori English are a result of a change in progress, with young Maori speakers using significantly higher mean pitch than young Pakeha speakers. Fig. 1. – taken from Szakay (2006) – demonstrates the interaction between age and ethnicity as predictors of mean pitch values ($p<.05$) in New Zealand.

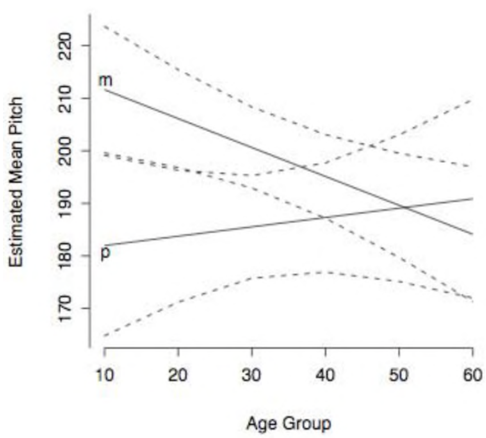


Fig. 1. Regression model predictions for estimated mean F_0 (Hz) for Maori speakers (m) and Pakeha speakers (p). Values adjusted to female speakers. Dashed lines indicate 95% confidence intervals.

The present perception study is part of a larger research project using innovative techniques to isolate the precise prosodic features that listeners might tune into in ethnic dialect identification in the New Zealand context. The study set out to investigate whether naïve listeners are aware of the differing suprasegmental features of the two ethnic dialects and whether they can identify the dialects if there are no segmental cues available in the speech signal. The results reported here relate to the perception of mean pitch only. For an extensive coverage of the perceptual relevance of other prosodic cues consult Szakay (2008).

2. METHOD

107 listeners (52 Maori, 55 Pakeha) performed a forced-choice dialect identification task. Based on the speech samples of 20 speakers (10 Maori, 10 Pakeha), seven speech conditions were created, each keeping different suprasegmental information in the speech signal (e.g. low-pass filtered condition, resynthesized rhythm-only at mean pitch condition, intonation-only condition). Condition Three was created as a monotonous speech rhythm only condition. Each consonant and pause was replaced by silence, while vowels were replaced by a tone complex created in Praat as a sum of a number of cosine waves with equidistant frequencies at a sampling frequency of 8000Hz. It was created at the mean pitch across all speakers according to gender (118Hz for males, 188Hz for females). A sample spectrogram of Condition Three is given in Fig. 2.

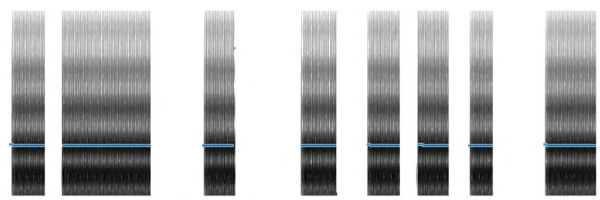


Fig. 2. Sample spectrogram for Condition Three: Rhythm only at mean pitch across speakers.

3. RESULTS

The results of logistic regression analyses in each condition indicate that listeners do rely on the F_0 characteristics to identify speaker ethnicity. In particular, Condition Three revealed a significant interaction between participant ethnicity and speaker mean pitch as predictors of perceived speaker ethnicity ($p<.05$). Fig. 3. illustrates the predictions of the linear regression model, where the y-axis shows the probability of perceived speaker ethnicity as Pakeha. Higher values indicate a Pakeha response, while low values indicate a Maori response. Pakeha participants identify a speaker with high mean pitch as Pakeha, and a speaker with low pitch as Maori. Maori participants do not rely on mean pitch in this condition. This result is at odds with the production results, where Maori speakers in fact produce significantly higher mean F_0 values.

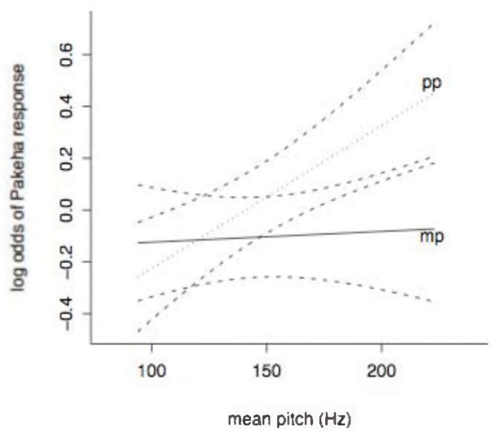


Fig. 3. Regression model predictions of perceived speaker ethnicity by speaker mean pitch and participant ethnicity (pp=Pakeha participant, mp=Maori participant). Dashed lines indicate 95% confidence intervals.

4. DISCUSSION

The results indicate that participants do rely on the mean pitch of a speaker in Condition Three, where many of the other cues have been eliminated from the speech signal. However, they perceive a lower mean pitch to be a characteristic of a Maori speaker and a higher mean pitch to indicate a Pakeha speaker. This is completely different from what is happening in production, where Maori speakers in fact use a significantly higher mean pitch. As this seems to be a change over time, it is plausible that listeners are not yet consciously aware of this new feature of Maori English. When people are overtly asked whether they think Maori speakers have a higher or lower mean pitch than Pakeha speakers, they tend to reply 'lower pitch' without hesitation. This might be the result of certain physical stereotypes held about Maori being big and bulky. However, when they are asked to imitate a Maori speaker, they almost always use a higher pitch in doing so. This suggests that subconsciously they might be aware of the ongoing change in Maori English pitch and, if such is the case, then we might expect that with time, perception results regarding mean pitch will adjust as listeners become actively aware of this new prosodic feature of the ethnolect.

This suggests that stereotypes might have a stronger effect on speech perception than the actual change in progress in this ethnolect of New Zealand English. Perception studies in the US yielded similar results with regards to the pitch characteristics of a speaker, where lower F_0 levels were associated with African Americans and higher F_0 with European Americans (Hawkins 1992, Foreman 2000, Thomas & Lass 2005).

Accounting for stereotypes – or ideologies – about certain speaker groups causes problems for current exemplar theoretic frameworks of speech production and perception. In an exemplar based model, categories are made up of a large set of remembered exemplars from a wide range of speakers, and the auditory properties that distinguish speakers are retained in these exemplars (e.g. Pierrehumbert 2001, Johnson 1997). The fact that stereotypes can get activated in speech perception instead of actually encountered exemplars with stored phonetic and social detail suggests that the framework should be revised accordingly.

REFERENCES

Britain, D. (1992), 'Linguistic Change in Intonation: The Use of High Rising Terminals in New Zealand English', *Language Variation and Change* **4**(1), 77–104.

Foreman, C. (2000), 'Identification of African-American English from Prosodic Cues', *Texas Linguistic Forum* **43**, 57–66.

Hawkins, F. D. (1992), Speaker Ethnic Identification: The Roles of Speech Sample, Fundamental Frequency, Speaker and Listener Variations, PhD thesis, University of Maryland at College Park.

Johnson, K. (1997), Speech perception without speaker normalization: An exemplar model, in K. Johnson and J. W. Mullennix, eds, 'Talker Variability in Speech Processing', San Diego: Academic Press, pp. 145–165.

Pierrehumbert, J. (2001), Exemplar dynamics: Word frequency, lenition and contrast, in J. Bybee and P. Hopper, eds, 'Frequency effects and the emergence of lexical structure', John Benjamins, Amsterdam, pp. 137–157.

Szakay, A. (2006), Rhythm and pitch as markers of ethnicity in New Zealand English, in P. Warren & C. Watson, eds, 'Proceedings of the 11th Australasian International Conference on Speech Science & Technology', University of Auckland, pp. 421–426.

Szakay, A. (2008), *Ethnic Dialect Identification in New Zealand: The Role of Prosodic Cues*. VDM Verlag Dr Müller.

Thomas, E. R. & Lass, N. J. (2005), 'Cues for Identification of African American and European American Speakers', Poster presented at the Annual Convention of the American Speech-Language-Hearing Association.

Warren, P. (1998), 'Timing Patterns in New Zealand English Rhythm', *Te Reo: Journal of the Linguistic Society of New Zealand* **41**, 80–93.

ACKNOWLEDGEMENTS

Special thanks to Jen Hay, Alex D'Arcy, Paul Warren and Margaret Maclagan. The funding provided by the University of Canterbury, New Zealand is also gratefully acknowledged.

PROCESSING OF JAPANESE PITCH ACCENT BY NATIVE JAPANESE AND ENGLISH LISTENERS

Jung-yueh Tu, Xianghua Wu[1], and Yue Wang
Dept. of Linguistics, Simon Fraser University, 8888 University Dr., Burnaby, BC, Canada V5A1S6

1. INTRODUCTION

1.1 Background

It is generally believed that the left hemisphere is dominant for phonemic processing and the right hemisphere is more involved in processing prosodic features [1, 2, 3]. More recent studies [4, 5, 6], however, indicate that lateralization of different levels of prosody varies with their functional load as well as listeners' linguistic experience. For example, whereas processing of monosyllabic lexical tone is left hemisphere dominant for native, but not non-native listeners; that of emotional intonation in phrasal or sentential length is right hemisphere dominant for both native and non-native listeners. The processing of Japanese pitch accent poses an interesting question, since pitch accent in Japanese, same as lexical tones in Mandarin, is used to differentiate word meanings. The domain where pitch accent realizes, however, is usually larger than Mandarin tones; thus reduces its functional load [4].

1.2 The current study

To investigate the effects of language background on processing of pitch accent, the current study includes not only native listeners of a pitch accent language, Japanese, but also nonnative listeners without pitch accent language background, English. A dichotic listening paradigm is adopted to examine how pitch accent in Japanese is processed in the brain by Japanese and English listeners. A right ear advantage (REA, i.e., left hemisphere dominance) is expected for the Japanese listeners, whereas for the English listeners, a left ear advantage (LEA, i.e., right hemisphere dominance) is anticipated.

2. METHOD

2.1 Participants

A total of 32 young adults participated in this experiment, including 16 native Japanese and 16 native English listeners. The Japanese listeners all had no other tonal language experience. The English listeners had no knowledge of Japanese and any other tonal languages. All participants were right-handed. None of them had hearing impairment.

2.2 Stimuli

The stimuli were 30 Japanese disyllabic words with three pitch accent patterns: H*L, LH* and LH[2]. Among them, 24 were minimal triplets, superimposed to eight sets; and the other six were minimal pairs, superimposed to three sets. Dichotic pairs were created such that in each pair, the two words have the same segment but differ only in the pitch accent pattern, e.g. H*L and LH* pairs such as /hana/ (H*L) 'a female name' and /hana/ (LH*) 'flower' or LH* and LH pairs such as /hana/ (LH*) 'flower' and /hana/ (LH) 'nose.' These dichotic pairs were constructed and edited using Audacity 1.2.6 where one word in each pair was imported into the left channel and the other into the right channel. Each pair was normalized for intensity and duration.

2.3 Procedure

There were three sections of the test: familiarization, identification, and dichotic listening. First, all the participants were familiarized with the 3 different pitch accent patterns. In the identification test, they were requested to identify the pitch accent pattern for 21 disyllabic words presented binaurally and no feedback was given after each response. Only those participants whose accuracy of responses was higher than 60% could continue to take the dichotic listening test. In the dichotic test, the stimuli were randomized into 4 blocks (i.e. 4 repetitions), with 21 dichotic pairs of tokens each. Each pair was presented to the participants simultaneously, one to the left ear and the other one to the right ear. The participants were asked to identify both stimuli.

2.4 Data analysis

Two measures were calculated in this study: one is the % errors of identification of each pitch accent pattern in each ear, and the other is the percentage of errors (POE), defined as [PL/ (PR+PL)]*100, where PL is percentage of errors in the left ear and PR is percentage of errors in the right ear [6]. If the percentage of errors for left ear exceeds 50% (equivalent to no ear preference), it indicates that the listener shows an REA and vice versa.

3. RESULTS

3.1 Percentage of Errors

The distribution of left ear errors and right ear is shown in Figure 1. The average POE for the Japanese listeners was 47%, whereas that for English listeners was 42%. One-way ANOVA, with POE as the dependent variable (DV) and Listener Group (English and Japanese) as the independent variable (IV):, showed that the difference in ear preference between the English and Japanese listeners was not statistically significant [$F(1, 30) = 3.632$, $p > .066$]. The left ear errors are significantly less than right ear errors for both groups [Japanese: $F(1, 30) = 4.826$, $p < .036$, and English: $F(1, 30) = 29.948$, $p < .001$].

3.2 Individual pitch accent pattern

Table 1 displays the mean % errors of identification of each

[1] The first two authors have made the same contributions to this study.
[2] H and L stand for high and low pitch, respectively; and '*' stands for accent.

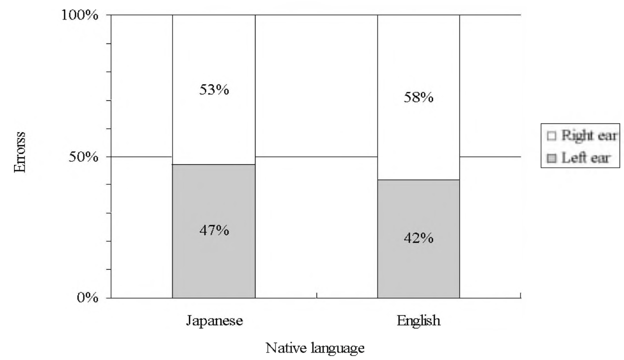


Fig.1. Distribution of left ear errors and right ear errors (in %) by the native Japanese (n=16) and native English (n=16) listeners.*: The difference in right-ear and left-ear error is significant at *p < .05.*

pitch accent pattern in each ear for each group. Overall, the English group made more errors than the Japanese group. Among the three patterns, LH* is the hardest for the Japanese listeners, and LH is the most difficult for the English listeners, whereas H*L is the easiest for both groups. A three-way ANOVA (DV: % errors; IV: Group*Ear*Pitch accent pattern) shows a significant interaction of group, ear, and pitch accent pattern [F (2, 2676) =3.851, p<.021]. One-way ANOVA (DV: % errors; IV: Ear) shows that identification of pitch accent pattern 1 (H*L) in the left ear is better than that in the right ear for both the Japanese group [F (1,446), p<.040], and the English group [F (1,446) =41.348, p<.001], indicating a right hemisphere dominance. Identification of pitch accent 2 (LH*) in the left ear, however, is significantly poorer than that in the right ear for the native Japanese group [F (1, 446) = 8.656, p < .003], indicating a left hemisphere dominance, but no ear preference was found for the non-native English group. For pitch accent pattern 3 (LH), no significant effects on ear were found for either group.

Table 1: Mean % errors of identification of each pitch accent pattern in each ear (Left, Right) for the Japanese and English groups.

	Japanese			English		
	Left	Right	Pooled	Left	Right	Pooled
H*L	28	35	31	31	48	40
LH*	56	46	51	50	45	47
LH	44	46	44	60	56	58

4. DISCUSSION AND CONCLUSIONS

The overall results show that for both the Japanese and English listeners, the percentage of errors for the left ear exceeds that for the right ear, indicating an LEA (i.e., right hemisphere dominance). The English group shows a greater degree of LEA than the Japanese group.

The native Japanese group did not reveal left hemisphere dominance for pitch accent, as previously found for linguistic tone processing by native listeners [6]. This finding supports Van Lancker's (1980) [4] "functional hypothesis," exhibiting that the functional load of pitch accent is lower than that of lexical tone for the native listeners. Therefore, the Japanese listeners revealed LEA instead of REA.

Although an overall LEA was obtained for both groups in the identification of pitch accent pattern H*L, REA was found only in the Japanese group when perceiving pitch accent pattern LH*, the hardest among the three patterns. This echoes Wang et al. (2004) [6] in that the native listeners tended to have REA in the perception of harder lexical prosodic features. It implies that native listeners relied more on the analytic linguistic information carried by harder prosodic features than that by easier ones. The nonnative group did not show a similar perception pattern, i.e. no REA was obtained when perceiving the pitch accent pattern, LH, which was the hardest for them. Thus, this study proposes that the difficulty of prosodic feature affects lateralization for native listeners.

These findings suggest that linguistic function differentially affects the hemispheric specialization of different domains of prosodic processing. Unlike lexical tone, pitch accent in Japanese has lower functional load which leads to LEA for the native listeners. Nevertheless, in certain situation where a prosodic feature is hard to distinguish by its acoustic properties, an REA is likely to occur. In contrast, the non-native listeners showed stronger LEA than the native listeners due to the lack of pitch accent language background. Therefore, linguistic function of target prosody interacting with listeners' language experience affects lateralization of Japanese pitch accent by native and non-native listeners.

REFERENCES

[1] T.G. Bever. (1975). Cerebral asymmetries in humans are due to the differentiation of two incompatible processes: Holistic and analytic. Annals of the New York Academy of Sciences, 163, 251-262.

[2] D. Kimura. (1961). Cerebral dominance and the perception of verbal stimuli. Canadian Journal of Psychology, 15, 166-171.

[3] D. Kimura. (1964). Left-right differences in the perception of melodies. Quarterly Journal of Experimental psychology, 16, 335-358.

[4] D. Van Lancker, (1980). Cerebral lateralization of pitch cues in the linguistic signal. Papers in Linguistics: International Journal of Human Communication, 13, 201–277.

[5] J. Gandour, M. Dzemidzic, D. Wong, M. Lowe, Y. Tong, L. Hsieh, N. Satthamnuwong & J. Lurito. (2003). Temporal integration of speech prosody is shaped by language experience: An fMRI study. Brain and Language, 84, 318-336.

[6] Y. Wang , D. Behne, A. Jongman and J.A. (2004). The role of linguistic experience in the hemispheric processing of lexical tone. Applied Psycholinguistics, 25, 449-46

[7] M. P. Bryden. (1988). An overview of the dichotic listening procedure and its relation to cerebral organization. In K. Hugdahl (Ed.), Handbook of dichotic listening: Theory, methods and research. Chichester, UK: Wiley.

ACKNOWLEDGMENTS

This work was supported by a Discovery grant from the National Sciences and Engineering Research Council of Canada (NSERC).

THE MCGURK EFFECT AFFECTED BY THE RIGHT EAR ADVANTAGE

Mark Scott
Department of Linguistics, University of British Columbia, Lower Mall, Vancouver, Canada V6T 1Z4

1. INTRODUCTION

It has been known since Broadbent (1954; cited in Hugdahl 2000) that, for most people, speech is better perceived in the right ear than the left. This is known as the Right Ear Advantage (REA). The explanation for this advantage is that most people are left-hemisphere dominant for speech processing and since this hemisphere receives most of its input from the right ear, the right ear dominates for speech. It should be noted that the strong tendency for the left-hemisphere to be dominant for language is less pronounced in left-handers (DeLeon 2005).

The most common demonstration of the REA is the dichotic listening test (e.g. Kimura 1961). In this test, subjects are simultaneously played two different syllables, one to each ear, and must report what syllable they perceive. Subjects tend to report the syllable that was played to their right ear. This test has been used as a simple behavioural test of hemispheric dominance for speech processing; however, the accuracy of this test is not impressive. Wada tests show that about 90-95% of right-handers are left-hemisphere dominant for language but only 80-85% show an REA in dichotic studies (DeLeon 2005). This is presumably due to the structure of dichotic listening tests, which, of necessity, bring in issues of word and sound frequency, attention and other possible confounds.

The current experiment provides a new behavioural measure of hemispheric dominance using the McGurk effect (McGurk and MacDonald 1976). The McGurk effect is an audiovisual illusion produced when auditory and visual sources of information give contradictory cues about the identity of a speech sound. For example, audio of a voice saying /aba/ dubbed over a video clip of a face saying /aga/ will typically be perceived as /ada/. The McGurk effect is dependent on the audio component being somewhat ambiguous – the more ambiguous the audio, the easier it is for the visual information to have an influence. Since the right ear is better at perceiving speech the McGurk effect should be harder to induce when the auditory component is played to the right ear. This experiment tests that hypothesis. If correct, this test may serve as a new diagnostic for hemispheric dominance.

2. METHOD

Subjects viewed audiovisual clips on a continuous loop. These clips consisted of a voice pronouncing /aba/ dubbed over video of a face saying /aga/, and so, with sufficient noise in the audio component, these clips would induce an illusory perception of /ada/. The audio was either delivered to subjects' left ear, their right ear, or both ears. The 'both-ears' condition was included as a foil.

The level of white-noise masking was low at the start of each clip, so subjects would not initially experience the McGurk effect. Subjects raised the level of white-noise masking in the audio until their perception of the sound changed from /aba/ to /ada/, thus giving a measure of the relative amount of noise needed to induce the illusion in the left versus right ears.

2.1. Stimuli

Two speakers were recorded (audio and video) saying the disyllables /aba/ and /aga/. These sounds were chosen because the McGurk effect is harder to induce in the environment of low vowels (Green and Norrix 1997), and it is important to avoid ceiling effects (in which subjects always hear the illusory McGurk percept because the audio integrates so easily with the video).

One audio token of /aba/ and one video token of /aga/ from each speaker was chosen. The audio /aba/ tokens were normalized to 70dB and then dubbed over their respective video tokens of /aga/. Three versions of each of these stimuli were created. In one, audio was presented only to the left ear; in the second, audio was in the right ear; and in the third, audio was presented to both ears. In total, this means there were 6 different stimuli created (3 types of audio x 2 speakers). 12 copies of each of these stimuli were made to produce 72 experimental tokens.

2.2. Procedure

The experiment was conducted in a quiet room, with subjects wearing 'Extreme Isolation'® headphones. The stimuli were presented on a Dell Inspiron Laptop computer. The software used to run the experiment was developed by the author.

There were two stimulus-types of interest: monaural presentation to the left and right ear. These were not segregated into separate blocks, but were interleaved with each other and with the foil condition (presentation to both ears) in random order.

At the commencement of each audiovisual token, white-noise was delivered to the same ear that was receiving the auditory component of the token. Because the white-noise was very quiet, subjects typically perceived each token as

/aba/ when it first appeared on the computer screen. They then used the arrow buttons on the keyboard to raise the level of white-noise until they stopped perceiving the audiovisual token as /aba/ and started to perceive it as the illusory /ada/. When they had achieved the /ada/ percept, subjects used the mouse to click a button on the computer screen marked 'Done', and proceeded to the next token. There were 72 tokens in total (presented in random order) and subjects took a break after every 18 tokens.

2.3. Subjects

There were 13 subjects, 11 right-handed and 2 left-handed. All reported normal hearing and normal or corrected-to-normal vision. Subjects were paid ten dollars each for their participation.

3. RESULTS

Of the 13 subjects, 10 showed a Right Ear Advantage (REA). A separate t-test was conducted for each subject, and this REA was significant (at $p < 0.05$) for 5 of them.

9 of the 11 right-handers showed an REA and a sign test on these pooled results found that this was significant ($p = 0.032$), indicating that this REA is significant for the right-handers as a group.

Table 1. Subject handedness & whether they show an REA.

	Subject	Handedness	REA?	p-value of within-subject results
1	DD	Right	No	0.6856635
2	JD	Right	Yes	0.0273787
3	KJ	Right	Yes	0.3263987
4	JM	Right	Yes	0.0001427
5	KM	Right	No	0.0000001
6	MO	Right	Yes	0.2805566
7	GS	Right	Yes	0.0219312
8	KS	Right	Yes	0.3439654
9	GT	Right	Yes	0.0423531
10	JV	Right	Yes	0.7669372
11	JW	Right	Yes	0.2783214
12	HD	Left	No	0.0263171
13	ST	Left	Yes	0.0175788

It should be noted that while 2 of the right-handers showed a left ear advantage instead of a right ear advantage, only the results for one of these (subject 5: KM) was significant. The other right-handed subject without an REA (subject 1: DD) had a p-value of 0.685, suggesting that his lack of an REA was probably due to chance. If this is correct, then the rate of REA found in this sample is at least superficially consistent with the 90-95% rate of left-hemisphere dominance in the general population.

4. DISCUSSION

This experiment demonstrates that the REA makes the McGurk effect harder to induce when the audio is presented to the subject's right ear. This may serve as a new behavioural measure of hemispheric dominance for language. This test has the advantage of being easy to prepare and administer and furthermore avoids the confounds of word and sound frequency that interfere with the dichotic listening test.

Previous experiments using monaural presentation have typically only found a small REA and could only induce the REA when subjects were forced to make speeded responses (DeLeon 2005). The current experiment found a strong REA and did not require subjects to make speeded responses, and thus appears to be a reliable way to induce an REA with monaural presentation.

Some researchers (e.g. Kinsbourne 1970) have suggested that the REA is at least partially due to attentional factors (i.e. that people's attention tends to be drawn to sounds in their right ear). This experiment does not support that hypothesis since attentional factors are unlikely to have played a role here – the stimuli were presented monaurally and were played on a continuous loop, so subjects had plenty of time to attend to the relevant ear and had no sound in the opposite ear competing for attention.

The current study is too small to determine if the REA found using this methodology matches the 90-95% rate of left-hemisphere dominance in the general population, but the results look promising.

REFERENCES

DeLeon, R. 2005. EEG analysis of ear advantage. Ph.D. Thesis: University of Houston.

Green, K. and L. Norrix. 1997. Acoustic cues to place of articulation and the McGurk effect: The role of release bursts, aspiration, and formant transitions. Journal of Speech & Hearing Research 40: 646-665.

Hugdahl, Kenneth. 2000. What can be learned about brain function from dichotic listening? Revista Espanola de Neuropsicologia 2(3): 62-84.

Kimura, D. 1967. Functional asymmetry of the brain in dichotic listening. Cortex 3: 163–168.

Kinsbourne, M. 1970. The cerebral basis of lateral asymmetries in attention. Acta Psychologica 33: 193-201.

McGurk, H. and J. MacDonald. 1976. Hearing lips and seeing voices. Nature 264: 746-748.

AKNOWLEDGEMENTS

This research was funded by an NSERC Discovery Grant to Bryan Gick at the University of British Columbia.

Analysing Coarticulation in Scottish English Children and Adults: an Ultrasound Study

Natalia Zharkova, Nigel Hewlett, and William J. Hardcastle

Speech Science Research Centre, Queen Margaret University, Edinburgh, UK nzharkova@qmu.ac.uk

1. INTRODUCTION

One of the gaps in our knowledge about developmental paths taken by children to adult-like motor control of speech concerns the development of coarticulation. There have been a number of studies which compared coarticulatory patterns in children and adults, but these studies have produced conflicting results: some show that children exhibit less coarticulation than adults (e.g., [2]); a similar amount (e.g., [6]); or more (e.g., [4]; [3]). A greater within-speaker variability in articulatory patterns exhibited by children than by adults may have contributed to the equivocal results. Another factor may be that most previous studies relied heavily on acoustic analysis, which provides only indirect evidence of articulatory movements, and is particularly problematic in child speech, because of the high fundamental frequency and consequent difficulties with formant tracking. Possibly as a result of the relative unavailability of suitable articulatory instrumental techniques, developmental studies of coarticulation comparing adults' and children's productions using articulatory data are very few (e.g., an EMA study reported in [1]). An advantage of ultrasound over EMA is that it is non-invasive, and it registers the movement of the whole midsagittal section of the tongue, including the tongue root.

This study used articulatory measures derived from ultrasound imaging for comparison of coarticulation in children and adults. The research questions were:

1) Do children demonstrate a significant difference from adults in coarticulatory patterns, and if there is a significant difference, what is the direction of the difference?

2) Do children exhibit significantly greater within-speaker variability than adults in their patterns of coarticulation?

2. METHOD

The data were the syllables /ʃi/, /ʃu/ and /ʃa/, in the carrier phrase "It's a … Pam" (ten repetitions). The participants, all native speakers of Standard Scottish English, were three normally developing children aged 6 to 8 years (C1 male aged 8;4, C2 female aged 6;10, C3 male aged 6;4), and three adults. Synchronised ultrasound and acoustic data were collected using the Queen Margaret University ultrasound system ([7]).

A new methodology for analysing ultrasound data (see [8]) was used. Ultrasound frames at two time points, the middle of the consonant and the middle of the vowel, were identified in each of the different CV sequences, based on the acoustic data. At each time point, a cubic spline was automatically (with subsequent manual correction) fitted to the tongue surface contour. Each spline was defined in terms of xy coordinates, and these coordinates were used for comparing tongue curves.

Tongue curve comparison was based on nearest neighbour calculations (e.g., [5]). Magnitude of Coarticulation (MC) for the consonant in each of the three pairs of vowel environments was calculated, separately for each subject. The following formula for calculating MC was developed:

$$MC_C = \frac{V1 - V2}{(C_{V1} - V1) + (C_{V2} - V2)}$$

In the formula, C is the target consonant; V1 and V2 are two vowel phonemes providing the alternative conditioning environments; C_{V1} is C in the environment of V1; C_{V2} is C in the environment of V2. This measure of coarticulation expresses the ratio of the difference between the vowel contours (which is proportionate to the possible degree of consonantal adaptation offered by the two vowel contexts) and the sum of the consonant-vowel differences in each vowel environment (which is in inverse proportion to the degree of consonantal adaptation to the vowel contexts). The greater the MC value, the stronger is the coarticulatory effect produced on a given consonant by the two vowels.

For each speaker, for the consonant in each pair of vowel contexts, MC values and Coefficients of Variation across ten tokens were obtained. MC values and Coefficients of Variation were compared across age group and vowel pair.

3. RESULTS

Table 1 presents MC values and Coefficients of Variation for each subject. The Univariate ANOVA showed a significant main effect of age group on MC ($F = 369.49$; $df = 1$; $p < 0.001$). On average, MC was greater in children (mean MC of 1.00 in children versus mean MC of 0.80 in adults). Note also a difference in MC across vowel pairs (F = 1372.45; df = 2; p < 0.001), with the pair /i/-/u/ affecting the consonant the least, and the pair /a/-/i/ producing, on average, the greatest effect.

Table 1. MC values and Coefficients of Variation (CV) in three children (first three rows) and three adults (last three rows). Standard Deviations for MC are in brackets.

	MC, a/i	MC, a/u	MC, i/u	CV
C1	1.40 (*0.21*)	1.36 (*0.26*)	0.39 (*0.15*)	23.9
C2	1.30 (*0.27*)	0.81 (*0.27*)	0.93 (*0.14*)	23.5
C3	1.05 (*0.10*)	1.11 (*0.11*)	0.69 (*0.24*)	13.1
S1	0.97 (*0.07*)	0.91 (*0.09*)	0.50 (*0.07*)	10.5
S2	1.02 (*0.07*)	1.18 (*0.11*)	0.24 (*0.05*)	12.8
S3	1.09 (*0.06*)	0.94 (*0.11*)	0.38 (*0.05*)	8.8

Figure 1 presents tongue curves for /ʃi/ and /ʃa/ in a child participant C2 and in an adult participant S1. The figure shows relatively small distances between the consonant and the vowel in C2, compared to S1; this difference has contributed to the greater MC in this child than in this adult.

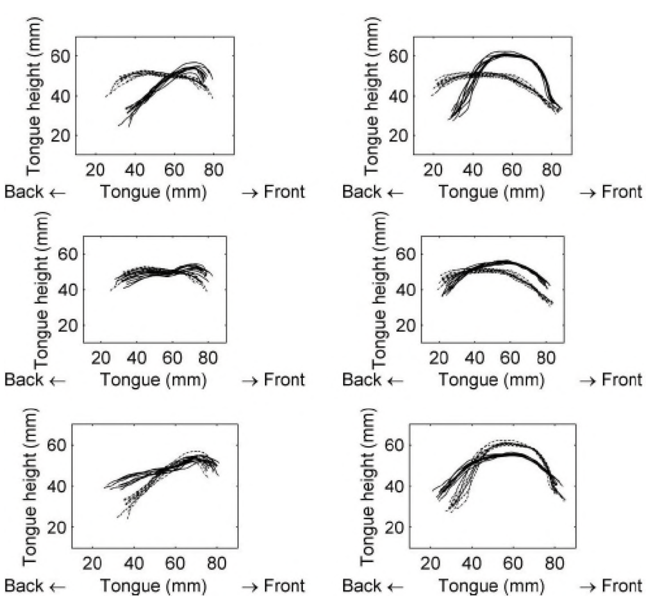


Fig. 1. Tongue contours for /ʃa/ and /ʃi/ in C2 (left) and S1 (right). Row 1: /i/ from /ʃi/ (solid) and /a/ from /ʃa/ (dashed). Row 2: /ʃ/ (solid) and /a/ (dashed) from /ʃa/. Row 3: /ʃ/ (solid) and /i/ (dashed) from /ʃi/. Lines for 10 repetitions are presented.

An independent t-test demonstrated a significant difference in the Coefficient of Variation between adults and children (t = 2.64; df = 16; p < 0.05). Table 1 illustrates greater values of the Coefficient of variation in children.

4. DISCUSSION

In this study, children showed a significantly greater amount of anticipatory lingual coarticulation than adults. This finding is in agreement with [3] and [4], but it contradicts [2] and [6]. Coefficients of variation were significantly greater in children than in adults. This shows that adults and children differ in the degree of within-speaker variability in coarticulatory patterns, children being more variable than adults. These results agree with existing literature (e.g., [3]). We are planning to conduct an acoustic analysis of these data, to find out whether acoustic results will corroborate our articulatory findings. A qualitative examination of patterns for each speaker will also be conducted, to establish whether individual results contribute disproportionately to a group picture.

REFERENCES

[1]. Katz, W. F. & Bharadwaj, S. (2001). Coarticulation in fricative-vowel syllables produced by children and adults: a preliminary report. *Clinical Linguistics and Phonetics*, 15, 139-143.

[2]. Kent, R.D. (1983). The segmental organization of speech. In P. F. MacNeilage (Ed.), *The Production of Speech*. New York: Springer-Verlag, 57-89.

[3]. Nijland, L., Maassen, B., Van der Meulen, S., Gabrieëls, F., Kraaimaat, F. W. & Schreuder, R. (2002). Coarticulation patterns in children with developmental apraxia of speech. *Clinical Linguistics and Phonetics*, 16, 461-483.

[4]. Nittrouer, S., Studdert-Kennedy, M. & Neely, S. T. (1996). How children learn to organize their speech gestures: further evidence from fricative-vowel syllables. *Journal of Speech and Hearing Research*, 39, 379-389.

[5]. Sankaranarayanan, J., Samet, H. & Varshney, A. (2007). A fast all nearest neighbour algorithm for applications involving large point-clouds. *Computers & Graphics*, 31, 157-174.

[6]. Sereno, J. A., Baum, A. R., Cameron Marean, G. & Lieberman, P. (1987). Acoustic analyses and perceptual data on anticipatory labial coarticulation in adults and children. *JASA*, 81, 512-519.

[7]. Vazquez Alvarez, Y. & Hewlett, N. (2007). The trough effect: an ultrasound study. *Phonetica*, 65, 105-121.

[8]. Zharkova, N. (2007). Quantification of coarticulatory effects in several Scottish English phonemes using ultrasound. *QMU Speech Science Research Centre Working Papers*, WP-13.

ACKNOWLEDGEMENTS

We are grateful to Alan Wrench and Steve Cowen for technical help. This work was supported by an ESRC research grant, RES-000-22-2833.

A FORMANT FREQUENCY ESTIMATOR FOR NOISY SPEECH BASED ON CORRELATION AND CEPSTRUM

Shaikh Anowarul Fattah, Wei-Ping Zhu, and M. Omair Ahmad
Dept. of Electrical and Computer Engineering
Concordia University, 1455 De Maisonneuve Blvd. W., Montreal, Quebec, Canada H3G 1M8

1. INTRODUCTION

Formant is one of the most informative speech features which helps in interpreting the mechanisms of human speech production. Formant frequency estimation of speech has numerous applications, such as, speech recognition, synthesis, and compression. As an acoustic feature, formant offers phonetic reduction in speech recognition [1]. Formants are associated with peaks in the smoothed power spectrum of speech. Among different formant estimation techniques, linear predictive coding (LPC) based methods are most commonly used [1]. Cepstrum features are also used in the formant estimation. Most of the formant frequency estimation methods, so far reported, deal only with the noise-free environments. The estimation performance of correlation based formant estimators deteriorates noticeably in the presence of noise. Formant estimation from noisy speech is a difficult but essential task as far as practical applications are concerned. Recently in [2] and [3], methods have been proposed in order to handle noisy environments. The method proposed in [2] is based on an adaptive band-pass filter bank (AFB) where the estimation accuracy depends on initial estimates. The formant frequency estimation approach, proposed in [3], is based on a correlation model of voiced speech signals.

The objective of this paper is to develop a formant estimation scheme combining the advantageous features of correlation and cepstral domains, which is capable of handling the adverse effect of observation noise. We propose a ramp cepstrum model of a once-repeated autocorrelation function (ORACF) of voiced speech signals in terms of formant parameters. It has been shown that in comparison to the conventional ACF, the ORACF can drastically reduce the effect of additive noise. A residue based least-squares optimization technique based on a model-fitting approach is introduced in order to obtain formant frequencies from noisy observations. Simulations are carried out to estimate formant frequencies from synthetic and natural speech signals under noisy conditions.

2. PROPOSED METHOD

2.1 A Ramp Cepstrum Model of ORACF of Speech

The human vocal-tract system can be represented by a P-th order AR system with a transfer function given by

$$H(z)=\frac{G}{\prod_{k=1}^{P}(1-p_k z^{-1})} \qquad (1)$$

where G is the gain factor and p_k is the system pole. In order to model each formant, a pair of complex conjugate poles is required. Formant frequency (F_k) and bandwidth (B_k) can be computed from the pole magnitude r_k, angle ω_k, and sampling frequency F_S as

$$F_k=\omega_k(F_S/2\pi)\ ; \qquad B_k=-(F_S/\pi)\ln(r_k) \qquad (2)$$

The complex-cepstrum of the impulse-response $h(n)$ of the vocal-tract filter is defined as

$$c_h(n)=F^{-1}\{\ln(H(e^{j\omega}))\} \qquad (3)$$

where $F^{-1}\{\cdot\}$ denotes the inverse Fourier transform (FT). From (1) and (3), assuming a minimum phase system, $c_h(n)$ can be expressed in terms of system poles as

$$c_h(n)=\sum_{k=1}^{P}\frac{p_k^n}{n}, \quad n>0 \qquad (4)$$

During a short duration of time, an observed speech $x(n)$ can be assumed to be stationary. The FT of the ACF of $x(n)$, $r_x(\tau)$, can be written as

$$R_x(e^{j\omega})=\left|H(e^{j\omega})\right|^2 R_u(e^{j\omega}) \qquad (5)$$

where $R_u(e^{j\omega})$ is the FT of the ACF of the input excitation $u(n)$. In the cepstum-based speech analysis, generally, cepstral coefficients are computed from observed speech or the estimate of its non-parametric power spectral density (PSD). The cepstral coefficients thus estimated, especially in the presence of observation noise, provide poor estimation accuracy. Hence, we propose to utilize a once-repeated ACF (ORACF) $\varphi_x(\tau)$ that exhibits a higher noise immunity. The FT of $\varphi_x(\tau)$ can be expressed as

$$\Phi_x(e^{j\omega})=\left|R_x(e^{j\omega})\right|^2 \qquad (6)$$

Using (5) and (6), cepstrum corresponding to $\Phi_x(e^{j\omega})$ can be expressed as

$$c_{\Phi_x}(n)=F^{-1}\{\ln(\Phi_x(e^{j\omega}))\}=c_h{}'(n)+c_u{}'(n) \qquad (7)$$

For a voiced speech, the input excitation $u(n)$ is treated as a periodic impulse-train with period T and it is found that $c_u'(n)$ exhibits significant values at origin and multiples of the period. Thus $c_u'(n)$ vanishes in the region $0 < n < T$. It can be shown that $c_h'(n) = 2c_h(n)$ in this region and thus (7) reduces to

$$c_{\Phi_x}(n)=2\sum_{i=1}^{P}\frac{p_i^n}{n}, \qquad 0<n<T \qquad (8)$$

It is clear that the cepstrum decays as $1/n$, which is difficult to handle. Hence, we propose a ramp-cepstrum $\psi_x(n)$, which for real values of $x(n)$ can be expressed as

$$\psi_x(n)=nc_{\Phi_x}(n)=\sum_{i=1}^{K}\eta(\omega_i)r_i^n\cos(\omega_i n),\ \ 0<n<T \qquad (9)$$

where K = number of real poles + the number of complex conjugate pole pairs, and $\eta(\omega_i) = 2$ if $\omega_i = 0$ or π, otherwise $\eta(\omega_i) = 4$. Equation (9) is the ramp-cepstrum model of voiced speech signals. Note that each of the K components in (9) for $0 < \omega_k < \pi$, corresponds to a particular formant. Next we will develop a model-fitting algorithm to estimate formants.

2.2 Formant Estimation Algorithm in Noise

The noise-corrupted speech signal is given by

$$y(n) = x(n) + v(n) \tag{10}$$

where, the additive noise $v(n)$ is assumed to be zero mean with variance σ_v^2. The ACF of $y(n)$ can be expressed as

$$\begin{aligned} r_y(\tau) &= r_x(\tau) + r_w(\tau) \\ r_w(\tau) &= r_v(\tau) + r_{vx}(\tau) + r_{xv}(\tau) \end{aligned} \tag{11}$$

The effect of noise term $r_w(\tau)$ on $r_x(\tau)$ is relatively less pronounced since, the crosscorrelation terms in (11) are negligible and $r_v(\tau)$ mostly affects only the zero lag. Hence, in the ORACF of $y(n)$, i.e., the ACF of $r_y(\tau)$, the effect of noise term will be drastically reduced. It can be shown that the ORACF, like the ACF, preserves the poles of the vocal-tract AR system. The ramp-cepstrum computed from the ORACF of $y(n)$ can be expressed as

$$\psi_y(n) = \psi_x(n) + \psi_n(n), \; 0 < n < T \tag{12}$$

Due to the error term $\psi_n(n)$, it is difficult to estimate $\psi_x(n)$ from $\psi_y(n)$ in a noisy condition. In order to overcome this problem, we exclude $r_y(0)$ in computing $\psi_y(n)$ at a low SNR, which significantly reduces the strength of noise term $\psi_n(n)$. Resulting noise-reduced $\psi_y(n)$ is then used to extract the ramp-cepstrum model parameters through a residue-based least-square optimization. The parameters of each component $F_l(n)$ of the ramp-cepstrum model (10) are determined such that the total squared error between the $(l-1)$th residual function and an estimate of $F_l(n)$ is minimized. The lth residual function can be defined as

$$\Re_l(n) = \Re_{l-1}(n) - F_l(n); \; \Re_0(n) = \psi_y(n); \; l = 1,2,\ldots,K-1 \tag{13}$$

and the objective function for the minimization is formed as

$$J_l = \sum_{n=1}^{M-1} \left| \Re_{l-1}(n) - F_l(n) \right|^2, \quad l = 1,2,\ldots,K; \; M < T \tag{14}$$

Values of $\hat{r}_l$ and $\hat{\omega}_l$, corresponding to the global minimum of J_l, are selected to compute the estimate of the lth formant. Thus different formant frequencies are sequentially determined. In the poposed optimization technique, r_l and ω_l are searched within a certain search space. The region of formants (both frequency and bandwidth) is available in literature and utilized to restrict the search space [1], [3]. Further reduction in the frequency search space is achieved by using an initial estimates obtained through smoothed spectral peaks of the zero lag excluded ACF.

In our implementation, we perform the formant estimation every 10 ms with a 20 ms window applied to overlapping speech segments. In addition to signal pre-emphasis a FFT pre-filtering is performed to remove very low frequencies (<100 Hz) which are not in our interest.

3. SIMULATION RESULTS

The proposed formant frequency estimation algorithm has been tested using various synthetic vowels synthesized using the Klatt synthesizer [1] and some natural vowels extracted from the TIMIT and North-Texas standard databases with corresponding reference values [1], [4]. For the performance comparison, the 12th order LPC [1] and the AFB methods [2] are considered and the percentage root-mean-square error (RMSE) at different noise levels are computed where each noise level consists of 20 independent trials of noisy environments. In the proposed optimization algorithm, the search range of r_l is chosen as $0.8 \le r_l \le 0.99$, and that for ω_l is $\pm 0.1\pi$ near initial estimates [3]. In Table 1, the estimated %RMSE (Hz) is shown for two synthesized vowels at SNR = 0 dB and 5 dB. It is clearly observed that the proposed method provides lower %RMSE for both male and female speakers. In Table 2, estimation accuracy in terms of %RMSE (Hz) for natural vowels /a/ and /i/ (contained in the words "hod" and "heed") are shown. It is found that the proposed method provides better estimation accuracy at both conditions.

Table 1. %RMSE (Hz) for Synthetic Vowels

Vowels			0 dB			5 dB		
			Prop.	LPC	AFB	Prop.	LPC	AFB
Male	/a/	F1	9.78	23.63	31.29	7.02	15.87	11.74
		F2	9.65	27.78	34.82	3.12	15.53	9.51
		F3	8.39	19.28	19.34	5.43	13.19	8.23
	/i/	F1	19.27	28.53	28.16	12.51	19.28	16.25
		F2	3.95	9.68	4.27	2.72	7.54	3.33
		F3	4.93	13.29	7.75	3.71	7.82	3.97
Female	/a/	F1	11.15	17.76	16.76	4.84	9.76	15.67
		F2	9.01	19.43	14.39	4.51	8.68	7.49
		F3	4.54	9.26	4.57	2.08	3.18	2.34
	/i/	F1	21.34	39.81	32.27	13.59	28.27	19.14
		F2	9.56	26.21	13.78	4.81	12.43	6.19
		F3	2.89	15.83	3.78	2.12	7.63	2.78

Table 2. %RMSE (Hz) for Natural Vowels

Vowels			0 dB			5 dB		
			Prop.	LPC	AFB	Prop.	LPC	AFB
Male	/a/	F1	9.85	14.33	13.93	6.72	9.57	8.54
		F2	14.52	44.93	28.76	11.26	28.27	16.29
		F3	12.25	38.01	23.34	10.39	23.67	18.31
Female	/i/	F1	9.84	21.19	16.84	5.03	8.61	5.95
		F2	10.25	23.78	19.49	5.96	11.28	10.21
		F3	10.11	31.29	24.82	5.05	21.92	14.58

4. CONCLUSION

A formant frequency estimation scheme based on a new ramp cepstrum model has been developed which is capable of efficiently handling the noisy environment. In the development of the ramp cepstrum model, the once-repeated ACF is employed which can significantly reduce the effect of noise in the correlation domain. It has been shown that even at a low level of SNR, the proposed residue based least-squares optimization algorithm for the model-fitting provides an accurate estimation of the formant frequencies. From experimental results on synthetic and natural speech signals under a noisy condition, it has been found that the proposed method provides an accurate formant frequency estimate even at a low level of SNR.

REFERENCES

[1] D. O'Shaughnessy, (2000). Speech Communications: Human and Machine (2nd ed.). *IEEE Press, NY*.

[2] K. Mustafa and I. C. Bruce, (2006). Robust formant tracking for continuous speech with speaker variability. *IEEE Trans. Audio Speech Lang. Processing*, 14, 435–444.

[3] S. A. Fattah, W. -P. Zhu, and M. O. Ahmad, (2007). An approach to formant frequency estimation at low signal-to-noise ratio. *ICASSP'07*, 4, 469–472.

[4] J. M. Hillenbrand, L. A. Getty, M. J. Clark, and K. Wheeler, (1995). Acoustic characteristics of American English vowels. *J. Acoust. Soc. Am.*, 97, 3099–3111.

QUANTITATIVE ANALYSIS OF SUBPHONEMIC FLAP/TAP VARIATION IN NAE

Donald Derrick[1] and Bryan Gick[1, 2]
[1]Dept. of Linguistics, University of British Columbia, Vancouver, 2613 West Mall, British Columbia, Canada V6T1Z4
[2]Haskins Laboratories, 300 George St., 9th Fl., New Haven, CT 16511, USA

1. INTRODUCTION

Extreme 'covert' categorical subphonemic variation has been thought to occur only in rare cases such as American English 'r' [1]. The present study demonstrates that English flaps/taps are produced using up to four distinct kinematic variations: up-flaps, down-flaps, alveolar taps and postalveolar taps. Surface distinctions between up-flaps and down-flaps, and between alveolar taps and post-alveolar taps, have not been previously described for any language. Our research expands on preliminary research by Gick [2,3] to include B/M mode ultrasound measures that capture details of flap kinematics with higher temporal resolution.

Based on our pilot work, we expect that in words with one flap, speakers will produce categorically distinct kinematic alternations primarily based on resolution of articulatory conflict. Articulatory conflict in flaps/taps arises based on the tongue positions required for surrounding vocalic sounds. Vowels are produced with the tongue tip below the alveolar ridge, while vocalic 'r's are produced with the tongue tip above the alveolar ridge.

Therefore, when a flap is preceded by a vocalic 'r' and followed by a vowel, as in the word 'Berta', we expect the flap to be produced by the tip of the tongue coming from above the alveolar ridge, hitting the ridge and continuing down. That is, we expect a down-flap.

Similarly, when a flap is preceded by a vowel and followed by a vocalic `r', as in the word 'otter', we expect the flap to be produced by the tip of the tongue coming from below the alveolar ridge, hitting the ridge and continuing up. That is, we expect an up-flap.

When a flap is preceded and followed by a vowel, as in the word `autumn', we expect the tap to be produced by the tip of the tongue coming from below the alveolar ridge, hitting low on the ridge and returning. That is, we expect an alveolar tap, like in Spanish [4].

When a flap is preceded and followed by a vocalic `r', as in the word 'murder', we expect the tap to be produced by the tip of the tongue coming from above the alveolar ridge, hitting above the ridge and return. That is, we expect a post-alveolar tap.

If speakers do not produce taps, we expect speakers to produce flaps favoring a suitable tongue position for the end of the word, so we expect down-flaps for `autumn' and up-flaps for 'murder'.

2. METHOD

2.1 Participants

Twenty-four participants were recorded, 12 female and 12 male. The present paper reports results from the first 4 of these.

2.2 Stimuli

38 unique stimuli were recorded in twelve randomized blocks. These were presented in carrier phrases designed to contain labial and glottal consonants only (except at phrase end) and induce stress on the first syllable of the stimuli phrases. Of these 38, 10 phrases contained a single flap. The stimuli with vowel-flap-vowel context included `autumn', 'edit the', 'audit the', 'edify', 'audify', 'vomit a' and 'acerbity'. The stimulus with vocalic `r'-flap-vocalic `r` context included 'murder'. The stimulus with a vocalic 'r'-flap-vowel sequence was 'Berta'. The stimulus with a vowel- flap-vocalic 'r' sequence was 'otter'.

2.3 Experiment procedure

Participants were seated comfortably in an American Optical Co. (1953) opthalmic chair with a twin-cup headrest. A 180^{o} EV tranducer, attached to an Aloka ProSound SSD-5000 ultrasound machine, was placed under the chin of the participant. A Sennheiser MKH-416 short shotgun microphone attached to an M-Audio DMP3 via XLR cable was used to record audio. The ultrasound and audio were channeled into an ADVC110 Canopus A/D video converter and recorded on a MacPro using iMovie HD (2006).

Ultrasound data were recorded in B/M mode. B-mode was set to record the tongue along the mid-sagittal plane. Three parallel M-mode lines were set following the line of the anterior palate, so as to maximally intersect movement of the tongue tip through the alveolar region (see Figure 1).

Stimuli was presented on an LCD monitor using PXLab [5] in two groups of six blocks of 38 stimuli each, with a break between the two groups.

2.4 Analysis procedure

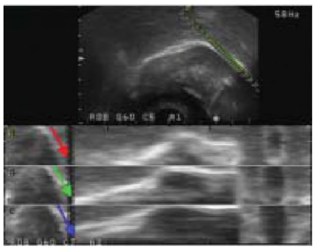
The ultrasound audio and video were separated to allow for accurate acoustic boundary identification. The audio recordings were labeled and segmented using PRAAT [6]. Audio segmenting was imported in ELAN [7] and the type of tap/flap was labeled and transcribed using both the audio and ultrasound video recordings. Results were exported back to PRAAT textgrids and extracted using PERL scripts into statistics tables used in JMP 5.1 [8].

3. RESULTS

3.1 Flap-tap Identification

Using the M-mode information allowed for a relatively straightforward identification and distinction of the four kinematic variations of taps and flaps. The flaps are the easiest to identify. In down-flaps, a fuzzy white line, indicating the tongue surface trajectory, in the M-mode ultrasound image moves downward, as seen in Fig. 1a and highlighted by the red, green and blue arrows.

Similarly, in up-flaps the tongue-surface trajectory moves upward, as seen in Fig. 1b.
As positions on the hard palate are rarely identifiable in ultrasound images, the two taps must be identified relative to each other. In alveolar taps the tongue-surface trajectory is flat and lower than the top of a down-flap, as seen in Fig. 1c. In post-alveolar taps the tongue-surface trajectory is flat and higher than the trajectory of the alveolar tap, as seen in Fig. 1d.



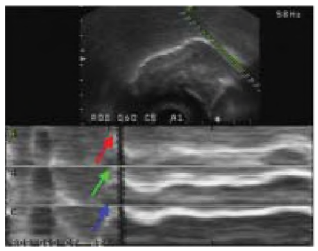
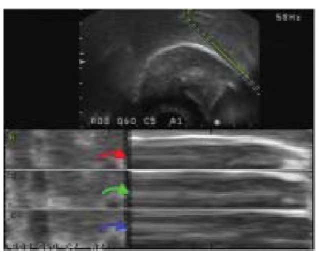


Fig. 1a. 'Berta' – down-flap Fig. 1b. 'otter' – up-flap



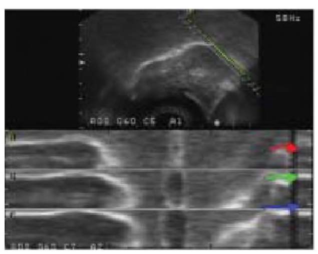


Fig. 1c. 'autumn' – alveolar tap Fig. 1d. 'murder' – PA tap

3.2 Kinematic variation frequency by token

The results show that there is a highly significant relationship between the categorical kinematic variation of the tap or flap produced and the vocalic context for all four subjects individually and as a group. The within-subject chi-square test likelihood ratio was a χ^2 of 732.4, $P < 0.001$, with an $R^2 = 81.9\%$.

Context	Dn-flap	Hi-tap	Low-tap	Up-flap	Exclude
RR	0	46	2	0	0
RV	47	0	0	0	1
VR	0	0	0	48	0
VV	1	0	246	18	23

Table 1. Kinematic variation frequency by token
(Excluded tokens = speech errors and dropped frames)

The most variation was found in the vowel-flap-vowel context, and occurred only with the tokens containing the words 'edit-the' and 'audit-the'. In all of `edit-the' and 'audit-the' tokens and for all of the participants, the body of the tongue raised higher in the mouth than in other vowel-vowel context tokens. For 18 of the tokens, all from the same subject, there was an up-flap instead of the more common low-tap.

4. DISCUSSION

The results of the experiment strongly support the hypothesis that there are four categorical kinematic variations of flaps and taps in English. All four of the participants demonstrated all four variations during the experiment. The results also support the hypothesis that in words with one flap, articulatory conflict resolution, constrained by vowel and vocalic 'r' context, has the largest effect on the kinematic variation. We expect there to be more variation as we measure more of our participants. In particular, based on previous work, we expect speech errors to influence variation. We also have participants whose data has not been measured yet who do not produce all four kinematic variants either at all or with as much frequency.

This research also contains multiple measures of words and phrases with two consecutive flaps, such as 'Saturday'. In these cases, we believe and have already observed that participants use preferred tap sequences that will interact with articulatory conflict, such that in words like 'auditor' and 'editor', the sequence is 'up-flap, post-alveolar tap' rather than 'alveolar-tap, up-flap', to avoid ending a sequence on an up-flap.

REFERENCES

[1] Westbury, J. R., Hashi, M., and Lindstrom, M. J. (1999). Differences among speakers in lingual articulation for American English `r'. *Speech Communication*, 26:203–226.
[2] Gick, B. (2002). Exercises in 'absurdity': Interactions between flapping and tongue position in American English with implications for speech motor planning. *Journal of the Acoustical Society of America*, 112(5),pt.2:2416.
[3] Gick, B. (2006). Errors and strategy shifts in speech production indicate multiple levels of representation. *Journal of the Acoustical Society of America*, 119(5),pt.2:3302.
[4] Ladefoged, P. and Maddieson, I. (1996). The Sounds of the World's Languages. Blackwell, Oxford
[5] Irtel, H. (2007). PXLab: The Psychological Experiments Laboratory [online]. Version 2.1.11. Mannheim (Germany): University of Mannheim. [cited 2007 June 19]. Available HTTP: http://www.pxlab.de
[6] P. Boersma and D. Weenik, (2008). "Praat: Doing phonetics by computer (version 5.0.27)" [Computer program], [cited 2008 Jul 05], Available HTTP: http://www.praat.org/
[7] Max-Planck-Institute. (2008). ELAN Linguistic Annotator [online]. Version 3.5.0. Nijmegen (The Netherlands): Max-Plank-Institute [cited 2008 Jul 05]. Available HTTP: http://www.lat-mpi.eu/tools/elan/
[8] JMP, Version 5.1. (2003). SAS Institute Inc., Cary, NC, 1989-2003.

ACKNOWLEDGMENTS

Special thanks to Aislin Stott for labeling and transcribing audio recordings and for help in conducting experiments. This research was funded by NSERC.

NUMERICAL MODEL OF A THERMOACOUSTIC ENGINE

Omar Hireche[1], Catherine Weisman[1,2], Diana Baltean-Carlès[1,2], Maurice X. François[1,2], Patrick Le Quéré[1] and Luc Bauwens[3]

[1]Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur (LIMSI-CNRS), BP 133, F-91403, Orsay, France

[2]Université Pierre et Marie Curie, 4 Place Jussieu, F-75252, Paris Cedex 05, France

[3]Department of Mechanical and Manufacturing Engineering, University of Calgary, Calgary, Canada T2N 1N4

1. INTRODUCTION

An asymptotically consistent small Mach number model of a complete thermoacoustic engine has been developed. The model couples linear acoustics in the resonator with a low Mach number flow model for the acoustically compact assembly stack + heat exchangers.

2. PHYSICAL MODEL

Both linear acoustics and incompressible flow are low Mach number approximations to the equations of gas dynamics. However the former assumes length and time in a ratio of the order of the speed of sound while the latter assumes a ratio comparable with the fluid velocity. For a common time scale, the two models assume length scales in the order of the Mach number M. Additionally [1-3], for the latter, a low Mach number approximation can be derived allowing for arbitrary spatially uniform pressure fluctuations that are superimposed to the dynamic correction of order M^2 that characterizes incompressible flow. An asymptotically consistent model of a complete thermoacoustic engine can be constructed matching these two approximations. Geometry is described in Fig. 1.

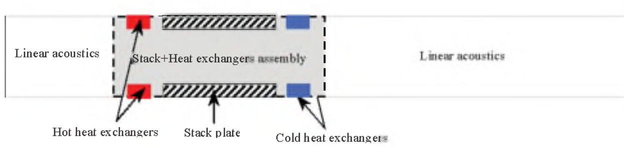


Fig. 1. System geometry

3. ACOUSTICS

The Riemann variables $L(x,t)=p'-\rho uc$ and $R(x,t)=p'+\rho uc$, in which u is the longitudinal velocity, T and ρ, leading order temperature and density, and p' the order M acoustic pressure correction are introduced. L and R are constant respectively on characteristics moving left and right at the speed of sound c. In the resonator, u and p' vary in space and time, while p is constant. ρ and T are constant but with different values on the two sides of the heat exchangers. Taking the outer boundary conditions into account, L at the left end of the heat exchanger section at time t is related to R at the same location at t minus the round-trip time. Likewise R at the right end of the heat exchangers depends upon the previous value of L at that end. This results in two relationships between acoustic pressure in the heat exchangers to the velocities at the ends.

4. STACK AND HEAT EXCHANGERS

In the stack and heat exchangers, calling $\boldsymbol{u}$ the velocity vector, and $p''(\mathbf{x},t)$ the order M^2 dynamic pressure correction, the multidimensional flow is represented by the conservation equations for mass, momentum and energy:

$$\frac{\partial\rho}{\partial t}+\nabla\cdot(\rho u)=0$$

$$\rho\frac{\partial u}{\partial t}+\rho u\cdot\nabla u=-\nabla p''+\nabla\cdot\tau$$

$$\frac{\partial p}{\partial t}+\gamma\nabla\cdot(up)=\nabla\cdot(k\nabla T)$$

where k is the thermal conductivity, γ is the ratio of specific heats and τ is the stress tensor, related to velocity in the usual way for Newtonian fluids. Pressure, density and temperature are related by the equation of state, $p=\rho(\mathbf{x},t)RT(\mathbf{x},t)$. While this formulation supports time-dependent pressure at leading order, the linear acoustics model in the resonator will only support fluctuations p' at order M, to which the heat exchanger section is transparent. Leading order pressure p is an absolute constant, which results in further simplification in the energy equation. The heat exchanger model is completed by conservation of energy in the stack, which is governed by a standard heat equation. At the boundary, temperature and the heat flux vector are continuous.

The solution is obtained using a code derived from [4], that is second order accurate in time and space. Viscous and conductive terms are dealt with using an implicit formulation, while advective terms are explicit. Time integration uses a predictor-corrector formulation.

5. MATCHING

Acoustics provide two equations relating acoustic pressure in the compact heat exchanger section to the velocities at the two interfaces. Integration of the energy equation in the heat exchangers (above) over the heat exchangers computational domain results in a third one, relating the velocities at the two ends to heat transfer. Thus, while it is transparent to the acoustic pressure, the heat exchanger section appears in the acoustics as a point source of volume. Solving the three equations completes the acoustic formulation and, at each time step, provides boundary conditions to the numerical low Mach number flow model.

6. RESULTS

Results were obtained for a resonator length of 8 m and a stack 0.15 m long, located at 1 m from the warm end. Cold end temperature was 293 K. The fluid was helium at 1 MPa.

The stack was made of stainless steel 304L, with thickness 0.2 mm and distance between plates, 0.77 mm. Temperature ratios T_{hot}/T_{Cold} of 1.3 and 2 were considered. Figure 2 shows the pressure history, while Fig. 3 shows velocities and Fig. 4, the phase relationships, for temperature ratio 1.3 and a standing wave as initial condition. Figures 5 and 6 refer to a temperature ratio of 2, respectively for a standing wave and for resonant noise as initial conditions. A resolution of 34 by 1024 grid points was used in space and of 628 time steps/period.

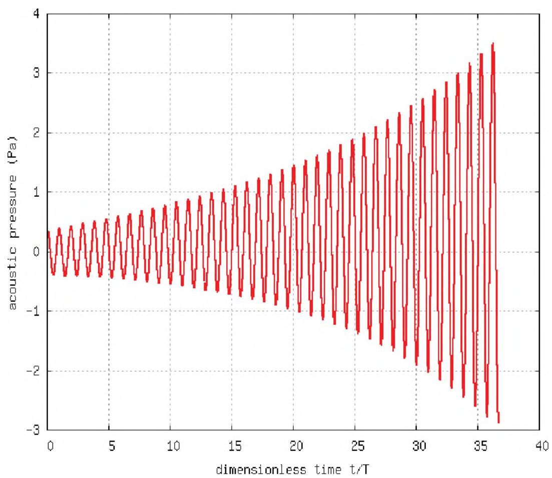


Fig. 2. Temperature ratio 1.3 – Stack pressure history

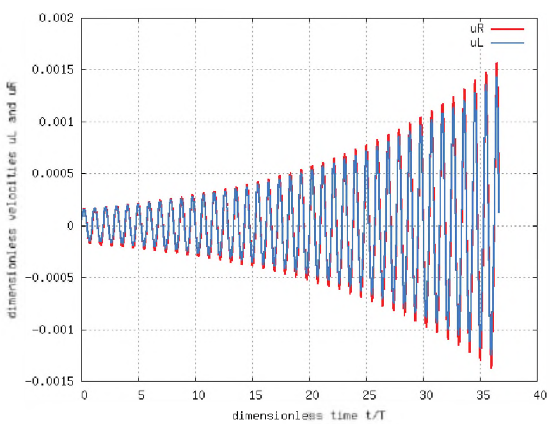


Fig. 3. Temperature ratio 1.3 – Stack end velocities

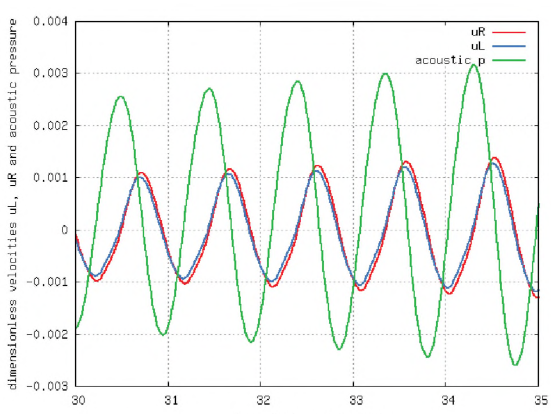


Fig. 4. Temperature ratio 1.3 – Phase relationships.

The results on Figs. 2 to 4 show that for lower temperature ratio, the lowest resonant mode is the most unstable; growth rate is 3.2 s^{-1}. However, for the higher temperature ratio, the second mode manifests itself and grows faster, regardless of whether the system was excited by random noise or by the resonant mode. Starting with noise, growth rates are respectively 8 s^{-1} and 13.5 s^{-1}.

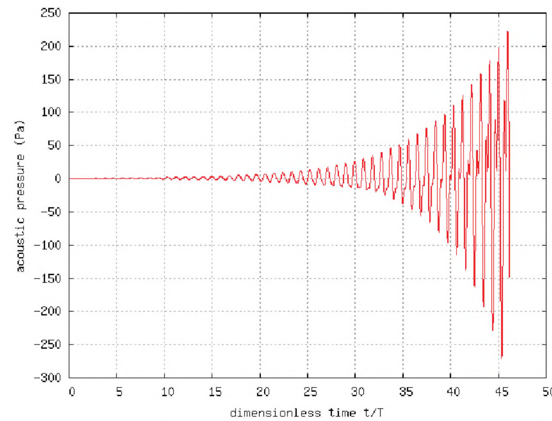


Fig. 5. Temperature ratio 2 – Initiation with standing wave.

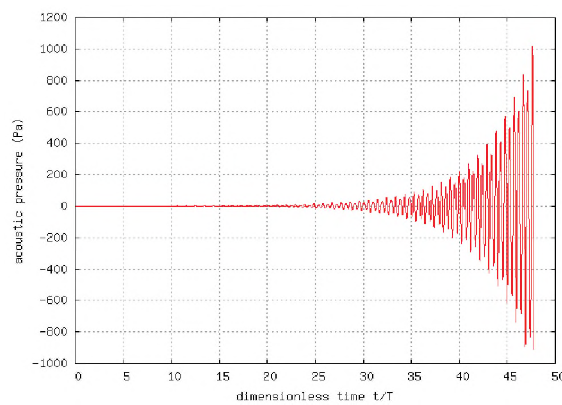


Fig. 6. Temperature ratio 2 – Initiation with random noise.

REFERENCES

[1]S. Paolucci. (1982) On the filtering of sound from the Navier-Stokes equations. Sandia National Laboratories report SAND82-8257.

[2]A. Majda and J.A. Sethian. (1984) The derivation and numerical solution of the equations for zero Mach number combustion. *Combust. Sci. Tech.* 42, 185-205.

[3]L. Bauwens. (1996). Oscillating flow of a heat-conducting fluid in a narrow tube. *J. Fluid Mech.*, 324, 135-161.

[4]P. Le Quéré, C. Weisman, H. Paillère, J. Vierendeels, E. Dick, R. Becker , M. Braack and J. Locke. (2005) Modelling of Natural Convection Flows with large Temperature Differences: A Benchmark Problem for Low Mach Number Flow, Part 1. Reference Solutions. ESAIM: M2AN, 39 (3), 609-616.

ACKNOWLEDGMENTS

Support from the Natural Science and Engineering Research Council of Canada is gratefully acknowledged.

The computations were performed on the NEC SX8 of the CNRS-IDRIS Computing Center.

A SEMI-ANALYTICAL APPROACH TO THE STUDY OF THE TRANSIENT ACOUSTIC RESPONSE OF CYLINDRICAL SHELLS

Cédric Leblond[1], Jean-François Sigrist[2] and Serguei Iakovlev[3]
[1]LEPTIAB, Université de la Rochelle, Av. Michel Crépeau, 17042 La Rochelle, Cedex 1, France, cedric.leblond@univ-lr.fr
[2]DCNS Propulsion, STS, 44620 La Montagne, France, jean-francois.sigrist@dcnsgroup.com
[3]Dept. of Engineering Mathematics and Internetworking, Dalhousie University, Halifax, Canada B3J 2X4, serguei.iakovlev@dal.ca

1. INTRODUCTION

A semi-analytical method related to the effects of a weak shock wave on a submerged evacuated cylindrical elastic structure is proposed. The circular shell/acoustic medium interaction problem has already been tackled in the frequency domain with a full elastic model by Pathak and Stepanishen [1]. The purely transient case has only been achieved with simplified thin shell models based on the Love-Kirchhoff hypotheses for the structural dynamics, see for instance Iakovlev [2] and the references therein. The resulting radiated pressure field displays some discrepancies related to the A_0/S_0 waves when compared to the experimental data obtained by Ahyi et al [3]. Since the thin shell models are known to be restricted to the low frequency domain (the wavelengths in the structure and in the fluid must be larger than the shell thickness), and a weak shock wave may contain some high frequency components, the discrepancies should be overcome by the use of a full elastic model. This is done in this paper in a two-dimensional framework. The approach is based on the methods of Laplace transform in time, Fourier series expansions and separation of variables in space. It is shown that the previously reported drawbacks related to the A_0/S_0 waves are eliminated and the new approach results in much more realistic images of the radiated acoustic field when compared to experiments.

2. METHOD

We consider a two-dimensional elastic shell of constant thickness h, external radius R_s, density ρ_s and longitudinal and transversal wave velocities c_l and c_t, respectively. It is submerged in an infinite fluid medium of density ρ_f and subjected to a weak incident acoustic excitation. A schematic of the problem is shown in Figure 1. In the following, the variables are written in a dimensionless form: the lengths are normalized by R_s, the time by R_s/c_f with c_f being the sound speed in the fluid, and the pressure by $\rho_f c_f^2$.

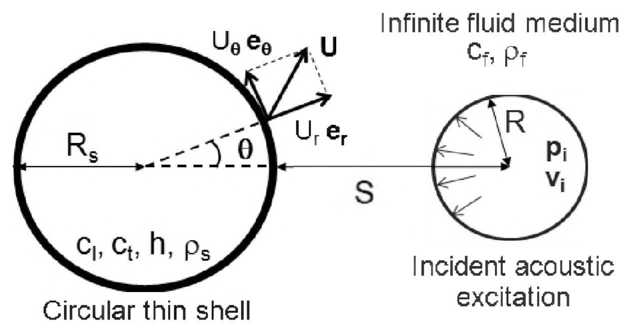


Fig. 1. Submerged two-dimensional circular cylindrical shell subjected to an incident acoustic excitation.

The displacements **U** into the shell are expressed in the two-dimensional theory of elasticity by the scalar potentials ϕ and ψ such that $\mathbf{U} = \nabla\phi + \nabla \times (\psi \mathbf{e_z})$. The displacement potentials satisfy the dimensionless wave equations in the Laplace domain: $\nabla^2\hat{\phi} - s^2\Omega_l^2\hat{\phi} = 0$ and $\nabla^2\hat{\psi} - s^2\Omega_t^2\hat{\psi} = 0$, with s the Laplace variable and Ω_l and Ω_t respectively given by $\Omega_l = c_f/c_l$ and $\Omega_t = c_f/c_t$. Performing the Fourier series expansions $\hat{\phi}(r,\theta,s) = \sum_{n=0}^{\infty}\hat{\phi}_n(r,s)\cos(n\theta)$ and $\hat{\psi}(r,\theta,s) = \sum_{n=1}^{\infty}\hat{\psi}_n(r,s)\sin(n\theta)$, and using the method of separation of variables, the potentials components are given by $\hat{\phi}_n(r,s) = A_n(s)\,\mathrm{J}_n(i\Omega_l sr) + B_n(s)\,\mathrm{Y}_n(i\Omega_l sr)$ and $\hat{\psi}_n(r,s) = C_n(s)\,\mathrm{J}_n(i\Omega_t sr) + D_n(s)\,\mathrm{Y}_n(i\Omega_t sr)$, where J_n and Y_n denote the classical Bessel functions. The potential coefficients A_n, B_n, C_n and D_n have to be determined with the boundary conditions: $\hat{\sigma}_{rr}(1,\theta,s) = -\hat{p}|_{r=1}$, $\hat{\sigma}_{r\theta}(1,\theta,s) = 0$, $\hat{\sigma}_{rr}(r_i,\theta,s) = 0$ and $\hat{\sigma}_{r\theta}(r_i,\theta,s) = 0$. Here r_i denote the inner dimensionless shell radius and p the fluid pressure. By virtue of the acoustic problem linearity, it is classically divided into three components: the incident pressure p_i which is a given data, the diffracted pressure p_d in order to balance the normal component of the incident wave velocity on the body surface, $\mathbf{v_i}\cdot\mathbf{e_r}$, and the last component p_r, representing the pressure radiated by the deformations of the submerged body. Using the classical elasticity relations and Fourier series expansions, the boundary conditions are expressed as functions of the potential coefficients, which yields a linear algebraic system of size 4×4 for each Fourier mode, very similar to that obtained by Pathak and Stepanishen (1994) for the harmonic problem. The solution is straightforward: it consists of simple matrix inversions for each point of the Laplace variable and Fourier mode, numerical inversions of the Laplace transforms [4] and Fourier series summations. Once the shell displacements are obtained, the pressure in the fluid domain is derived employing the analytical response functions available for circular geometries, see for instance Iakovlev [2].

3. RESULTS

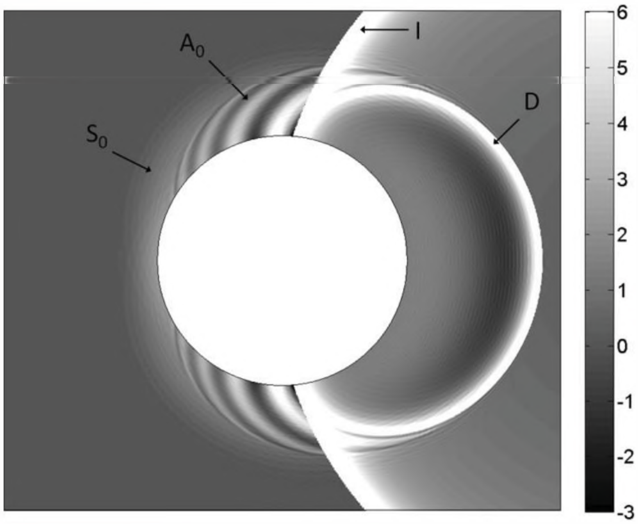


Fig. 2. Pressure field in the fluid domain in MPa, obtained with the elastic model for the shell dynamics described in Section 2.

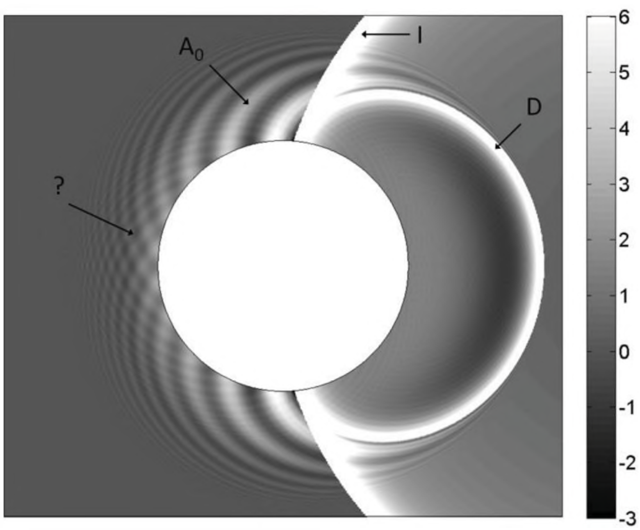


Fig. 3. Pressure field in the fluid domain in MPa, obtained with a thin shell model based on the Love-Kirchhoff hypotheses as in Iakovlev (2008).

The method is illustrated by computing the shell response to a weak underwater explosion. For this purpose, the similitude relation for the far-field shock-wave pressure profile provided by Price [5] is used. It consists of the hyper-acoustic pulse, $P(R,t) = P_c \left[\frac{a_c}{R}\right]^{1+A} F\left(\left[\frac{a_c}{R}\right]^B \frac{v_c}{a_c} t\right)$, with F a known function given by $F(t) = 0.8251 e^{-1.338t} + 0.1749 e^{-0.1805t}$ valid for $t \leq 7$, and R the distance from the explosive charge of radius a_c. The constants depend on the explosive and take the following values for the TNT: $P_c = 1.67\ GP_a$, $v_c = 1010\ m/s$, $A = 0.18$ and $B = 0.185$. As in Geers and Hunter [6], it is assumed that this pressure profile propagates in the far-field at the linear acoustic speed of sound and the incident velocity field is deduced through acoustic relations. Finally, the incident pressure and velocity fields required for the fluid-structure interaction problem, respectively p_i and $\mathbf{v_i}$, are expressed on the shell surface and derived analytically in the Laplace domain. Here, the incident fields are obtained by the detonation of 1 kg of TNT located at a dimensionless stand-off of $S = 2$.

The resulting pressure field in the fluid domain is illustrated in Figure 2 for a shell of thickness $h/R_s = 0.06$ (with material parameters $\rho_s = 7800$ kg/m^3, $c_l = 5800$ m/s, $c_t = 3100$ m/s, $\rho_f = 1000$ kg/m^3 and $c_f = 1470$ m/s). The pressure field derived with a thin shell model is shown in Figure 3. The incident (I) and diffracted (D) waves are obviously the same since they are independent of the shell dynamics. Some discrepancies arise for the shell-induced waves, i.e. the pseudo-Rayleigh wave (A_0) and the Lamb wave (S_0). With the elastic shell model, these two waves are well distinguished and the agreement with experimental data provided by Ahyi et al [3] seems to be excellent (see figure 6(a) of that paper). As for the pressure field obtained with the thin shell model, the streaky pattern induced by the symmetric A_0 can be recognized just ahead the incident wave, but the phase velocity of its high frequency component is overestimated. The S_0 and A_0 waves cannot in this case be identified.

REFERENCES

[1] A. G. Pathak & P. R. Stepanishen (1994). Acoustic harmonic radiation from fluid-loaded infinite cylindrical elastic shells using elasticity theory. *JASA.*, 99, 573-582.

[2] S. Iakovlev (2008). Interaction between a submerged evacuated cylindrical shell and a shock wave – Part I: Diffraction-radiation analysis. *J. Fluids Struct.*, in press.

[3] A. C. Ahyi, P. Pernod, O. Gatti, V. Latard, A. Merlen & H. Uberall (1998). Experimental demonstration of the pseudo-Rayleigh wave. *JASA.*, 104, 2727-2732.

[4] L. Brancik (2001). Utilization of quotient-difference algorithm in FFT based numerical ILT method. *Proceedings 11th Radioelektronica*, Brno, Czech Republic, 352-355.

[5] R. S. Price (1979). Similitude equations for explosives fired underwater. *Technical Report* R 80-299, NSWC.

[6] K. S. Hunter & T. L. Geers (2004). Pressure and velocity fields produced by an underwater explosion. *JASA.*, 115, 1483-1496.

ACKNOWLEDGMENTS

The first author gratefully acknowledges the financial support provided by the company DCNS Propulsion when he was a PhD student. The third author acknowledges financial support of the Killam Trusts and the Natural Sciences and Engineering Research Council of Canada.

SHOCK WAVE REFLECTION AND FOCUSING PHENOMENA IN FLUID-INTERACTING SHELL SYSTEMS

Serguei Iakovlev [1], Garrett Dooley[2] and Kyle Williston [1]
[1] Dept. of Engineering Mathematics and Internetworking, Dalhousie University, Halifax, Nova Scotia, Canada B3J 2X4
[2]Dept. of Mechanical Engineering, University, Halifax, Nova Scotia, Canada B3J 2X4

1. INTRODUCTION

Submerged circular cylindrical shells containing fluids are common in ocean engineering and naval architecture applications (underwater oil pipelines for example), but compared to their evacuated counterpart, have received relatively little attention [1, 2, 3]. The present work concerns with the analysis of the hydrodynamic fields induced in such systems by an external shock loading. Particular attention is paid to the case of two different fluids inside and outside the shell as it is the most practically interesting scenario. The effect of the different phenomena observed in the fluid on the stress-strain state is of definite practical interest as well.

2. MATHEMATICAL APPROACH

The model of irrotational, inviscid, and linearly compressible fluid is used, and the shell is assumed to be thin enough for the linear theory of shells to apply; additionally, Love-Kirchhoff hypothesis is assumed to hold true as well [4]. The fluids and the shell are coupled through the dynamic boundary condition on the interface.

A semi-analytical solution has been developed, and the separation of variables was used in combination with the Laplace transform to obtain the hydrodynamic pressure in modal form. The finite difference technique was employed to obtain the harmonics of the shell displacements. The simulated images based on the solution developed were compared to the available experimental ones for some of the pressure components [5], and a very good agreement was observed.

3. RESULTS AND CONCLUSIONS

A steel shell was considered with the thickness and radius of 0.005 m and 0.5 m, respectively. The interaction with a cylindrical incident wave [5] with the rate of exponential decay of 0.0001314 s, and the pressure in the front of 10 kPa, was analyzed. Three scenarios of fluid contact were addressed: identical fluids (ς=1.00), the internal fluid with the acoustic speed lower than that in the external one (ς=0.50), and the other way around (ς=1.50), ς being the ratio of the internal and external acoustic speeds.

Fig. 1 shows the dynamics of the acoustic field when the fluids are identical. Of primary interest to us here are the reflection and focusing, thus we note the Mach stems clearly visible inside the shell at t=1.80 (also seen in the experiments [5] as well), reflected wave at t=2.10, and the focusing that occurs after the reflection takes place. This is the classical "reflection-focusing" pattern observed earlier for reflection in cylindrical cavities [5], but it is of interest to note how the phenomena in the internal fluid manifest themselves in the external fluid (the elastic interface makes such transition possible).

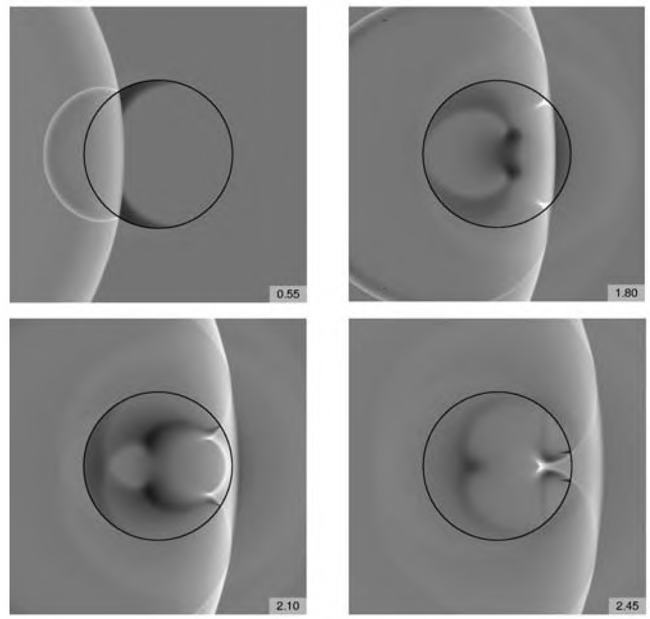


Figure 1. The dynamics of the acoustic field for the case of two identical fluids (ς=1.00).

Fig. 2 shows the dynamics of the acoustic field when the acoustic speed in the internal fluid is lower than that in the external one, ς=0.50. We observe that the hydrodynamic pattern in this case is dramatically different – not only the shape of the internal wave is no longer convex, there is no "reflection-focusing" sequence observed anymore. Instead, we observe the pre-reflection focusing at t=3.90 followed by the reflection shortly after, and then the secondary, post-reflection focusing. Thus, in the case of ς<1.00 a new, "focusing-reflection-focusing" sequence is possible which means that the case of two different fluids can be not only qualitatively, but also phenomenologically different. We note that the pre-reflection focusing occurs shortly before the internal wave falls on the back wall, and that could have important implications in terms of the stress-strain state. We also note that the internal shock wave is no longer a geometrical continuation of the external one.

Fig. 3 shows the dynamics of the acoustic field when the acoustic speed in the internal fluid is higher than that in the external one, ς=1.50. In this case, the internal shock wave is not a geometrical continuation of the external one either, but in a different way, i.e. it has an even higher curvature than the incident wave. The internal wave reaches the tail region ahead of the external one, and the Mach stems of the internal reflection are not only clearly visible inside the shell, but also manifest themselves in the external fluid. In terms of the sequence of the reflection and focusing phenomena, they are the same as in the case of two identical fluids, but are shifted in time and occur earlier than in that case. The appearance of the reflection and focusing patterns is very similar as well. This scenario where the internal acoustic speed is higher than the external is, therefore, less unique than its $\varsigma<1.00$ counterpart. We also note what could be referred to as the "leaking" of the internal shock wave into the external fluid, t=1.40.

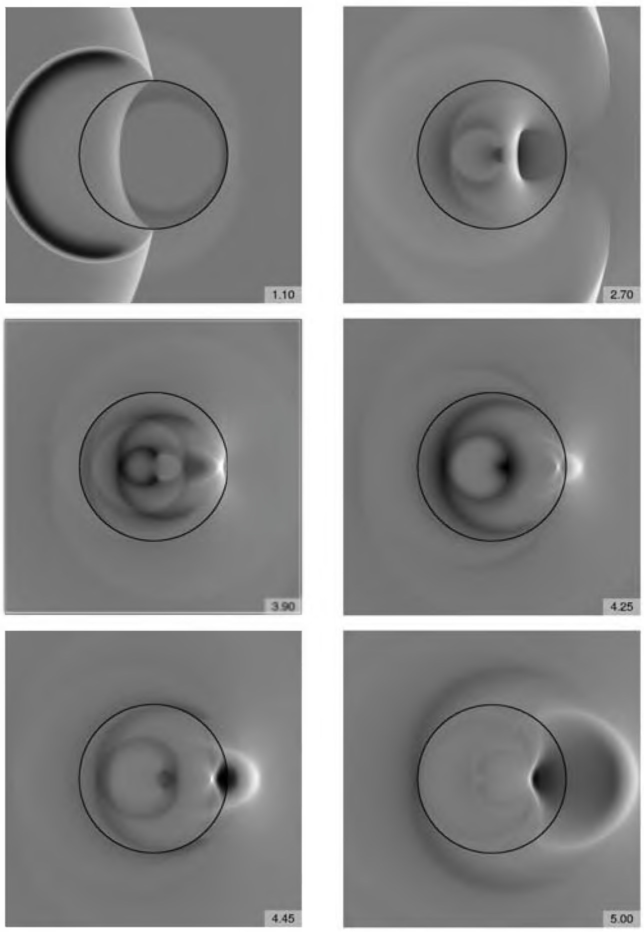


Figure 2. The dynamics of the acoustic field for the case of $\varsigma=0.50$.

We can therefore conclude that the classical "reflection-focusing" [5] pattern is not always present in the fluid-contacting shell systems, and other scenarios are possible, depending on the ratio of the internal and external acoustic speeds. In particular, a more complex and practically interesting "focusing-reflection-focusing" scenario could occur.

Another aspect we would like to comment on is the effect of the shock wave reflection and focusing effects on the stress-strain state of the shell. The case of two identical fluids was addressed in some detail in [6], but not that of the different fluids. When $\varsigma>1.00$, the effects observed in the $\varsigma=1.00$ case are simply shifted in time, as were the reflection and focusing themselves. When $\varsigma<1.00$, however, the pre-reflection focusing in some cases occurs very close to the shell surface thus causing a high tensile stress in the tail region very late in the interaction. For example, when $\varsigma=0.50$, this peak stress can be up to 60% of the highest compressive stress observed in the shell. This effect is of noticeable practical importance, and is being currently further investigated.

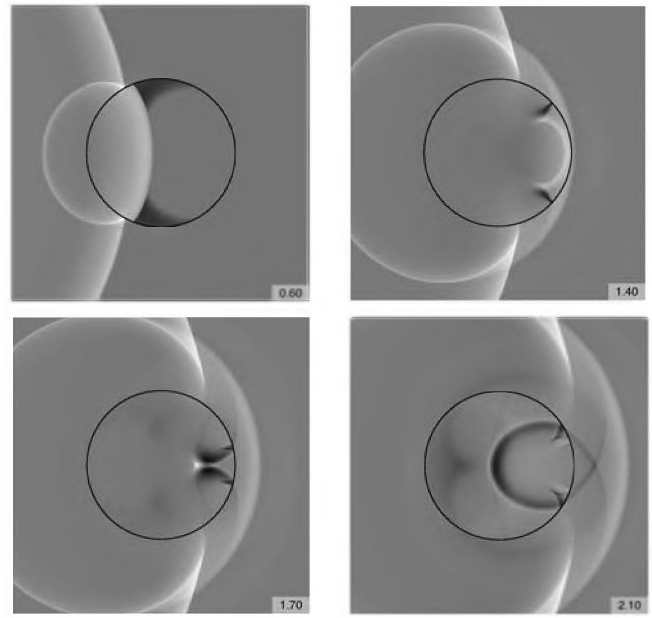


Figure 3. The dynamics of the acoustic field for the case of $\varsigma=1.50$.

REFERENCES

[1] Huang, H. (1979) Transient response of two fluid-coupled cylindrical elastic shells to an incident pressure pulse, *Journal of Applied Mechanics* 46, 513-518.

[2] Zhang, P. and Geers, T. L. (1993) Excitation of a fluid-filled, submerged spherical shell by a transient acoustic wave, Journal of the Acoustical Society of America 93, 696-705.

[3] Neubauer, W. G. and Dragonette, L. R. (1970) Observation of waves radiated from circular cylinders caused by an incident pulse, *Journal of the Acoustical Society of America* 48, 1135-1149.

[4] Junger, M. C. and Feit, D. (1972) *Sound, structures, and their interaction*, MIT Press, Cambridge, USA.

[5] Sun, M. and Takayama, K. (1996) A holographic interferometric study of shock wave focusing in a circular reflector. *Shock Waves* 6, 323-336.

[6] Iakovlev, S. (2007) Submerged fluid-filled cylindrical shell subjected to a shock wave: Fluid-structure interaction effects, *Journal of Fluids and Structures* 23 (1), 117-142.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the financial support of the Natural Sciences and Engineering Research Council (NSERC) of Canada, the Killam Trusts, and the Faculty of Engineering, Dalhousie University.

ARRAY ELEMENT LOCALIZATION OF A BOTTOM-MOUNTED HYDROPHONE ARRAY USING SHIP NOISE

Gordon R. Ebbeson[1], I2ETA Laetitia Thierion[2] and Garry J. Heard[1]
[1]DRDC Atlantic, PO Box 1012, 9 Grove St., Dartmouth, NS, Canada, B2Y 3Z7, gordon.ebbeson@drdc-rddc.gc.ca
[2]ENSIETA, 2, rue François Verny 29806 Brest Cedex 9 – France

1. INTRODUCTION

The DRDC Atlantic Rapidly Deployable Systems (RDS) project was a major research effort whose purpose was to develop an array system that could be deployed in a few minutes and was capable of detecting and localizing sources of acoustic and electromagnetic energy traveling on or underneath the sea surface. For this system to be functional, the locations of the deployed sensors must be known with considerable accuracy. The three-dimensional sensor positions are obtained using a technique referred to as Array Element Localization (AEL). The AEL process is based on the linearized inversion of the measured arrival time data from a series of controlled impulse sources activated in a pattern around the array [1]. Traditionally, imploding light bulbs are used as the sources. Recently, researchers at the University of Victoria in BC [2] have been investigating the use of ship noise as a source of broadband energy for AEL. Encouraged by these results, ship noise was used to carry out AEL on two RDS bottom-mounted horizontal arrays that were previously localized using light-bulb pops. This paper describes the ship AEL method and presents the results for those two arrays.

2. ARRAY ELEMENT LOCALIZATION

AEL acoustic surveys conducted after the array has been deployed involve measuring the arrival times of the signals transmitted from a series of impulsive sources to the hydrophones that are to be localized. Given the sound speed profile in the ocean and the positions of the sources, these arrival times can be inverted to produce estimates of the array hydrophone locations. Ideally, the AEL inversion should address all sources of error in the environment and use all physical *a priori* information about the solution. For a bottom-mounted array, the exact positions and depths of both the sources and the hydrophones are unknown. However, *a priori* information about these parameters, that is, their nominal deployment locations, are often available. In addition, it is quite likely that the array was laid in a smooth curve. This constraint adds further information to the AEL process. Finally, any uncertainties in the sound speed profile can also be included in the inversion. If all of this information is taken into consideration and the corresponding uncertainties are kept as small as possible, the AEL technique will yield an accurate result.

3. Light-Bulb AEL

The AEL process with light bulbs utilizes the arrival that travels directly from the imploding light bulb to the hydrophone as the source of timing. The direct arrival time at each hydrophone is measured relative to some arbitrary start time. This measurement is easy to carry out as the direct arrival is always the first to arrive at the hydrophone [3].

Figure 1 shows the AEL results for a 22-sensor horizontal RDS array. The black triangles show the locations of the 14 light bulbs that were used for the localization. The light-bulb depths, which were measured with a depth recorder that was attached to the bulb breaker, were all very close to 42 m. The black squares of the figure show the sensor locations used to start the AEL process, which are based on the GPS positions of the start and finish of the array deployment. The hydrophone depths were taken from the measured bathymetry around the array and were all estimated to be 67.7 m. The uncertainties in the source positions in Easting, Northing and depth were set to 5, 5 and 2 m, respectively. Those for the hydrophones were set to 150, 150 and 2 m, respectively. In addition, the uncertainty in the sound speed profile was assumed to be 1 m/s. To ensure convergence, an uncertainty of 0.1 ms in the direct arrival times was required. After 20 iterations of the AEL inversion, the hydrophones were localized to the locations shown by the grey triangles of the figure. The AEL result shows the curved shape that was the goal of the array deployment.

A longer, 34-sensor RDS array was localized in a similar manner and the results are shown in Fig. 2. In this case there were 16 light bulbs (black triangles) that were imploded at depths near 40 m. Once again, the black squares show the starting array locations and the hydrophones were all assumed to have a depth of 67.7 m. The uncertainties in the source and hydrophone locations and depths, and the sound speed profile were the same as for the shorter array. An uncertainty in the direct arrival times of 0.1 ms was required to ensure convergence. After 11 iterations, the AEL inversion localized the hydrophones to the slight “S” shape shown by the grey triangles of the figure. This was the shape that was intended at the time of the deployment.

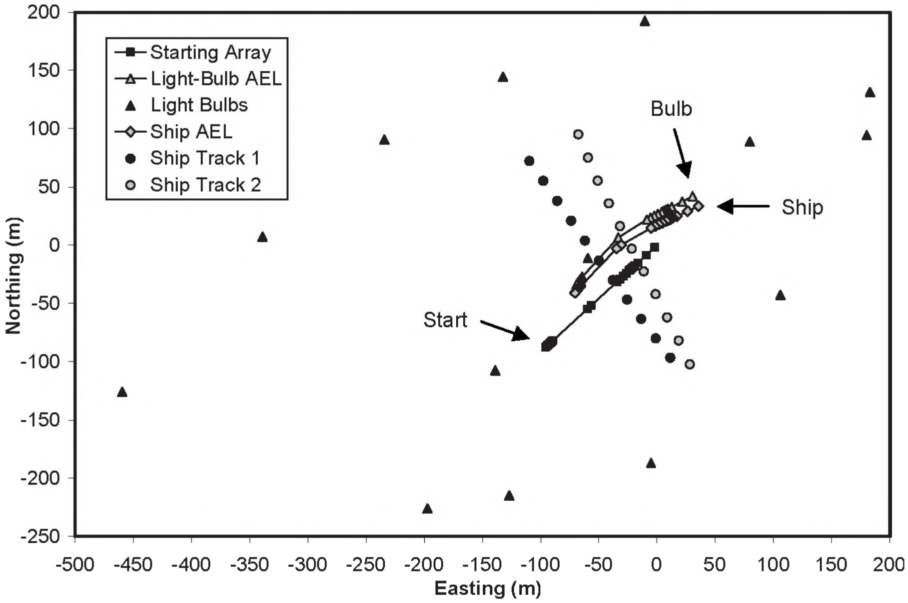


Fig. 1. AEL results for the 22-sensor array.

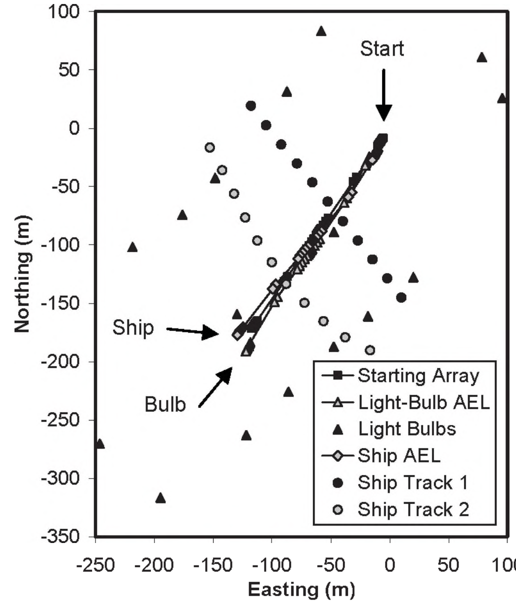


Fig. 2. AEL results for the 34-sensor array.

4. Ship AEL

The AEL process with broadband ship noise also uses the direct arrivals that reach the hydrophones. However, since the ship noise is continuous in time, the ship-AEL method relies on measuring the time delays between the first hydrophone in the array and the remaining hydrophones. This is done by taking the cross-correlation between these hydrophones. Unfortunately, determining a time delay for the direct arrival is not easy as it is often bracketed in time by the delays for the other arrivals. Tracking the delays with ship movement helps in the arrival identification. To increase the measurement accuracy of the time delays, the time series of the ship noise were interpolated in time before carrying out the cross-correlation. In addition, a frequency domain cross-correlation technique was used in order to suppress the narrow band lines in the ship noise [4].

Figure 1 shows the results of the ship AELT for the 22-hydrophone array. The black and grey circles mark the 22 segments of the ship tracks that were used for the localization. Each segment contained 4.1 s of broadband noise. The depth of the ship's propeller, which was 1.2 m, was taken as the source depth. The array starting locations and depths were the same as for the light-bulb AEL (black squares) as were the uncertainties in the source and hydrophone locations and depths, and the sound speed profile. For the AEL process to converge, an uncertainty of 0.13 ms in the direct arrival times was required. After 15 iterations of the AEL inversion, the hydrophones were localized to the locations shown by the grey diamonds in the figure. These AEL results compare extremely well with the light-bulb AEL results (grey triangles).

The ship AEL results for the 34-hydrophone array are shown in Fig. 2. Once again, 22 segments of ship noise, each 4.1 s long, were chosen for the noise sources (black and grey circles). Also, the array starting locations and depths were the same as for the light-bulb AEL (black squares) as were the uncertainties in the source and hydrophone locations and depths, and the sound speed profile. With an uncertainty of 0.37 ms in the arrival times, the inversion converged after 9 iterations and localized the hydrophones to the locations given by grey diamonds of the figure. Once again, these ship-AEL results are very close to the light-bulb AEL results (grey triangles).

5. Summary

Initial efforts using broadband ship noise for AEL have provided encouraging and fast localization results. Proper identification of the direct arrival time differences is the most difficult aspect of the ship noise technique.

REFERENCES

[1] Dosso, S.E. and Ebbeson, G.R. (2006) "Array Element Localization Accuracy and Survey Design", Can. Acoust., Vol. 34, No. 4, 3-13.

[2] Morley, M.G., Dosso, S.E. and Chapman, N.R (2007) "Array element localization using ship noise", http://pruac.apl.washington.edu/abstracts/Morley.pdf.

[3] Ebbeson, G.R., Heard, G.J., Desharnais, F., Matthews, M.N.R. and Dosso, S.E. (2007) "Array Element Localization of a Rapidly Deployable System of Sensors", DRDC Atlantic TM 2007-009.

[4] Knapp, C.H. and Carter, G.C. (1976) "The Generalized Correlation Method for Estimation of Time Delay," IEEE Trans. Acoust., Speech and Signal Processing, Vol. ASSP-24, No. 4.

COMPARISON OF MEASURED AND MODELLED TRANSMISSION LOSS IN EMERALD BASIN

Brian H. Maranda and Nicole E. Collison
Defence Research and Development Canada – Atlantic, P.O. Box 1012, Dartmouth, NS, Canada B2Y 3Z7
Brian.Maranda@drdc-rddc.gc.ca and Nicole.Collison@drdc-rddc.gc.ca

1. INTRODUCTION

A key determinant of sonar performance is the propagation of sound in the ocean, which hence remains a subject of applied interest. On 28 July 2007, acoustic data were collected by DRDC Atlantic in Emerald Basin, an open-ocean site near Nova Scotia, for the purpose of measuring the transmission loss (TL) at frequencies relevant to multistatic active sonar. Comparing the measured TL results with theoretical predictions from propagation models was part of the analysis plan. A previous investigation in the same ocean area is described in [1].

2. THE EXPERIMENT

The experiment took place in Emerald Basin near 43° 50′N and 63°W. An acoustic projector was towed at 50-m depth by the research ship CFAV *Quest*; the transmitted signal consisted of 11 continuous-wave (CW) tones spaced 100 Hz apart from 1010 to 2010 Hz. As *Quest* steered a straightline course, sonobuoys were periodically launched from the ship's stern. The sonobuoy receivers were set for a depth of either 60 m or 120 m, and the received acoustic signals were relayed to the ship over a radio-frequency (RF) telemetry link. The maximum range was 14 km, as determined by the RF link. Although the bathymetry varied slightly along the ship's track, it will be represented as a flat bottom at 265 m depth. A sound-speed profile from the day of the experiment shows a pronounced duct with an axis close to the source depth (Fig. 1).

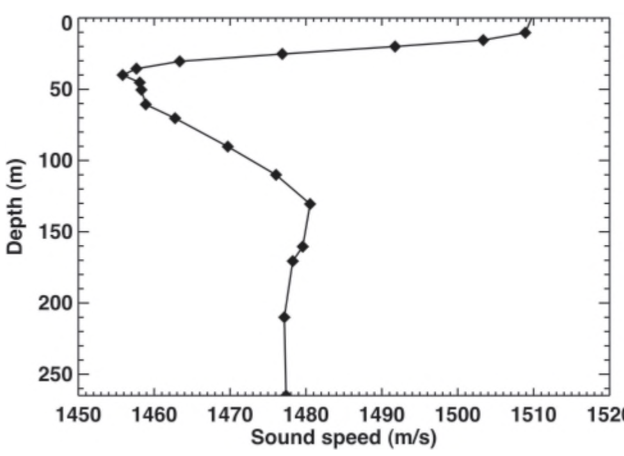


Fig. 1. Sound speed profile taken on 28 July 2007.

3. MEASURED TRANSMISSION LOSS

The data collected during the sea trial were later processed to yield estimates of the transmission loss via the equation TL = SL – SPL, where SL is the projector source level and SPL is the sound pressure level at the receiver (both measured in dB re 1 μPa). The main step in the processing was to perform spectral analysis of the acoustic data to obtain the SPL of each tone at the sonobuoys. The projector SL was measured at the Acoustic Calibration Barge, a facility owned and operated by DRDC Atlantic.

The plots in Fig. 2 show measured TL curves versus range for all 11 frequencies; the gaps in the curves are the result of temporary interruptions that occurred in data acquisition. The curves at the different frequencies cluster quite closely together, but the TL is markedly less for the 60-m receiver than for the 120-m receiver. This latter effect is explained by the fact that the sound propagating to the receiver at 60-m depth is almost entirely trapped in the duct, whereas for the deeper receiver the sound interacts with the waveguide boundaries.

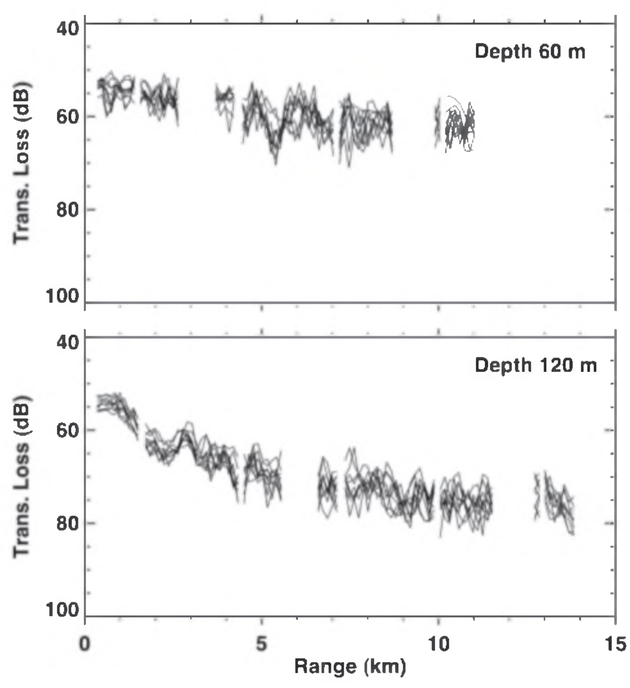


Fig. 2. Measured TL at two receiver depths (60 and 120 m). The results for all eleven frequencies are shown.

4. MODEL RESULTS

The next step was to compare the measured TL curves with theoretical predictions made using the propagation model PECan (Canadian Parabolic Equation model) [2] and Bellhop [3], a model developed by Michael Porter. PECan is based on a finite-difference solution of the parabolic approximation to the wave equation, whereas Bellhop uses Gaussian beams [4]. The geo-acoustic bottom was modelled as a 20-m layer of sediment (density 1.6 g/cm^3, speed 1521.7 m/s, absorption 0.5 dB/ λ) overlying an infinite half-space (density 2.1 g/cm^3, speed 1846.7 m/s, absorption 0.5 dB/ λ).

PECan and Bellhop were used to calculate theoretical TL values on a 5-m range grid. However, the experimental TL values were in effect averaged over 120 m in range due to source motion during the analysis intervals. The model results were therefore averaged in range using a 120-m-long boxcar window. The PECan results for all 11 frequencies are shown in Fig. 3. The TL curves for the 60-m receiver display the expected low loss, without much range structure. There is much greater loss for the 120-m receiver, and an interference pattern is evident. The same general pattern appears in the Bellhop results (not shown).

5. COMPARISON

The experimental and modelled TL results are compared at a single frequency in Fig. 4. The experimental results, earlier plotted as lines in Fig. 2, are now plotted as discrete points superimposed on the theoretical curves. First comparing PECan and Bellhop between themselves, we

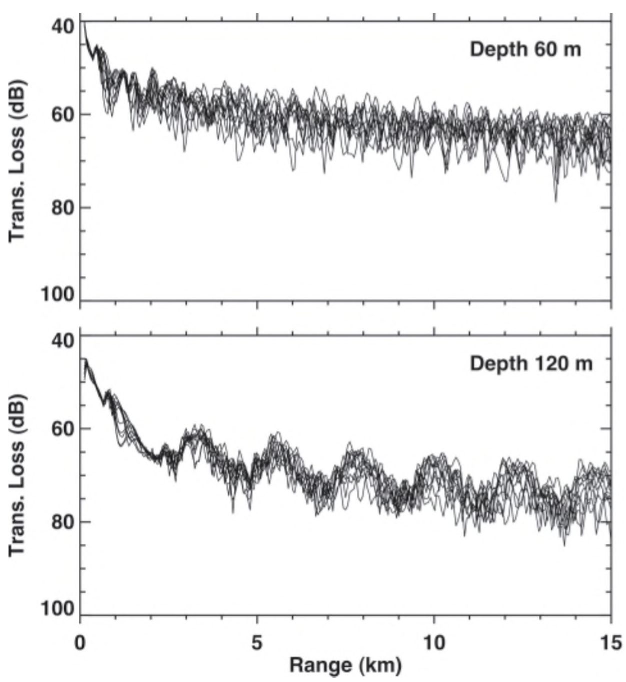


Fig. 3. TL at all 11 frequencies as modelled by PECan.

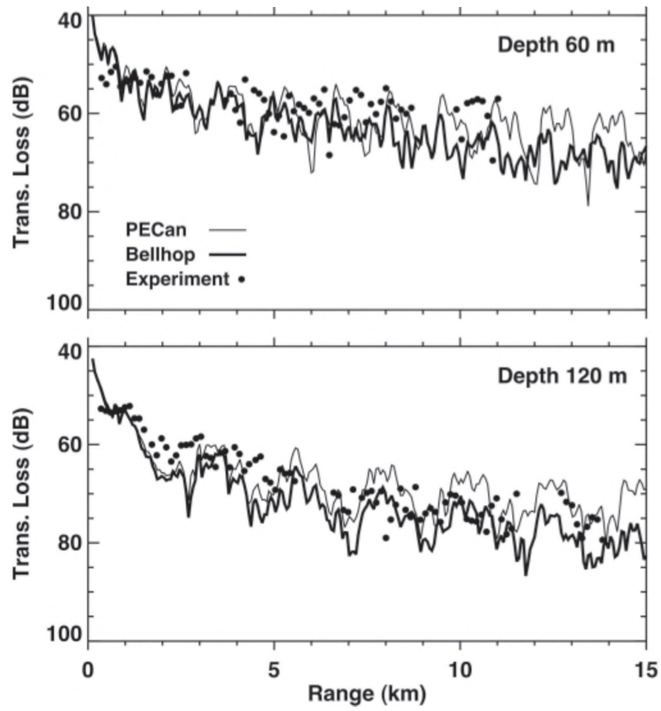


Fig. 4. Comparison of measured and modelled transmission loss at a single frequency (1010 Hz).

find reasonable agreement, particularly for the 120-m receiver, although Bellhop shows greater attenuation with range. (Volume attenuation and surface scattering were omitted from the PECan model runs.) Next comparing the model predictions with the measured TL results, we observe good agreement at the frequency shown (1010 Hz). The agreement at the other frequencies is usually as good as in the case shown here; where systematic discrepancies exist, buoy-to-buoy comparisons suggest the presence of frequency-dependent gain errors in the buoy receivers, which are not intended for the accurate measurement of acoustic level. In conclusion, the results of this paper clearly support the use of the two numerical models as tools for predicting sonar performance in the 1 – 2 kHz band.

ACKNOWLEDGEMENTS

Sean Pecknold, Gary Brooke, and David Thomson provided advice on propagation modelling, and Mark Radcliffe produced the plots.

REFERENCES

[1] Pecknold, S.; Theriault, J.; and Hazen, D. (2002) Propagation loss measurements for low frequency sonar in Emerald Basin and Exuma Sound. Canadian Acoustics, vol. 30, no. 3, 84–85.

[2] Brooke, G.H.; Thomson, D.J.; and Ebbeson, G.R. (2001) PECAN: A Canadian parabolic equation model for underwater sound propagation. J. Comp. Acoustics, vol. 9, 69–100.

[3] Available at http://oalib.hlsresearch.com/Rays/index.html.

[4] Porter, M.B. and Bucker, H.P. (1987) Gaussian beam tracing for computing ocean acoustic fields. J. Acoust. Soc. Am., vol. 82, no. 4, 1349 – 1359.

Bayesian Source Tracking and Environmental Inversion

Stan E. Dosso
School of Earth and Ocean Sciences, University of Victoria, Victoria BC Canada V8W 3P6, sdosso@uvic.ca

1. INTRODUCTION

This paper describes a Bayesian approach to two important and related inverse problems in underwater acoustics: localizing/tracking an acoustic source when ocean environmental properties are unknown, and determining environmental properties using acoustic data from an unknown (moving) source. The goal is not simply to estimate values for source and environmental parameters, but to determine parameter uncertainty distributions, quantifying the information content of the inversion. A common formulation is applied for both problems in which source parameters (location and spectrum) and environmental parameters are considered unknown random variables constrained by noisy acoustic data and by prior information on parameter values (e.g., physical limits for environmental properties) and on inter-parameter relationships (limits on horizontal and vertical source speed). Given the strong nonlinearity of the inverse problem, marginal posterior probability densities are computed numerically using efficient Markov-chain Monte Carlo importance sampling methods. Source tracking results are represented by joint marginal probability distributions over range and depth, integrated over unknown environmental parameters. The approach is illustrated with synthetic examples representing tracking a quiet submerged source and geoacoustic inversion using noise from an unknown ship-of-opportunity.

2. THEORY

Let $\mathbf{m}$ represent the model vector containing the unknown source locations and environmental parameters, and $\mathbf{d}$ represent the data vector containing measured acoustic fields, with the elements of both vectors considered random variables that obey Bayes rule, which may be written

$$P(\mathbf{m}\,|\,\mathbf{d}) \propto L(\mathbf{m},\mathbf{d})P(\mathbf{d}).$$

In the above equation, $P(\mathbf{m}|\mathbf{d})$ represents the PPD which quantifies the information content for the model parameters given both data information, represented by the likelihood function $L(\mathbf{m},\mathbf{d})$, and prior information $P(\mathbf{m})$. The likelihood can typically be written $L(\mathbf{m},\mathbf{d}) \propto \exp[-E(\mathbf{m},\mathbf{d})]$, where E represents the data misfit (log likelihood) function. The multi-dimensional PPD is typically characterized in terms of properties representing parameter estimates, uncertainties, and inter-relationships. Considered here are one- and two-dimensional marginal probability distributions, defined

$$P(m_i\,|\,\mathbf{d}) = \int \delta(m_i - m_i')\,P(\mathbf{m}'\,|\,\mathbf{d})\,d\mathbf{m}',$$
$$P(m_i, m_j\,|\,\mathbf{d}) = \int \delta(m_i - m_i')\,\delta(m_j - m_j')\,P(\mathbf{m}'\,|\,\mathbf{d})\,d\mathbf{m}'.$$

For nonlinear problems, such as acoustic localization and geoacoustic inversion, analytic solutions to the above integrals are not available, and numerical methods must be employed. Here integration is carried out using the method of fast Gibbs sampling (FGS), which applies Markov-chain Monte Carlo importance sampling methods in a principal-component parameter space [1, 2].

Let the complex acoustic pressure fields measured at an array of N sensors for F frequencies and S source positions be given by $\mathbf{d} = \{\mathbf{d}_{jk},\ j=1,S;\ k=1,F\}$. Assuming the data errors are complex, circularly-symmetric Gaussian-distributed random variables, the likelihood function is

$$L(\mathbf{m},\mathbf{d}) \propto \prod_{j=1}^{S}\ \prod_{k=1}^{F} \exp\left\{|\,\mathbf{d} - A_{jk}e^{i\theta_{jk}}\mathbf{d}_{jk}(\mathbf{m})\,|^2/\sigma_{jk}^2\right\}$$

where $\mathbf{d}_{jk}(\mathbf{m})$ is the modelled acoustic pressure, A and θ represent the unknown source spectrum (amplitude and phase) and σ is the standard deviation Maximizing the likelihood (analytically) with respect to A , θ and σ leads to misfit function

$$E(\mathbf{m},\mathbf{d}) = N\sum_{j=1}^{S}\ \sum_{k=1}^{F} \log_e\left\{\mathbf{d}_{jk}^T\,\mathbf{d}_{jk} - \frac{|\,\mathbf{d}_{jk}^T\,\mathbf{d}_{jk}(\mathbf{m})\,|^2}{|\,\mathbf{d}_{jk}^T(\mathbf{m})\,\mathbf{d}_{jk}(\mathbf{m})\,|}\right\}$$

3. EXAMPLES

The first example illustrates tracking a quiet submerged source in shallow water with little knowledge of environmental parameters. The environment and source parameters are illustrated in Fig. 1. Seabed geoacoustic parameters include the thickness h of an upper sediment layer with sound speed c_s, density ρ_s, and attenuation α_s, overlying a semi-infinite basement with sound speed c_b, density ρ_b, and attenuation α_b. The water depth is D, and the water-column sound-speed profile is represented by four parameters c_1–c_4 at depths of 0, 10, 50, and D m. Wide uniform prior distributions (search intervals) are assumed for all parameters. Acoustic data are measured at 300 Hz at

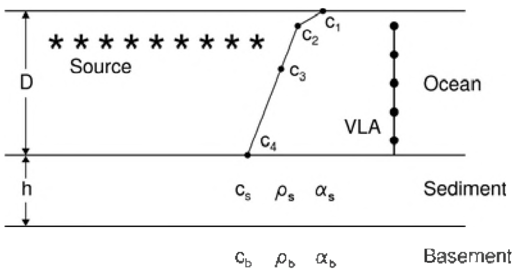


Fig. 1. Experiment geometry and model parameters.

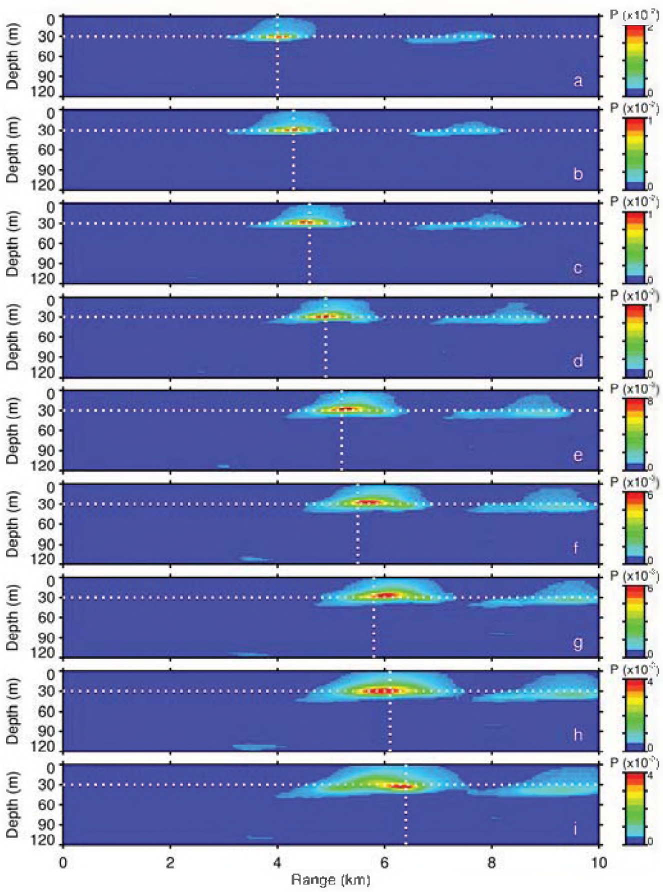


Fig. 2. PASs for the source-tracking example. Dotted lines indicate the true source depth and range

a vertical array consisting of 24 sensors at 4-m spacing from 26- to 118-m depth (simulated acoustic fields are computed using a normal-mode model). The track consists of an acoustic source at 30-m depth moving away from the array at a constant radial velocity of 5 m/s (~10 kts). Acoustic data are collected at the array once per minute for 9 minutes, corresponding to source-receiver ranges of 4.0, 4.3,…, 6.4 km. Random complex-Gaussian errors are added to the data to achieve a signal-to-noise ratio (SNR) that varies from –2 to –8 dB with increasing range along the track. Figure 2 shows probability ambiguity surfaces (PASs), which consist of joint marginal probability distributions for source range and depth integrated over all unknown environmental parameters via FGS, with the additional constraint (prior information) of a maximum source velocity of 10 m/s in the radial and 0.06 m/s in the vertical. The PASs have a strong maximum near the true source location, although weaker secondary maxima are also evident.

The second example simulates geoacoustic inversion using noise lines emanating from a moving ship-of-opportunity of unknown location. The environmental parameters, sensor array, and source track are identical to the tracking example, except that the source depth is 6 m with prior bounds of 2-10 m (i.e., prior knowledge that the source is a surface ship); source range is considered unknown over 0-10 km. Acoustic data are considered at two frequencies of 300 and 350 Hz, with SNR of 4 dB at the shortest source range decreasing with range to approximately –3 dB. Figure 3 shows marginal probability distributions computed for the environmental parameters, indicating a good resolution of the seabed sound-speed structure (h, c_s, c_b) despite the lack of knowledge of source location and water-column sound-speed profile.

REFERENCES

[2] Dosso, S. E. and P. L. Nielsen (2002). Quantifying uncertainty in geoacoustic inversion II: A fast Gibbs sampler approach. J. Acoust. Soc. Am., vol 111, 143-159.

[2] Dosso, S. E., P. L. Nielsen and M. J. Wilmut (2006). Data error covariances in matched-field geoacoustic inversion. J. Acoust. Soc. Am., vol 119, 208-219.

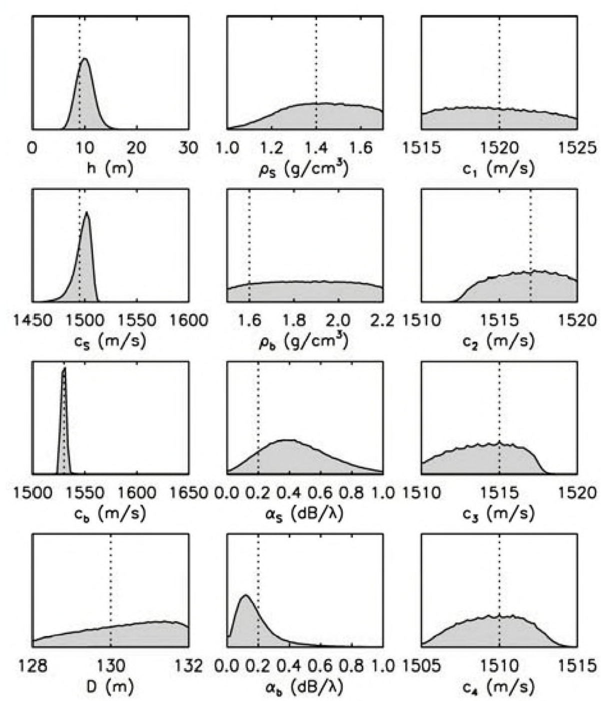


Fig.3. Posterior marginal probability distributions for the ship-noise geoacoustic inversion example.

Two Approaches to Source Tracking in an Unknown Ocean Environment

Michael J. Wilmut and Stan E. Dosso
School of Earth and Ocean Sciences, University of Victoria, Victoria BC Canada V8W 3P6, mjwilmut@uvic.ca

1. INTRODUCTION

This paper compares two approaches, within a Bayesian context, to localizing and tracking a low-level acoustic source in the ocean when environment properties are poorly known. Optimization is based on determining the source and environmental parameters that maximize the multi-dimensional posterior probability distribution (PPD). Marginalization integrates the PPD over nuisance environmental parameters to obtain marginal probability distributions over source range and depth, and the optimal track is determined from these marginal distributions. The question addressed here is which method yields track estimates that are, on average, closer to the true track.

2. THEORY

Let **m** and **d** represent the model and data vectors, respectively, with elements considered random variables that obey Bayes rule, which may be written

$$P(\mathbf{m}\,|\,\mathbf{d}) \propto L(\mathbf{m},\mathbf{d})P(\mathbf{d}).$$

In the above equation, $P(\mathbf{m}|\mathbf{d})$ represents the PPD which quantifies the information content for the model parameters given both data information, represented by the likelihood function $L(\mathbf{m},\mathbf{d})$, and prior information $P(\mathbf{m})$. The likelihood can typically be written $L(\mathbf{m},\mathbf{d}) \propto \exp[-E(\mathbf{m},\mathbf{d})]$ where E represents an appropriate data misfit function.

The multi-dimensional PPD is typically characterized in terms of parameter estimates and uncertainties, such as the maximum *a posteriori* (MAP) model and marginal probability distributions, defined by

$$\hat{\mathbf{m}} = \mathrm{Arg}_{\max}\{P(\mathbf{m}\,|\,\mathbf{d})\}$$

$$P(m_i, m_j\,|\,\mathbf{d}) = \int \delta(m_i - m_i')\,\delta(m_j - m_j')P(\mathbf{m}'\,|\,\mathbf{d})\,d\mathbf{m}'.$$

Optimization seeks the source track and environmental parameters that minimize the misfit to acoustic data (i. e., the model MAP estimate) For efficiency, the optimization is carried out only over environmental parameters, since the most probable source track (subject to source velocity constraints) given the environmental parameters can be calculated using the Viterbi algorithm [1]. Optimization is a generalization of the focalization technique [2] to source tracking. Both adaptive simplex simulated annealing [3] and differential evolution have been used for optimization with good success.

Marginalization requires integration over the environmental parameters [4] to obtain track marginal distributions. Here integration is carried out using the method of fast Gibbs sampling (FGS), which applies Markov-chain Monte Carlo importance sampling methods in a principal-component parameter space. The Viterbi algorithm can then be applied to determine the optimal (most probable) track from the marginal distributions.

3. RESULTS

The optimization and marginalization approaches are compared here with a synthetic example illustrated in Fig. 1. The geoacoustic parameters include the thickness h of an upper sediment layer with sound speed c_s, density ρ_s, and attenuation α_s, overlying a semi-infinite basement with sound speed c_b, density ρ_b and attenuation α_b. The water-column sound speed profile is represented by four unknown sound speeds c_1-c_4 at depths of 0, 10, 50, and D m, where D is the water depth. A low-level 300 Hz source travels at constant depth of 20 m and at a constant velocity of 5 m/s. Acoustic fields from this source are recorded at a 24-sensor vertical array (VLA) once every minute for nine minutes, during which the source moves from 4-km to 6.4-km range. For the inversions, wide prior bounds are applied for the geoacoustic and water column parameters, and the source horizontal and vertical velocities are constrained to be less than 10 m/s and 0.07 m/s, respectively.

For the study carried out here, acoustic data are considered at six different signal-to-noise ratios (SNRs), with average SNRs along the track ranging from –11 to –3 dB (the SNR decreases along the track by approximately 6 dB due to the increasing range). At each SNR, 20 different noisy data sets were inverted using both optimization and marginalization. Figure 2 shows the probability of an acceptable track (PAT), defined as mean absolute depth and range errors less than 10 m and 500 m, respectively, for

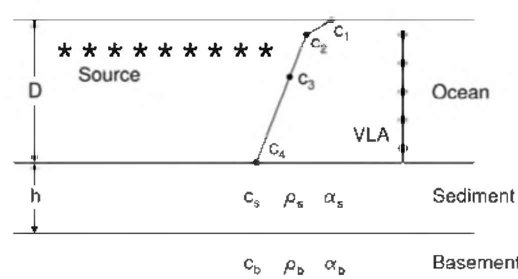


Fig. 1. Geometry and model parameters.

optimization and marginalization, including one standard-deviation error bars. Also included in the figure for reference are PATs computed via the Viterbi algorithm using either exact knowledge of the environmental parameters or environmental parameters drawn at random from the prior bounds. The key result is that marginalization gives significantly better average results than optimization, particularly for SNRs of –9 to –3 dB. Marginalization has the added benefit that the marginal distributions used to calculate the optimal track provide a measure of track uncertainty. Figure 3 shows an example at SNR = –6 dB with one mean-deviation uncertainties about the optimal track.

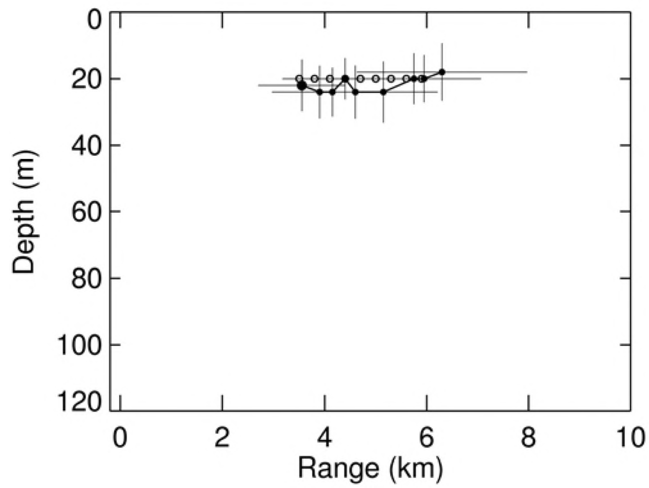


Fig. 3. Example of track estimate and uncertainty. The true track is indicated by open circles and the estimated track by filled circles with mean-deviation uncertainties.

4. SUMMARY

The study carried out here indicates than marginalization significantly outperformed optimization for source tracking in an unknown ocean environment. In addition, marginalization also provides a measure of the track uncertainty. However, the integrations required in marginalization require greater computational time effort than optimization.

REFERENCES

[1] Viterbi, A. J. (1967) Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. IEEE Trans. Inf. Theory IT-13, 268-269.

[2] Collins, M. D., and W. A. Kuperman (1991). Focalization: Environmental focusing and source localization. J. Acoust. Soc. Am., vol 90, 1410-1422.

[3] Dosso, S. E., M. J. Wilmut and A. L. Lapinski (2001). An adaptive hybrid algorithm for geoacoustic inversion. IEEE. J. Ocean. Eng., vol 26, 324-336.

[4] Dosso, S. E. and P. L. Nielsen (2002). Quantifying uncertainty in geoacoustic inversion II: A fast Gibbs sampler approach. J. Acoust. Soc. Am., vol 111, 143-159.

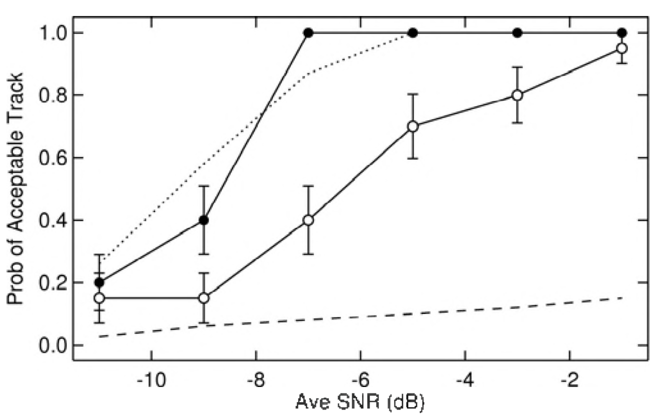


Fig. 2. Comparison of optimization (open circles) and marginalization (filled circles) results for tracking. Dotted and dashed lines indicate results using exact and random environmental parameters, respectively.

GEOACOUSTIC INVERSION OF NOISE FROM SHIPS-OF-OPPORTUNITY WITH UNKNOWN POSITION

Dag Tollefsen[1,2] and Stan E. Dosso[2]
[1]Norwegian Defence Research Establishment (FFI), Maritime Systems Division, NO-3191 Horten, Norway
[2]School of Earth and Ocean Sciences, University of Victoria, Victoria, BC, V8W 3P6

1. INTRODUCTION

The use of noise from ships-of-opportunity for geoacoustic inversion [1-3] can provide several practical advantages: it allows for unobtrusive geoacoustic characterization, a dedicated ship is not required, and additional acoustic sources are not introduced to the marine environment. Use of ship-noise, however, does not allow for control of factors such as source frequency content and position, and supporting data on ship positions may be unavailable or inaccurate. This paper presents results from matched-field geoacoustic inversion (MFI) of noise from a quiet research ship (position known) and from a merchant ship (position unknown) recorded on a horizontal line array (HLA) deployed on the seafloor in shallow water. A Bayesian inversion method [4] is employed; this provides quantitative estimates of model parameters and their uncertainties, and allows for meaningful comparisons of results from different data sets. Ship-noise inversion results are compared with previous results from inversion of controlled-source data collected in the same experiment.

2. METHOD

Statistical properties of model parameters **m** can be obtained from the posterior probability density (PPD)

$$P(\mathbf{m}\mid\mathbf{d}) \propto P(\mathbf{m})\exp[-E(\mathbf{m})]$$

where $E(\mathbf{m})$ is the data mismatch function for (fixed) measured data **d** and model **m**, and $P(\mathbf{m})$ the prior information. The multi-dimensional PPD is typically interpreted in terms its integral quantities, such as the marginal probability distributions, and 95% highest probability density credibility intervals. The integrals are evaluated by the method of fast Gibbs sampling [4]. For acoustic data at N sensors, F frequencies and J segments, $\mathbf{d}=\{\mathbf{d}_{fj},\ f=1,F;\ j=1,J\}$, the standard assumptions of uncorrelated complex-Gaussian distributed errors, unknown source amplitude and phase, and unknown error variance lead to the data mismatch function [3]

$$E(\mathbf{m}) = N\sum_{f=1}^{F}\sum_{j=1}^{J} \log_e B_{fj}(\mathbf{m}),$$

where $B_{fj}(\mathbf{m})$ is the Bartlett mismatch defined by

$$B_{fj}(\mathbf{m}) = \mathrm{Tr}\{\mathbf{C}_{fj}\} - \left[\mathbf{d}^{\dagger}_{fj}(\mathbf{m})\mathbf{C}_{fj}\mathbf{d}_{fj}(\mathbf{m})\right] \Big/ \left|\mathbf{d}_{fj}(\mathbf{m})\right|^2 .$$

Here Tr{•} represents the matrix trace, the dagger represents conjugate transpose, $\mathbf{d}_{fj}(\mathbf{m})$ is the replica acoustic field and $\mathbf{C}_{fj}$ is the data cross-spectral density matrix (CSDM) at the fth frequency and jth data segment (defined by the ensemble average over K time-series snapshots [3]).

With a priori unknown ship position, the ship track is first estimated by simultaneous optimization (minimization of $E(\mathbf{m})$) over environment parameters and source positions. The ASSA hybrid search algorithm [5] is used, with source positions searched over a range-depth grid, and track constraints applied to ship velocity. Bayesian MFI is subsequently employed, with small a priori source position uncertainties centred on the optimal track.

3. RESULTS

Acoustic data were collected using the FFI research array (a 900-m HLA with 18 sensors spaced at 10-m to 160-m intervals) deployed on the seabed at water depth 280 m in a relatively flat area of the Barents Sea. First considered is noise from the R/V H U SVERDRUP II in transit along a track (speed 5 kn) starting at the north end of the array and extending radially outward to range 6 km at a bearing of 30° relative to the array endfire-north. Data from a controlled source towed along this track have previously been used for inversion [6].

3.1 Research-ship noise

Noise from the research ship was processed at three frequency lines within 40-145 Hz at signal-to-noise ratio (SNR) of 5 dB and lower. CSDM estimates were formed for five data segments each at two ranges, each 18-s segment from ten 50%-overlapping data snapshots. Inversions were run for a two-layer model of Quaternary sediment (constant-gradient sound speed upper layer over homogenous lower layer) with seven unknown geoacoustic parameters. Small a priori uncertainties were applied to water depth, source depth, and ship range and bearing (offsets from known track). The ORCA normal-mode model [7] was used to compute replica fields.

Fig. 1 shows marginal PPDs for four geoacoustic model parameters from inversion of controlled-source data and ship-noise at source-array ranges of 1.5 km, Figs. 1(a) and 1(b), and 5.1 km, Figs. 1(c) and 1(d). Figs. 1(a) and 1(b) show in general good consistency between marginal PPDs for the geoacoustic parameters from the controlled-source data and ship-noise, with parameters well defined by both data types. Mean parameter estimates (with mean-deviation uncertainties) from controlled-source data and ship-noise are 1510±21 m/s and 1507±23 m/s for sound speed at top of sediment (c_1), 1753±13 m/s and 1737±43 m/s for sound speed of the lower layer (c_2), and 2.0±0.2 g/cm^3 and 1.7±0.2 g/cm^3 for upper layer density (ρ_1). These values compare well with reference geophysical data. The 5.1-km ship-noise, Fig. 1(d), resolved only an average sound speed of the upper layer. (For further discussion of results, see [3].)

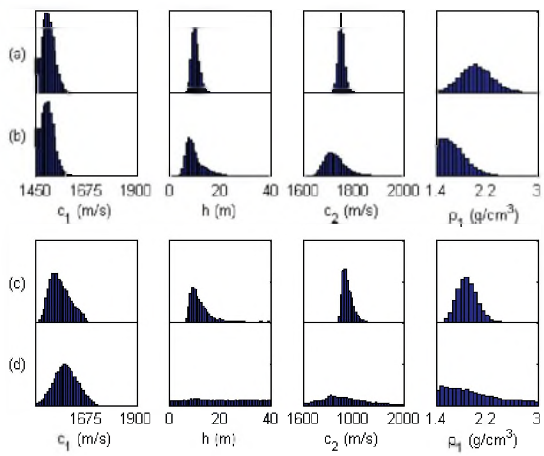


Fig. 1. Results from inversions of controlled-source data and research-ship noise. (See text for explanation).

3.2 Merchant-ship noise

Noise from a 20,000-ton product tanker in transit through the area of experiment was processed at three frequency lines within 40-120 Hz, with SNR in excess of 30 dB. Only one observation of the ship position was logged during the experiment, thus ship track parameters required estimation from the acoustic data. The optimization used eight data segments (total time 2.5 min), and established a track in the array endfire-south direction at 7.0-7.5 km range with ship velocity 15 kn. Subsequent Bayesian geoacoustic inversion used the two-layer seabed model described above, with upper sediment layer thickness constrained to 0-120 m. Results are displayed in Fig. 2 in terms of marginal PPDs for geoacoustic model parameters for inversions of data (two segments in each inversion) at range 7.4 km, Figs. 2(a) and 2(c), and 7.2 km, Figs. 2(b) and 2(d). The data did not resolve distinct upper-layer structure of the seabed, and parameter estimates represent the entire column of Quaternary sediment. Mean values (with mean-deviation uncertainties) are 1716±23 m/s for sound speed at top (c_1) and 1858±47 m/s at bottom (c_2) of sediment with an average sound speed (c_{AVE}) of 1788±14 m/s (consistent results for the two ranges).

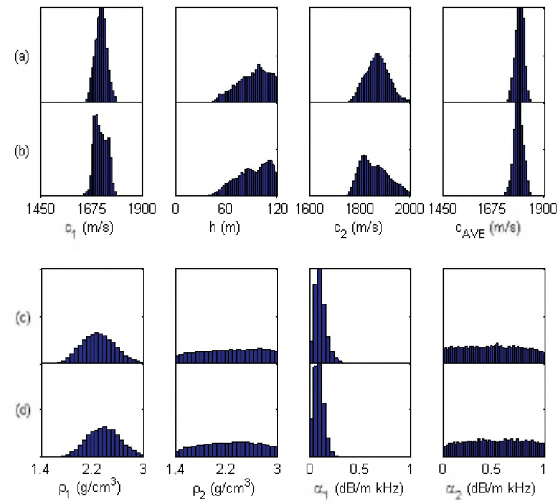


Fig. 2. Results from inversions of merchant-ship noise.

These relatively high-SNR, long-range data also resolved sediment density and attenuation reasonably well, with mean values of 2.4±0.2 g/cm^3 for density (ρ_1) and 0.11±0.04 dB/m/kHz for attenuation (α_1).

4. SUMMARY

Bayesian MFI has been applied to low-frequency narrowband ship-noise recorded on a HLA deployed on the seafloor in shallow water. Seabed geoacoustic model parameter estimates compared well with results from inversion of controlled-source data and with prior geophysical data from the experiment site. Further research is in progress to simultaneously quantify uncertainties in source track and seabed geoacoustic parameters using a Bayesian approach.

REFERENCES

[1] N. R. Chapman, R. M. Dizaij, and R. L. Kirlin, "Geoacoustic inversion using broad band ship noise," in *Proceedings of the Fifth ECUA*, ed. M. E. Zakharia, Lyon, France, pp. 787-792, 2000.

[2] R. A. Koch and D. P. Knobles, "Geoacoustic inversion with ships as sources", *J. Acoust. Soc. Amer.*, vol. 117, pp. 626-637, 2005.

[3] D. Tollefsen and S. E. Dosso, "Bayesian geoacoustic inversion of ship noise on a horizontal array," *J. Acoust. Soc. Amer.*, vol. 124, No. 2, 2008.

[4] S. E. Dosso, "Quantifying uncertainty in geoacoustic inversion. I. A fast Gibbs sampler approach", *J. Acoust. Soc. Amer.*, vol. 111, pp. 129-142, 2002.

[5] S. E. Dosso, M. J. Wilmut, and A. L. Lapinski, "An adaptive hybrid algorithm for geoacoustic inversion," *IEEE J Oceanic Eng.*, vol. 26, pp. 324-336, 2001.

[6] D. Tollefsen, S. E. Dosso, and M. J. Wilmut, "Matched-field geoacoustic inversion with a horizontal array and low-level source," *J. Acoust. Soc. Amer.*, vol. 120, pp. 221-230, 2006.

[7] E. K. Westwood, C. T. Tindle, and N. R. Chapman, "A normal mode model for acousto-elastic ocean environments," *J. Acoust. Soc. Amer.*, vol. 100, pp. 3631-3645, 1996.

Presentations Without Summary Papers

Acoustical Consulting

Addressing the Effects of Overtopping Vegetation on the Perfromance of Highway Noise Barriers
C. W. Wakefiled, Wakefield Associates, Victoria, BC

Architectural and Classroom Acoustics

Ideal Maximum Noise Levels for Elementary School Classrooms
John Bradley, NRC, Ottawa, ON

Ray-Tracing Predictions of Sound-pressure fields in empty and fitted rooms
Murray Hodgson, Gary Chan, and Owen Cousins, UBC, Vancouver, BC

Fire Resistance and and Noise Control in Multi-family buildings
David Quirt and T.R.T. Nightingale

Auditory Scene Analysis

Aging and the cocktail party problem: where is the problem?
Claude Alain, University of Toronto, Toronto, Ontario

The perception of auditory continuity with interrupted masking sounds
Valter Ciocca and Julie Chang, UBC, Vancouver, BC

Psychophysical and neuroelectric evidence of context effects in auditory stream segregation
Joel Snyder, University of Nevada, Las Vegas, USA

The role of attention in the development of auditory scene analysis
Elyse S. Sussman, Albert Einstein College of Medicine, Bronx, NY, USA

Biomedical Acoustics

Finite element model of a SAW Sensor for hydrogen detection
Mohamed M. El-Gowini and Walied A. Moussa, University of Alberta

Environmental Noise/Vibration

Case study: development of a high performance acousitcal pipe lagging system
Corey D. Kinart, HGC Engineering, Mississauga, Ontario

The role of source motion on noise generated by wind turbines
Werner Richarz, Aercoustics Engineering, Toronto, Ontario

Propagation of wind turbine noise in a boundary layer
Werner Richarz, Aercoustics Engineering, Toronto, Ontario

Low noise pavement technologies in the Canadian context – a review of recent studies conducted in Alberta and Ontario
Peter VanDelden and Scott Penton, RWDI Associates, Guelph, Ontario

The Harmonoise/Imagine method
Hans van Leeuwen and Erwin Hartog van Banda, DGMR Industry, The Hague, The Netherlands

Implementing noise prediction standards in calculation software - the various sources of uncertainty
Erwin Hartog van Banda and Hardy Stapelfeldt, DGMR Industry, The Hague, The Netherlands

First Nations Languages Acoustics

Perturbations of pitch by ejectives in Upriver Halkomelem
Jason Brown and James J. Thompson, UBC, Vancouver, BC

Laryngealized vowels in Quiaviní Zapotec: from phonetic realization to phonological contrast
Mario E. Chávez-Peón, UBC, Vancouver, BC

Conflict resolution in SENĆOTEN: uvulars - /e/ sequences
Sonya Bird and Janet Leonard, University of Victoria, Victoria, BC

On the phonetics of schwa in Sliammon (M. Comox Salish): implications of rhe representations of Salish vowels
Susan Blake and Kimary Shahin, UBC, Vancouver, BC

Tonogenesis in two Mayan languages: a contextual acoustic analysis
Francisco Arellanes Arellanes, Fidencio Briceño Chel, Mario E. Chávez-Peón, and H. Antonio Garcia Zuñiga

Tsilhqut'in ejectives: a descriptive phonetic study
SooYoun Ham, University of Victoria, Victoria, BC

A Perception Study of Glottalization in Gitksan Resonants
John Lyon, UBC, Vancouver, BC

'Green'-Building Acoustics

The acoustics of sustainable buildings
Max Richter, Zohreh Razavi, Murray Hodgson, and Khaleghi, Stantec and UBC, Vancouver, BC

Measurement, Analysis and Musical Acoustics

Measuring the displacement of engines using acoustics
Richard Peppin, Scantek, Maryland, USA

Vibrational characteristics of harp soundboards
Chris Waltham, UBC, Vancouver, BC

Occupational Noise / Standards

Ocupational noise exposure assessment using perceived and quantitative measures
Richard Neitzel, William Daniell, Lianne Sheppard, Hugh Davies, Noah Seixas, University of Washington, Seattle

Psychological Acoustics

Effect of frequeny in virtual rumble strips
A. Russo, Jeffrey A. Jones and Mohammad Abdoli-Eramaki, Ryerson University, Toronto, Ontario

Second Language Acquisition Acoustics

An acoustic study of stress in L2 production of German and Spanish
Viola Miglio and Dorothy Chun, University of California, Santa Barbara, CA, USA

Sound Absorbing Materials

Sound absorption of a microperforated panel backed by an irregular-shaped cavity
C.Q. Wang and L. Cheng, J. Pan, The Hong Kong Polytechnic University

Establishing relationships between acoustic and physical properties of shoddy-based fibre absorbers for acoustic modelling
John Manning and Raymond Panneton, Bauer Industries, Waterloo, Ontario

Speech Perception and Automatic Speech Recognition

Acoustic modeling in automatic speech recognition
Li Deng, Microsoft Research, Richmond, Washington, USA

The perceptual basis of velar epenthesis in Costa Rican Spanish
Bethany McLeod, University of Toronto, Toronto, Ontario

Spectral peaks as acoustic correlates to speech perception
Michael Kiefte and Tara Collins, Dalhousie University, Halifax, Nova Scotia

The temporal coordination of recited and spontaneous speech
Adriano V. Barbosa, Barry Ross, Johanna Tan and Eric Vatikiotis-Bateson, UBC, Vancouver, BC

Speech Production and Speech Disorders

Acoustic testing for phonologization
Kimary Shahin, Simon Fraser University, Burnaby, BC

Tongue root retraction and tongue tip recoil in Xhosa alveolar click
Amanda Miller, UBC, Vancouver, BC

Thermoacoustics

A perfectly matched layer technique for Lattice Boltzmann Method
Alireza Najafi-Yazdi and Luc Mongeau, McGill University, Montreal, Quebec

Performance characterization of a small-capacity thermoacoustic cooler for air-conditioning applications
Insu Paek, Luc Mongeau and James E. Braun, Kangwon National University, Korea and McGill University,Montreal

Underwater Acoustics

An efficient 3D PE propagation model with tessellation
Melanie Austin, Ross Chapman and David Hannay, University of Victoria, Victoria, BC

Vibroacoustics

Influence of dilatational propagating motion on diffuse field transmission loss of orthotropic sandwich composite panels
Sebastian Ghinet and Noureddine Atalla, University of Sherbrooke, Sherbrooke, Quebec

FE modeling of aircraft fuselage structure treated with viscoelastic material
Esen Cintosun, Noureddine Atalla and Tatjana Stecenko, University of Sherbrooke, Sherbrooke, Quebec

Introducing
For Digital Recorders
For USB A/D Systems
PHANTOM POWER
7052PH
Measurement Mic System
7052H Type 1.5TM
Titanium Diaphragm
3Hz to >20 kHz
<20 dBA > 140 dBSPL
MK224 (<14 dBA to >134 dBSPL) Optional
4048 Preamp
Superior
IEC 1094 Type 1
Long-term Stability
Temperature and Humidity
Performance
Now in Stock
Phantom
to IEPE/ICP
Adaptor
Supplies 3-4 mA
Power
Accelerometers
Microphones
ICP1248
ACOustics Begins with ACO
TM
MATTtm
Family
MATT TM
Mic
Attenuator
Handle Higher Sound Pressure Levels
ACO Pacific, Inc., 2604 Read Ave., Belmont, CA 94002
Tel:(650) 595-8588 FAX: (650) 591-2891 E-Mail: sales@acopacific.com
Web Site: www.acopacific.com

NEWS / INFORMATIONS

CONFERENCES

If you have any news to share with us, send them by mail or fax to the News Editor (see address on the inside cover), or via electronic mail to stevenb@aciacoustical.com

2008

6-8 October. Canadian Acoustical Association Acoustics Week in Canada. Vancouver, BC. Web: www.caa-aca.ca

14-15 October: Underwater Noise Measurement, Impact and Mitigation. Southampton, U.K. Web: http://underwaternoise2008.lboro.ac.uk

21-23 October. 13th Conference on Low Frequency Noise and Vibration. Tokyo, Japan. Web: www.lowfrequency2008.org

21-24 October. Acustica 2008. Coimbra, Portugal. Web: www.spacustica.pt

26-29 October: Internoise 2008, Shanghai, China. Web: www.internoise2008.org

01-05 November. IEEE International Ultrasonic Symposium. Beijing, China. Web: www.ieee-uffa.org/ulmain.asp?page=symposia

05-07 November: Iberoamerican Acoustics Congress (FIA 2008). Buenos Aires, Argentina. Web: www.fia2008.com.ar

10-14 November. 156th Meeting of the Acoustical Society of America, Miami, Florida, USA. Web: www.asa.aip.org

24-26 November. Australian Acoustical Society National Conference. Victoria, Australia. Web: www.acoustics.asn.au

17-19 December: Symposium on the Acoustics of Poro-Elastic Materials (SEPAM). Bradford, UK. Web: http://sapem2008.matelys.com

2009

05-08 April: Noise and Vibration: Emerging Methods (NOVEM2009). Oxford, UK. Web: www.isvr.soton.ac.uk/NOVEM2009

13-17 April: 2nd International Conference on Shallow Water Acoustics. Shanghai, China. Web: www.apl.washington.edu

19-24 April. International Conference on Acoustics, Speech, and Signal Processing. Taipei, R.O.C. Web: icassp09.com

18-22 May. 157th Meeting of the Acoustical Society of America, Portland, Oregon, USA. Web: www.asa.aip.org

05-09 July: 16th International Congress on Sound and Vibration (ICSV16). Krakow, Poland. Web: www.icsv16.org

23-26 August: Internoise 2009, Ottawa, Canada.

23-27 August: International Confress on Acoustics 2010. Sydney, Australia. Web: www.acoustics.asn.au

06-10 September: Interspeech 2009. Brighton, UK. Web: www.interspeech2009.org

19-23 September: IEEE 2009 Ultrasonics Symposium. Rome, Italy. E-mail: pappalar@uniroma3.it

26-28 October: Euronoise 2009. Edinburgh, UK. Web: www.euronoise2009.org.uk

CONFÉRENCES

Si vous avez des nouvelles à nous communiquer, envoyez-les par courrier ou fax (coordonnées incluses à l'envers de la page couverture), ou par courriel à stevenb@aciacoustical.com

2008

6-8 octobre. Canadian Acoustical Association Acoustics Week in Canada. Vancouver, BC. Web: www.caa-aca.ca

14-15 octobre: Underwater Noise Measurement, Impact and Mitigation. Southampton, U.K. Web: http://underwaternoise2008.lboro.ac.uk

21-23 octobre. 13th Conference on Low Frequency Noise and Vibration. Tokyo, Japan. Web: www.lowfrequency2008.org

21-24 octobre. Acustica 2008. Coimbra, Portugal. Web: www.spacustica.pt

26-29 Octobre: Internoise 2008, Shanghai, China. Web: www.internoise2008.org

01-05 novembre. IEEE International Ultrasonic Symposium. Beijing, China. Web: www.ieee-uffa.org/ulmain.asp?page=symposia

05-07 novembre: Iberoamerican Acoustics Congress (FIA 2008). Buenos Aires, Argentina. Web: www.fia2008.com.ar

10-14 novembre. 156th Meeting of the Acoustical Society of America, Miami, Florida, USA. Web: www.asa.aip.org

24-26 novembre. Australian Acoustical Society National Conference. Victoria, Australia. Web: www.acoustics.asn.au

17-19 decembre: Symposium on the Acoustics of Poro-Elastic Materials (SEPAM). Bradford, UK. Web: http://sapem2008.matelys.com

2009

05-08 avril: Noise and Vibration: Emerging Methods (NOVEM2009). Oxford, UK. Web: www.isvr.soton.ac.uk/NOVEM2009

13-17 avril: 2nd International Conference on Shallow Water Acoustics. Shanghai, China. Web: www.apl.washington.edu

19-24 avril. International Conference on Acoustics, Speech, and Signal Processing. Taipei, R.O.C. Web: icassp09.com

18-22 mai. 157th Meeting of the Acoustical Society of America, Portland, Oregon, USA. Web: www.asa.aip.org

05-09 juillet: 16th International Congress on Sound and Vibration (ICSV16). Krakow, Poland. Web: www.icsv16.org

23-26 août: Internoise 2009, Ottawa, Canada.

23-27 août: International Confress sur Acoustics 2010. Sydney, Australia. Web: www.acoustics.asn.au

06-10 septembre: Interspeech 2009. Brighton, UK. Web: www.interspeech2009.org

19-23 septembre: IEEE 2009 Ultrasonics Symposium. Rome, Italy. E-mail: pappalar@uniroma3.it

26-28 octobre: Euronoise 2009. Edinburgh, UK. Web: www.euronoise2009.org.uk

2010

19-24 March. International Conference on Acoustics, Speech, and Signal Processing. Dallas, TX, USA. Web: icassp2010.org

23-27 August: International Confress on Acoustics 2010. Sydney, Australia. Web: www.acoustics.asn.au

26-30 September: Interspeech 2010. Makuhari, Japan. Web: www.interspeech2010.org

11-14 October: IEEE 2010 Ultrasonics Symposium. San Diego, California, USA. E-Mail: b.potter@vectron.com

2010

19-24 mars. International Conference on Acoustics, Speech, and Signal Processing. Dallas, TX, USA. Web: icassp2010.org

23-27 août: International Confress sur Acoustics 2010. Sydney, Australia. Web: www.acoustics.asn.au

26-30 septembre: Interspeech 2010. Makuhari, Japan. Web: www.interspeech2010.org

11-14 octobre: IEEE 2010 Ultrasonics Symposium. San Diego, California, USA. E-Mail: b.potter@vectron.com

NEWS

We want to hear from you! If you have any news items related to the Canadian Acoustical Association, please send them. Job promotions, recognition of service, interesting projects, recent research, etc. are what make this section interesting.

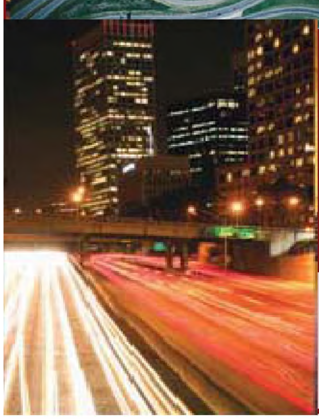
Dalimar Instruments Inc.

À votre service / at your service depuis / since 1986

831 sound level meter/real time analyzer

- Consulting engineers
- Environmental noise monitoring
- Highway & plant perimeter noise
- Aircraft noise
- General Surveys
- Community noise



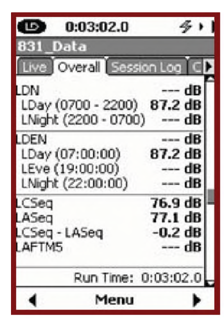



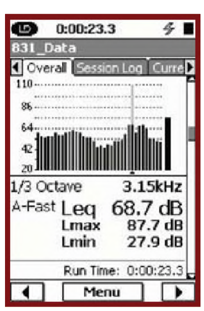
FEATURES

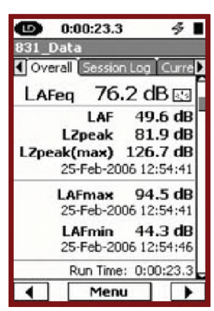
- Class 1/Type 1 sound level meter
- Small size with large display. Ergonomic
- User friendly operator interface
- 120MB standard memory expandable up to 2GB
- Single measurement range from 20 to 140 dB SPL
- Up to 16 hours of battery life
- Provided with utility software for instrument set-up and data download
- Field upgradeable
- AUX port for connection to USB mass storage & cellular modems


MEASUREMENT CAPABILITIES

- Real time 1/1 & 1/3 octave frequency analysis
- Simultaneous display of several noise measurements—ANY DATA (Leq, Lmax, Spectra, etc
- Automatic logging of user selectable noise measurements (Leq, Lmax, Spectra, etc…)
- Exceedance logging with user selectable trigger levels
- Audio and voice recording with replay







Larson Davis
A PCB Group Co.

WWW.DALIMAR.CA INFO@DALIMAR.CA
QUEBEC 450-424-0033 ONTARIO 905-707-9000 ALBERTA 403-288-4416

INSTRUCTIONS TO AUTHORS FOR THE PREPARATION OF MANUSCRIPTS

Submissions: The original manuscript and two copies should be sent to the Editor-in-Chief.

General Presentation: Papers should be submitted in camera-ready format. Paper size 8.5" x 11". If you have access to a word processor, copy as closely as possible the format of the articles in Canadian Acoustics 18(4) 1990. All text in Times-Roman 10 pt font, with single (12 pt) spacing. Main body of text in two columns separated by 0.25". One line space between paragraphs.

Margins: Top - title page: 1.25"; other pages, 0.75"; bottom, 1" minimum; sides, 0.75".

Title: Bold, 14 pt with 14 pt spacing, upper case, centered.

Authors/addresses: Names and full mailing addresses, 10 pt with single (12 pt) spacing, upper and lower case, centered. Names in bold text.

Abstracts: English and French versions. Headings, 12 pt bold, upper case, centered. Indent text 0.5" on both sides.

Headings: Headings to be in 12 pt bold, Times-Roman font. Number at the left margin and indent text 0.5". Main headings, numbered as 1, 2, 3, ... to be in upper case. Sub-headings numbered as 1.1, 1.2, 1.3, ... in upper and lower case. Sub-sub-headings not numbered, in upper and lower case, underlined.

Equations: Minimize. Place in text if short. Numbered.

Figures/Tables: Keep small. Insert in text at top or bottom of page. Name as "Figure 1, 2, ..." Caption in 9 pt with single (12 pt) spacing. Leave 0.5" between text.

Line Widths: Line widths in techincal drawings, figures and tables should be a minimum of 0.5 pt.

Photographs: Submit original glossy, black and white photograph.

Scans: Should be between 225 dpi and 300 dpi. Scan: Line art as bitmap tiffs; Black and white as grayscale tiffs and colour as CMYK tiffs;

References: Cite in text and list at end in any consistent format, 9 pt with single (12 pt) spacing.

Page numbers: In light pencil at the bottom of each page.
Reprints: Can be ordered at time of acceptance of paper.

DIRECTIVES A L'INTENTION DES AUTEURS PREPARATION DES MANUSCRITS

Soumissions: Le manuscrit original ainsi que deux copies doivent être soumis au rédacteur-en-chef.

Présentation générale: Le manuscript doit comprendre le collage. Dimensions des pages, 8.5" x 11". Si vous avez accès à un système de traitement de texte, dans la mesure du possible, suivre le format des articles dans l'Acoustique Canadienne 18(4) 1990. Tout le texte doit être en caractères Times-Roman, 10 pt et à simple (12 pt) interligne. Le texte principal doit être en deux colonnes séparées d'un espace de 0.25". Les paragraphes sont séparés d'un espace d'une ligne.

Marges: Dans le haut - page titre, 1.25"; autres pages, 0.75"; dans le bas, 1" minimum; latérales, 0.75".

Titre du manuscrit: 14 pt à 14 pt interligne, lettres majuscules, caractères gras. Centré.

Auteurs/adresses: Noms et adresses postales. Lettres majuscules et minuscules, 10 pt à simple (12 pt) interligne. Centré. Les noms doivent être en caractères gras.

Sommaire: En versions anglaise et française. Titre en 12 pt, lettres majuscules, caractères gras, centré. Paragraphe 0.5" en alinéa de la marge, des 2 cotés.

Titres des sections: Tous en caractères gras, 12 pt, Times-Roman. Premiers titres: numéroter 1, 2, 3, ..., en lettres majuscules; sous-titres: numéroter 1.1, 1.2, 1.3, ..., en lettres majuscules et minuscules; sous-sous-titres: ne pas numéroter, en lettres majuscules et minuscules et soulignés.

Equations: Les minimiser. Les insérer dans le texte si elles sont courtes. Les numéroter.

Figures/Tableaux: De petites tailles. Les insérer dans le texte dans le haut ou dans le bas de la page. Les nommer "Figure 1, 2, 3,..." Légende en 9 pt à simple (12 pt) interligne. Laisser un espace de 0.5" entre le texte.

Largeur Des Traits: La largeur des traits sur les schémas technique doivent être au minimum de 0.5 pt pour permettre une bonne reproduction.

Photographies: Soumettre la photographie originale sur papier glacé, noir et blanc.

Figures Scanées: Doivent être au minimum de 225 dpi et au maximum de 300 dpi. Les schémas doivent être scannés en bitmaps tif format. Les photos noir et blanc doivent être scannées en échelle de gris tifs et toutes les phoots couleurs doivent être scannées en CMYK tifs.

Références: Les citer dans le texte et en faire la liste à la fin du document, en format uniforme, 9 pt à simple (12 pt) interligne.

Pagination: Au crayon pâle, au bas de chaque page.
Tirés-à-part: Ils peuvent être commandés au moment de l'acceptation du manuscrit.

The Canadian Acoustical Association
l'Association Canadienne d'Acoustique

Application for Membership

CAA membership is open to all individuals who have an interest in acoustics. Annual dues total $65.00 for individual members and $25.00 for Student members. This includes a subscription to ***Canadian Acoustics***, the Association's journal, which is published 4 times/year. New membership applications received before August 31 will be applied to the current year and include that year's back issues of ***Canadian Acoustics***, if available. New membership applications received after August 31 will be applied to the next year.

Subscriptions to *Canadian Acoustics* *or* Sustaining Subscriptions

Subscriptions to ***Canadian Acoustics*** are available to companies and institutions at the institutional subscription price of $65.00. Many companies and institutions prefer to be a Sustaining Subscriber, paying $300.00 per year, in order to assist CAA financially. A list of Sustaining Subscribers is published in each issue of ***Canadian Acoustics***. Subscriptions for the current calendar year are due by January 31. New subscriptions received before August 31 will be applied to the current year and include that year's back issues of ***Canadian Acoustics***, if available.

Please note that electronic forms can be downloaded from the CAA Website at **caa-aca.ca**

Address for subscription / membership correspondence:

Name / Organization ______________________________

Address ______________________________

City/Province ______________ Postal Code _________ Country ______________

Phone ______________ Fax ______________ E-mail ______________

Address for mailing Canadian Acoustics, if different from above:

Name / Organization ______________________________

Address ______________________________

City/Province ______________ Postal Code _________ Country ______________

Areas of Interest: (Please mark 3 maximum)

1. Architectural Acoustics
2. Engineering Acoustics / Noise Control
3. Physical Acoustics / Ultrasound
4. Musical Acoustics / Electro-acoustics
5. Psychological / Physiological Acoustic
6. Shock and Vibration
7. Hearing Sciences
8. Speech Sciences
9. Underwater Acoustics
10. Signal Processing / Numerical Methods
11. Other

For student membership, please also provide:

_________ _________ _________ _________

(University) (Faculty Member) (Signature of Faculty Member) (Date)

I have enclosed the indicated payment for:
[ ] CAA Membership $ 65.00
[ ] CAA Student Membership $ 25.00

[ ] Institutional Subscription $ 65.00
plus mailing surcharge outside Canada:
[ ] $8 to USA, [ ] $15 other International

[ ] Sustaining Subscriber $ 300.00
includes subscription (4 issues /year)
to ***Canadian Acoustics.***

Payment by: [ ] Cheque
[ ] Money Order
[ ] VISA credit card *(Only VISA accepted)*

For payment by VISA credit card:

Card number ______________________

Name of cardholder ______________________

Expiry date ______________________

____

(Signature) (Date)

Mail application and attached payment to:

D. Quirt, Secretary, Canadian Acoustical Association, PO Box 74068, Ottawa, Ontario, K1M 2H9, Canada

l'Association Canadienne d'Acoustique
The Canadian Acoustical Association

Formulaire d'adhésion

L'adhésion à l'ACA est ouverte à tous ceux qui s'intéressent à l'acoustique. La cotisation annuelle est de 65.00$ pour les membres individuels, et de 25.00$ pour les étudiants. Tous les membres reçoivent ***l'Acoustique Canadienne***, la revue de l'association. Les nouveaux abonnements reçus avant le 31 août s'appliquent à l'année courante et incluent les anciens numéros (non-épuisés) de ***l'Acoustique Canadienne*** de cette année. Les nouveaux abonnements reçus après le 31 août s'appliquent à l'année suivante.

Abonnement pour la revue *Acoustique Canadienne* et abonnement de soutien

Les abonnements pour la revue ***Acoustique Canadienne*** sont disponibles pour les compagnies et autres établissements au coût annuel de 65.00$. Des compagnies et établissements préfèrent souvent la cotisation de membre bienfaiteur, de 300.00$ par année, pour assister financièrement l'ACA. La liste des membres bienfaiteurs est publiée dans chaque issue de la revue ***Acoustique Canadienne***. Les nouveaux abonnements reçus avant le 31 août s'appliquent à l'année courante et incluent les anciens numéros (non-épuisés) de ***l'Acoustique Canadienne*** de cette année. Les nouveaux abonnements reçus après le 31 août s'appliquent à l'année suivante.

Pour obtenir des formulaires electroniques, visitez le site Web: **caa-aca.ca**

Pour correspondence administative et financière:

Nom / Organisation ____

Adresse ____

Ville/Province ____________________ Code postal __________ Pays ____________________

Téléphone ______________ Téléc. ____________ Courriel ____________________

Adresse postale pour la revue Acoustique Canadienne

Nom / Organisation ____

Adresse ____

Ville/Province ____________________ Code postal __________ Pays ____________________

Cocher vos champs d'intérêt: (maximum 3)

1. Acoustique architecturale
2. Génie acoustique / Contrôle du bruit
3. Acoustique physique / Ultrasons
4. Acoustique musicale / Electro-acoustique
5. Physio / Psycho-acoustique
6. Chocs et vibrations
7. Audition
8. Parole
9. Acoustique sous-marine
10. Traitement des signaux /Méthodes numériques
11. Autre

Prière de remplir pour les étudiants et étudiantes:

__________ ______________ ________________ __________

(Université) (Nom d'un membre du corps professoral) (Signature du membre du corps professoral) (Date)

Cocher la case appropriée:
[ ] Membre individuel 65.00 $
[ ] Membre étudiant(e) 25.00 $
[ ] Abonnement institutionnel 65.00 $
Surtaxe d'envoi à l'extérieur du Canada :
[ ] 8 $ vers les États-Unis
[ ] 15 $ tout autre envoi international
[ ] Abonnement de soutien 300.00 $
(comprend l'abonnement à *L'acoustique Canadienne*)

Méthode de paiement:
[ ] Chèque au nom de l'Association Canadienne d'Acoustique
[ ] Mandat postal
[ ] VISA *(Seulement VISA)*

Pour carte VISA: Carte n° ____________________

Nom ____________________

Date d'expiration ____________________

____

(Signature) (Date)

Prière d'attacher votre paiement au formulaire d'adhésion. Envoyer à :
Secrétaire exécutif, Association Canadienne d'Acoustique, Casier Postal 74068, Ottawa, K1M 2H9, Canada

SUSTAINING SUBSCRIBERS / ABONNES DE SOUTIEN

The Canadian Acoustical Association gratefully acknowledges the financial assistance of the Sustaining Subscribers listed below. Their annual donations (of $300.00 or more) enable the journal to be distributed to all at a reasonable cost.

L'Association Canadienne d'Acoustique tient à témoigner sa reconnaissance à l'égard de ses Abonnés de Soutien en publiant ci-dessous leur nom et leur adresse. En amortissant les coûts de publication et de distribution, les dons annuels (de $300.00 et plus) rendent le journal accessible à tous nos membres.

ACI Acoustical Consultants Inc.
Mr. Steven Bilawchuk - (780) 414-6373
stevenb@aciacoustical.com - Edmonton, AB

ACO Pacific Inc.
Mr. Noland Lewis - (650) 595-8588
acopac@acopacific.com - Belmont, CA

Acoustec Inc.
Dr. J.G. Migneron - (418) 834-1414
courrier@acoustec.qc.ca - St-Nicolas, QC

Aercoustics Engineering Ltd
Mr. John O'Keefe - (416) 249-3361
aercoustics@aercoustics.com - Rexdale, ON

Bruel & Kjaer North America Inc.
Mr. Andrew Khoury - (514) 695-8225
andrew.khoury@bksv.com - Pointe-Claire, QC

Dalimar Instruments Inc.
Mr. Daniel Larose - (514) 424-0033
daniel@dalimar.ca - Vaudreuil-Dorion, QC

Earth Tech Canada Inc.
Frank Babic - (905) 747-7411
frank.babic@earthtech.ca - Markham, ON

Eckel Industries of Canada Ltd.
Mr. Bruce Allan - (613) 543-2967
eckel@eckel.ca - Morrisburg, ON

ECORE International
Mr. Paul Downey - (416) 440-1094
pcd@ecoreintl.com - Toronto, ON

H.L. Blachford Ltd.
Mr. Dalton Prince - (905) 823-3200
amsales@blachford.ca - Mississauga, ON

Hatch Associates Ltd.
Mr. Tim Kelsall - (905) 403-3932
tkelsall@hatch.ca - Mississauga, ON

HGC Engineering Ltd.
Mr. Bill Gastmeier - (905) 826-4044
info@hgcengineering.com - Mississauga, ON

Hydro-Quebec
M. Blaise Gosselin - (514) 840-3000x5134
gosselin.blaise@hydro.qc.ca - Montréal, QC

Integral DX Engineering Ltd.
Mr. Greg Clunis - (613) 761-1565
greg@integraldxengineering.ca - Ottawa, ON

J.E. Coulter Associates Ltd.
Mr. John Coulter - (416) 502-8598
jcoulter@on.aibn.com - Toronto, ON

J.L.Richards & Assoc. Ltd.
Mr. Terry Vivyurka, P.Eng. - (613) 728-3571
mail@jlrichards.ca - Ottawa, ON

Jade Acoustics Inc.
Ms. Dalila Giusti - (905) 660-2444
dalila@jadeacoustics.com - Concord, ON

JASCO Research Ltd.
Mr. Scott Carr - (902) 405-3336
scott@jasco.com - Halifax, NS

Mc SQUARED System Design Group
Mr. Wade McGregor - (604) 986-8181
info@mcsquared.com - North Vancouver, BC

MJM Conseillers en Acoustique Inc.
MJM Acoustical Consultants Inc.
M. Michel Morin- (514) 737-9811
mmorin@mjm.qc.ca - Montréal, QC

Novel Dynamics Test Inc.
Mr. Andy Metelka - (519) 853-4495
ametelka@cogeco.ca - Acton, ON

Owens-Corning Canada Inc.
Mr. Salvatore Ciarlo - (800) 988-5269
salvatore.ciarlo@owenscorning.com -
St.Leonard, QC

OZA Inspections Ltd.
Mr. David Williams - (800) 664-8263x25
oza@ozagroup.com - Grimsby, ON

Peutz & Associés
M. Marc Asselineau +33 1 45230500
m.asselineau@peutz.fr
Paris, FRANCE

Michel Picard
(514) 343-7617; FAX: (514) 343-2115
michel.picard@umontreal.ca - Brossard, QC

Pinchin Environmental Ltd.
Ms. Robin Connley - (905) 363-0678
info@pinchin.com - Mississauga, ON

Pyrok Inc.
Mr. Howard Podolsky - (914) 777-7770
info@pyrokinc.com - Mamaroneck, NY

RWDI AIR Inc.
(519) 823-1311; FAX: (519) 823-1316
peter.vandelden@rwdi.com - Guelph, ON

Scantek Inc.
(410)-290-7726; FAX: (410) 290-9167
peppinr@scantekinc.com - Columbia, MD

SILEX Innovations Inc.
Mr. Mehmood Ahmed - (905) 612-4000
mehmooda@silex.com - Mississauga, ON

SNC/Lavalin Environment Inc.
M. Jean-Luc Allard, (450) 651-6710
jeanluc.allard@snclavalin.com - Longueuil, QC

Soft dB Inc.
M. André L'Espérance - (418) 686-0993
contact@softdb.com, Sillery, QC

SounDivide Inc.
C.W. Ray Bakker - (877) 816-5435
ray.bakker@SounDivide.com - Calgary, AB

Spaarg Engineering Ltd.
Dr. Robert Gaspar - (519) 972-0677
gasparr@kelcom.igs.net - Windsor, ON

State of the Art Acoustik Inc.
Dr. C. Fortier - (613) 745-2003
sota@sota.ca - Ottawa, ON

Swallow Acoustic Consultants Ltd.
Mr. John Swallow - (905) 271-7888
jswallow@jsal.ca - Mississauga, ON

Tacet Engineering Ltd.
Dr. M.P. Sacks - (416) 782-0298,
mal.sacks@tacet.ca - Toronto, ON

Valcoustics Canada Ltd.
Dr. Al Lightstone - (905) 764-5223
solutions@valcoustics.com
Richmond Hill, ON

Vibro-Acoustics
Mr. Tim Charlton - (800) 565-8401
tcharlton@vibro-acoustics.com
Scarborough, ON

Wakefield Acoustics Ltd.
Mr. Clair Wakefield - (250) 370-9302
clair@wakefieldacoustics.com - Victoria, BC

West Caldwell Calibration Labs
Mr. Stanley Christopher - (905) 595-1107
info@wccl.com - Brampton, ON

Wilrep Ltd.
Mr. Don Wilkinson - (905) 625-8944
info@wilrep.com - Mississauga, ON