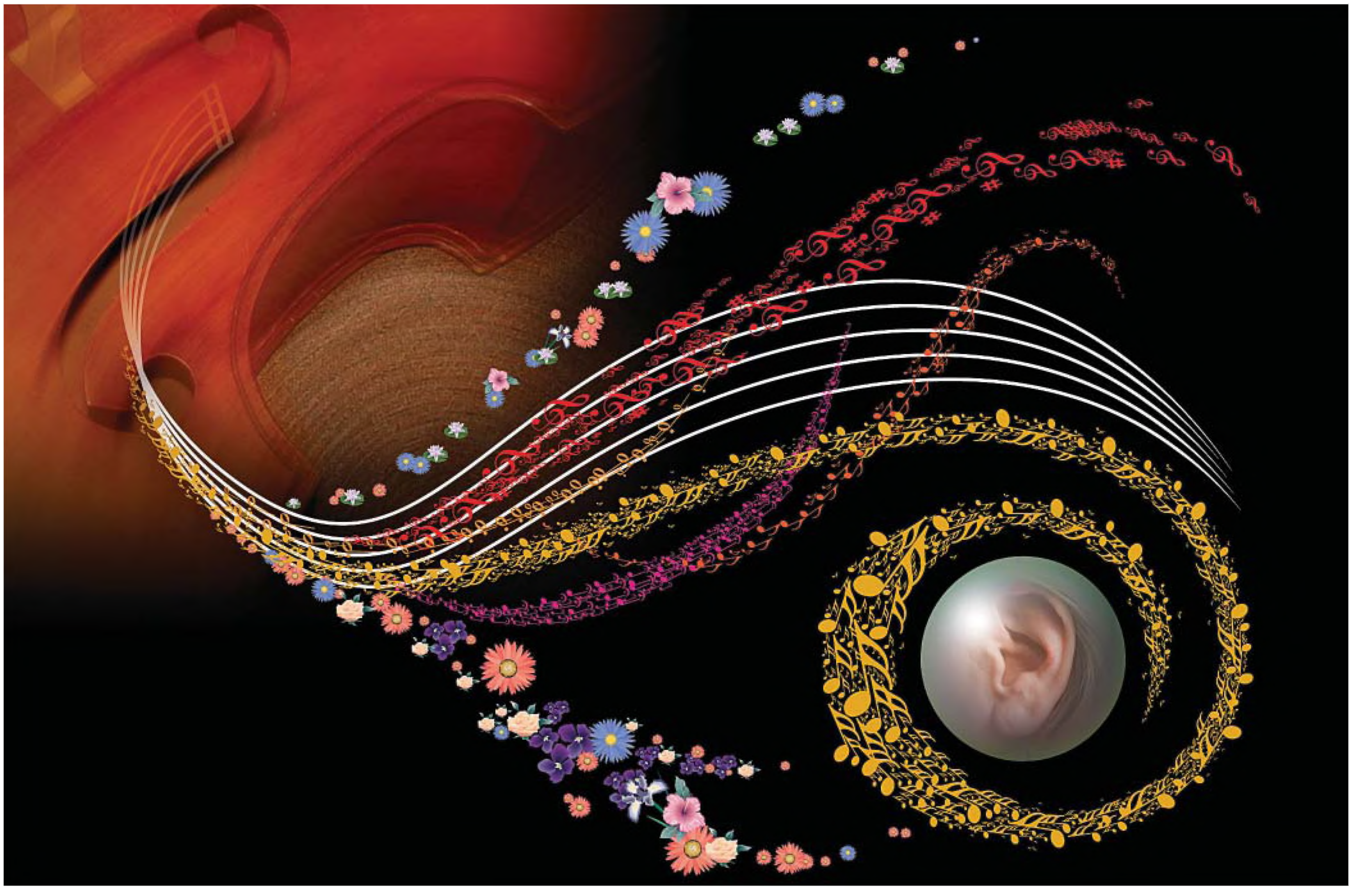
ISSN 0711-6659

canadian acoustics
acoustique canadienne

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

MARCH 2009 MARS 2009
Volume 37 -- Number 1 Volume 37 -- Numéro 1



canadian acoustics

THE CANADIAN ACOUSTICAL ASSOCIATION
P.O. BOX 1351, STATION "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.

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: $30 (student); $70 (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

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

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.

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: $30 (étudiant); $70 (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.

GUEST EDITORIAL / ÉDITORIAL INVITÉ

Canadian researchers have been and continue to be at the forefront of research in music cognition. By my count, there are currently 12 labs in Canada that are engaged in some aspect of music cognition, and in any given year, work from these labs accounts for no less than 15% of the peer-reviewed literature in the field. Although the roots of music cognition can be found in well-controlled psychoacoustic paradigms, researchers have increasingly branched out to consider questions that touch on aspects that run closer to our everyday experience with music. Consistent with this trend, there has been a growing interest in the manner by which separate physical dimensions of music including those that span modalities are registered, compared, and in some cases integrated. This special issue features four papers that investigate the multidimensional and multimodal nature of music and related questions of integration.

The first paper, by Hall & Beauchamp, concerns a classic question in music cognition: What are the primary dimensions that determine timbre perception? Their approach, however, is novel focusing on the common listening goal of identification rather than discrimination. Utilizing synthetic timbres made to vary systematically along spectral and temporal continua, they show that the spectral envelope dominates identification judgements and that formant structure has relatively more influence than does spectral centroid. The second paper, by Schutz and Kubovy, presents a convincing albeit surprising example of multimodal integration. The study demonstrates that perception of tone duration can be dominated by visual information even when that visual information is highly degraded. This finding is surprising because audition is normally thought of as the dominant modality for temporal discrimination. This finding also represents a formidable challenge for the optimal integration theory, a widely accepted account of multimodal integration. The third paper, by Ceaser, Thompson & Russo, examines auditory and visual influences on the perception of tonal closure. The authors provide evidence that tonal closure can be perceived on the basis of visual cues alone, and in some cases, on the basis of expressive non-pitch acoustic features. The final paper, by Prince, Schmuckler & Thompson, examines perception of cross-modal similarity. The study demonstrates that observers are sensitive to similarities between contours presented as melodies and contours presented as line drawings. Similarity judgements appear to be based on surface structure similarity as well as in deep structure similarity obtained by something akin to Fourier analysis.

Collectively, the research described in this issue suggests that although it is possible to attend in an analytic manner to the isolated dimensions of music, our typical experience tends to be much more holistic, involving registration and comparison of dimensions that sometimes span modalities. This research adds to a growing body of evidence concerning the pervasiveness of multidimensional and multimodal integration in music. In some cases, a subset of dimensions appears

Les chercheurs canadiens continuent d'être aujourd'hui au premier plan de la recherche en cognition musicale. D'après mon dénombrement, il y a présentement 12 laboratoires au Canada qui se consacrent à certains aspects de la cognition musicale, et l'ouvrage émis annuellement par ces laboratoires constitue non moins de 15% des articles évalués par les pairs dans ce domaine. Malgré le fait que les origines de la cognition musicale se trouvent dans les paradigmes de recherche psychoacoustique contrôlée, les chercheurs se sont éloignés progressivement de ces paradigmes afin de considérer des questions portant sur des aspects qui se rapprochent à notre expérience musicale quotidienne. Conformément à ces changements, il y eu aussi de plus en plus d'intérêts à la manière dans laquelle les différentes dimensions physiques de la musique incluant celles de toutes modalités sont perçues, comparées, et dans certains cas intégrées. Cette édition spéciale présente quatre articles qui examinent les propriétés multidimensionnelles et multimodales de la musique ainsi que certaines questions portant à l'intégration.

Le premier article, par Hall et Beauchamp, adresse une question classique dans le domaine de la cognition musicale : Quelles sont les dimensions principales qui déterminent la perception du timbre? Leur approche, cependant, est une concentration atypique sur l'objectif d'écoute commun d'identification au lieu de celui de discrimination. En utilisant des timbres synthétiques variant systématiquement selon leur emplacement sur les continuums spectral et temporel, ils démontrent que les jugements d'identification sont principalement basés sur l'enveloppe spectrale et que la structure du formant a relativement moins d'influence sur ces jugements que les centroïdes spectraux. Le deuxième article, par Schultz et Kubovy, présente un exemple convaincant, bien que surprenant, d'intégration multimodal. Leur étude démontre que la perception de la durée d'un son musical peut être principalement influencée par l'information visuelle, même quand cette information visuelle est très dégradée. Ce résultat est surprenant puisque l'audition est normalement conçue comme étant la modalité dominante pour la discrimination temporelle. Ce résultat représente aussi un défi redoutable pour la théorie d'intégration optimale, une théorie sur l'intégration multimodale couramment acceptée. Le troisième article, par Ceaser, Thompson et Russo, examine les influences auditives et visuelles sur la perception de la conclusion tonale. Les résultats provenant de cette étude suggèrent que la conclusion tonale peut être perçue à l'aide d'indicateurs visuelles seulement, et dans certains cas à l'aide de caractéristiques acoustique expressives qui ne sont pas reliées à la hauteur du son. Le dernier article, par Prince, Schmuckler et Thompson, examine la perception de similarité entre différentes modalités. Cette étude démontre que les observateurs sont sensibles aux similarités entre les contours présentés sous forme de mélodies et les contours présentés sous forme de dessins au trait. Il semblerait que les jugements de similarités sont basés sur la similarité de la to overwhelm the others, whereas in other cases the perceptual system appears to prefer a compromise. Quite possibly, music may be a particularly apt stimulus for observing multidimensional and multimodal integration because of its inherent subjectivity and its structural complexity.

Although the four papers in this issue are limited to consideration of music presented through audio and/or visual modalities, I would be remiss to not make some mention of the notion that important information about music can also be conveyed through vestibular and vibrotactile input. Vestibular information is readily available through dance, a close cousin of music, and vibrotactile information is available in many acoustic instruments (e.g., flute) as a bi-product of performance. Recent work in my lab has shown that when vestibular and vibrotactile information are lawfully related to an incoming musical acoustic signal, they can have influence over auditory judgements of music. These novel examples of multimodal integration should provide unique opportunities for future research, including the exploration of tri- and even quad-modal integration (auditory, visual, vestibular and vibrotactile).

Frank Russo
Department of Psychology, Ryerson University

NOTE: We'd like to acknowledge the assistance of Stéphanie Marion for the French translations of the Editorial and Abstracts in this issue.

structure superficielle ainsi que la similarité de la structure profonde obtenue par une procédure ressemblant une analyse de Fourier.

Collectivement, la recherche détaillée dans cette édition suggère que malgré la possibilité de capter de manière analytique les dimensions individuelles de la musique, notre expérience habituelle a tendance à être beaucoup plus holistique, impliquant la perception et la comparaison de dimensions sur une modalité ou plus. Cette recherche contribue à la masse croissante de preuves concernant la prédominance de l'intégration multidimensionnelle et multimodale de la musique. Dans certains cas, un sous-ensemble de dimensions semblent accabler les autres, tandis que dans d'autres cas le système perceptif semble préférer un compromis. Il est bien possible que la musique soit un stimulus particulièrement approprié pour observer l'intégration multidimensionnelle et multimodal grâce à sa subjectivité et complexité structurelle.

Bien que ces quatre articles présentés dans cette édition soient limités à la considération de la musique présentée à travers les modalités auditive et visuelle, je m'en voudrais de ne pas mentionner l'idée que des informations importantes au sujet de la musique peuvent aussi être transmises à travers les indications vestibulaires et vibrotactiles. L'information vestibulaire est facilement accessible à travers la dance, une cousine de la musique, et l'information vibrotactile est accessible comme sous-produit de la performance de plusieurs instruments acoustique (ex : la flûte). Des recherches récemment entreprises dans mon laboratoire indiquent que lorsque les informations vestibulaires et vibrotactiles sont rigoureusement associées à un signal acoustique musical, ils peuvent influencer les jugements auditifs de la musique. Ces nouveaux exemples de l'intégration multimodal devraient fournir de nouvelles occasions de recherche uniques, incluant l'exploration de l'intégration de trois ou même quatre modalités (auditive, visuelle, vestibulaire et vibrotactile).

Frank Russo
Département de Psychologie, Ryerson University

WHAT'S NEW in Canada ??

QUOI DE NEUF en Canada??

Promotions
Deaths
New jobs
Moves

Retirements
Degrees awarded
Distinctions
Other news

Promotions
Décès
Offre d'emploi
Déménagements

Retraites
Obtention de diplômes
Distinctions
Autres nouvelles

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

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

Jeremie Voix - Email: voix@caa-aca.ca

CLARIFYING SPECTRAL AND TEMPORAL DIMENSIONS OF MUSICAL INSTRUMENT TIMBRE

Michael D. Hall[1], James W. Beauchamp[2]
[1]Department of Psychology, MSC 7704, James Madison University
Harrisonburg, VA 22807, U.S.A., hallmd@jmu.edu
[2]School of Music and Dept. of Electrical & Computer Engineering, University of Illinois at Urbana-Champaign, Illinois 61801, U.S.A., jwbeauch@uiuc.edu

ABSTRACT

Classic studies based on multi-dimensional scaling of dissimilarity judgments, and on discrimination, for musical instrument sounds have provided converging support for the importance of relatively static, spectral cues to timbre (e.g., energy in the higher harmonics, which has been associated with perceived brightness), as well as dynamic, temporal cues (e.g., rise time, associated with perceived abruptness). Comparatively few studies have evaluated the effects of acoustic attributes on instrument identification, despite the fact that timbre recognition is an important listening goal. To assess the nature, and salience, of these cues to timbre recognition, two experiments were designed to compare discrimination and identification performance for resynthesized tones that systematically varied spectral and temporal parameters between settings for two natural instruments. Stimuli in the first experiment consisted of various combinations of spectral envelopes (manipulating the relative amplitudes of harmonics) and amplitude-vs.-time envelopes (including rise times). Listeners were most sensitive to spectral changes in both discrimination and identification tasks. Only extreme amplitude envelopes impacted performance, suggesting a binary feature based on abruptness of the attack. The second experiment sought to clarify the spectral dimension. Listener sensitivity was compared for a) modifications of spectral envelope shape via variation of formant structure and b) spectral changes that minimally impact envelope shape (using low-pass filters to match the centroids of the formant-varied envelopes). Only differences in formant structure were easily discriminated and contributed strongly to identification. Thus, it appears that listeners primarily identify timbres according to spectral envelope shape. Implications for models of instrument timbre are discussed.

RESUME

Plusieurs études classiques se servant d'une mise en échelle multidimensionnelle pour mesurer des jugements de dissemblance, et de discrimination, pour des sons d'instruments de musique, stipulent un appui concourant à l'importance d'indicateurs de timbre spectrales qui sont relativement statiques (ex., l'énergie dans les harmoniques de haute fréquence, qui a été associée à la brillance perçue), ainsi que d'indicateurs temporels dynamiques (ex., le temps de montée d'une salve , qui est associé à la soudaineté perçue). Comparativement moins d'études ont évaluées les effets de caractéristiques acoustiques sur l'identification d'instruments, malgré le fait que la reconnaissance du timbre est un objectif important de l'écoute. Pour évaluer la nature, et la prédominance, de ces indicateurs de reconnaissance de timbre, deux expériences ont été conçues pour comparer la performance de l'identification et de la discrimination de sons musicaux resynthétisés, dans lesquelles des paramètres spectraux et temporels entre ajustements ont été modifiés systématiquement pour deux instruments naturels. Les stimuli utilisés dans la première expérience étaient composés de plusieurs combinaisons d'enveloppes spectrales (en manipulant les amplitudes relatives des harmoniques) et d'enveloppes d'amplitudes-vs-temps (incluant les temps de montée de slaves). La sensibilité des auditeurs la plus élevée était celle envers les changements spectraux, et ce pour les tâches de discrimination et d'identification. Seules les enveloppes d'amplitude extrêmes ont influencées la performance, ce qui suggère une caractéristique binaire basée sur la soudaineté de l'attaque. La deuxième expérience avait comme objectif de clarifier la dimension spectrale. La sensibilité des auditeurs a été évaluée contre 1) les modifications de forme des enveloppes spectrales par la modification de la structure du formant et 2) les changements spectraux qui ont un effet minime sur la forme de l'enveloppe (en utilisant des filtres passe-bas pour accorder les centroïdes des enveloppes à formants variés). Seules les différences dans la structure du formant étaient facilement discriminées et contribuaient considérablement à l'identification. Ainsi, il paraît que les auditeurs identifient principalement le timbre selon la forme de l'enveloppe spectral. Les implications des résultats pour des modèles de timbre d'instruments sont discutées.

1. INTRODUCTION

Timbre can be thought of as the collection of perceived qualities that help to identify a given sound source. Thus, timbre is what distinguishes a particular person's voice, or what makes a musical instrument, such as a piano, sound like itself. Traditionally, timbre has been "defined" by the exclusion of properties. For example, the American Standards Association (1960) defines timbre as "that attribute of auditory sensation in terms of which a listener can judge that two sounds similarly presented and having the same loudness and pitch are dissimilar".

Since this definition's inception, considerable gains have been made in research to define timbre by inclusion, that is, by what it is rather than what it is not. Yet, the admittedly inadequate definition by exclusion is still often used today. This is due in large part to the fact that timbre has been repeatedly demonstrated to be multi-dimensional. Much of this evidence has come from research with sounds produced by musical instruments, including multi-dimensional scaling (MDS) of timbre dissimilarity judgments (e.g., Caclin et al. 2005; Grey, 1977; Krumhansl, 1989; Miller and Carterette, 1975; Winsberg et al. 1995), as well as from discrimination tasks in response to manipulations of one or more potentially relevant acoustic dimensions (e.g., Grey and Moorer, 1977; McAdams et al. 1999).

By comparison, relatively little research has been aimed towards the study of timbre identification, even though such a task most closely resembles a typical perceiver goal — to identify the sound source. The current investigation consists of experiments that directly compare data from a timbre identification task with discrimination performance in order to further evaluate the relative contributions of some acoustic dimensions that have been argued to be critical to musical instrument timbre. Prior to summarizing these experiments, a brief overview of some classic methodologies and their corresponding findings will be provided, including the relevant perceptual dimensions that have been identified by that research. Within this overview, potential limitations of traditional methodologies will be discussed, and further justification that timbre identification data is necessary for a more thorough evaluation of the relative importance of timbre dimensions will be given.

1.1. Evidence for critical dimensions of timbre

In typical MDS evaluations of instrument timbre, listeners are presented with all possible instrument pairings from a limited set of stimuli. Their task is to provide estimates of the perceptual distance between each stimuli-pair by rating them on a scale of dissimilarity (usually from *1*-7). All ratings are then submitted to a computer model [e.g., INDSCAL (see Carroll and Chang, 1970), CLASCAL (Winsberg and De Soete, 1993), or CONSCAL (Winsberg and De Soete, 1997)] in order to generate a best-fitting model of perceptual (timbre) space using a minimum number of dimensions. The addition of a given perceptual dimension will significantly increase the proportion of variance in ratings data that is explained by the MDS solution to the degree that the dimension was relied upon by listeners. As a result, MDS is typically argued to reveal the relative strength of contributions from each dimension (e.g., Miller and Carterette, 1975; also see Caclin, et al. 2005).

Probably the most important step in the MDS process comes last, when researchers attempt to correlate the resulting positions of the stimuli along each perceptual axis of the timbre space with changes of a particular physical measure. High correlations between acoustic and perceptual ratings/dimensions are used to infer that a particular source of acoustic variation was likely used by listeners to distinguish between timbres. Using this approach, several acoustic parameters have repeatedly been shown to be closely related to dimensions in timbre scaling solutions. These commonly include not only relatively static characteristics of the spectral envelope, but also more dynamic/temporal attributes, as shown in Krumhansl's (1989) 3-dimensional timbre space.

The primary aspect of the spectral envelope that strongly correlates with dimensions in MDS solutions, as in Ehresman and Wessel (1978), is the spectral centroid, formed by weighting the harmonic frequencies of a musical tone by the amplitudes corresponding to them and by normalizing and adding the resulting values. Thus,

$$f_c = \frac{\sum_{k=1}^{K} f_k A_k}{\sum_{k=1}^{K} A_k}, \qquad [1]$$

where f_c is the spectral centroid (in Hz), k is harmonic number, f_k is harmonic frequency, K is the number of harmonics, and A_k is the amplitude of the k^{th} harmonic. A similar dimension was obtained in the seminal work of Grey (1977), who found a correlation with the distribution of spectral energy as relatively narrow (reflecting a concentration of low-frequency energy) or broad (reflecting contributions from higher harmonics). Insofar as this measure reflects the relative presence of high- or low-frequency energy in the signal, it is frequently argued that the corresponding dimension of a timbre space reflects the perception of brightness (i.e., a tone is perceived as brighter given more high-frequency energy, or less low-frequency energy).

A solution by Krimphoff et al. (1994) also identified a temporal dimension, rise time, which can be defined as the time interval from tone onset to when the most intense portion of the tone is reached. Rise times are typically calculated as the time difference between where a priori criterion values are first reached for very low and high percentages of the signal's maximum amplitude (in the current study, between 10 and 90 percent of peak amplitude). Strong correlations with both the spectral centroid and (log) rise time were later confirmed by Winsberg et al. (1995) using a subset of a stimulus set devised by Krumhansl (1989) and was also used by Krimphoff et al. (1994).

The importance of both spectral and temporal properties has been confirmed by MDS procedures involving stimulus sets that reflect direct manipulations of potentially relevant acoustic characteristics. For instance, timbre space has been shown to be systematically altered when spectral envelopes have been exchanged across instruments, and dimensions from the resulting solutions still correlated with changes in spectral energy distribution (Grey and Gordon, 1978). Artificial spectral manipulations (number of harmonics) and temporal manipulations (rise time conveyed by linear ramps of amplitude) also have been found to correlate with separate dimensions in MDS solutions, thereby presumably indicating separate spectral and temporal contributions to timbre (Samson, Zatorre, and Ramsay, 1997). Furthermore, distortions of these MDS solutions in listeners with right temporal lobe lesions suggest involvement of those brain regions in spectral processing (Samson, Zatorre, and Ramsay, 2002). MDS has been coupled with direct manipulation of stimulus values to confirm the importance of the spectral centroid and attack time as relevant dimensions (Caclin, et al. 2005).

Combined spectrotemporal properties of timbre also have been proposed from MDS solutions. One commonly proposed spectrotemporal dimension (e.g., see Krumhansl, 1989) is spectral flux, which characterizes variation in the shape of the spectral envelope over time. A similar dimension also was described by Grey's (1977) timbre space for attack transients. However, recent MDS evidence has not provided strong support for the dimension of spectral flux, and other alternative dimensions have been proposed. Krimphoff et al. (1994) demonstrated that the dimension originally identified by Krumhansl (1989) as corresponding to spectral flux was better predicted by spectral irregularity, a measure of the relative jaggedness of the spectrum. Caclin, et al. (2005) demonstrated a reduced contribution of spectral flux with increases in the number of concurrently manipulated dimensions. They also suggested that the relative amplitude of even- to odd-numbered harmonics was an important factor in spectral envelope shape. Further evidence for even-odd importance was reported by Beauchamp et al. (2006). Using a set of both sustained and percussion instruments which were normalized with respect to spectral centroid and rise time, Beauchamp and Lakatos (2002) attempted to correlate some of these measures with MDS solutions.

Discrimination performance can additionally indicate what minimal acoustic manipulations are audible, and therefore, which dimensions could potentially constitute a basic feature of timbre. If a given simplification of an instrument tone is easily discriminated from the original or resynthesized version of that tone, then this would indicate that a relevant dimension of timbre was affected. For example, Grey and Moorer (1977) revealed the relevance of attack transients by obtaining evidence of very accurate discrimination of tones from versions of those tones with their attack removed (e.g., Grey and Moorer, 1977). This approach also has been used to identify several potentially relevant, spectrotemporal properties in tones where spectral flux was controlled. These include spectral envelope irregularity (i.e., the relative jaggedness of spectrum, possibly due to a relative emphasis on odd-numbered harmonics), and particularly, amplitude envelope incoherence, the degree to which each harmonic has a unique amplitude envelope shape. For example, McAdams, et al. (1999) showed, using a discrimination method, that both spectral incoherence and spectral irregularity are important for musical instrument tone perception. It should be noted that amplitude envelope incoherence (also known as spectral incoherence) is equivalent to spectral flux (fluctuation in spectral envelope shape) as defined by Krumhansl (1989).

1.2. Need for timbre identification research and the utility of timbre interpolation

While discrimination, and particularly, MDS, procedures have been very informative at revealing several potentially critical dimensions of timbre, the information gathered from any single task in necessarily limited. In this case, data from either of these tasks must be combined with timbre identification data in order to permit strong conclusions about acoustic parameters that listeners utilize for instrument recognition. Although the MDS approach can reveal strong correlations between acoustic parameters and salient dimensions, such correlations do not prove causal relationships. It is possible that other acoustic parameters will better correlate with the MDS dimensions, and that these parameters more closely model what listeners rely on in making their responses. This limitation was indicated effectively by Caclin, et al. (2005), who noted that

> "MDS studies are thus presumed to highlight the most perceptually salient timbre parameters that are likely to be of importance in a variety of situations (voice recognition, music listening). Nevertheless they have a common drawback: given the multiplicity of acoustical parameters that could be proposed to explain perceptual dimensions, one can never be sure that the selected parameters do not merely covary with the true underlying parameters." (p. 472)

Identification performance also need not be directly predicted by discrimination performance, as previously indicated by McAdams (2001):

> "The extent to which an event can be simplified without affecting identification performance is the extent to which the information is not used by the listener in the identification process, even if it is discriminable." (p. 161)

Thus, it is possible that discrimination of variation along one or more acoustic dimensions could be very accurate, and yet this variation might not be sufficiently large for listeners to perceive a change in instrument category. After all, listeners frequently claim to readily perceive differences in timbre between examples of a particular instrument. For example, electric guitarists often discuss their individual preferences for an instrument's signature sound, such as the twang of a *Fender Stratocaster* or the warmth of a large *Gretsch* hollowbody. Furthermore, differences between sets of instruments given supra-threshold levels of stimulus variation could reflect learning of cues to specific timbres. In other words, it is possible that a particular stimulus parameter might prove to be characteristic to a particular instrument (e.g., the presence of inharmonic energy in a piano tone), or instruments, and yet, not particularly informative for others.

Despite the importance of identification data for gaining a more complete understanding of timbre, comparatively few research studies have focused on timbre identification. Most of these studies were early investigations that highlighted the importance of information in the attack to timbre recognition by either eliminating or altering attack transients. For example, Saldanha and Corso (1964) showed that timbre identification performance is reduced for tones whose attack transients have been deleted. A similar reduction/alteration in identification performance was demonstrated in response to swapping of attack transients across instruments (Thayer, 1974).

Many of the existing studies of instrument timbre that rely upon identification tasks (and also, frequently, studies that have used discrimination and MDS procedures) involve a stimulus set that includes some form of synthesized interpolation between natural instrument timbres. Since interpolation requires systematic control of potentially relevant physical parameters, inclusion of interpolated tones should help the researcher in evaluating the relative contribution of those parameters to task performance, and thus, presumably, timbre recognition. Anecdotally, timbre interpolation has had a long history. Any imitation of one instrument or voice by another instrument or voice can be considered timbre interpolation in that aspects of a target instrument are superimposed on a source instrument. Generally this means either using unorthodox manipulation of an instrument's excitation (e.g., vocal folds, reed vibration) or its body resonances (e.g., vocal track or violin body resonances).

To these authors' knowledge, the first instance of systematic timbre interpolation using a computer was accomplished by John Grey as reported in his PhD dissertation (1975, pp. 75-95). In his method the time-varying amplitudes of the individual harmonics were cross-faded between two instruments before resynthesis. The method was tantamount to cross-fading between two signals except for two differences: 1) harmonic phases were aligned and frequencies were flattened; 2) segments before and after amplitude maxima occuring in the endpoint spectra were aligned and interpolated to produce smooth transitions. A series of tones were presented to subjects where the crossfade parameter gradually changed the timbres from a source to a target. Conclusions were that there is a strong hysteresis effect according to the direction of transition but that there is no sharp boundary for identification of which of the two instruments was heard.

Recently, more advanced timbre interpolation has been described by Haken, et al. (2007). In order to align times between eight sounds, a time dilation function is defined for each sound which reveals when prominent points in the sound's structure occur. Then a time envelope is defined which interpolates amongst the eight sounds' time dilations, and this in turn is applied to functions governing amplitudes, frequencies, and noises of each harmonic of each sound which in turn are combined (mixed) to form the additive synthesis control functions prior to final resynthesis.

The current investigation was motivated by an interest in using timbre interpolation to assess the relative contributions of spectral and temporal properties to the ability to identify musical instrument sounds. Such an assessment requires data from a timbre identification task that can be compared with data from procedures that have traditionally been used to evaluate timbre dimensions (e.g., discrimination or MDS). Also required is a direct manipulation of spectral and temporal parameters while excluding any spectrotemporal variation (i.e., spectral flux) that could interact with the perceptual dimensions of interest. Note that these parameters are really vectors, in that their definitions generally require many numerical values.

The current investigation was intended to provide such an assessment. We restricted our focus to manipulating only the most commonly identified timbral parameters — that is, spectral envelope (epitomized by the single-valued spectral centroid) and amplitude-vs.-time envelope (epitomized by the single-valued rise time). (For the sake of brevity, "amplitude-vs.-time envelope" will henceforth be referred to as "amplitude envelope".) Manipulation of each parameter was accomplished by synthesizing a set of hybrid stimuli whose parameters were interpolated between those of two instruments, namely an A_4 violin and an A_4 trombone, whose spectral and temporal properties differed considerably. The nature and relative salience of these two parameters were evaluated in an experiment (Experiment 1) that compared discrimination and timbre identification performance for the hybrid stimuli. A follow-up experiment (Experiment 2) was designed to further clarify whether listeners were using the complex spectral envelope or merely the spectral centroid for instrument recognition.

2. EXPERIMENT 1: SPECTRAL ENVELOPES V. AMPLITUDE ENVELOPES

Experiment 1 was designed to evaluate the relative contribution of a static property, the spectral envelope (epitomized by spectral centroid), and a dynamic property, the amplitude envelope (epitomized by rise time), to timbre identification and discrimination. Sixteen hybrid tones interpolated between two reference tones, a violin and a trombone, both pitched at A_4, were generated for this purpose. Each tone was constructed by amplitude-

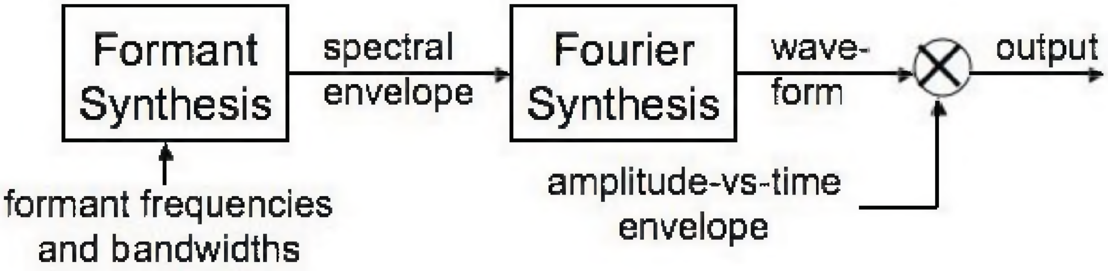


Figure 1. Conceptual diagram of the synthesis procedure in Experiment 1 and their products.

modulating a static waveform formed by interpolation between the reference tone spectra with an amplitude envelope interpolated between those of the reference tones. Based on existing timbre literature (e.g., MDS (Grey, 1977; McAdams et al. 1995; Caclin et al. 2005), discrimination (McAdams et al. 1999)), it was expected that both spectral and temporal parameters would contribute to identification and discrimination. Whether one or the other parameter impacted task performance to a significantly greater degree was left as an open question for the research to address.

2.1. Participants

Seven students from James Madison University participated in exchange for extra credit counted towards a psychology course. All listeners were between 18 and 40 years of age, and none reported having any known hearing deficits.

In neither experiment were participants screened for performance training on a musical instrument. However, some information was collected about extent and type of training via questionnaire. The participants in Experiment 1 could generally be characterized as having some musical experience. Participants had a mean of 3.2 years of performance training on a musical instrument (with a standard error of 0.9 years) and a range of 0 to 6 years. Two listeners had no musical training at all; the remaining 5 listeners had some formal musical training. No listener had previous performance experience with one of the target instruments from the experiment. Only 1 participant was occasionally playing an instrument at the time of testing; no other participant had played an instrument in the preceding 4 years.

2.2. Stimuli

Sixteen stimuli of 500 ms duration were generated that orthogonally combined 4 levels of spectral envelope with 4 levels of amplitude envelope. This was accomplished using a source-filter synthesis method that involved a series of operations. Figure 1 shows a conceptual equivalent block diagram for this process. In reality the Praat program was used to process a sawtooth waveform by a bank of band-pass filters to yield a waveform whose harmonic amplitudes form a static spectral envelope. The block "Formant Synthesis" in Figure 1 represents this process. An important aspect of the spectral envelope is the formants that correspond to the filter resonances. Another important aspect is the general negative slope of the spectral envelope that corresponds to the inverse frequency characteristic of the sawtooth waveform. (Similar kinds of formant-based estimations of spectral envelopes have been previously applied successfully to musical instrument tones; for a detailed summary of the general utility of such estimations, including a detailed description of several related synthesis methods, see Rodet and Schwarz (2007).) The waveform which corresponded to the spectral envelope was then multiplied by an amplitude contour to impose an instrument-specific (or hybrid) amplitude envelope on the waveform, and thereby obtain the stimulus (labeled "output"). This approach to synthesis, in which a time-varying amplitude envelope was supplied separately from a static spectral envelope, eliminated all inharmonicity, fundamental frequency (F_0) deviations, and spectral flux, thus allowing a direct assessment of the relative perceptual contributions of the spectral and temporal parameters.

To aid in understanding the synthesis model for this experiment, let the resynthesized violin tone (henceforth called *Vn*) be characterized by its amplitude envelope $A_V(t)$ and spectral envelope $S_V(f)$. Likewise, let the resynthesized trombone tone (henceforth called *Tr*) be characterized by its amplitude envelope $A_T(t)$ and spectral envelope function $S_T(f)$. Then the interpolated signal is given by

$$s(t) = \{\alpha A_V(t) + (1-\alpha) A_T(t)\} \times \sum_{k=1}^{K} A(\beta, S_V(kf_0), S_T(kf_0)) \cos(2\pi k f_0 t + \theta_k), \quad [2]$$

where t = time, k = harmonic number, K = number of harmonics, α = interpolation value for amplitude, β = interpolation value for the spectral envelope, f_0 = fundamental frequency, and θ_k = phase of harmonic k. Note that the time-varying amplitude in front of the summation sign does not depend on frequency. Likewise, the $A()$ function weights in front of the cos functions (that give the harmonic sinusoidal variations) do not depend on time, but rather only on the frequency of the corresponding harmonic.

The original violin and trombone tones, both pitched at A_4 (440 Hz), were taken from McGill University Master Samples (MUMS) library (Opolko and Wapnick, 1987). These instruments were selected to share a similar total number and distribution of harmonics, while differing significantly with respect to both spectral centroid and rise

	Vn	Vn Hybrid	Tr Hybrid	Tr
Rise Time (ms)	404	398	335	236
Centroid (Hz)	1082	984	953	922
F1	492	600	715	839
F2	2159	1899	1660	1441
F3	3651	3205	2802	2438
F4	4457	4767	5094	5440
F5	7006	7177	7352	7531
F6	8465	8833	9216	9613
B1	277	311	347	383
B2	193	400	643	928
B3	222	507	857	1290
B4	953	1694	2717	4128
B5	744	823	906	993
B6	955	944	932	921

Table 1. Formant center frequencies (F) for each stimulus in Experiment 1 (Vn = violin; Tr = trombone) and their bandwidths (B). Also provided are corresponding measures of spectral centroid for each spectral envelope, as well as measured rise time values for each amplitude envelope.

time. The open string production of the violin tone was selected to eliminate vibrato that was otherwise present throughout the chromatic series of recordings. Measured envelope values for the violin tone and trombone tone represented endpoints along the temporal and spectral dimensions.

The computer program Praat was used to determine the average spectra of the natural violin and trombone tones, as well as to construct all stimuli/spectra for our experiment (Boersma and Weenink, 2007). The source-filter synthesis model in Praat that was used is similar to that described for speech by Klatt and Klatt (1990); a vibrational source (sawtooth wave) was submitted to a bank of band-pass filters specifying six formants. Formants were derived from an LPC analysis (extended over an 11,025 Hz range, with a .05 s analysis window length, and based upon a 30 dB range for each measured formant). Artificial spectra for *Vn* and *Tr* were determined by combining the formants whose mean center frequencies and corresponding average bandwidths were measured over the initial 500 ms of each tone's steady-state (i.e., immediately after the tone's peak amplitude, reflecting completion of its attack).

Formant center frequencies and bandwidths for *Vn* and *Tr* (as well as for hybrid stimuli) are provided in Table 1. As Table 1 reveals, extremely wide bandwidths were measured and synthesized for the mid-frequency range of the trombone tone. These very broad spectral peaks reflect the smooth regions that naturally occur in instrument spectra. Mean measures of spectral centroid for each stimulus also are included in Table 1. Previous MDS studies (e.g., Krumhansl, 1989; McAdams et al. 1995) have established that the range of spectral centroid variation across these instruments is moderate relative to other pairs of instruments.

Formant frequencies and bandwidths for two hybrid spectral envelopes were chosen to be in-between the natural instrument (endpoint) values and were spaced for equal perceptual distance (with respect to frequency discrimination) from each endpoint stimulus. The frequency spacings were obtained by a log/Mel scale transformation of the formant center frequencies and their corresponding bandwidths according to the approximation proposed by Fant (1973),

$$M = 1000 \log(1 + f/1000)/\log(2), \qquad [3]$$

where M is mels and f equals frequency in *Hz*. Hybrid values along each dimension that were closest to the natural violin and trombone values will henceforth be referred to as *Vn Hybrid and Tr Hybrid*, respectively (see Table 1 for the center frequencies and bandwidths of their formants). Spectral envelopes for the four stimuli, *Vn*, Vn *Hybrid*, *Tr Hybrid*, and *Tr*, are shown in Figure 2.

Filtered stimuli representing each level of spectral envelope were multiplied separately by each level of amplitude envelope in Praat to complete synthesis of the stimulus set. Amplitude envelope estimates for the attack transients of the violin and trombone tones were determined from measurements in Praat of waveform maxima (in relative dB) taken every 2 ms over the first 500 ms of the original tone production. Hybrid values were obtained by interpolation to form equal steps between violin and trombone amplitudes. Additionally, amplitude was down-ramped linearly over the final 20 ms of each tone to avoid the perception of abrupt offsets. A depiction of each amplitude envelope is provided in Figure 3. Measures of rise time, the time interval from tone onset during which the signal moved from 10 to 90 percent of its peak amplitude, also are included in Table 1. Rise times of the endpoint stimuli reveal that the stimulus set reflected a reasonably broad range (236 – 404 ms) along this characteristic.[1]

Several aspects of the stimuli and stimulus presentation were shared with Experiment 2. All stimuli were 500 ms in duration, and were synthesized with a 44.1 kHz (16-bit) sampling rate. Furthermore, all tones were presented through a low-pass (Butterworth) anti-aliasing filter with a cutoff frequency of 11 kHz, and the peak sound level (to the nearest dB) of the presented stimuli was 80 dB[A]. Both experiments were conducted in a quiet room, and all stimuli were delivered over Sennheiser HD 280 earphones.

2.3. Procedure

Listeners completed two tasks—a timbre identification task and a tone discrimination task. These tasks were counterbalanced across participants, and participants were afforded a brief rest break in between the two tasks. A few procedures were shared across tasks, as well as with Experiment 2. First, stimulus delivery and the collection of responses across tasks were controlled by Music Experiment Development System software (v. 2002-B-1; Kendall, 2002). Also, within each task, a 500 ms inter-trial interval followed any given response prior to stimulus delivery for the next trial.

Discrimination task

In the discrimination task listeners were instructed to respond whether the two tones presented on each trial were either physically identical (i.e., "same") or different. On any given trial either the *Vn* tone or the *Tr* tone was used as a "standard" stimulus ($p = 0.5$ for each instrument). The other stimulus on each trial was either identical to the standard ($p = 0.5$ across trials), or alternatively, a different stimulus. As in the discrimination task of the subsequent experiment, a 250 ms inter-stimulus interval separated the pair of tones on each discrimination trial.

To enable an evaluation of the relative contribution of the spectral and amplitude envelope parameters to timbre, stimulus comparisons on different trials could involve a manipulation of either parameter in isolation, or alternatively, both parameters together. Discrimination would be expected to improve with increasing distance along a given perceptual dimension. Therefore, step size was manipulated across trials. Included were *1-step comparisons* (i.e., *Vn⇔VnHybrid*, *Tr⇔TrHybrid*), *2-step* comparisons (i.e., *Vn⇔TrHybrid*, *Tr⇔VnHybrid*), and *3-step* comparisons (*Vn⇔Tr*) along each dimension, or combination of dimensions. There were 17 such "different" pairs of stimuli.

Participants used a laptop keyboard to indicate their response on each trial. Listeners were instructed to press the ***1*** key if the two tone stimuli on a trial were perceived to be identical or the ***3*** key if the tone stimuli on a trial were perceived to be different.

A brief familiarization period preceded the discrimination task. During this familiarization, listeners were presented with the 14 tone stimuli (once in random order) that they would subsequently hear during discrimination trials so that they would have a clear sense of the range of stimulus differences that they would be exposed to during the task. As with each subsequently described familiarization procedure, a 1-sec inter-trial-interval separated each tone. No responses were made during familiarization. Listeners could request to repeat the familiarization sequence as needed to feel comfortable with the differences between stimuli before proceeding with the task; no such request was made.

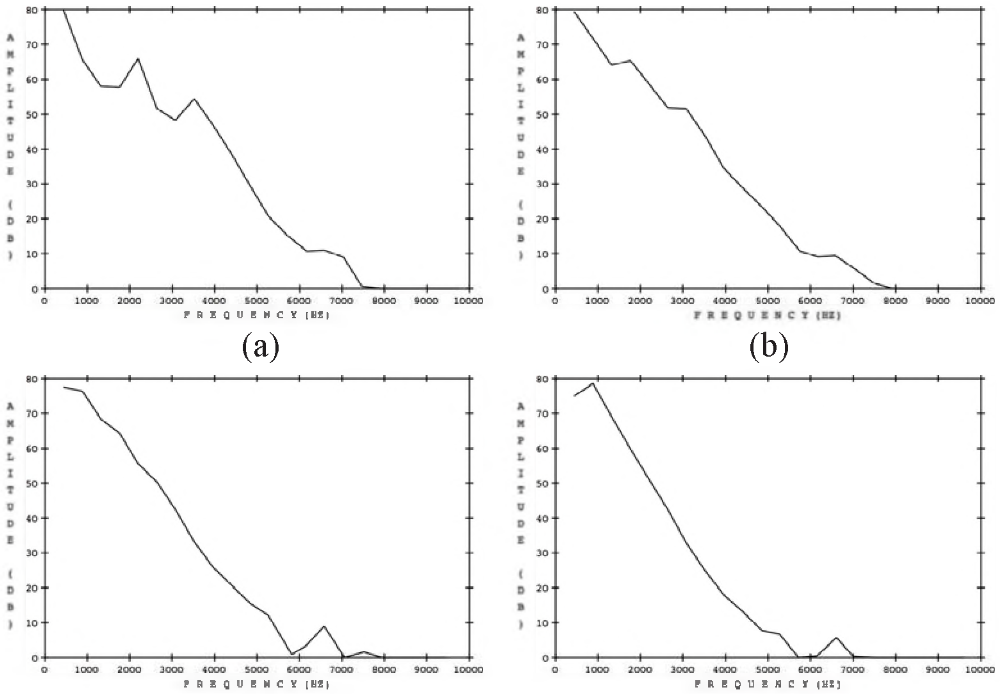
The familiarization period was immediately followed by two blocks of 272 randomized discrimination trials (i.e., 544 total trials). Each block of trials consisted of 8 repetitions for each of the 17 pairs of different stimuli (4 repetitions for each ordering of the standard and comparison tones). For the "same" trials, which were half of the trials within each block, *Vn* was presented on an equivalent number of trials as the *Tr* stimulus.

Timbre identification

In the timbre identification task listeners were instructed to indicate whether the tone they heard on a given trial corresponded to a violin, a tenor trombone, or the timbre of a different instrument. The participants also were informed that the tones that they would be hearing had been resynthesized based on naturally occurring parameters for the violin and tenor trombone, and that they may sound quite artificial as a result of being simplified in several ways. They were told that some trials would contain a tone that was a simplified version of a natural instrument tone (either *Vn* or *Tr*), and that other trials would contain a hybrid tone that resulted from a combination of attributes that differed from those of a natural instrument. Participants were asked simply to categorize the timbre of each instrument tone to the best of their ability.

Responses on each trial were made by using the computer's mouse to click on a small bitmap image on the laptop screen corresponding to the perceived instrument. Each image consisted of a verbal label for the instrument in white lettering against a blue rectangular background (e.g., *violin* or *trombone*). The background of each image was lighter on the edges so that the collection of images appeared as a series of buttons from a button box. In addition, the response category of *other* was included so that listeners could indicate if any of the stimuli (particularly, the hybrid tones) sounded like they were derived from an instrument other than the violin or tenor trombone.

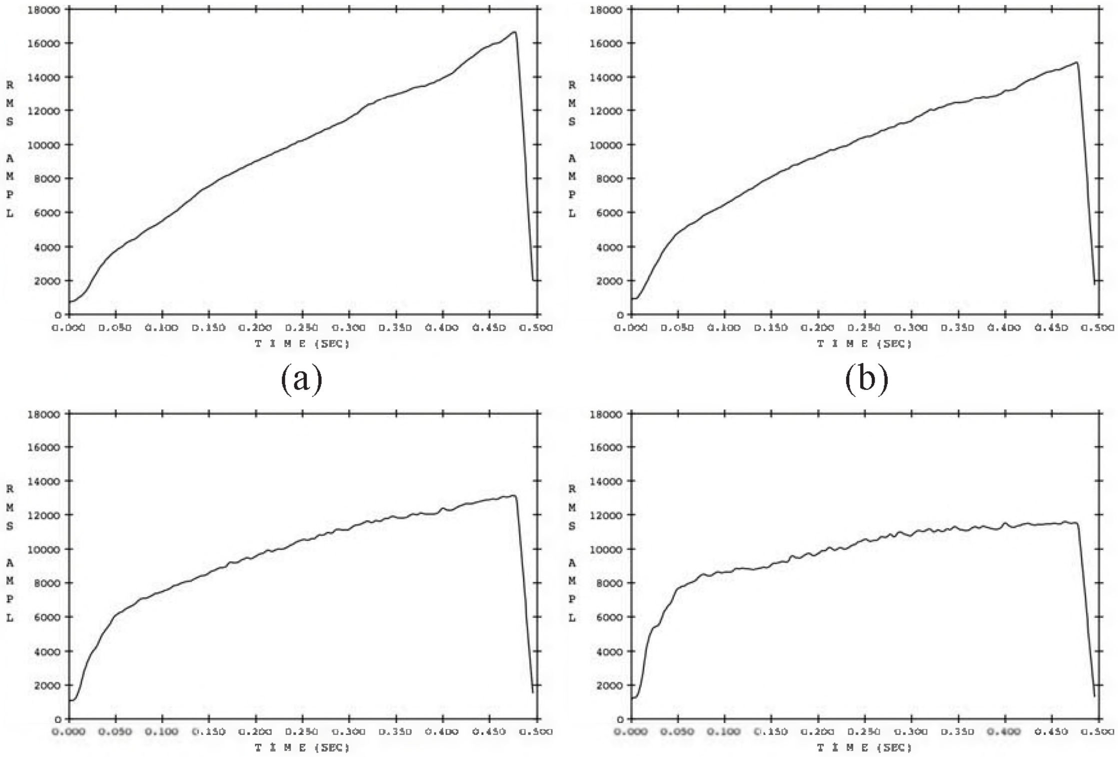
Before proceeding with the task, listeners first were familiarized with the tones that most closely approximated the original violin and trombone tones. Ten tones were presented in non-random order, comprising five repetitions of the *Vn* tone followed by the *Tr* tone. No responses were made during this familiarization period. Rather, participants just closely listened to each sound to get a better sense of the violin-like and trombone-like sounds in the stimulus set. Listeners were permitted to repeat the familiarization procedure again if they felt that they had any trouble recognizing basic timbre differences between the target timbres, but, in fact, no listener requested to repeat the procedure. This familiarization period was immediately followed by a block of 320 randomized experimental trials consisting of 20 repetitions of each tone stimulus.



(a) (b)

(c) (d)

Figure 2. Spectral envelopes used in Experiment 1 to synthesize a violin tone, *Vn* (panel a), a tenor trombone, *Tr* (panel d), and two hybrid tones, *Vn Hybrid* (b) and *Tr Hybrid* (c), based on resonances equally spaced in *Mels* between *Vn*, *Tr*, and each other.



(a) (b)

(c) (d)

Figure 3. Average (RMS) amplitude-v.-time (in seconds, 0 to 0.5 s) displays corresponding to each amplitude envelope in Experiment 1. Envelopes ranged from the onset of a violin tone, *Vn* (panel a), to that of the tenor trombone, *Tr* (panel d), including hybrid envelopes with instantaneous amplitudes that were equally spaced from these natural instrument values and each other, *VnHybrid* (panel b) and *TrHybrid* (panel c).

2.4. Results and Discussion

Discrimination task

Discrimination performance was assessed using d′, a theoretically bias-free measure of sensitivity from Signal Detection Theory. Sensitivity (d′) was calculated according to an Independent Observations Model, which assumes an individual assessment of each stimulus on a trial (see Macmillan and Creelman, 2005). The probability of false alarms was obtained from performance on same trials for the given standard tone (i.e., violin or trombone). The resulting measures of sensitivity were submitted to a 3x2x3 3-way repeated measures ANOVA with timbre dimension(s) (spectral, temporal, both), standard instrument (*Vn*, *Tr*), and step size (1, 2, or 3) as factors. Corresponding mean calculations of d′ (along with standard error bars) are provided in Figure 4 for each timbre dimension (displayed as differently shaded bars), standard stimulus (displayed to the left for *Vn*, and to the right for *Tr*), and step size (varying along the horizontal axis).

As can be seen in Figure 4, listeners were readily able to discriminate changes in spectral envelope, but not amplitude envelope. This corresponded to a main effect of dimension, $F(2,12) = 207.285$, $p = .001$. Post-hoc pair-wise comparisons of means via Tukey HSD tests further revealed that discrimination of different spectral envelopes was significantly better than for amplitude envelopes ($p < .05$), and that no further performance gains were obtained when both dimensions varied across tone stimuli. In fact, ceiling levels of discrimination performance were approached for discrimination of spectral envelope differences. In contrast, d′ never exceeded criterion levels of performance (d′ = 1) for the amplitude envelope dimension.

As expected, discrimination performance also generally improved with increasing distance along perceptual dimensions, as indicated by a significant main effect of step size, $F(2,12)= 9.305$, $p < .01$. The fact that sensitivity was essentially at ceiling across step sizes for comparisons involving the *Vn* standard (see Figure 4, left) also contributed to a standard instrument x step size interaction, $F(2,12) = 8.142$, $p < .01$. No other effects approached significance ($p > .10$).

Identification task

For each listener the probabilities of each type of timbre identification response (i.e., *violin*, *trombone*, and *other*) were calculated as a function of each combination of spectral envelope and amplitude envelope. In order to assess the relative contribution of each parameter to timbre identification, the probabilities of each type of response were submitted to a separate 4x4 2-way repeated measures ANOVA with spectral envelope level and amplitude envelope level as factors.

Mean response probabilities (and standard error bars) across participants are displayed in Figure 5 for the *violin* (displayed to the left) and *trombone* responses (right), with spectral envelopes displayed along the horizontal axis and amplitude envelopes as differently shaded bars. Corresponding means for *other* responses are not shown because they were rarely used, occurring on less than 1.5 percent of trials. Anecdotal reports from several participants indicated that they only felt it necessary to use the *other* response category in instances when they were indecisive about the instrument that produced the perceived timbre. No significant effects were obtained from ANOVA and post-hoc analyses involving *other* responses (all F's < 1).

As can be seen in Figure 5, spectral envelopes strongly affected timbre identification, whereas amplitude envelopes did not. Specifically, the incidence of *violin* responses decreased, and *trombone* responses increased, as spectral envelopes were systematically altered from *Vn* to *Tr* (i.e., from left to right for either side of Figure 5). This trend resulted in a significant main effect of spectral envelope for both types of responses [$F(3,18) = 54.378$, $p < 0.0001$ and $F(3,18) = 59.546$, $p < 0.0001$ for *violin* and *trombon*e responses, respectively]. Post-hoc pair-wise comparisons of means (Tukey HSD tests) further revealed that response probabilities for tones with the *TrHybrid* and *Tr* spectral envelopes did not significantly differ. However, a significant increase in *violin* responses (plus a corresponding decrease in *trombone* responses), was obtained for the *VnHybrid* spectral envelope, and another such increase (or decrease for *trombone* responses) was obtained for the *Vn* spectral envelope ($p < .05$). In contrast, the probabilities of neither *violin* responses nor *trombone* responses significantly changed as a function of amplitude envelope, as indicated by the absence of a main effect of amplitude envelope [$F(3,18) = 1.213$, $p > 0.33$ and $F(3,18) = 1.161$, $p > 0.35$, for *violin* and *trombone* responses, respectively].

Cross-task comparisons

It is clear that listeners in Experiment 1 relied more heavily on the static spectral envelope manipulation than the dynamic amplitude envelope manipulation in both timbre identification and tone discrimination tasks. The fact that amplitude envelope did not contribute significantly to timbre identification appears to be attributable to the fact that the range of variation along that dimension was not really discriminable in the context of roving spectral envelopes across trials. Insofar as a quite broad (70%), naturally occurring range of rise times was used in Experiment 1, it is unlikely that this difference in performance across dimensions is due to reliance on a truncated range of amplitude envelopes. Furthermore, insofar as spectral centroids varied by only 160 Hz (17%), it also seems unlikely that listeners' reliance on spectral envelopes for making judgments was simply due to an exaggerated range of spectral centroid variation. We will elaborate on this argument in the general discussion section.

Based on the findings from Experiment 1, it could potentially be claimed that common spectral aspects of timbre, including perceived brightness due to shifts in the spectral centroid, could frequently be more salient than a

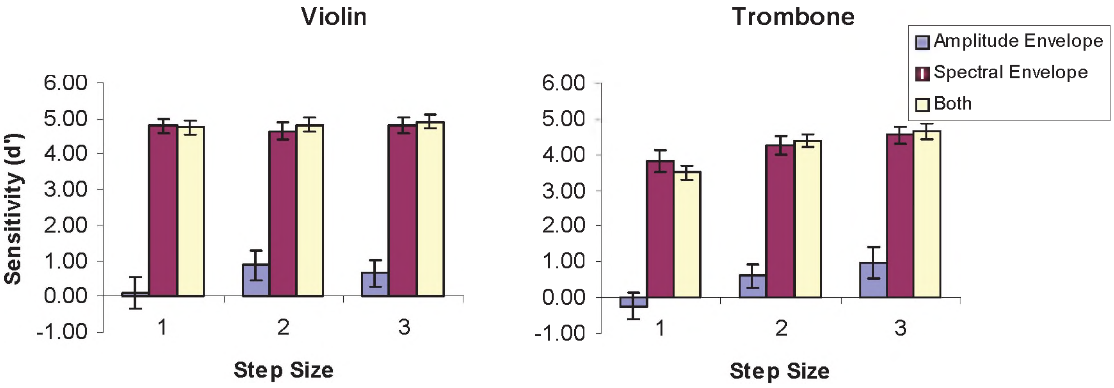


Figure 4. Mean sensitivity and corresponding standard errors for tone discrimination in Experiment 1 as a function of acoustic dimension (temporal, spectral, or both dimensions), standard instrument [violin (left) or trombone (right)] and physical distance (step size along the stimulus continuum).

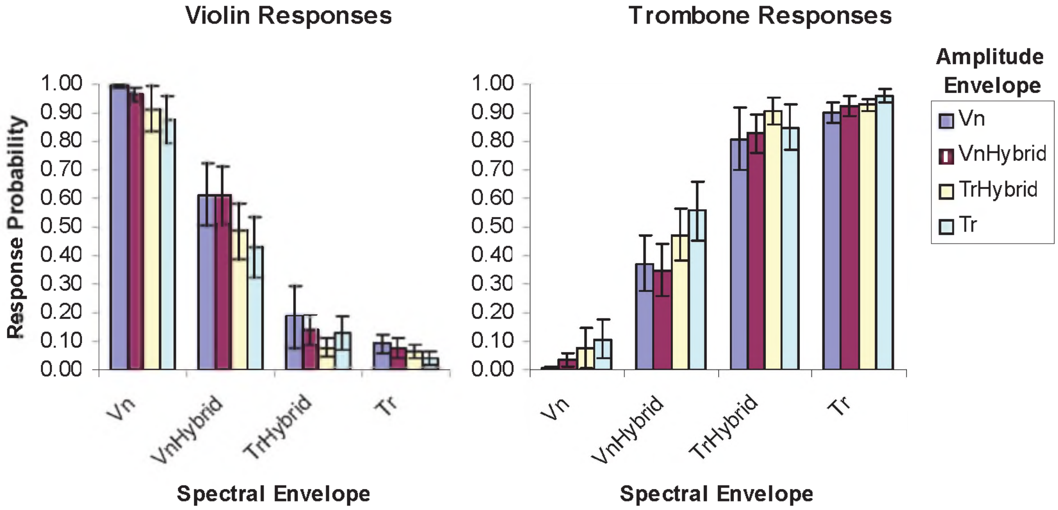


Figure 5. Mean probabilities of ***violin*** responses (left) and ***trombone*** responses (right), along with corresponding standard errors, for timbre identification in Experiment 1. Results are displayed for each combination of spectral envelope (varying along the horizontal axis) and amplitude envelope (distinguished by differently shaded bars).

commonly identified dynamic aspect of timbre, rise time. After all, the entirety of the original attack functions for the violin and trombone were retained within the amplitude envelopes used in this study.

3. EXPERIMENT 2: FORMANT STRUCTURE V. SPECTRAL CENTROID

While it is clear that listeners in Experiment 1 relied on static spectral information more than dynamic amplitude information, the nature of that spectral cue required additional clarification. Even though traditional interpretations of MDS results have suggested that listeners likely respond to differences in perceived brightness, as indicated by the spectral centroid, there is another possibility: detailed formant structure.

Manipulation of the spectral envelope in Experiment 1 was accomplished by shifting spectral peaks between naturally occurring values, thereby creating hybrid envelopes. As a result, shifts in the spectral centroid were confounded with a corresponding shift in formants. Spectral centroids gradually decreased from the *Vn* value (1082) to the *Tr* value (922). Furthermore, the center frequencies for second and third formant also decreased systematically from the *Vn* to the *Tr* spectrum (although it is obvious from Figure 2 that the formants virtually disappear in *Tr*'s spectral envelope). It is therefore possible that listeners responded to the spectral envelope shape, rather than changes in spectral centroid (brightness), in both discrimination and timbre identification tasks.

Distinguishing between the potential contribution of spectral centroid and spectral envelope structure to instrument timbre is complicated by the fact that they are very closely related properties. Raising (or lowering) the center frequency of a formant should increase (or decrease) the spectral centroid, particularly when that formant is high in amplitude to begin with. This close relationship is also seen in the timbre literature, where the same property could be argued to reflect either envelope structure or centroid. For example, one dimension from an MDS solution for a set of FM-synthesized tones was labeled by Krumhansl (1989) as indicating "spectral envelope". However, the same dimension was later found to be more strongly correlated with spectral centroid. This was demonstrated through subsequent acoustic analyses (Krimphoff, 1993; Krimphoff et al. 1994), as well as through additional MDS data involving a large subset of Krumhansl's (1989) stimuli (McAdams et al. 1995; for a summary of findings, see Donnadieu, 2007).

Despite this strong correspondence between variables, there is evidence that general information about the shape of the spectral envelope and the spectral centroid correspond to distinct perceptual properties. For example, discrimination of harmonic series according to differences in spectral slope, a characteristic of natural sources of vibration, is relatively unaffected by changes in the number of spectral peaks, a filter characteristic (see Li and Pastore, 1995). This finding is particularly relevant to the current study insofar as spectral slope manipulations should impact the perceived brightness of tones, whereas peaks in the spectral envelope determine formant structure.

Teasing apart which spectral cue listeners relied upon more heavily in Experiment 1—spectral centroid or spectral envelope/formant structure—constituted the goal of Experiment 2. Addressing this issue required manipulation of the spectral centroid in a manner that was largely independent of formant structure (and thus, minimized its impact on the basic shape of the spectral envelope). This was accomplished by using low-pass filtering to match the spectral centroids of three Experiment 1 tones that had a different formant structure from *Vn*.

There are indications from the literature on timbre recognition by machine that a cepstral coefficient measure (which accounts for formant structure) consistently leads to significantly more accurate classification of orchestral timbres than reliance solely on the spectral centroid (e.g., Brown, Houix and McAdams, 2001). Thus, there were reasons to anticipate that listeners in Experiment 2 would rely more heavily on the shape of spectral envelopes in making their judgments. It was therefore hypothesized that these listeners would exhibit greater sensitivity in timbre discrimination based upon differences in spectral envelope detailed structure rather than simply differences in spectral centroids, and that as a result, timbre identification also would be primarily affected by manipulations of this structure.

3.1. Participants

Eighteen students from introductory psychology courses at James Madison University participated in partial fulfillment of course requirements. All listeners were between 18 and 40 years of age, and none reported having any known hearing deficits.

As with Experiment 1, participants in Experiment 2 typically had not received much prior training on a musical instrument. Participants had a mean of 3.1 years of training (with a standard error of 0.6 years), ranging in experience from 0 to 8 years. Two listeners had no musical training. Of the remaining listeners, 14 had received some formal musical training, with 11 of them receiving 3 years or less. One listener was a former violinist who had not played the instrument for the preceding 6 years. None of the participants had continued practicing their instrument(s) at the time of testing, and only 4 had actively practiced in the past 4 years.

3.2. Stimuli

Seven tone stimuli were used in Experiment 2. All of them shared the violin's amplitude envelope. One of these (stimulus 1) was the *Vn* tone of Experiment 1, from which all of the remaining stimuli were derived. Note that *Vn* combined both the average spectrum of a violin tone and its amplitude envelope during its attack. It had no spectral or frequency variations.

Three other tones (stimuli 2 - 4) that were taken from Experiment 1 provided manipulations of the spectral envelope's formant structure (i.e., *VnHybrid*, *TrHybrid*, and *Tr*). The remaining three tones (stimuli 5 - 7) were produced by submitting *Vn* to a first order (i.e., shallow slope) low-pass filter to match the spectral centroids of stimuli 2 – 4; this was accomplished by setting the filter's cutoff frequencies to 3300, 2640, and 2165 Hz, respectively. In this way, the general shapes of the spectral envelopes for stimuli 5 - 7 were minimally impacted (relative to *Vn*) by the changes in centroid. Henceforth in this experiment, the labels *VnHybrid*, *TrHybrid*, and *Tr* will be used to refer to stimuli with either the corresponding manipulation of spectral centroid via filtering or a particular spectral envelope/formant structure.

3.3. Procedure

Both a timbre identification task and a tone discrimination task were used to assess the relative contribution of each manipulated dimension (formant structure v. spectral centroid alone) to timbre. These tasks were closely modeled after the corresponding tasks in Experiment 1. Unless otherwise noted below, remaining aspects of the procedure, including the timing of stimulus presentation, as well as the means of making responses and collecting data, were as described for that task in Experiment 1.

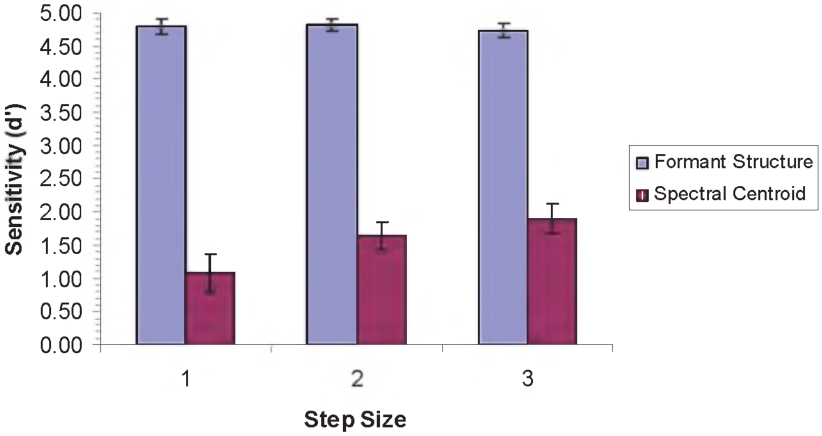


Figure 6. Mean sensitivity and corresponding standard errors for tone discrimination performance in Experiment 2 as a function of acoustic dimension (formant structure v. spectral centroid alone) and physical distance (step size along the stimulus continuum) from the standard stimulus.

Discrimination task

The discrimination task of Experiment 2 was very similar to that of Experiment 1. Listeners again were instructed to respond whether the two tones presented on each trial were either physically identical (i.e., "same") or were different. One fundamental difference in discrimination procedures was that there was only a single standard stimulus in Experiment 2. This standard was the *Vn* tone, which was presented on every trial. The remaining stimulus on any given trial was either a second presentation of the standard ($p = 0.5$) or one of the six alternative tone stimuli. In other words each "different" trial included a manipulation of either formant structure or spectral centroid. There were three step-sizes for both the formant structure and the spectral centroid conditions. Comparison stimuli for either the formant structure or spectral centroid conditions were as follows: *1-step* comparisons *(VnHybrid)*, *2-step* comparisons *(TrHybrid)*, and *3-step* comparisons (*Tr*).

Before proceeding with experimental trials, listeners first were familiarized with the stimuli that would be presented in the task. This was accomplished by randomly presenting each of the seven tones once. The listeners could request repetition of the familiarization sequence until they felt comfortable with the range of perceived differences in timbre. No listener requested that the tones be repeated.

A single block of 240 randomized experimental trials followed the familiarization procedure. Within this block of trials there were 120 "same" trials, in which the standard constituted both stimuli on a trial. The remaining 120 trials were "different" trials, which paired the standard with a different comparison tone. Each of the six comparison tones were provided on twenty trials. On 10 of these 20 trials the standard was presented first; the standard was the second stimulus on the remaining 10 trials.

Identification task

As in Experiment 1, listeners in the timbre identification task of Experiment 2 were instructed to indicate whether the tone they heard on a given trial was that of a violin, a trombone, or a different instrument. Before proceeding with this task, listeners first were familiarized with the tones that most closely approximated the natural violin and trombone timbres. During this familiarization procedure, ten tones were presented in non-random order, including five repetitions of *Vn* followed by *Tr*. No listener requested that the familiarization procedure be repeated before proceeding immediately to experimental trials. A single block of 140 randomized experimental trials was given, including 20 repetitions of each individual of the 7 tone stimuli.

3.4. Results and Discussion

Discrimination task

Sensitivity (d′) for each stimulus condition in the discrimination task was calculated in the manner described for Experiment 1. The resulting measures of sensitivity were submitted to a 2x3 2-way repeated measures ANOVA with manipulated dimension (i.e., formant structure v. spectral centroid alone) and step size (1, 2, or 3) as factors. Corresponding mean calculations of d′ (along with standard error bars) are provided in Figure 6 for each timbre dimension (displayed as differently shaded bars) and step size (varying along the horizontal axis).

As can be seen in Figure 6, discrimination was much easier for any manipulation of spectral envelope shape compared to corresponding manipulations of spectral centroid alone via low-pass filtering. In fact, discrimination of spectral envelope shape approached ceiling levels for each step size, whereas mean d′ calculations for the spectral centroid manipulation did not exceed 2.0. This difference contributed to a main effect of timbre dimension, $F(1,17) = 182.522$, $p < .001$. Pair-wise comparisons of means (using adjusted Bonferroni values) further confirmed that sensitivity to changes in spectral envelopes/formant

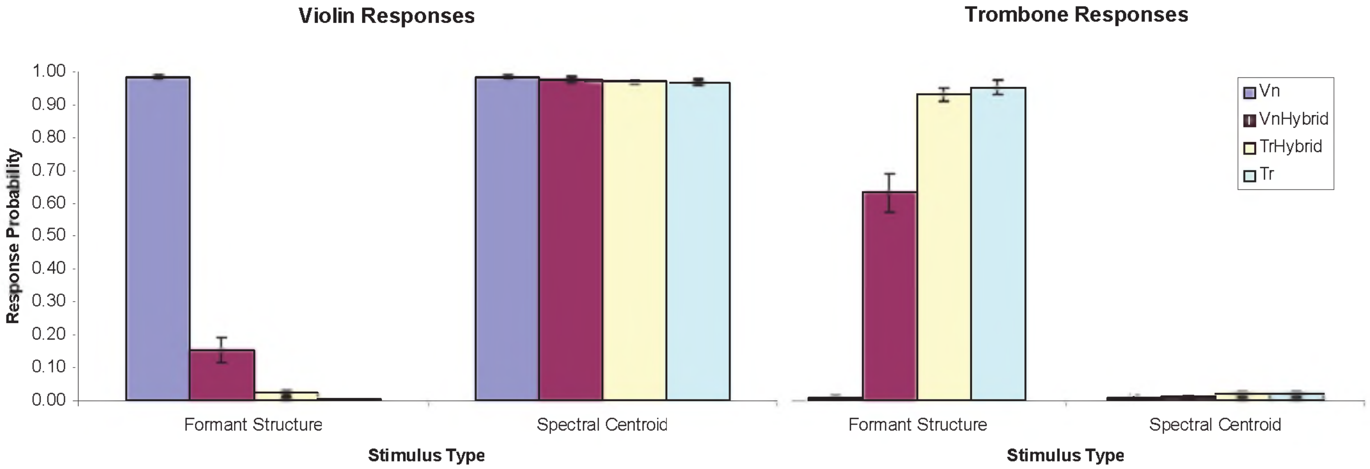


Figure 7. Mean probabilities of *violin* responses (left) and *trombone* responses (right), along with corresponding standard errors, for timbre identification in Experiment 2. For each type of response, results are displayed individually for each manipulation of spectral envelope shape (formant structure, shown to the left) and filtering (to have a corresponding effect on the spectral centroid, shown to the right), progressing gradually from spectral characteristics of a natural violin (*Vn*) to that of a trombone (*Tr*).

structures was significantly higher than for corresponding changes in spectral centroid at each step size ($p < .001$).

Discrimination performance also improved with increasing acoustic distance between stimuli, as revealed by a main effect of step size, $F(2,34) = 4.593$, $p < .05$. However, examination of the means across stimulus conditions in Figure 6 reveals that this effect of step size was driven solely by improvements in performance across spectral centroid conditions. Consistent with this interpretation, pair-wise comparisons of means revealed that the average sensitivity to 1-step changes in spectral centroid alone was significantly lower than for 2- and 3-step comparisons ($p < .05$), whereas sensitivity to changes in spectral envelopes did not significantly differ across the different step sizes. This difference in the effect of step size across (spectral envelope v. centroid) manipulations also contributed to a significant dimension x step size interaction, $F(2,34) = 7.577$, $p < .01$. No other effects approached significance ($p > .10$).

Identification task

For every listener the probabilities of each type of timbre identification response (i.e., *violin*, *trombone*, and *other*) were determined individually for the seven timbre stimuli. The resulting probabilities for each response category were submitted to a separate 1-way repeated measures ANOVA with the 7 levels of stimuli as the sole factor.[2] The decision to collapse these analyses to single-factor ANOVAs permitted the inclusion of timbre identification data for the *Vn* tone without violating assumptions of the statistical test. Pair-wise comparisons of means (according to Bonferroni adjustments) additionally were used to assess the relative contribution of spectral envelope shape and spectral centroid to timbre identification.

Figure 7 displays mean probabilities (and standard error bars) across listeners for the *violin* responses (displayed to the left), as well as the *trombone* responses (right). Within the graph for each response category, average responses to manipulations of formant structure are depicted as the left set of bars, whereas responses to corresponding changes to the spectral centroid alone are depicted to the right. Mean identification data for the *Vn* tone is displayed as the leftmost bar within both formant structure and spectral centroid displays. This duplication of data is intended to simplify visual comparisons with the mean response probabilities that were obtained from each dimension.

The pattern of results displayed in Figure 7 shows that timbre identification performance was strongly impacted by changes in spectral envelope shape (formant structure), but not by changes solely in the spectral centroid. The mean probabilities of *violin* responses decreased, and *trombone* responses increased, as spectral envelopes varied from that of *Vn* to *Tr* (i.e., from left to right for either side of the figure). In fact, *violin* responses approached a mean probability of 1.0 when listeners were given the spectral envelope of *Vn*, and *trombone* responses approached a similar maximum when listeners were presented with the spectral envelope of *Tr*. This trend resulted in very robust main effects of formant structure for *violin* responses [$F(6,102) = 1027.102$, $p < .001$] as well as for *trombone* responses [$F(6,102) = 378.098$, $p < .001$]. Pair-wise comparisons of means further revealed that a significantly greater probability of a *violin* response (and a corresponding reduced probability of a *trombone* response) was obtained for the *Vn* spectral envelope relative to each of the alternative formant structures ($p < .001$). Additionally, the mean probability of a *violin* response also was greater for the stimulus with the *VnHybrid* formant structure than for either the *TrHybrid* or the *Tr* formant structure, ($p < .05$). A corresponding reduction in the probability of a *trombone* response was likewise obtained for the stimulus reflecting the *VnHybrid* spectral envelope relative to the alternative formant structures ($p < .001$).

In contrast, changes to the spectral centroid through low-pass filtering had a negligible effect on timbre identification. The probabilities of neither *violin* responses nor *trombone* responses significantly changed as a function of low-pass filtering; the probabilities of *violin* responses remained near maximum across the different filter settings, and the probabilities of *trombone* responses remained near minimum values (see the mean probabilities displayed above the *Spectral Centroid* label in Figure 7). This was apparent within pair-wise comparisons of means, which revealed a distinct lack of variation in the obtained probabilities for either response category as centroid alone was varied relative to the *Vn* tone ($p > .87$).

The response of *other* was utilized more frequently in Experiment 2 than in Experiment 1. The incidence of *other* responses differed across the stimulus set, as reflected by a significant main effect of stimulus, $F(6,102) = 10.680$, $p < .001$. An examination of the mean probabilities across stimuli reveals that *other* responses were generally reserved for the tone with the *VnHybrid* formant structure ($M = 0.22$, with a standard error of 0.06). *Other* responses were rarely used to identify the remaining stimuli, occurring on 5 percent of trials involving tones with alternative formant structures (*TrHybrid* and *Tr*; standard errors = 0.02) and on only 1 percent of the trials for each of the other 4 stimuli (standard errors $\leq$ 0.01). Pair-wise comparisons of means confirmed that the mean probability of *other* responses for the *VnHybrid* formant structure was significantly greater than for the corresponding manipulation of spectral centroid alone ($p < .05$), and was marginally greater than the mean probability obtained for any other stimulus ($p < .10$). Anecdotal reports from participants additionally indicated that they used the *other* response to indicate when a particular stimulus was perceived ambiguously with respect to the violin and trombone categories. Clearly, the tone with the *VnHybrid* formant structure was often perceived as such a stimulus. The perceived ambiguity of this particular timbre stimulus also was likely heightened relative to the corresponding stimulus in Experiment 1 due to the reduced number of stimuli (7) in Experiment 2. As a result, listeners in Experiment 2 probably were able to store some information about each stimulus in working memory for the duration of the identification task.

Cross-task comparisons

The major findings from both tasks in Experiment 2 provide supporting evidence for the general hypothesis that listeners rely more heavily on information about detailed spectral envelope shape rather than perceived brightness in making timbre judgments. Not only were listeners able to maximally discriminate any change in formant structure, but such changes in the spectral envelope strongly affected instrument identification as well. In contrast, while isolated changes to the spectral centroid through low-pass filtering still produced moderate levels of discrimination performance, such changes had almost no impact on timbre identification.

It also may be inappropriate to attribute the moderate levels of discrimination performance at the larger step size in the centroid conditions (d′ approaching 2.0) to changes in spectral centroid. The manipulation of centroid using a first-order low-pass filter was done to closely match the spectral centroid of formant-manipulated stimuli (to the nearest Hz). At the larger step size this necessarily required that the cutoff frequency be moved much lower than for the other comparisons, substantially reducing the intensities of higher-frequency components within the original waveform. This raises the possibility that the filter's cutoff frequency could have been sufficiently low that filtering also might have strongly impacted higher resonances, and thus, begun to affect the general shape of the spectral envelope. This was confirmed by follow-up acoustic analyses. Figure 8 displays spectral envelopes for the *Vn* tone both before (panel a) and after (panel b) low-pass filtering to match the *Tr* spectral centroid. This side-by-side spectral comparison reveals that the lower-amplitude resonances (F5 and F6) that are present in the original tone are virtually absent after filtering. It thus appears that listeners were really only sensitive to changes in spectral centroid when such changes also began to impact the shape of the spectral envelope.

When taken collectively, the results of Experiment 2 indicate that, by itself, brightness was not a particularly salient attribute of instrument timbre. Some important caveats to this conclusion are necessary. For example, in the current investigation judgments about centroid manipulations were made in the context of other information about the shape of the spectral envelope. Thus, it is possible, even likely, that the perceived effect of filtering could have been greater in the absence of other salient spectral envelope structure (e.g., including obvious formants). Spectral centroid also was varied across only two instruments in the current investigation. While these instruments were selected to reflect a reasonably wide difference in centroids, it also is acknowledged that more support for the role of spectral centroid might have been obtained using a wider array of instruments, and therefore, greater variation in spectral slopes. However, it is noteworthy that such increased variation also would further increase the difficulty in isolating changes in spectral centroid from the shape of the spectral envelope.

4. GENERAL DISCUSSION

4.1. Assessment of amplitude envelope contributions to timbre

In Experiment 1 it was expected that both spectral envelope and amplitude envelope manipulations would contribute significantly to timbre identification and discrimination performance. Thus, it was somewhat surprising that the various amplitude envelopes were not only less salient than our spectral envelope manipulation insofar as they contributed minimally to timbre

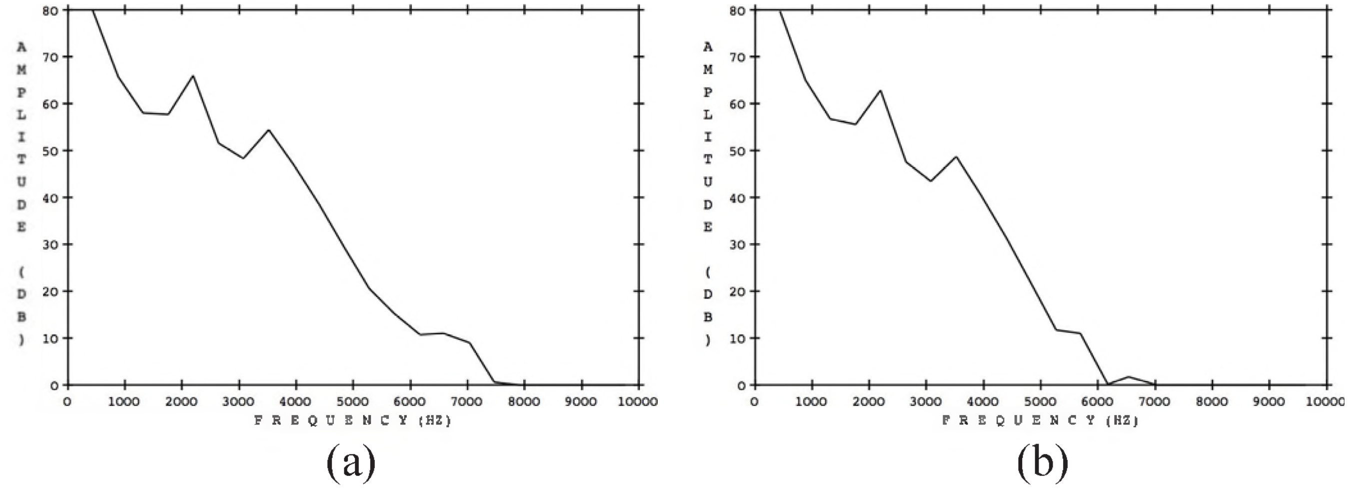


Figure 8. Spectral envelopes [amplitudes (in dB) v. frequency (in Hz)] for the resynthesized violin tone before (panel a) and after (panel b) low-pass filtering to match the spectral centroid of a tenor trombone tone.

identification, but also were not very discriminable from each other. This suggests that timbre identification performance was probably a reflection of basic psychoacoustic limits that were demonstrated by performance within the discrimination task.

Given that the amplitude envelopes in Experiment 1 were distinguished by rise time, these findings seem initially contradictory to several MDS studies (e.g., see Grey and Gordon, 1978; Iverson and Krumhansl, 1993; Krimphoff, et al. 1994; McAdams, et al. 1995) that have concurred that rise time is a primary dimension of instrument timbre. What factor or factors are likely responsible for this apparent discrepancy in results across studies?

A few potential explanations can probably be considered less likely. For example, it could be argued that the amplitude envelope dimension was weaker in the current study because entire amplitude envelopes were not included; natural decay portions of the tones were replaced with 10 ms linear ramps to tone offset. Furthermore, part of the steady-state portion of the original trombone tone was missing because it was truncated to 500 ms. Steady-states were almost absent from the tones derived from violin since the attack constituted the majority of the tone. It is therefore possible that these restrictions of the amplitude envelope limited its contribution to timbre identification and discrimination performance. However, it is the amplitude envelope of the *attack* that has typically been argued to be a critical timbre dimension, and this part of the envelope (epitomized by rise time) was retained. Furthermore, some MDS results have indicated a greater reliance on basic spectral information (number of harmonics) when it was manipulated along with certain amplitude envelope types (horn, string, or trapezoidal), although it is noteworthy that attack length was fixed (Miller and Carterette, 1975).

It also is acknowledged that only a single fundamental frequency was used throughout the current investigation. It therefore could be argued that different results might have been obtained had different pitches, from different registers, been included. For example, it has been demonstrated that identification of some timbres can change depending on pitch register, raising the possibility of “characteristic pitches” for an instrument, as well as a dependence upon training and exposure to a wide array of pitches to truly understand the overall timbre of any instrument (e.g., see Sandell and Chronopoulos, 1997). However, it should be noted that the A_4 pitch was selected to be well within the range that is typically produced by both instruments. Thus, there does not seem to be much reason to expect different outcomes with different pitches unless samples were selected to include atypically high or low pitches for one or both instruments.

It also could be suggested that the impact of amplitude envelope variation might have been different had the tones been presented within melodic sequences rather than in isolation. While this possibility is acknowledged, available evidence suggests that the contribution of amplitude envelope to timbre would actually be expected to be further reduced in such sequences. For example, Grey (1978) demonstrated that simplifications to the attacks of trumpet and clarinet sounds were more poorly discriminated in musical contexts, whereas the discrimination of simplifications to the spectral envelopes of bassoon sounds was unaffected by context. Likewise, Kendall (1986) found that the presence/absence of attack transients did not aid instrument recognition across pairs of short melodic sequences.

Additionally, it is acknowledged that the current investigation does not take into account known visual influences on instrument timbre. Specifically, there have been demonstrated shifts in timbre (i.e., the report of plucked vs. bowed strings) depending upon the presence of congruent/incongruent visual information (the synchronous movie of a musician plucking or bowing the instrument; see Saldaña & Rosenblum, 1993). While corresponding shifts would be likely to occur given corresponding visual productions for the tones in the experiments reported here, there is no reason to expect either categorically different responding or increased/decreased contributions from a given acoustic dimension in the presence of incongruent visual information.

Another possible explanation was raised at the end of Experiment 1, where we indicated that poor discrimination performance for the amplitude envelope parameter could be argued to be due to a truncated range of amplitude envelopes relative to the spectral envelope dimension. We pointed out that the instruments were selected to reflect a fairly wide distribution of values across both spectral and temporal dimensions, including rise times (see Table 1). Thus, it is not likely that our results are simply due to a lack of physical variation along the temporal dimension.

Further support for this interpretation comes from the results of pilot experiments for the current investigation. We had collected identification and discrimination data for tones derived from a larger set of instruments (piano, vibraphone, clarinet, and violin). Although the amplitude envelopes for these preliminary tones included linear onset ramps, and thus were not as natural as those used in the experiments reported here, they also essentially had maximally different rise time values across the set (40 ms for piano v. 458 ms for violin; the remaining rise times were 67 and 111 ms for the vibraphone and clarinet, respectively). In that pilot study listeners were reasonably sensitive (with mean values of d′ between 1 and 2) to differences in amplitude envelopes only for conditions involving the longest rise time, and mean values of d′ approached 2 only for the largest difference in rise times (an inordinately long 418 ms difference). In light of these results, it is unlikely that discrimination performance in the current investigation would have improved much unless extreme differences in rise times were used, likely needing some of the largest rise time differences that occur in natural instruments. Thus, it appears that, despite our reliance on a reasonably broad range of rise times within the included amplitude envelopes, listeners in Experiment 1 were not able to perceive much variation along that dimension.

A more reasonable explanation for the relatively poor discrimination performance with respect to amplitude envelope, as well as for the minimal contribution of this parameter to timbre identification, may be the possibility that rise time acts like a binary feature characterized by the presence or absence of a very abrupt attack. In other words, listeners would categorize amplitude envelopes during the attack as either abrupt or not. This argument does not require that rise times be categorically perceived, although there have been debates about whether or not categorical perception occurs along rise time continua (e.g., see Cutting, 1982; Donnadieu, McAdams, and Winsberg, 1996; Rosen and Howell, 1981, 1983). Rosen and Howell (1983) provided evidence that discrimination performance was at a maximum at the short rise time end of their continuum, and thereafter linearly decreased with increasing rise times. Thus, distinct performance differences along a rise time dimension are possible in the absence of a fixed category boundary between instruments with abrupt and gradual onsets.[3] This finding also is consistent with the notion that an abrupt stimulus with a particularly short rise time could act as a type of perceptual anchor against which all other stimuli are evaluated. If so, poor discrimination performance could be obtained despite large physical differences in rise time when the distribution of rise times does not include very short values. This was indeed the case in our Experiment 1, where the shortest rise time exceeded 150 ms.

A closer look at some classic MDS results that report rise time as a critical timbre dimension also lends further support for regarding rise time as a binary feature. One such study comes from McAdams, et al. (1995), who collected timbre dissimilarity ratings for pairs of tones taken from a larger stimulus set (that was previously developed by Wessel, Bristow and Settel, 1987 and used in a frequently cited MDS study by Krumhansl, 1989). Perceptual dimension 1 of their scaling solution (in their Figure 1), which is strongly correlated with rise time, shows a very large gap in the middle along with a clustering of instruments to either side of the dimension, particularly for instruments lacking abrupt onsets. Thus, despite a broad range of physical differences in rise time, listeners grouped instruments into perceptually abrupt (e.g., for vibraphone, guitar, piano, harp, and harpsichord) versus other values (e.g., bowed string, bassoon, English horn), and all non-abrupt rise times were perceived as quite similar to each other.

A similar conclusion can be reached upon an examination of MDS results from Iverson and Krumhansl (1993), which, like the current investigation, were based upon samples from the MUMS database, including the violin and trombone. They found similar MDS solutions based upon pair-wise ratings obtained for complete tones, their onsets only, or the remainder of the tones, leading to the conclusion that attributes used in making similarity judgments were present throughout the entire tone rather than being confined to the attack. Furthermore, in the horizontal dimension of their scaling solution for tone onsets (their Figure 6), which was labeled as relating to dynamic attributes of timbre, there was a perceptual clustering of all instruments except those with abrupt onsets (tubular bells, piano, vibraphone, and cello). Thus, very few instrument tones were judged as very dissimilar along that perceptual dimension, and those that were distinctly perceived were tones with an abrupt attack.

When these classic findings from MDS and categorical perception are taken together with the lack of a strong contribution from rise time to either discrimination or identification performance in the current investigation, they collectively suggest that rise time is only likely to permit reliable instrument identification when very short values along the dimension are contrasted with longer values. This suggestion should not be viewed as contradictory to seminal timbre research that reveals that timbre identification is negatively impacted when the attack is excised (e.g., see Saldanha and Corso, 1964). After all, removal of the attack effectively alters all amplitude envelopes to have abrupt onsets. Such a transformation should be easily perceived if rise time is perceptually evaluated as the presence or absence of abruptness. In fact, such results should be expected because relatively few sustained tone instruments' attacks approximate immediate onsets.

The idea of heightened salience for abrupt attacks also is consistent with other findings. For example, temporal order judgments are more accurate for tones with short rise times (e.g., see Bregman, Ahad and Kim, 1994; also see Pastore, Harris and Kaplan, 1981). This suggests that listeners may have difficulty detecting, or attending to, the temporal locations of intense portions of tones that have more gradual onsets.

Finally, while spectrotemporal variation was necessarily eliminated from the current investigation in order to focus on the respective contributions of spectral envelope and amplitude envelope to timbre, it is quite possible that one or more spectrotemporal dimensions could correlate highly with rise time as well, and thus enhance perception of attack transients in tones produced by natural instruments. As alluded to in the introduction, several important spectrotemporal parameters related to instrument timbre have been defined, including spectral centroid variation, spectral incoherence, and spectral irregularity (Beauchamp and Lakatos, 2002). Traditionally, in MDS studies naturally occurring spectrotemporal variations have been retained in the (attacks of) tones used to evaluate timbre. Under such stimulus conditions it is conceivable that spectral variation could contribute to, or even explain, findings for a perceptual dimension based on rise times that are overall measures of complex spectrotemporal phenomena (e.g., spectral flux or other spectrotemporal variables that are functions of rise time). This would be consistent with spectrotemporal changes that occur in a relatively systematic way when moving from instruments with abrupt to more gradual onsets. While beyond the scope of the current investigation, the relative salience of the spectral envelope and weakness of the amplitude envelope information in Experiment 1 leaves open the possibility that spectral variation could contribute in cases where a strong perceptual relation to rise time is found and spectrotemporal variation is not controlled. This possibility warrants future investigation to permit a more complete assessment of the contribution of amplitude envelope to timbre.

4.2. Clarifying the role of the spectral envelope

The advantage observed in Experiment 1 for the spectral envelope dimension relative to the amplitude envelope dimension does not appear to be due to a reliance on brightness perception. Experiment 2 was designed to directly evaluate this possibility for discrimination and timbre identification tasks. Spectral centroid, which is generally regarded as an acoustic measure related to the perceptual dimension of brightness, was equivalently manipulated in Experiment 2 by either altering the center frequencies of formants or by low-pass filtering tones. The latter manipulation preserved most of the original shape of the spectral envelope; only moderate discrimination performance was obtained in response to such variation, and virtually no effect on timbre identification was observed. Had spectral centroid been the primary cue that listeners used to evaluate timbre, then performance in the filtering conditions should have instead approached the near-ceiling levels of discrimination and sharp changes in timbre identification that were observed for the spectral envelope conditions.

The conclusion that listeners in both our experiments relied upon perceptual information about the shape of the spectral envelope, rather than brightness, should not be considered surprising. After all, it is the entire spectral envelope that reflects the natural resonances of the instrument body. In contrast, brightness reflects much less information for the listener, indicating the relative contribution of components within the spectral envelope having higher or lower frequencies.

This conclusion also is consistent with several findings from the timbre literature. This includes existing evidence that machine recognition of instrument timbres is significantly improved when supplying information about formant structure (via cepstral coefficients; see Brown, et al. 2001) rather than simply supplying data about the spectral centroid. The major findings from Experiment 2 also could be regarded as further evidence for the perceptual separability of spectral slope, which directly impacts the spectral centroid, and the shape of the spectral envelope (see Li and Pastore, 1995). This suggestion comes from the fact that our filtering manipulation, which was essentially a manipulation of spectral slope, resulted in much poorer discrimination and timbre identification performance than our formant-based manipulation of the spectral envelope. Finally, the conclusion for listeners' greater reliance on spectral envelope information also is consistent with the initial interpretations of spectral dimensions in some early MDS solutions (e.g., see Krumhansl, 1989).

So what should be made of the results from the current investigation in light of several classic MDS studies that demonstrate that the spectral centroid is the primary spectral measure that strongly correlates with an obtained perceptual dimension (e.g., Ehresman and Wessel, 1978; Krimphoff, 1993; Krimphoff, et al. 1994; McAdams, et al. 1995)? Given the lack of another purely spectral dimension within the MDS solutions from these studies, it is likely that the researchers initially attributed perceptual effects of spectral envelope shape to brightness. The latter perceptual dimension presumably reveals corresponding changes in the physical signal that could be summarized by the spectral centroid.

Then, how can we account for the very high correlation coefficients that have been obtained between spectral centroid measures and perceptual dimensions in these MDS solutions? It is important to remember that the spectral centroid is just an acoustic measure, one that reflects the center of the distribution of energy across the spectrum. The spectral centroid should be expected to be highly correlated with uniform shifts in peaks within the spectral envelope. For example, a typical vibrational sound source produces a prominent spectral peak near the sound's fundamental frequency, and the amplitudes of other harmonics are eventually reduced with increasing frequency (indicating a negative spectral slope or rolloff). Harmonics are also typically grouped within spectral envelopes or resonances. As these peaks, or formants, are distributed more widely, the spectral centroid also should increase. Likewise, displacement of any formant higher in frequency should also increase the spectral centroid. Both of these cases should result in a brighter timbre. Insofar as this description reflects natural acoustic consequences of different resonance patterns within musical instruments, there should be a systematic relationship between the spectral centroid and the shape of the spectral envelope. Perceived brightness based on changes in the spectral centroid should still be expected to contribute heavily to timbre in instances where there is a particularly strong or weak contribution from higher frequency components. However, the latter stimulus conditions are likely to also drastically reduce energy across the spectrum, and thus, minimally specify the spectral envelope.

It should not be too difficult to disentangle these properties. For the sake of simplicity, let us assume that we are presented with a signal consisting of just two formants. The same spectral centroid should result regardless of whether the center frequencies of both formants are compressed to the middle of the spectrum or are carefully adjusted in opposing frequency directions. However, clearly these two sounds have drastically different spectral envelopes, and therefore, they should be perceived as drastically different. After all, this difference between formant structures is the very distinction that exists between vowels based on changes in tongue height and position; modal productions of low back vowels like /a/ tend to have a compact spectrum based on F1 and F2 center frequencies, whereas high front vowels like /i/ tend to have more diffuse center frequencies for F1 and F2 (e.g., see Klatt and Klatt, 1990).

Insofar as brightness can be distinguished from other timbre information, it also is possible that it could contribute to tone perception in unexpected ways. For example, there are numerous demonstrations of pitch judgments being impacted by changes in timbre, particularly in musically untrained listeners. Large individual differences in the weighting of tone height and chroma have been attributed to brightness (e.g., Demany and Semal, 1993). Consistent with this possibility, shifts in the position of a spectral envelope with a fixed (bell) shape spectral have been found to impact pitch judgments in the *tritone paradox* (Repp, 1997; also see Deutsch, 1987). Recently, it has been demonstrated that musically untrained listeners will often perceive a tone's pitch to change upon removal of its fundamental frequency (Seither-Preisler, Johnson, Seither and Lütkenhöner, 2007). Also, Pitt (1994) used speeded classification tasks to demonstrate that pitch and timbre are perceptually integral properties, and that, furthermore, non-musicians are frequently likely to confuse timbre variation for changes in pitch. The reported size of pitch intervals by listeners with little musical training has even been shown to depend upon differences in the relative weighting of amplitudes in the synthesized tones' (upper and lower) harmonics (Russo and Thompson, 2005; Warrier and Zatorre, 2002).

It is possible that these various demonstrations share a common basis. A possible explanation for these phenomena is that some (particularly musically untrained) listeners frequently attribute changes in brightness, presumably in response to changes in the spectral centroid, to changes in pitch. According to this view, pitch would frequently be perceived to increase with increases in the relative weighting of higher-frequency components.

Ongoing efforts in our laboratories are therefore comparing performance across these traditional demonstrations, coupled with our manipulations of spectral centroid, in order to determine the potential impact of brightness on pitch judgments. Then the extent to which purely spectral dimensions can contribute to tone perception will hopefully be better understood. For now, the results of the current investigation provide further indications of the relative salience of some important dimensions of timbre. These results also suggest that a more thorough understanding of the nature of critical timbre dimensions is likely to be attained by comparing performance across an array of tasks that includes timbre identification.

5. ACKNOWLEDGEMENTS

The authors would like to thank the reviewers for their helpful comments, as well as the following research assistants for all their help with data collection: Charlotte Claflin, Rachel Ledebuhr, and Benjamin Rodes.

6. REFERENCES

American Standards Association (1960). *American Standard Acoustical Terminology*. New York.

Beauchamp, J. W. and Lakatos, S. (2002). New spectrotemporal measures of musical instrument sounds used for a study of timbral similarity or risetime-and centroid-normalized musical sounds. *Proc. 7th Int. Conf. on Music Perception & Cognition*, Univ. of New South Wales, Sydney, Australia, 592-595.

Boersma, P. and Weenink, D. (2007). Praat: doing phonetics by computer (Version 4.6.01**)** [Computer program]. Retrieved May, 2007, from http://www.praat.org/.

Bregman, A. S., Ahad, P. and Kim, J. (1994). Resetting the pitch analysis system 2: Role of sudden onsets and offsets in the perception of individual components in a cluster of overlapping tones. *J. Acoust. Soc. Am.*, ***96***, 2694-2703.

Brown, J. C., Houix, O. and McAdams, S. (2001). Feature dependence in the automatic identification of musical woodwind instruments. *J. Acoust. Soc. Am.*, ***109***(3), 1064-1072.

Caclin, A., McAdams, S., Smith, B. K. and Winsberg, S. (2005). Acoustic correlates of timbre space dimensions: A confirmatory study using synthetic tones. *J. Acoust. Soc. Am.*, ***118***(1), 471-482.

Carroll, J. D. and Chang, J. J. (1970). Analysis of individual differences in multidimensional scaling via an n-way generalization of Eckart-Young decomposition. *Psychometrika*, ***35***, 283-319.

Cutting, J. E. (1982). Plucks and bows are categorically perceived, sometimes. *Perception & Psychophysics*, ***31***, 462-476.

Demany, L. and Semal, C. (1993). Pitch versus brightness of timbre: Detecting combined shifts in fundamental and formant frequency. *Music Perception*, ***11***(1), 1-13.

Deutsch D. (1987). The tritone paradox: Effects of spectral variables. *Perception & Psychophysics*, ***41***, 563-575.

Donnadieu, S. (2007). Mental representation of the timbre of complex sounds. In J. W. Beauchamp (Ed.), *Analysis, Synthesis, and Perception of Musical Sounds* (pp. 272-319). New York: Springer.

Donnadieu, S., McAdams, S. and Winsberg, S. (1996). Categorization, discrimination and context effects in the perception of natural and interpolated timbres. In B. Pennycook and E. Costa-Giomi (Eds.), *Proc. 4th Int. Conf on Music Perception and Cognition* (pp. 73-78). Montreal: McGill University.

Ehresman, D. and Wessel, D. (1978). Perception of timbral analogies. *Rapports IRCAM*, ***13***, Paris: IRCAM.

Fant, G. (1973). *Speech Sounds and Features*. Cambridge, Massachusetts: MIT Press.

Grey, J. M. (1975). *An exploration of musical timbre*. Doctoral dissertation, Dept. of Music, Report No. STAN-M-2. Stanford Univ., Stanford, CA.

Grey, J. M. (1977). Multidimensional perceptual scaling of musical timbres. *J. Acoust. Soc. Am.*, ***61***(5), 1270-1277.

Grey, J. M. (1978). Timbre discrimination in musical patterns. *J. Acoust. Soc. Am.*, ***64***(2), 467-472.

Grey, J. M. and Gordon, J. W. (1978). Perceptual effects of spectral modifications on musical timbres. *J. Acoust. Soc. Am.*, ***63***(5), 1493-1500.

Grey, J. M. and Moorer, J. A. (1977). Perceptual evaluations of synthesized musical instrument tones. *J. Acoust. Soc. Am.*, ***62***(2), 454-462.

Haken, L., Fitz, K. and Christensen, P. (2007). Beyond traditional sampling synthesis: Real-time timbre morphing using additive synthesis. In J. W. Beauchamp (Ed.), *Analysis, Synthesis, and Perception of Musical Sounds* (pp. 122-144). New York: Springer.

Iverson, P. and Krumhansl, C. L. (1993). Isolating the dynamic attributes of musical timbre. *J. Acoust. Soc. Am.*, ***94***(5), 2595-2603.

Kendall, R. A. (1986). The role of acoustic signal partitions in listener categorization of musical phrases. *Music Perception*, ***4***, 185-214.

Kendall. R. A. (2002). Music Experiment Development System (Version 2002-B) [Computer software]. Los Angeles, CA: Author,

www.ethnomusic.ucla.edu/systematic/Faculty/Kendall/meds.htm.

Klatt, D. H. and Klatt, L. C. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. *J. Acoust. Soc. Am.*, ***87***(2), 820-857.

Krimphoff, J. (1993). *Analyse acoustique et perception du timbre* . Unpublished DEA thesis. Université du Maine, Le Mans, France.

Krimphoff, J., McAdams, S. and Winsberg, S. (1994). Caractérisation du timbre des sons complexes. II : Analyses acoustiques et quantification psychophysique. *Journal de Physique*, ***4***(C5), 625-628.

Krumhansl, C. L. (1989). Why is musical timbre so hard to understand? In S. Nielsen and O. Olsson (Eds.), *Structure and Perception of Electroacoustic Sound and Music* (pp. 43-53). Amsterdam: Elsevier.

Li, X. and Pastore, R. E. (1995). Perceptual constancy of a global spectral property: Spectral slope discrimination. *Journal of the Acoustical Society of America*, ***98***(4), 1956-1968.

Macmillan, N. A. and Creelman, C. D. (2005). *Detection Theory: A User's Guide*. Mahwah, New Jersey: Lawrence Erlbaum Associates.

McAdams, S. (2001). Recognition of sound sources and events. In S. McAdams and E. Bigand (Eds)., *Thinking in Sound: The Cognitive Psychology of Human Audition* (pp. 146-198). New York: Oxford University Press.

McAdams, S., Beauchamp, J. W. and Meneguzzi, S. (1999). Discrimination of musical instrument sounds resynthesized with simplified spectrotemporal parameters. *J. Acoust. Soc. Am.*, ***105***(2), pt. 1, 882-897.

McAdams, S., Winsberg, S., Donnadieu, S., De Soete, G. and Krimphoff, J. (1995). Perceptual scaling of synthesized musical timbres: Common dimensions, specificities and latent subject classes. *Psychological Research*, ***58***, 177-192.

Miller, J. R. and Carterette, E. C. (1975). Perceptual space for musical structures. *J. Acoust. Soc. Am.*, ***58***(3), 711-720.

Opolko, F., and Wapnick, J. (1987). McGill University Master Samples [CD-ROM]. Montreal, Quebec, Canada: jwapnick@music.mcgill.ca.

Pastore, R. E. (1990). Categorical perception: Some psychophysical models. In S. Harnad (Ed.), *Categorical Perception: The Groundwork of Cognition* (pp. 29-52). New York: Cambridge University Press.

Pastore, R. E., Harris, L. B., and Kaplan, J. K. (1981). Temporal order identification: Some parameter dependencies. *J. Acoust. Soc. Am.*, ***71***, 430-436.

Pitt, M. A. (1994). Perception of pitch and timbre by musically trained and untrained listeners. *J. Exp. Psych.: Human Perception and Performance*, ***20***(5), 976-986.

Repp, B. H. (1997). Spectral envelope and context effects in the tritone paradox. *Perception*, ***26***, 645-665.

Rodet, X. and Schwarz, D. (2007). Spectral envelopes and additive + residual analysis/synthesis. In J. W. Beauchamp (Ed.), *Analysis, Synthesis, and Perception of Musical Sounds* (pp. 175-227). New York: Springer.

Rosen, S. M. and Howell, P. (1981). Plucks and bows are not categorically perceived. *Perception & Psychophysics*, ***30***(2), 156-168.

Rosen, S. and Howell, P. (1983). Sinusoidal plucks and bows are not categorically perceived, either. *Perception & Psychophysics*, ***34***(3), 233-236.

Russo, F. A. and Thompson, W. F. (2005). An interval size illusion: The influence of timbre on the perceived size of melodic intervals. *Perception & Psychophysics*, ***67***(4), 559-568.

Saldaña, H. M. and Rosenblum, L. D. (1993). Visual influences on auditory pluck and bow judgments. *Perception & Psychophysics*, ***54***(3), 406-416.

Saldanha, E. L. and Corso, J. F. (1964). Timbre cues and the identification of musical instruments. *J. Acoust. Soc. Am.*, ***36***, 2021-2026.

Samson, S., Zatorre, R. J. and Ramsay, J. O. (1997). Multidimensional scaling of synthetic music timbre: Perception of spectral and temporal characteristics. *Canadian J. Exp. Psych.*, ***51***(4), 307-315.

Samson, S., Zatorre, R. J. and Ramsay, J. O. (2002). Deficits of musical timbre perception after unilateral temporal-lobe lesion revealed with multidimensional scaling. *Brain*, ***125***, 511-523.

Sandell, G. J. and Chronopoulos, M. (1997). Perceptual constancy of musical instrument timbres; generalizing timbre knowledge across registers. P*roc. Third Triennial Conf for the European Society for the Cognitive Sciences of Music*, Uppsala, Sweden, 222-227.

Seither-Preisler, A., Johnson, L., Seither, S. and Lütkenhöner, B. (2007). Ambiguous tone sequences are heard differently by musicians and non-musicians [A]. *Program for the 8th Conf. of The Society for Music Perception and Cognition (SMPC)*, 7.

Thayer, R. C. (1974). The effect of the attack transient on aural recognition of instrumental timbres. *Psychology of Music*, ***2***(1), 39-52.

Warrier, C. M. and Zatorre, R. J. (2002). Influence of tonal context and timbral variation on perception of pitch. *Perception & Psychophysics*, ***64***(2), 198-207.

Wessel, D.L., Bristow, D. and Settel, Z. (1987). Control of phrasing and articulation in synthesis. *Proc. 1987 Int. Computer Music Conf* (ICMC), pp. 108-116. Computer Music Association, San Francisco.

Winsberg, S. and De Soete, G. (1993). A latent-class approach to fitting the weighted Euclidean model, CLASCAL. Psychometrika, ***58***, 315-330.

Winsberg, S. and De Soete, G. (1997). Multidimensional scaling with constrained dimensions: CONSCAL. *British J. of Mathematical and Statistical Psychology*, ***50***(1), 55-72.

7. NOTES

[1] Envelopes from both instruments had a very shallow increase in amplitude immediately prior to reaching (90 percent) criterion for the measurement of rise time, This was particularly true for the trombone tone. Adjusting the end measurement for rise time down by 1 dB, a decrease that should not be audible over an extended time period, produced rise time approximations of 345 ms and 156 ms for violin and trombone, respectively. The difference between these measures is 21 ms more than for our initial measures. Thus, it is quite possible that functional differences in rise times for the listeners in Experiment 1 were actually slightly greater than indicated by the values in Table 1.

[2] Although the assumption of sphericity appeared to have been violated bv the ANOVA for each identification response, all reported effects continued to be significant when relying instead upon the Greenhouse-Geisser correction procedure. Thus, critical findings remained the same regardless of which statistical procedure was used.

[3] The observed changes in instrument timbre identification might lead some readers to question whether or not the instrument timbres in the current investigation were categorically perceived. It should be noted that the experiments reported here were not designed to directly assess categorical perception. There were very broad acoustic differences between adjacent steps along continua composed of very few stimuli. This complicates any determination of whether true categorical boundaries were perceived (i.e., discrete changes in timbre given a relatively small physical change in the middle of an acoustic dimension). Furthermore, in contrast to what is typically done in studies of categorical perception, in the discrimination tasks not all adjacent stimuli were compared along a given dimension. Specifically, the two hybrid stimulus values were not directly compared.

Despite this apparent limitation, there are indications that the violin and trombone timbres were not categorically perceived. For example, discrimination performance for single-step comparisons was nearly perfect for within-category comparisons (based on timbre identification data) along the spectral envelope/formant structure dimension in both experiments, which contrasts with the expected troughs of within-category discrimination performance that is characteristic of categorically perceived events (see Figure 4). Furthermore, the changes in instrument identification that were observed along the formant structure dimension were clearly gradual in Experiment 1 (see Figure 5), which is inconsistent with a clear demonstration of a category boundary. Evidence from the amplitude envelope dimension also is inconsistent with arguments for categorical perception. Highly accurate levels of discrimination performance were never obtained for the amplitude envelope dimension (even when larger step sizes were used). Additionally, timbre identification in Experiment 1 (see Figure 5) was generally unaffected by this dimension. [For a review of criteria and stimulus conditions needed to effectively demonstrate categorical perception, see Pastore (1990)].

DECONSTRUCTING A MUSICAL ILLUSION:
POINT-LIGHT REPRESENTATIONS CAPTURE SALIENT PROPERTIES OF IMPACT MOTIONS

Michael Schutz and Michael Kubovy
Department of Psychology, University of Virginia, P.O. Box 400400, Charlottesville, VA 22904, USA;
E-mail: schutz@virginia.edu or kubovy@virginia.edu.

ABSTRACT

Although visual information affects auditory perception in a variety of tasks, audition is generally believed to be relatively immune from visual influence when judging tone duration. However, Schutz and Lipscomb (2007) report a musical illusion in which physical gestures influence the perceived duration of notes performed on the marimba. In order to better understand which aspects of these gestures are responsible for the illusion, we created a "schematic marimbist” consisting of either a four-point skeleton or a single moving dot. This schematic abstraction captured the essential properties of the gestures, replicating the effect under both conditions. Therefore, this illusion requires seeing only a sudden change in gesture direction — independent of the depiction of a struck object. As this finding means that it can be replicated with a minimum of visual information, it will be useful in facilitating future research aimed at uncovering the reason for this break with the widely accepted theory of ‘optimal integration’.

RESUME

Malgré les effets de l’information visuelle sur la perception auditive lors de plusieurs tâches variées, l’audition est généralement considérée relativement immune à l’influence visuelle lors d’un jugement de la durée d’un son musical. Cependant, Schutz et Lipscomb (2007) décrivent une illusion musicale dans laquelle les gestes physiques influencent la durée perçue de notes jouées sur un marimba. Pour mieux comprendre quelles caractéristiques de ces gestes peuvent être attribuées à cette illusion, nous avons créé un "marimbiste schématique” qui consiste soit d’un bras squelette à quatre points soit d’un seul point animé. Cette abstraction schématique a capté les caractéristiques essentielles des gestes puisque les effets visuels se sont reproduits lors des deux conditions schématiques. La seule condition requise de cette illusion est donc de voir un changement subit dans la direction d’un geste – indépendamment d’une représentation de l’objet frappé. Puisque ces résultats indiquent que l’illusion peut être reproduite avec un minimum d’information visuelle, ils pourront faciliter l’éventuelle recherche visée à la découverte de la cause de cette déviation de la théorie d’ "intégration optimale” couramment acceptée.

1. INTRODUCTION

1.0 Music: A multi-modal experience

Since the advent of recording technology, it has been tempting to think of music as a purely auditory phenomenon. It is now apparent, however, that non-auditory information such as kinesthetic feedback (Phillips-Silver & Trainor, 2005, 2007) and visual information (Schutz, 2008; Thompson, Graham, and Russo, 2005) play important roles. In fact, certain aspects of a musical performance can be communicated through purely non-auditory means.

For example, the relative sizes of sung musical intervals can be discerned from a singer’s lip movements (Thompson & Russo, 2007). Likewise, a performer’s emotional intentions can be inferred by watching their body movements, even in the absence of facial information (Dahl & Friberg, 2007). As both of these studies used vision alone, they effectively demonstrate that it *can* be a salient channel for communicating musical information. However, when determining the degree to which vision plays a meaningful role in the musical experience, it is important to examine whether it alters the listeners' experience. To this end, Thompson, Russo, and Quinto (2008) provide a convincing demonstration. They combined auditory and visual presentations of “happy” (major third) and “sad” (minor third) intervals, and asked participants to judge the emotional tenor (the affect) of each audio-visual pairing. Even though (1) participants were asked to judge auditory information alone and (2) they had to concurrently perform a distracter task designed to minimize “cross-talk” between the modalities, the type of visual information used (happy vs. sad) changed affect ratings.

Vision can affect virtually all aspects of the musical experience: evaluations of expressivity (Davidson, 1993, 1994), audience interest (Broughton & Stevens – in press), judgments of musical tension and phrasing (Vines, Krumhansl, Wanderley, & Levitin, 2006), and even assessments of performance quality (Wapnick, Darrow, Mazza, & Dalrymple, 1997, Wapnick, Mazza, & Darrow, 1998). It can also affect the perception of low-level attributes such as pitch (Thompson et al., 2005, Gillespie, 1997), loudness (Rosenblum & Fowler, 1991), note duration (Schutz & Lipscomb, 2007), and timbre (Saldaña & Rosenblum, 1993). Furthermore, it can improve lyric comprehension (Hidalgo-Barnes & Massaro, 2007), as well as affect judgments of musical dissonance (Thompson et al., 2005, experiment 1), and interval size (Thompson et al., 2005, experiment 3)

Although many musical instruments have been used to study visual influences, percussion offers a particularly rich domain for exploring such issues. Likely, this reflects the relatively large physical motions used by percussionists, and the clear causal relationship between their gestures and sounds (Schutz, 2008). In particular, the marimba (a tuned, wooden bar-percussion instrument similar to the xylophone) has received a great deal of research attention in recent years, including studies showing the importance of visual information with respect to the communication of emotional intention (Dahl & Friberg, 2007, Broughton & Stevens – in press), audience interest (Broughton & Stevens – in press) and note duration (Schutz & Lipscomb, 2007).

1.1 The Schutz-Lipscomb illusion

Schutz and Lipscomb (2007) report an audio-visual illusion in which an expert musician's gestures affect the perceived duration of a note without changing its acoustic length. To demonstrate this, they recorded world-renowned marimbist Michael Burritt (Professor of Percussion at the Eastman School of Music) performing single notes on the marimba using long and short gestures. They paired both types of sounds with both types of gestures, resulting in a combination of natural (i.e., congruent gesture–note pairs) and hybrid (i.e., incongruent gesture–note pairs) stimuli. They informed participants that some auditory and visual components had been mismatched, and asked them to judge tone duration based on the auditory component alone. Despite these instructions, the participants' duration ratings were strongly influenced by visual gesture information (i.e., notes were rated as longer when paired with long gestures than when paired with short gestures). This suggests that the integration of visible striking gestures with heard percussive sounds is perceptually obligatory.

1.2 Why this is puzzling: Previous work on audio-visual integration

These results contradict the view that judgments of tone duration are relatively immune to visual influence (Walker & Scott, 1981, Welch & Warren, 1980), i.e., in temporal tasks visual influence on audition is negligible. For example, audition affects judgments of light duration, but vision does not influence judgments of tone duration (Walker & Scott, 1981). Likewise, the rate of auditory flutter (i.e. number of tones per second) affects the perceived rate of concurrent visual flicker, but the rate of visible flicker either fails to affect the perceived rate of concurrent auditory flutter (Shipley, 1964) or affects it minimally (Welch, DuttonHurt, & Warren, 1986).

Generally, visual information dominates conflicting auditory information only when it is of higher quality — such as in source localization. In the ventriloquism illusion, for example, speech is heard to originate from the moving lips of a silent puppet (Jack & Thurlow, 1973), because the spatial resolution of the visual system is significantly better than that of the auditory system. This dominance is not limited to speech, as shown by similar effects involving non-speech sounds (Bertelson & Radeau, 1981, Bertelson, Vroomen, de Gelder, & Driver, 2000, Jackson, 1953, Thomas, 1941).

This pattern of results led to the formulation of the 'optimal integration hypothesis,' according to which intermodal conflicts are resolved by giving more weight to the modality providing the more reliable information (Ernst & Banks, 2002, Alais & Burr, 2004). This theory has been tested with many different types of cross-modal integration tasks, including visual-haptic (Gepshtein, Burge, Ernst, & Banks, 2005, Guest & Spence, 2003, Miller, 1972) and audio-visual (Alais & Burr, 2004, Ernst & Banks, 2002). It correctly predicts reversal of modality dominance when lowering the quality of information in the generally dominant modality. For example, although the rate of visible flicker does not generally influence the perceived rate of auditory flutter (Shipley, 1964, Welch et al., 1986), vision does exert such an influence when the quality of auditory information is low (Wada, Kitagawa, & Noguchi, 2003).

1.3 Causal relationships

Our previous research has shown, however, that the Schutz-Lipscomb illusion is not based on information quality, but rather on perceived causality. For example, the marimbist's gestures do not affect the perceived duration of non-percussive sounds (such as those produced by a french horn or a clarinet), but they do affect judgments of piano tones (which are also produced by an impact event — that of a hammer striking a string). Likewise, when the causal relationship between the auditory and the visual streams is temporally disrupted (e.g., the percussive sound precedes the visible impact) the gestures fail to influence auditory perception. Furthermore, this manipulation is asynchronous – sounds lagging the moment of impact continue to be influenced by the gestures despite the lack of influence when leading by equal amounts. This is consistent with the physical structure of our environment, in which the speed of sound is substantially less than that of light (Schutz & Kubovy, in press).

These results suggest a causal account of the illusion – gestures integrate (and therefore influence) only the sounds they could have caused. Additionally, they argue against a post-perceptual account (alternatively referred to as a response-bias or cognitive correction), by showing that the illusion is not an artifact of differential amounts of motion between the long and short gestures. Furthermore, they provide a clear way to test whether the illusion is in fact consistent with the optimal integration hypothesis, which in this paradigm predicts that visual influence is related to auditory ambiguity.

In the literature, stimulus ambiguity is operationally measured by response variability. To this end, we compared the variability of duration ratings for percussive sounds to those of the non-percussive sustained sounds (such as the french horn and clarinet) from the previous experiment. The results were inconsistent with the optimal integration hypothesis — ratings were no more variable for the visually influenced (percussive) than the non-influenced (sustained) tones. Furthermore, the visual influence on percussive tones was not related to decay time: slowly decaying sounds were no more influenced than quickly decaying ones (Schutz & Kubovy, in press).

Our interest in the role of causality in cross modal integration is not without precedent. Previous research on issues related to the "unity assumption" (Welch, 1972; see also: Spence, 2007, Welch and Warren, 1980, Welch, 1999, Vatakis and Spence, 2008, Vroomen, 1999) and the "identity decision" (Bedford, 2001a, 2001b, 2004) represent similar thinking. Such work explores the conditions under which cross-modal influences can occur, and suggests that integration requires a mechanism for inferring that auditory and visual information originate from the same event. We agree that such an "identity decision" is a requirement for integration, and posit that causality serves as one of the primary cues by which it can be triggered.

1.4 Motivation for this study

It is well known that point-light displays are capable of conveying biological motion (Johansson, 1973), and can be effectively used in studies of cross-modal interactions (Arrighi, Alais, & Burr, 2006, Saygin, Driver, & de Sa, 2008). Therefore, we were curious whether they could be used within this paradigm as well. Our motivations were two-fold: (1) to better understand the particular visual cues driving this illusion, and (2) to facilitate future research by establishing the feasibility of using abstractions offering obvious methodological advantages with respect to manipulability and control.

2. METHODS AND PROCEDURE

The original videos showed percussionist Michael Burritt performing notes using long and short gestures on the marimba. They were made with a Cannon GL1 video camera and Audio-Technica AT4041 microphones. Here, we used two types of gestures (long and short) taken from the original stimuli, presenting them at one of three levels of abstraction (shown in Figure 1) for a total of six visual stimuli. The conditions were:

1. Video. Movies displaying the marimbist performing single notes at 3 pitch levels using either long or short gestures.
2. Skeleton. A point-light skeleton version of these videos consisting of 4 white dots connected by line-segments against a black background. The 4 dots tracked the shoulder, elbow, hand, and mallet head.
3. Dot. A reduced version of the point-light skeleton showing one dot, tracking the mallet head.

2.1 Stimuli

We made the animations by tracking the horizontal and vertical location of four key points (shoulder, elbow, hand, and mallet tip) from the original videos on a frame-by-frame basis, using the program GraphClick[1]. Each of these joint locations was rendered as a white dot, with consecutive joints connected by white lines (bottom left panel of Figure 1). These animations were generated in real time on a trial-by-trial basis using custom designed software[2]. The animations contained no further information, and therefore the struck object (originally a marimba bar) was not represented. However, from the motion of the striking implement alone it was clear to us (and to participants in a pilot experiment) that the motion represented an impact event.

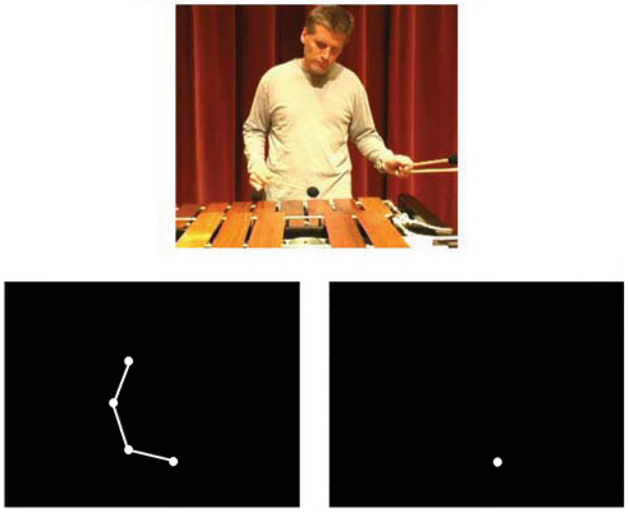


Figure 1:
Top: The video stimuli, consisting of videos displaying a marimbist playing with long and short gestures.
Bottom left: The skeleton stimuli, consisting of 4-point versions of the video stimuli. The points tracked the performer's shoulder, elbow, hand, and mallet head.
Bottom right: The single point dot stimuli, tracking only the position of the mallet head.

For auditory stimuli we used natural and damped marimba tones from the three pitch levels in the original experiment. For the natural tones, the marimbist allowed the sound to decay naturally, whereas for damped tones, he manually damped the bar after striking. One example of each type (damped, natural) was chosen for three pitch levels: E1 (~82 Hz), D4 (~587 Hz), and G5 (~1568 Hz), for a total of six auditory stimuli. We crossed (i.e., took the

---

[1] http://www.arizona-software.ch/graphclick

[2] Designed and implemented by Simeon Fitch of Mustard Seed Software http://www.mseedsoft.com

Cartesian product of) these six visual and six auditory streams, to produce 36 stimuli (samples of the stimuli used can be viewed online at http://sites.google.com/site/schutzresearch/virtualmarimbist).

2.2 Procedure

Thirty-eight participants[3] (receiving course credit) saw the stimuli on a computer screen, and rated the duration of the tone under two conditions: audio-visual and audio-alone. They rated the duration of each using an on-screen slider with endpoints labeled “Short” and “Long.” We translated their rating for each trial into an integer in the [0, 100] interval, where 0 stood for short. In the audio-visual condition, they were instructed to base their judgments on the auditory information alone.

To discourage participants from ignoring the screen, we asked them to respond to a second question concerning the level of agreement between the auditory and visual components, using a second slider with endpoints “Low agreement” and “High agreement.” Previous research has shown that asking participants to provide agreement ratings does not interfere with the primary task (Rosenblum & Fowler, 1991, Saldaña & Rosenblum, 1993), a finding that has held throughout our previous work (Schutz & Lipscomb, 2007, Schutz & Kubovy, in press). As the purpose of this agreement rating was only to force attention to both modalities, they were not analyzed and will not be discussed further.

After a warm-up block containing a random selection of stimuli, we presented each of the 36 audio-visual stimuli three times. The 108 trials were organized into three blocks, with each block containing one instance of each stimulus (within-block trial order was randomized independently).

3. RESULTS AND DISCUSSION

We analyzed the duration ratings using mixed-effects linear models in which all manipulated variables were treated as fixed effects within participants. Several textbooks (Baayen, 2008, Kreft & Leeuw, 1998, Raudenbush & Bryk, 2002, Snijders & Bosker, 1999) present mixed-effects analyses, which have considerable advantages (Baayen, Davidson, and Bates, in press and Maxwell & Delaney, 2004, Part IV). We report each result in terms of an effect (and its standard error, SE, in parentheses), from which a Cohen effect size, d, can be obtained by dividing the effect by its SE. To these we added a 95% confidence interval, as well as a p-value for a test of the null hypothesis that the effect in question is 0.

The single-point animations replicated the illusion (Figure 2). Overall, visual information affected the duration ratings by an estimated 9.6 (±2.0) rating points (95% CI: [5.6, 13.6], $p \approx 0$[4]). We found no meaningful difference between the visual influence from the full videos and the influence from the 4-point animations (2.8±2.1, 95% CI: [-1.3, 6.8], p=0.2). Similarly, we found no meaningful difference between influence from the full videos and the 1-point animations (1.4±2.0, 95% CI: [-2.6, 5.4], p=0.5). As expected, natural notes were judged longer than damped (by 9.2±2.2, 95% CI: [4.9,13.5], p=0.004). However, this parameter had no effect on the degree of visual influence, as shown by the non-significant interaction between note type and visual stroke type (0.45±2.9, 95% CI: [−5.3, 6.1], p=0.9). This is consistent with the results of Schutz and Lipscomb (2007), as well as our subsequent investigations (Schutz & Kubovy, in press).

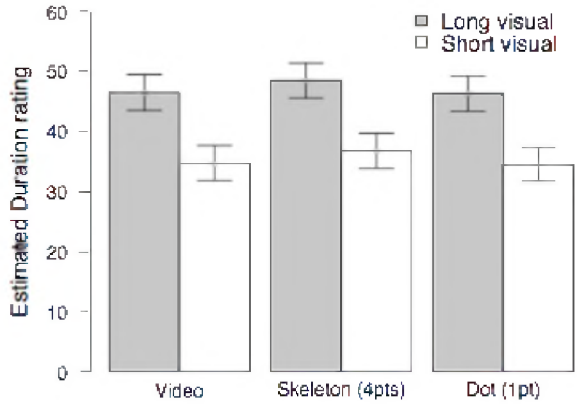


Figure 2: Visual influence was consistent across presentation conditions. The difference in ratings for notes paired with long impact gestures (grey) and short gestures (white) was similar across the video, 4-point skeleton animations, and 1-point dot animation conditions. Error bars indicate a 95% confidence interval.

4. CONCLUSIONS

Given the success of visual (Johansson, 1973) and audio-visual (Saygin et al., 2008) point-light displays in capturing biological motion, we were not surprised to see that the 4-point skeleton produced patterns of visual influence similar to those obtained in the video condition. However, we were pleased to discover that a single dot was sufficient to replicate the original effect (in a different experiment using only the single-dot animations, most participants reported informally that the motion appeared to depict an impact). These results attest to the robustness of the illusion, suggesting interesting possibilities for future research. Furthermore, they are consistent with our conjecture that a crucial component of the illusion is the perception of an audio-visual causal relationship.

We believe these results will be of potential interest for at least two reasons. First, because the illusion has already proven informative with respect to our understanding of

---

[3] Because previous research within this paradigm found no meaningful difference between musically trained (Schutz & Lipscomb, 2007) and musically untrained (Schutz & Kubovy, in press) participants, we did not record any information regarding years of musical training.

[4] Here we use ≈ 0 to indicate that within the computational precision available, for all practical purposes the value is not meaningfully different from 0.

sensory integration, we believe there will be interest in its further study. Consequently, reducing the visual component of the illusion to a single dot will allow us to carry out controlled investigations to discover which aspects of the visual stimulus are responsible for the illusion. This will prove valuable in understanding how the perceptual system integrates sensory information across modalities, and further explore the role of causality as a cue for triggering the identity decision.

Second, we believe that a greater understanding of and appreciation for the role of visual information in shaping the musical experience holds clear artistic value. Given that some expert musicians use vision to strategically enhance audience interest (Broughton & Stevens – in press), communicate musical structure (Vines et al., 2006), and manipulate evaluations of performance quality (Wapnick et al., 1998), deepening our understanding of this process will be helpful to both performers and audiences alike. Future experiments using abstract versions of the original striking gestures will help illuminate how to best use this information to enhance the quality of performer-audience communication.

ACKNOWLEDGMENTS

We thank S. Fitch of Mustard Seed Software for designing the program used to run the experiment, and J. Armontrout for his help in producing stimuli and running participants. This research was supported in part by a grant from NIDCD R01 DC 005636 (Michael Kubovy, PI).

REFERENCES

Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology, 14*, 257–262.

Arrighi, R., Alais, D., & Burr, D. (2006). Perceptual synchrony of audiovisual streams for natural and artificial motion sequences. *Journal of Vision, 6*, 260–268.

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics*. Cambridge, UK: Cambridge University Press.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*, 390–412.

Bedford, F. L. (2001a). Object identity theory and the nature of general laws. *Cahiers de Psychologie Cognitive—Current Psychology of Cognition, 20*, 277–293.

Bedford, F. L. (2001b). Towards a general law of numerical/object identity. *Cahiers de Psychologie Cognitive—Current Psychology of Cognition, 20*, 113–175.

Bedford, F. L. (2004). Analysis of a constraint on perception, cognition, and development: One object, one place, one time. *Journal of Experimental Psychology: Human Perception and Performance, 30*, 907–912.

Bertelson, P., & Radeau, M. (1981). Cross-modal bias and perceptual fusion with auditory-visual spatial discordance. *Perception & Psychophysics, 29*, 578–584.

Bertelson, P., Vroomen, J., de Gelder, B., & Driver, J. (2000). The ventriloquist effect does not depend on the direction of deliberate visual attention. *Perception & Psychophysics, 62*, 321–332.

Broughton, M., & Stevens, C. (in press). Music, movement and marimba: an investigation of the role of movement and gesture in communicating musical expression to an audience. *Psychology of Music.*

Dahl, S., & Friberg, A. (2007). Visual perception of expressiveness in musicians' body movements. *Music Perception, 24*(5), 433-454.

Davidson, J. (1993). Visual perception of performance manner in the movements of solo musicians. *Psychology of Music, 21.*

Davidson, J. (1994). Which areas of a pianist's body convey information about expressive intention to an audience? *Journal of Human Movement Studies, 26*, 279–301.

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature, 415*, 429–433.

Gepshtein, S., Burge, J., Ernst, M. O., & Banks, M. S. (2005). The combination of vision and touch depends on spatial proximity. *Journal of Vision, 5*, 1013–1023.

Gillespie, R. (1997). Ratings of violin and viola vibrato performance in audio-only and audiovisual presentations. *Journal of Research in Music Education, 45*, 212–220.

Guest, S., & Spence, C. (2003). Tactile dominance in speeded discrimination of textures. *Experimental Brain Research, 150*, 207–207.

Hidalgo-Barnes, M., & Massaro, D. W. (2007). Read my lips: an animated face helps communicate musical lyrics. *Psychomusicology, 19*, 3-12.

Jack, C. E., & Thurlow, W. R. (1973). Effects of degree of visual association and angle of displacement on the "ventriloquism" effect. *Perceptual and Motor Skills, 37*, 967–979.

Jackson, C. (1953). Visual factors in auditory localization. *Quarterly Journal of Experimental Psychology, 5*, 52–65.

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics, 2*, 201–211.

Kreft, I., & Leeuw, J. de. (1998). *Introducing multilevel modeling*. London, UK: Sage.

Maxwell, S. E., & Delaney, H. D. (2004). *Designing experiments and analyzing data: A model comparison perspective* (2nd ed.). Mahwah, NJ, USA: Erlbaum.

Miller, E. A. (1972). Interaction of vision and touch in conflict and nonconflict form perception tasks. *Journal of Experiment Psychology, 96*, 114–123.

Phillips-Silver, J., & Trainor, L. J. (2005). Feeling the beat: Movement influences infant rhythm perception. *Science, 308*, 1430.

Phillips-Silver, J., & Trainor, L. J. (2007). Hearing what the body feels: Auditory encoding of rhythmic movement. *Cognition*, *105*, 533-546.

Raudenbush, S. W., & Bryk, A. S. (2002). *linear mode: Applications and data analysis methods*. London, UK: Sage.

Rosenblum, L. D., & Fowler, C. A. (1991). Audiovisual investigation of the loudness-effort effect for speech and nonspeech events. *Journal of Experimental Psychology: Human Perception and Performance*, *17*, 976–985.

Saldaña, H. M., & Rosenblum, L. D. (1993). Visual influences on auditory pluck and bow judgments. *Perception & Psychophysics*, *54*, 406–416.

Saygin, A. P., Driver, J., & de Sa, V. R. (2008). In the footsteps of biological motion and multisensory perception. *Psychological Science*, *19*, 469–475.

Schutz, M. (2008). Seeing music? What musicians need to know about vision. *Empirical Musicology Review*, *3*, 83–108.

Schutz, M., & Kubovy, M. (in press). Causality and cross-modal integration. *Journal of Experimental Psychology: Human Perception and Performance*.

Schutz, M., & Lipscomb, S. (2007). Hearing gestures, seeing music: vision influences perceived tone duration. *Perception*, *36*, 888–897.

Shipley, T. (1964). Auditory flutter-driving of visual flicker. *Science*, *145*, 1328–1330.

Snijders, T., & Bosker, R. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. London, UK: Sage.

Spence, C. (2007). Audiovisual multisensory integration. *Acoustic Science & Technology*, *28*, 61–71.

Thomas, G. (1941). Experimental study of the influence of vision on sound localization. *Journal of Experimental Psychology*, *28*, 163–175.

Thompson, W. F., Graham, P., & Russo, F. A. (2005). Seeing music performance: Visual influences on perception and experience. *Semiotica*, *156*, 203–227.

Thompson, W. F., & Russo, F. A. (2007). Facing the music. *Psychological Science*, *18*, 756–757.

Thompson, W. F., Russo, F. A., & Quinto, L. (2008). Audio-visual integration of emotional cues in song. *Cognition & Emotion*, *22*, 1457–1470.

Vatakis, A., & Spence, C. (2008). Evaluating the influence of the 'unity assumption' on the temporal perception of realistic audiovisual stimuli. *Acta Psychologica*, *127*, 12–23.

Vines, B. W., Krumhansl, C. L., Wanderley, M. M., & Levitin, D. J. (2006). Cross–modal interactions in the perception of musical performance. *Cognition*, *101*, 80–113.

Vroomen, J. (1999). Ventriloquism and the nature of the unity decision. Commentary on Welch. In G. Aschersleben, T. Bachmann, & J. Müsseler (Eds.), *Cognitive contributions to the perception of spatial and temporal events* (pp. 389–393). Amsterdam: Elsevier.

Wada, Y., Kitagawa, N., & Noguchi, K. (2003). Audio-visual integration in temporal perception. *Psychophysiology*, *50*, 117–124.

Walker, J. T., & Scott, K. J. (1981). Auditory-visual conflicts in the perceived duration of lights, tones, and gaps. *Journal of Experimental Psychology: Human Perception and Performance*, *7*, 1327–1339.

Wapnick, J., Darrow, A.-A., Mazza, J. K., & Dalrymple, L. (1997). Effects of physical attractiveness on evaluation of vocal performance. *Journal of Research in Music Education*, *45*, 4270-479.

Wapnick, J., Mazza, J. K., & Darrow, A.-A. (1998). Effects of performer attractiveness, stage behavior, and dress on violin performance evaluation. *Journal of Research in Music Education*, *46*, 510–521.

Welch, R. B. (1972). The effect of experienced limb identity upon adaptation to simulated displacement of the visual field. *Perception & Psychophysics*, *12*, 453–456.

Welch, R. B. (1999). Meaning, attention, and the 'unity assumption' in the intersensory bias of spatial and temporal perceptions. In G. Aschersleben, T. Bachmann, & J. Müsseler (Eds.), *Cognitive contributions to the perception of spatial and temporal events* (pp. 371–387). Amsterdam, Netherlands: North-Holland/Elsevier Science Publishers.

Welch, R. B., DuttonHurt, L. D., & Warren, D. H. (1986). Contributions of audition and vision to temporal rate perception. *Perception & Psychophysics*, *39*, 294–300.

Welch, R. B., & Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin*, *88*, 638–667.

EXPRESSING TONAL CLOSURE IN MUSIC PERFORMANCE: AUDITORY AND VISUAL CUES

Donovon Keith Ceaser, William Forde Thompson[1], and Frank Russo[2]
1 - Department of Psychology, Macquarie University, Sydney, Australia
2 - Department of Psychology, Ryerson University, Toronto, Canada

ABSTRACT

We examined whether musical performers communicate tonal closure through expressive manipulation of facial expressions and non-pitch features of the acoustic output. Two musicians hummed two versions of Silent Night: one ended on the tonic of the scale and exhibited tonal closure; the other ended on the dominant and was therefore tonally unclosed. In Experiment 1, video-only recordings of the hummed sequences were presented to 15 participants, who judged whether the (imagined) melody was closed or unclosed. Accuracy was reliably above chance, indicating that the musicians expressed tonal closure in facial expressions and listeners decoded these cues. Experiment 2 was conducted to determine whether musicians also communicate tonal closure in acoustic attributes other than pitch. All tones in the hummed melodies were pitched-shifted to a constant mean value, but performances still differed in loudness, microtonal pitch variation, timing, and timbre. Participants judged whether audio-only recordings were closed or unclosed. Accuracy was not above chance overall, but was marginally above chance for judgement of one of the two singers. Results suggest that tonal closure can be mapped onto non-pitch aspects of performance expression, but is primarily restricted to the use of facial expressions.

RESUME

Nous avons examiné la communication de conclusion tonale chez les performeurs musicaux, et spécifiquement si elle se fait à travers les manipulations d'expressions faciales et de caractéristiques d'information acoustique autres que la hauteur du son. Deux musiciens ont fredonné une différente version de la mélodie de Sainte Nuit : l'une d'elle terminant sur la tonique de la gamme et donc ayant une conclusion tonale; et l'autre terminant sur la dominante et donc n'ayant pas de conclusion tonale. Lors de l'expérience 1, des extraits d'enregistrements vidéo sans sons des mélodies fredonnées ont été présentés à 15 participants, qui ont ensuite effectué un jugement selon la présence ou l'absence de conclusion tonale dans la mélodie (imaginée). Le niveau d'exactitude des participants excédait le niveau du hasard, indiquant que les musiciens exprimaient la conclusion tonale à travers de leurs expressions faciales et que ces indicateurs furent décodés par leurs auditeurs. L'expérience 2 a été effectuée afin de déterminer si les musiciens communiquaient aussi la conclusion tonale à travers les caractéristiques acoustiques autres que la hauteur du son. La hauteur de tous les sons musicaux dans les mélodies fredonnées a été changée à une hauteur moyenne constante, mais les performances étaient toutefois différentes en tant que force sonore, variations de hauteur microtonale, synchronisation, et timbre. Les participants ont effectué un jugement selon la présence ou l'absence de conclusion tonale dans des extraits d'enregistrement audio des mélodies. Dans l'ensemble, le niveau d'exactitude des participants n'excédait pas le niveau du hasard, mais l'exactitude des jugements envers l'un des deux musiciens excédait à peine le niveau du hasard. Ces résultats suggèrent qu'une conclusion tonale peut être établie à l'aide d'attributs d'expressions de performance autres que la hauteur des sons, mais qu'elle est généralement établie à l'aide d'expressions faciales.

1.0 Introduction

The performer is the middle link in a system of communication that extends from the composer who notates musical ideas onto a score to the listener who perceives, interprets, and experiences an acoustic realization of these ideas. The performer not only converts notated symbols (musical notes) into sound; they interpret the musical structure and convey this interpretation expressively (Palmer, 1997). That is, a critical goal of music performance is to communicate the structure of the music clearly to listeners (Gabrielsson, 1999).

A number of researchers have investigated how musicians manipulate acoustic attributes in order to elucidate musical structure. For example, musicians often manipulate timing, dynamics (changes in sound intensity) and tempo (musical speed) in order to communicate or emphasize the temporal structure of music (Clarke, 1982; 1985; 1988; Povel, 1977; Repp 1992; Sloboda, 1983, 1985; Todd, 1985). Some expressive actions, such as a crescendo (getting louder), are easier to detect than others, such as a diminuendo (getting softer)

(Nakamura, 1987; Senju & Ohgushi, 1987). More generally, listeners may not decode all aspects of musical structure, so the communicative intentions of performers may be only partially realized.

Recently, researchers have examined how expressive information from auditory and visual modalities is combined to communicate musical structure (for a review, see Schutz, 2008). In experiments that involve auditory-only, visual-only, and audio-visual conditions, participants show increased sensitivity to musical structure in audio-visual conditions (Krumhansl & Schenck, 1997; Vines et al., 2006). Indeed, visual information is often surprisingly informative. When participants are asked to make judgements about expressive intentions, participants may perform better in a visual-only condition than in an audio-only condition (Davidson, 1993). Thompson, Graham, and Russo (2005) reported that facial expression influence perceptions of interval size, dissonance, and emotional connotations. Thompson, Russo and Quinto (2008) confirmed that listeners unconsciously and automatically consider cues arising from the facial expressions of performers when forming an emotional interpretation of music. Schutz & Lipscomb (2007) reported that the judged duration of a marimba note was longer if accompanied by a video recording of a performer making a "long gesture" when striking the note than if accompanied by a video recording of a "short gesture." Finally, Thompson & Russo (2007) found that observers were able to differentiate the size of sung intervals on the basis of visual information alone. Judgements of size were well correlated with veridical interval size and appeared to be based on a combination of head movement, eyebrow raising and mouth opening. The findings underscore the importance of the visual modality for the perception of music.

In the current investigation, we examined the manner in which musicians modify their use of performance expression in order to communicate tonal closure, both in their facial expressions and the expressive manipulation of sound attributes other than pitch. Musical closure is a general term that refers to the perception that a musical phrase or section has come to a permanent or temporary end. Tonal closure refers to the contributions of tonality to this effect, and occurs when a musical phrase ends on a tonally stable pitch, such as the tonic note or chord (e.g., as in a perfect cadence). Although tonal closure is an aspect of pitch structure, performers may introduce non-pitch cues to reinforce or clarify the perception of tonal closure, including temporal cues (ritards, duration lengthening), cues related to sound intensity (diminuendos), and even visual cues associated with facial expressions and gestures. Two experiments were conducted to determine whether tonal closure is mapped onto such non-pitch aspects of performance expression. Experiment 1 considered expressive uses of facial expression; Experiment 2 considered the expressive manipulation of acoustic cues.

2.0 EXPERIMENT 1

Experiment 1 examined the role of facial expressions in the communication of tonal closure. If listeners can perceive the difference between the tonally closed and unclosed clips by viewing the faces of singers alone (i.e. with no acoustic signal), then such a finding would corroborate and extend previous reports that facial expressions of performers reflect and communicate pitch structure.

2.1 Method

Participants. Fifteen undergraduate students (7 males and 8 females) were recruited from the University of Toronto community, ranging in age from 18 to 27 years (mean = 22.36), and possessing between zero and 14 years of musical training (mean = 7.53 years). All participants reported normal hearing.

Stimuli. The stimuli consisted of sung recordings of the melody of Silent Night. This melody was selected because it is popular and highly recognizable. The final phrase of the song ("Sle-ep in hea-ven-ly peace") follows the diatonic scale pattern of 8-5-3-5-4-2-1 and implies a tonic-dominant-tonic (I-V-I) harmonic accompaniment.

Three experienced singers (one male and two females) were recruited to hum two versions of the final phrase of the melody. One version (original) ended on the tonic of the scale (as in the original melody) and was tonally closed; the other version ended on the dominant of the scale and was therefore tonally unclosed. More specifically, in the unclosed version, the last four notes of the original melody were reversed (1, 2, 4, 5), such that the final note of the melody (5) is the dominant. Note that the size of the final melodic interval (major second) was the same in both versions. Singers were instructed to stare directly into the camera as they hummed the tunes. Each singer hummed six examples of the closed and unclosed versions each.

From the audiovisual recordings of these performances, 3 examples were selected for each singer and each condition of closure, yielding 18 recordings (3 examples X 2 conditions of closure X 3 singers). The video recordings (no sound) were extracted for use in Experiment 1; audio recordings (no video) were used in Experiment 2. Each clip was edited so that it began with the singer's inhaling breath, and ended after the first beat of the final bar.

Procedure. Edited video clips were loaded into the Experiment Creator program. Each of the 18 clips was presented three times, yielding 54 trials presented in random order. Participants were tested in a sound-attenuated booth. They were told that they would be viewing video-only recordings of Silent Night, and that there would be two versions: a conventional version and a nonconventional version with an alternate ending. For each presentation, the task was to judge whether the imagined melody was closed or unclosed (forced choice). The experiment took between 20 and 30 minutes to complete. Following the experimental trials, participants were administered a questionnaire related to the cues that they used to make their judgements.

2.2 Results

The proportion of closed responses for closed and open clips was 0.61 and 0.37 respectively. ANOVA was conducted with repeated measures on two factors: Singer (3 levels) and Closure (2 levels). There was a significant effect of Closure, $F(1, 14) = 21.69$, $p < 0.001$, confirming that performers expressed musical closure in their facial expressions and participants successfully decoded these cues.

To further explore the role of facial expressions in the communication of musical closure, we examined the closed and unclosed clips with the highest decoding accuracy, focusing on three features: width of eye opening (squint), eyebrow height, and head height. These features were informally scrutinized because a post-experiment interview revealed that participants identified them as relevant to their judgements of closure. Measurements were taken from two video stills for each singer's clip, one at the onset of the first note (after the initial breath), and again at the end of the last note. Eye opening was measured from the top of the lower lid to the bottom of upper lid. Eyebrow position was measured from the top of the eyebrow to the bottom of upper lid. Head movement was measured by comparing the position of the chin at the beginning and end of the video clip. By subtracting the value obtained in the second still from the first, we obtained a measurement of the difference in facial position.

Table 1 displays the change in measurements for each feature from the beginning to the end of each clip. The amount of change is shown in centimeters. Positive values for eye opening indicate that the eyes were wider at the end of the clip than at the beginning of the clip; positive values for eyebrow height indicate that the eyebrows were raised to a greater degree at the end of the clip than at the beginning of the clip; positive values for head height indicate that the chin was at a higher position at the end of the clip than at the beginning of the clip. Note that the values in the Table indicate measurements from the computer monitor and understate actual movements. A comparison of values for closed and unclosed clips suggests that, in comparison with closed clips, unclosed melodic endings were associated with reduced widening of the eyes, raised eyebrows, and raised chin.

The greatest movements were observed for eyebrow height, consistent with findings on the use of facial expressions during music performance (Thompson & Russo, 2007). Specifically, for closed clips the eyebrows were raised more at the beginning of the clip than at the end of the clip, reflecting movement from instability to stability and closure; for unclosed clips the eyebrows were raised more at the end of the clip than at the beginning of the clip, reflecting the lack of closure at the end of the clip.

2.3 Discussion

Participants reliably differentiated between tonally closed and unclosed clips based on visual cues alone. This finding confirms that musicians can express musical closure in the use of facial expressions and viewers are able to decode these expressive actions. A comparison of movement analyses for closed and unclosed clips suggests that eyebrow and head movement may be especially relevant to this effect. In comparison with facial movements associated with closure, facial expressions associated with non-closure involved raised eyebrows and chin relative to initial positions. These differences in the relative height of facial features may reflect the degree to which each ending violated or confirmed expectancies. When expectancies were fulfilled with a closed ending, the chin and eyebrows were not raised, or were lowered. When expectancies were violated with an unclosed ending, the chin and eyebrows were raised.

3.0 EXPERIMENT 2

We next assessed whether listeners can decode tonal closure on the basis of the acoustic aspects of performance expression other than pitch. Stimuli consisted of the audio-only recordings of "Silent Night" taken from the recordings described in Experiment 1. Because tonal closure is determined by pitch structure, it was necessary to disentangle the direct cues provided by pitch structure from the indirect cues provided by performance expression. To this end, Protools software was used to shift all tones in each performance to a constant mean pitch level. This procedure retaining expressive variations in intensity, timbre, and timing.

3.1 Method

Participants. 25 undergraduate students (12 males and 13 females) ranging in age from 18 to 28 (mean = 19.44) years were recruited from the University of Toronto community. Participants had between zero and 18 years of musical training (mean = 3.66 years). All participants reported normal hearing.

Stimuli. Stimuli were a subset of the recordings described in Experiment 1. We selected three exemplars of closed and unclosed clips performed by a female singer (M) and two exemplars of closed and unclosed performed by a male singer (T), yielding 10 clips for use in the experimental trials. Each clip was presented 5 times, yielding 50 trials. Using Protools,

Table 1. Change in the position of facial features from initial to final notes of closed and unclosed clips (in cm).

Facial feature	Closed	Unclosed	Difference in change
Eye width	+0.64	+0.38	-0.26
Eyebrow height	-2.29	+0.38	+2.67
Head height	-0.64	+0.51	+1.15

each audio file was altered such that all tones were set to the same pitch. This procedure yielded a set of monotonic phrases that differed in performance expression. That is, they differed in terms of expressive variation in intensity, timing, and vocal timbre throughout the phrase. All pitches were adjusted to the mediant (third scale degree) of the key of the melody. This pitch was chosen so that the degree of pitch shifting required was approximately equal for the closed and unclosed conditions.

Procedure. Experiment Creator was used to control presentation of the 50 trials and collection of responses. Participants were seated in front of a Macintosh computer in a sound-attenuated booth. They were first introduced to one example of an original and altered (pitch shifted) version of a closed and unclosed exemplar from each singer (i.e., four familiarisation trials). These trials were used to familiarise participants with the nature of the pitch-shifted stimuli and were excluded from the experimental trials. They were then given 10 practice trials: 5 unclosed and 5 closed trials. After each presentation, they indicated whether the clip was derived from a musically closed or unclosed melody (forced choice). No feedback was provided. The 50 experimental trials were then presented in random order and took about 30 minutes to complete. Following the experimental trials, participants were administered a questionnaire related to the cues that they used to make their judgements.

3.2 Results

The proportion of closed responses for closed and open clips was 0.58 and 0.53 respectively. ANOVA was conducted with repeated measures on two factors: Closure (closed or unclosed) and Example (1-5). There was no significant main effect for Closure, $F(1, 24) < 1.0$, nor for Example, $F(1, 24) < 1.0$. The interaction between Closure and Example was significant, $F(1, 24) = 7.00$, $p < 0.05$. The interaction was explored by running separate analyses of variance for each of the two singers. These analyses revealed that there was a marginally significant effect of Closure for the female singer, $F(1, 24) = 4.16$, $p < 0.055$. For this singer, the proportion of closure responses was higher for trials derived from tonally closed sequences ($M = 0.64$, $SE = 0.059$) than for trials derived from tonally unclosed sequences ($M = 0.43$, $SE = 0.051$). There was no significant effect of Closure for the male singer, $F(1, 24) = 3.21$, n.s. This pattern of results implies that some singers are more capable than others at using non-pitch expressive cues to communicate tonal closure.

Follow-up acoustic analyses revealed that for the female singer, final tones of tonally open productions were more variable in pitch and amplitude that final tones of tonally closed productions (i.e., perturbations within tones were more frequent and extreme). For both singers, the final tones of tonally open productions were marginally longer (~ 125 msec) and possessed higher harmonicity (~ 1 to 2 dB; i.e., were less hoarse) than their closed counterparts.

3.3. Discussion

Across the two singers, participants were unable to differentiate closed and unclosed audio clips when pitch cues were absent. However, closure was communicated with marginal significance by one of the two singers. Because closed and unclosed versions were equivalent in nominal pitch, judgements of closure for this singer must have been based on expressive cues related to intensity, microtonal pitch variation, timing, and vocal timbre. Thus, even though tonal closure is defined and communicated by pitch stability, other expressive cues may have also been introduced by the performer to convey closure. It is unclear how well these cues were perceived by viewers. The data do not allow us to determine which, if any, of these cues were used to decode tonal closure. However, the post-experiment questionnaire suggested that the unclosed ending seemed “faster” than the closed ending (expressive timing); closed endings had more “flow” than unclosed endings (expressive phrasing); and the tonal quality (vocal timbre) differed between closed and unclosed endings.

4.0 GENERAL DISCUSSION

The results of this investigation indicate that performers effectively communicate tonal closure through the use of non-pitch cues. These expressive cues are reliably communicated in the facial expressions of performers, and may be conveyed by certain performers in the expressive manipulation of non-pitch acoustic features. The findings corroborate and extend the growing evidence that performers make use of both auditory and visual aspects of music, and that “listeners” attend not only to musical sounds; they are also “viewers” that integrate visual information with the acoustic signal. Surprisingly, the expressive use of facial expressions was more reliably effective at communicating tonal closure than the acoustic cues introduced by the performer. In particular, the performer’s capacity to communicate tonal closure through acoustic aspects of performance expression was surprisingly limited. Quite possibly, tonal closure may be most effectively communicated through musical structure and facial expressions, but to a limited extent through acoustic cues introduced by the performer.

It is important to acknowledge that closed and unclosed melodies differed in a number of features, including the tonal stability of the final tone (tonal closure), the overall pitch height of the final tone, and the contour of the final four tones. Any of all of these differences may have influenced the facial expressions adopted by performers, as well as expressive variation in singing. For example, melodic contour varied with our manipulation of closure (closed melodies had a downward contour, unclosed melodies had an upward contour). Along with differences in the tonal stability of the final note of the melody, such contour differences likely contributed to the facial expressions adopted by performers, as well as the expressive variation in their singing. Quite possibly, production constraints associated with differences between closed and unclosed melodies contributed to the mapping of closure onto facial expressions and expressive variation on singing (difficulty in singing upward versus downward melodies, certain scale notes being more stable in memory and easier to produce than others). Nonetheless, the aim of this investigation was to document the ability of viewers to detect closure from facial expressions and other expressive actions, and they were indeed successful with this task. As many cues contribute to tonal closure in the acoustic realization of music, it is reasonable to assume that multiple cues also contributed to the perception of tonal closure in our investigation.

More generally, it should be emphasized that musical closure invariably arises from a collection of convergent cues and it is not possible to manipulate one aspect of musical structure (e.g., tonal stability) without affecting other potential features (e.g., pitch height, pitch interval). As such, the relative contribution of potential cues to melodic closure cannot be determined from the current results. Nonetheless, because our participants explicitly judged closure, only those expressive actions that were relevant to closure should have influenced judgements, and therefore cannot be construed as an experimental confound. That is, the presence of multiple differences between conditions is compatible with the conclusion that melodic closure is mapped onto facial expressions and other expressive actions and communicated to perceivers.

Participants were still able to distinguish between closed and unclosed melodies when provided with visual information alone. The finding extends previous evidence that facial features reflect music structure, including melodic, rhythmic, and tonal structure (Thompson et al., 2005; Thompson & Russo, 2007). Self-reports by participants, corroborated by an informal analysis of two videos, suggested that eyebrow and head movements were associated with expressions of closure and non-closure. In particular, relative to closed endings, raised eyebrows and a higher head position signaled an unclosed ending. As argued, facial expressions reflected multiple ways in which closed and unclosed versions differed, but only relevant aspects of facial expressions should have influenced direct judgements of closure. Although not the focus of the current investigation, future research may include a more standardized set of facial measurements. More generally, research is needed to determine whether facial expressions reflect tonal closure in a way that is distinct from the manner in which they reflect other aspects of musical closure.

The cues used in expressing musical closure may reflect a general system of symbolic communication whereby facial and acoustic features express internal states. Across several cultures, lowered eyebrows signal increased dominance whereas raised eyebrows signal appeasement (Keating et al., 1981). Such cues are analogous to findings for musical and spoken stimuli. In music, melodies played at a higher pitch height are judged to be less aggressive and more submissive than melodies played at a lower pitch height (Huron, Kinney and Precoda, 2006). In speech, high or rising vocal pitch is associated with a lack of confidence, whereas low or falling vocal pitch is associated with confidence (Bolinger, 1964).

Familiarity with the music, along with sensitivity to closure, may have played a role in determining the current results. All participants were aware that the performer was singing Silent Night. This familiarity may have led participants to form a stable mental representation of the melody along with the facial expressions associated with sung versions of the phrase. Future research should be conducted to confirm that sensitivity to facial expressions of tonal closure is reliable for unfamiliar melodies.

The current results may be interpreted within a multimodal framework, whereby one of two strategies for introducing facial expressions may be invoked: reinforcement or compensation. Using a reinforcement strategy, performers may form a mental representation of musical structure and emphasize that structure through their use of facial expressions. In the current investigation, performers appeared to underscore tonal closure through movements of the eyebrows and head, complementing their expressive manipulations of intensity, timing, and vocal timbre. Using the categories of facial expression outlined by Kurosawa and Davidson (2005), such expressions may be characterized as illustrators because they function to reinforce or illustrate a structural feature of the music. In this case, raising the eyebrows and head at the end of the melody acted to reinforce the lack of tonal closure that was conveyed by the pitch structure.

Alternatively, performers may adopt a compensatory strategy. In this case, performers might evaluate the degree of tonal information available in pitch and rhythmic structure, compare this information to their expressive intentions, and evaluate the discrepancy between the two. Expressive actions are then introduced only to the extent that available cues become sufficiently clear to allow a typical listener to extract the structural information that the performer wishes to convey. For example, performers in this investigation appeared to compensate for a lack of tonal closure in open sequences through final tone lengthening. Again, this process of communicating musical structure may involve the expressive manipulation of acoustic cues but may also include facial expressions, gestures, and other body movements.

REFERENCES

Bolinger, D.L. (1964). Intonation across languages. In: J.H. Greenberg, C.A. Ferguson & E.A. Moravcsik (Eds.), *Universals of Human Language, Vol. 2: Phonology.* Stanford, CA: Stanford University Press, pp. 471-524.

Clarke, E. F. (1982). Timing in the performance of Erik Satie's 'Vexations'. *Acta Psychologica, 50*, 1-19.

Clarke, E. F. (1985). Structure and expression in rhythmic performance. In P. Howell, I. Cross, & R. West (Eds.), *Action and perception in rhythm and music* (pp. 19-33). Stockholm: Publication issued by the Royal Swedish Academy of Music, No. 55.

Clarke, E. F. (1988). Generative principles in music performance. In J. A. Sloboda (Ed.), *Generative processes in music: The psychology of performance, improvisation. and composition* (pp. 1-26). Oxford: Clarendon Press.

Davidson, J. W. (1993). Visual perception of performance manner in the movements of solo musicians. *Psychology of Music*, 21, 103-113.

Gabrielsson, A. (1999). The performance of music. In D. Deutsch (Ed.), *The Psychology of Music (Vol. xvi, pp. 807)*. San Diego, CA US: Academic Press, Inc.

Huron, D., Kinney, D., & Precoda, K. (2006). Influence of pitch height on the perception of submissiveness and threat in musical passages. E*mpirical Musicology Review, 1*, 170-177.

Krumhansl, C. L., & Schenck, D. L. (1997). Can dance reflect the structural and expressive qualities of music? A perceptual experiment on Balanchine's choreography of Mozart's Divertimento no. 15. *Musicae Scientiae, 1*, 63-85.

Kurosawa, K., & Davidson, J. W. (2005). Nonverbal behavior in popular music performance: A case study of The Corrs. *Musicae Scientiae, 9*, 111-133.

Nakamura, T. (1987). The communication of dynamics between musicians and listeners through musical performance. *Perception & Psychophysics, 41*, 525-533.

Palmer, C. (1997). Music Performance. *Annual review of Psychology, 48*, 115-158.

Povel, D. J. (1977). Temporal structure of performed music. Some preliminary observations. Acta Psychologica, 41, 309-320.

Repp, B. H. (1992). Diversity and commonality in music performance: An analysis of timing microstructure in Schumann's "Traeumerei." Journal of the Acoustical Society of America, 92, 2546-2568.

Senju, M., & Ohgushi, K. (1987). How are the player's ideas conveyed to the audience? *Music Perception, 4,* 311-324.

Schutz, M. (2008). Seeing music? What musicians need to know about vision. *Empirical Musicology Review, 3*, 83 –108.

Schutz. M., & Lipscomb, S. (2007). Hearing gestures, seeing music: Vision influences perceived tone duration. *Perception, 36*, 888-897.

Sloboda, J. A. (1983). The communication of musical metre in piano perfomance. *Quarterly Journal of Experimental Psychology, 39*, 273-293.

Sloboda, J. A. (1985). Expressive skill in two pianists: Metrical communication in real and stimulated performances. *Canadian Journal of Psychology, 39*, 273-293.

Thompson, W. F., & Russo, F.A. (2007). Facing the music. *Psychological Science, 18*, 756-757.

Thompson, W. F., Graham, P., & Russo, F. A. (2005). Seeing music performance: Visual influences on perception and influence. *Semiotica, 1(4)*, 177-202.

Thompson, W. F., Russo, F. A. & Quinto, L. (2008). Audiovisual integration of emotional cues in song. *Cognition & Emotion, 22*, 1457-1470.

Vines, B. W., Krumhansl, C. L., Wanderley, M. M., & Levitin, D. J. (2006). Cross-modal interactions in the perception of musical performance. *Cognition, 101*, 80-113.

Perception of Synthetic Vowels by Monolingual Canadian-English, Mexican-Spanish, and Peninsular-Spanish Listeners

Geoffrey Stewart Morrison
School of Language Studies, Australian National University, Canberra, ACT 0200, Australia
e-mail: geoff.morrison@anu.edu.au

Errata

The article published in Volume 34, Number 4 (December 2008), pp. 17–23 contained a number of typesetting errors.

In the English abstract, delete "νγ4105"

Throughout the article, except in the figures and in the first paragraph of page 21, the phonetic symbols were incorrect.

for /ι/ read /**i**/
for /I/ read /ɪ/
for /ε/ read /**e**/
for /Ε/ read /ɛ/
for /ει/ read /**ei**/
for /α/ read /**a**/
for /π/ read /**p**/
for /πα/ read /**pa**/
for /βVπ|/ read /**b**V**p**˺/
for /βιπα/, /βειπα/, and /βεπα/
read /**bipa**/, /**beipa**/, and /**bepa**/
for /βιπ↔/, /βΙπ↔/, /βεπ↔/, and /βΕπ↔/
read /**bipə**/, /**bɪpə**/, /**bepə**/, and /**bɛpə**/

CROSS-MODAL MELODIC CONTOUR SIMILARITY

Jon B. Prince[1], Mark A. Schmuckler[1], and William Forde Thompson[2]
1 - Department of Psychology, University of Toronto, Mississauga, Canada
2 - Department of Psychology, Macquarie University, Sydney, Australia

ABSTRACT

In two experiments participants rated the similarity of melodic contours presented as auditory (melodies) and visual (line drawings) stimuli. Longer melodies were assessed in Experiment 1 (M = 35 notes); shorter melodies were assessed in Experiment 2 (M = 17 notes). Ratings for matched auditory and visual contours exceeded ratings for mismatched contours, confirming cross-modal sensitivity to contour. The degree of overlap of the surface structure (the relative position of peaks and troughs), and the strength and timing of the cyclical information (the amplitude and phase spectra produced by a Fourier analysis) in the contours predicted cross-modal similarity ratings. Factors such as the order of stimulus presentation (auditory-visual or visual-auditory), melody length (long versus short), and musical experience also affected the perceived similarity of contours. Results validate the applicability of existing contour models to cross-modal contexts and reveal additional factors that contribute to cross-modal contour similarity.

RESUME

Au cours de deux expériences des participants ont estimé la similarité des contours mélodiques présentés comme stimuli auditifs (des mélodies) et visuels (des dessins au trait). Des mélodies longues (M = 35 notes) ont été évaluées dans la première expérience; des mélodies courtes (M = 17 notes) ont été évaluées dans la deuxième expérience. Les estimations de similarité des contours auditifs et visuels équivalents étaient plus élevées que les estimations de similarité des contours auditifs et visuels différents, ce qui confirme la sensibilité des participants aux contours représentés par des modalités sensorielles différentes. Le degré de chevauchement de la structure superficielle (la position relative des crêtes et des cuvettes), et la force et le rythme de l'information cyclique (les spectres d'amplitude et de phase obtenus par analyse de Fourier) dans les contours ont prédit pour les modalités sensorielles différentes des estimations de similarité élevées. Certains facteurs tels que l'ordre de la présentation des stimuli (auditif-visuel ou visuel-auditif), la durée de la mélodie (longue ou courte), et l'expérience musicale ont aussi affecté la similarité perçue des contours. Ces résultats déclarent valide l'applicabilité des modèles de contours existants aux différents contextes de modalités sensorielles et dévoilent des facteurs additionnels qui contribuent à la similarité des contours dans ces modalités.

1 INTRODUCTION

Contour, or the overall pattern of ups and downs, is a basic attribute of auditory and visual stimuli. In the case of audition, pitch contour plays an important role in two forms of auditory information: language and music. In language, contour is a primary attribute of speech intonation and contributes to the supralinguistic dimensions of speech. Speech intonation provides cues about emphasis, emotional attitude, and syntactic structure, and it may also facilitate the processing of verbal content in tonal and non-tonal languages (‘t Hart, Collier, & Cohen, 1990; Lieberman, 1967; Pierrehumbert & Hirschberg, 1990; for a review, see Cutler, Dahan, & van Donselaar, 1997). Contour also plays a crucial role in music cognition, providing one of the most important cues for melody recognition and melodic similarity (Dowling, 1978; Dowling & Harwood, 1986; for a more thorough review see Schmuckler, 1999).

1.1 Contour in music cognition

Listeners can recognize familiar melodies even when the intervals of a melody (the specific pitch distance between successive notes) are severely distorted as long as the contour of the melody, or the relative pattern of rises and falls in pitch, remains intact (Deutsch, 1972; Dowling & Hollombe, 1977; Idson & Massaro, 1978; White, 1960). Moreover, contour is critical for discrimination between (Watkins, 1985) and memory of (Dowling, 1978) novel melodies, especially when there is no tonal framework to aid in constructing a representation of the melody in memory (Dowling, 1991; Dowling & Fujitani, 1971; Francès, 1988; Freedman, 1999). Children and infants also preferentially use contour over more specific, local information when listening for changes in melodies (Chang & Trehub, 1977; Morrongiello, Trehub, Thorpe, & Capodilupo, 1985; Pick, Palmer, Hennessy, & Unze, 1988; Trehub, Bull, & Thorpe, 1984).

Research has elucidated how listeners segment melodies into meaningful units, store this information in memory and subsequently use it for recognition. Pitch accents created by contour reversals (i.e., changes in pitch direction) contribute to the perceptual segmentation of both melodies and speech (Bregman & Campbell, 1971; Deutsch & Feroe, 1981; Frankish, 1995; Thomassen, 1982), and also direct attention to important notes within a melody (Boltz & Jones, 1986; Boltz, Marshburn, Jones, & Johnson, 1985; Jones & Boltz, 1989; Jones & Ralston, 1991; Monahan, Kendall, & Carterette, 1987). Indeed, alterations to a melody are more obvious when they involve a contour reversal (Dyson & Watkins, 1984; Jones & Ralston, 1991; Monahan et al., 1987; Peretz & Babaï, 1992), and recognizing novel melodies is more challenging as the contour becomes more complex (Boltz et al., 1985; Cuddy, Cohen, & Mewhort, 1981; Morrongiello & Roes, 1990; but see Croonen, 1994). According to Narmour's (1990) implication-realization model, contour reversals represent a crucial feature of melodic structure and listeners expect them to occur after large melodic leaps.

Contour also plays a critical role in melodic similarity. Eiting (1984), for instance, found that similarity judgements of short (3-note) melodic sequences depended primarily on contour. Contour also contributes significantly to similarity judgements of 7-note melodies (Quinn, 1999) and 12-note melodies (Schmuckler, 1999). Categorization of 7-note melodies varying in contour, rhythm, timbre and loudness is almost exclusively determined by the contour (Schwarzer, 1993). More generally, contour is a salient feature in naturalistic passages of music (Halpern, Bartlett, & Dowling, 1998; Lamont & Dibben, 2001).

1.2 Cross-modal melodic contour

Melodic contour can be represented in both auditory and visual modalities. Notated music exemplifies visual depictions of melodic contour. In a musical staff, higher and lower pitches correspond to higher and lower spatial positions on the musical score, allowing a visual analogue of pitch contour. Musical notation in many cultures perpetuates this analogy (and implied relation) by representing pitch in the vertical spatial dimension. Even gross simplifications of Western musical notation preserve this relation – composer and theorist Arnold Schoenberg's (1967) line drawings of Beethoven piano sonatas notated pitch contours in terms of ups and downs based on the frequencies of the notes.

The spatial mapping of pitch height is a pervasive and robust phenomenon. The human auditory system translates the sensation of frequency of vibration (caused by the fluctuations in air pressure from a sound-emitting object) into the psychological construct of pitch. Whether through cultural learning or innate bias, we experience notes of higher and lower pitch according to higher and lower frequencies of vibration, respectively. Pitch is described as having "height" (Bachem, 1950; Ruckmick, 1929; Shepard, 1982), and pitch relations, which form the basis for contours, are described as moving "up" and "down," despite the fact that pitch itself is a function of time (i.e., vibrations per second) not space. In other words, listeners automatically represent pitch height spatially, such that they perceive higher pitches to be above (in a spatial sense) lower pitches. For example, in a pitch height comparison task, congruency between the spatial organization of response keys and the relative pitch height of isolated tones improves listeners' reaction time (incongruency is detrimental), regardless of the degree of musical expertise (Lidji, Kolinsky, Lochy, & Morais, 2007). Furthermore, both musicians and untrained listeners exhibit activation in visual cortex while attending to pitches within a melody (Démonet, Price, Wise, & Frackowiak, 1994; Perry et al., 1999; Zatorre, Evans, & Meyer, 1994; Zatorre, Perry, Beckett, Westbury, & Evans, 1998). Thus, there is direct physiological evidence that under certain circumstances pitch can be represented spatially. Such spatial representations of pitch are not fully understood, but it is clear that listeners can activate a visual representation of melodic contour. It is possible that this auditory-visual mapping may instantiate a more general and complex process of structure mapping (cf. Gentner, 1983; McDermott, Lehr, & Oxenham, 2008). However, the goal of the present research was not to propose the existence of a unitary mechanism or module by which this transfer occurs. Instead, the primary objective of these studies is to explore the information that listeners use when they consciously compare melodic contours across the auditory and visual modalities.

1.3 Mechanisms of cross-modal contour perception

Despite the connection between pitch height and spatial height, there is little work specifying how listeners transfer contour information from one modality to the other. What information are listeners using in their mental representation of a melodic contour? In the mapping between auditory pitch height and visuospatial coordinates, what is the nature of the information that listeners use to construct a spatial representation of contour? Is contour represented as a sequence of upward and downward directions between adjacent events, or are relative heights also encoded with respect to nonadjacent events, or even all other events in a sequence? Addressing such questions requires the development of a quantitative model of cross-modal melodic contour perception. Existing models of auditory contour perception may help to account for the cross-modal perception of melodic contour.

Several contour models adopt a reductive approach by condensing contours to a small number of salient events, such as reversal points (changes in the direction of movement) or the location of the highest and lowest (pitch) event. Reductive models have been proposed to account for contour in both speech (e.g., Ladd, 1996; Pierrehumbert & Beckman, 1988; Xu, 2005) and music (Adams, 1976; Dyson & Watkins, 1984; Morris, 1993). Although reductive models provide a parsimonious description of contour, it is questionable whether they provide a complete and accurate characterization of the psychological representation of contour, as they discard important information through their selective focus.

A number of more elaborate models of contour have been developed. These models go beyond simple descriptions such as reversal points and consider (to varying extents and by various statistical means) the relative heights of both adjacent and non-adjacent events. Within the speech domain, several techniques of describing the similarity of two pitch contours have been developed, such as tunnel measures, root mean square distance, mean absolute difference, and a correlation coefficient (Hermes, 1998b). Hermes (1998a) asked phoneticians to provide similarity ratings for pairs of auditory or visual contours derived from the pitch contour of spoken sentences. Ratings were then compared with the above contour similarity measures (Hermes, 1998b). Of the various measures, the best predictor of rated similarity was obtained by calculating the correlation between piecewise-linear approximations of the pitch contours (reproducing the contour with a concatenation of line segments representing the original shape). As such, a simple correlation measure (hereafter referred to as surface correlation) holds great promise for predicting melodic contour similarity.

1.3.1 Music-specific contour models

There are also contour models developed within the musical domain. One such approach, called CSIM, is based on a combinatorial model of contour (Friedmann, 1985; Marvin & Laprade, 1987; Polansky & Bassein, 1992; Quinn, 1999) in which each pitch event within a melody is coded as either higher or same/lower than every other pitch, resulting in a matrix of pitch relations. Calculating the number of shared elements between the matrices of two melodies quantitatively determines the CSIM contour similarity. In an experimental test of this model, Quinn (1999) found that contour relations between adjacent and non-adjacent notes predicted musicians' similarity ratings of diatonic, 7-note melody pairs. Interestingly, recent work by Shmulevich (2004) suggests that the CSIM measure is algebraically equivalent to surface correlation measures, such as Kendall's tau or Spearman's rho, thus generalizing the surface correlation measure used in speech research (Hermes, 1998b) to music.

An alternative model of contour characterizes melodies through a Fourier analysis of their pitch relations. Fourier analysis represents the cyclic nature of a signal by breaking it down into a set of harmonically related sine waves. Each sine wave is characterized by a frequency of oscillation, an amplitude and phase. The amplitude measure of each frequency represents how strongly that particular sine wave contributes to the original signal, and the phase describes where in its cycle the sine wave starts. This technique efficiently describes the complete contour rather than discarding potentially important cues that a reductive approach might ignore. Using this procedure, Schmuckler (1999, 2004) proposed a model of melodic contour in which a melody is coded into a series of integers; this series is then Fourier analyzed, producing amplitude and phase spectra for the contour. These spectra thus provide a unique description of the contour in terms of its cyclical components. Comparing the amplitude and phase spectra from different melodies gives a quantitative measure of predicted contour similarity. Schmuckler (1999) provided initial support for this model, demonstrating that listeners' perceptions of contour complexity for both atonal and tonal 12-note melodies were consistently predictable based on amplitude (but not phase) spectra similarity. More recently, Schmuckler (2004) described a further test of this model in which similarity judgements of longer, more rhythmically diverse folk melodies were also predictable based on amplitude spectra correspondence. Together, these findings support the idea that the relative strengths of underlying frequency components can characterize the internal representation of a contour.

1.4 Experimental goals

Testing how well these contour models can predict the similarity of auditory and visual contours is a straightforward way of investigating how listeners convert melodic contour between modalities. There is already some work on cross-modal melodic contour perception (Balch, 1984; Balch & Muscatelli, 1986; Davies & Jennings, 1977; Messerli, Pegna, & Sordet, 1995; Mikumo, 1997; Miyazaki & Rakowski, 2002; Morrongiello & Roes, 1990; Waters, Townsend, & Underwood, 1998). Although these studies represent a wide range of research questions, they all address some aspect of how contour contributes to the perception and production of music in both the auditory and visual modalities. Of this work, the most directly relevant for the current purposes are studies by Balch (1984; Balch & Muscatelli, 1986). Balch and Muscatelli (1986), for instance, tested the recognition of six-note melodies using all possible cross-modal combinations of auditory and visual contours, specifically auditory-auditory (AA), auditory-visual (AV), visual-visual (VV) and visual-auditory (VA). In this work, participants experienced pairs of auditory and/or visual contours, and indicated whether the second contour matched the first. Of the four possible cross-modal combinations produced by this design, Balch and Muscatelli (1986) found that overall, performance was best in the VV condition, worst in the AA condition, and intermediate in the cross-modal (AV and VA) conditions. However, speed of presentation influenced recognition; performance in all but the AA condition suffered with increasing speed such that all conditions performed equally at the fastest rate. These findings suggest that it is more difficult to abstract melodic contour information from the auditory than the visual modality, but also generally validate the viability of a direct cross-modal matching procedure.

The goal of current investigation was to examine the cross-modal evaluations of melodic contour similarity. Music is often a multimodal experience and involves frequent transfer of information across modalities. Accordingly, the main theoretical interest is to gain understanding of the transfer across modalities of one of the most salient features in music – melodic contour. Tasks such as reading music, transcribing melodies, and online monitoring of performance accuracy rely on the ability to successfully transfer melodic contours between the visual and auditory modalities.

This research focuses on two primary questions of cross-modal melodic contour. First, can listeners with various levels of musical expertise recognize cross-modal melodic contour similarity? If so, then second, what forms of information can they use? Of particular interest is if listeners use the cyclic nature of pitch height oscillations (as measured by Fourier analysis) and/or more surface-based information (as measured by a correlation coefficient) when comparing melodic contours cross-modally.

Therefore, the current studies tested if established quantitative models of contour similarity within modalities can predict cross-modal similarity of melodic contours, by directly comparing auditory and visual contours. This procedure should illuminate the features of contour that listeners use to transfer melodic contours across modalities, and shed light on the processes by which melodic and visual contours are mapped onto one another.

2 EXPERIMENT 1

In Experiment 1, participants judged the similarity between melodic and visual contours. On each trial, some listeners heard a melody followed by a visual contour (the auditory-visual, or AV condition); others experienced the opposite order (visual-auditory, or VA condition). Although simultaneous presentation of melodic and visual contours is possible, it is problematic as it allows participants to use a simple element by element matching strategy. In contrast, by presenting only one contour at a time, listeners must extract and represent in memory the information from the first contour and subsequently compare it with the second. Hence, the design highlights the mental representation of contour, and whether theoretical characterizations of a contour are relevant in similarity judgements. Cross-modal presentation of contours also circumvents the impact of an array of potentially confounding auditory factors (e.g., tonal influences, rhythmic and metrical factors) and visual factors (e.g., spatial extent, spatial density, colour) that might arise when using solely melodic or visual stimuli.

If participants can make use of Fourier analysis and surface correlation information then they should judge visual and auditory contours as being similar according to their theoretical degree of similarity, as judged by these models.

2.1 Method

Participants

All participants were undergraduate students in an introductory psychology course at the University of Toronto Scarborough, and received course credit for their participation. There was no prerequisite or exclusion based on participants' level of musical training. There were 19 participants in the AV condition, with an average age of 19.4 years (SD = 1.5), and an average of 5.2 years (SD = 5.9; range = 0 to 13 years) of formal musical instruction. For the VA condition, there were 23 participants, with an average age of 20 years (SD = 1.6), and an average of 4.8 years (SD = 4.4, range = 0 to 15 years) of formal musical instruction.

Stimuli

Twenty-five tonal melodies composed by Bach, Mozart, Beethoven, Schubert and Brahms were selected from a sight singing textbook (Ottman, 1986) for this study. All of these melodies remained in a single key. The average length of the melodies was 35 notes (SD = 8) and the average duration was 14 s (SD = 3). In these melodies the tempo (the level of the metric pulse) was 120 beats per minute (.5 s per beat), and the timbre was an acoustic grand piano MIDI patch. A series of integers represented the fundamental frequency of each pitch in the melodies, where the lowest note had a value of 0 and the highest note a value of n-1, with n equal to the number of unique notes in the melody (see Schmuckler, 1999). The integer series was graphed as a stair plot, whereby each step of the stair represents a discrete pitch in the melody. Stair plots were then saved as a graphics file (jpeg) to serve as the "matching" visual contour. Figure 1 displays a sample melody from this study (in musical staff notation) and its matching visual contour.

Along with the matching visual contour, a family of mismatching visual contours was created for each melody by randomly reordering the values in the original series. There were some restrictions on these mismatched series. First, the initial two and final three numbers of the original series were the same for all related mismatches so as to prevent participants from relying exclusively on beginning (i.e., primacy) or ending (i.e., recency) information in their similarity judgements. Second, the number of intervals in the mismatches that were bigger than three steps could not vary more than 5% from the number of such intervals in the original. Lastly, no interval in the mismatched series could be larger than the largest interval in the original. This final restriction ensured that the mismatched series did not contain any distinctive features that obviously differentiated them from the original.

For each original sequence, there were initially nine mismatched sequences, with these mismatches varying in their theoretical similarity relation to the original series. Specifically, both Fourier analysis and surface correlation techniques assessed the theoretical similarity between contours. For the Fourier analysis measure, the amplitude and phase spectra of each integer coding were calculated. The amplitude spectra for these contours were then converted to percent variance (technically, the energy spectra), which normalizes the relative strengths of the various sine wave components. For simplicity, this measure will be referred to as amplitude spectra (as they are essentially a normalized derivative of this information). As phase spectra are, by definition, already normalized, there is no need to modify these values. Correlating the amplitude spectra between the original series and the mismatch series determined the amplitude similarity; phase spectra were not considered given the earlier results suggesting that amplitude, not phase information is critical for auditory contours (Schmuckler, 1999, 2004). There were nine mismatched sequences because there was one sequence for each tenth of amplitude similarity between mismatch and original from 0 and 1. In other words, there was one mismatch with an amplitude spectra correlation with the original between 0 and .1, another between .1 and .2, and so on up to between .8 and .9.

For the nine mismatches, the surface correlation similarity was derived by calculating the correlation coefficient of the original (the integer code representing the coded pitch height of the notes in the original melody) with each mis-

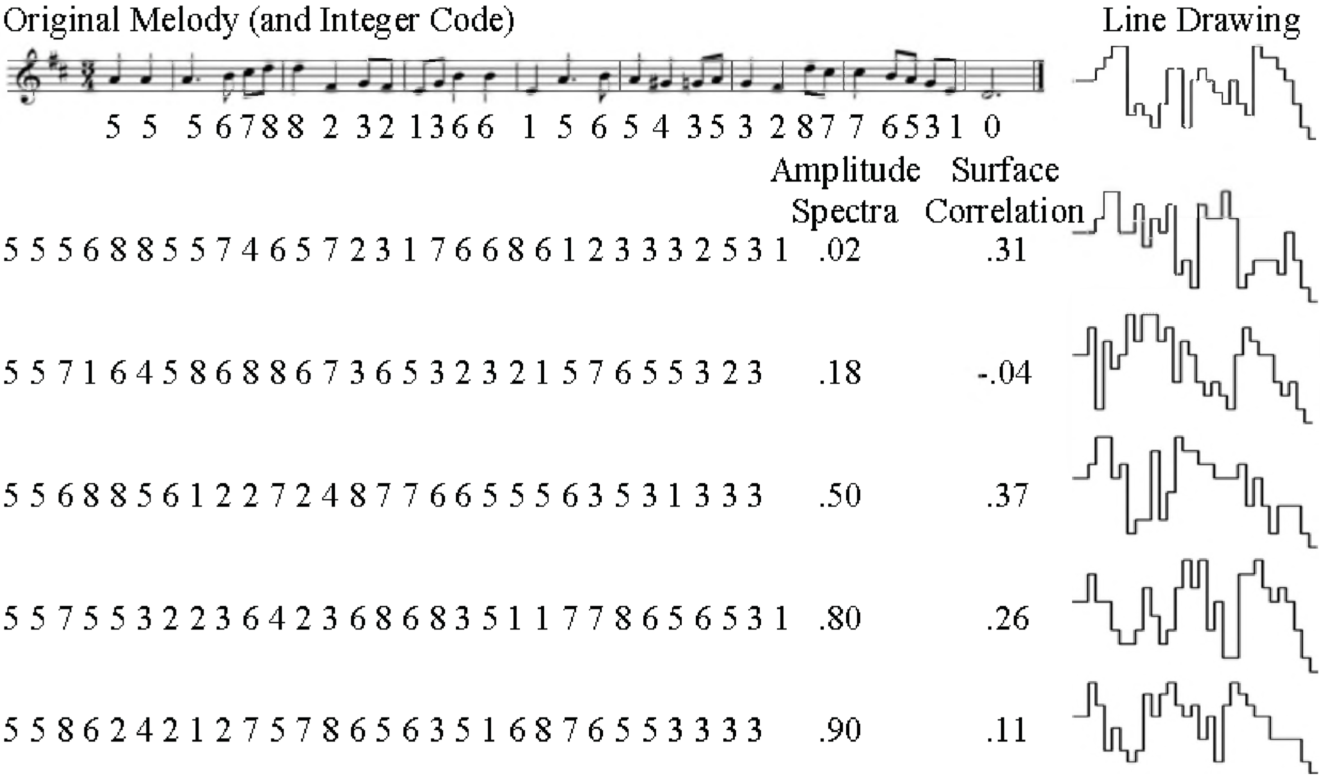


Figure 1. Sample stimulus melody (in musical notation), its integer coding, and line drawing. Below, the integer codes for the final (chosen) five mismatches as well as their line drawings are shown. Measures of similarity between each mismatch and the original are also listed, specifically the correlation of the amplitude spectra and the surface correlation.

match. Ultimately, five of these mismatches were chosen for presentation to participants, selected by choosing the five series with the lowest surface correlation with the original, in an effort to empirically separate (as much as possible) the potential effect of surface correlation and amplitude spectra similarity. Although this attempt to disentangle amplitude spectra and surface correlation did so by minimizing surface correlations, both measures nonetheless produced a fairly wide (and equivalent) range of correlation coefficients with the original series (Fourier analysis: .01 to .9; surface correlation: -.49 to .52). The final five mismatched series were graphed as line drawings and saved as graphics files in the same manner as the matching visual contour. Figure 1 also displays the five mismatched integer series for the corresponding sample melody, along with the amplitude spectra correlations and surface correlations with the original, and the line drawing resulting from these series. Combined with the matching stimulus, this procedure yielded six possible visual contours for comparison with each auditory melody.

Apparatus

Generation of random mismatched series, and analyses of all (original and mismatch) sequences were performed using code written in MATLAB, in conjunction with the midi toolbox (Eerola & Toiviainen, 2004). Presentation of the stimuli and the experimental interface were programmed with MATLAB 7.0 using Cogent 2000 (developed by the Cogent 2000 team at the FIL and the ICN and Cogent Graphics developed by John Romaya at the LON at the Wellcome Department of Imaging Neuroscience). Two Pentium(R) 4 computers (3.0 and 1.7 GHz) were used for running the experiment. Auditory stimuli for this study were generated using Audigy Platinum Soundblaster sound cards, and were presented to listeners over Audio Technica ATH-M40fs or Fostex T20 RP Stereo Headphones, set to a comfortable volume for all participants. Visual stimuli appeared on either a Samsung 713V or LG Flatron L1710S 15” monitor.

Procedure

Participants in the auditory-visual (AV) condition heard a melody, followed by a picture that represented the shape of a melody, and then rated the similarity between them. Each trial for the AV participants began with the phrase “Listen carefully to the melody” displayed on the computer monitor while the melody played. After the melody finished, the computer loaded and displayed the graphics file as quickly as possible (due to hardware limitations, this was not immediate, however the delay was always less than one second). This contour remained present until listeners entered a response, at which point the monitor was blank for 250 ms, until the beginning of the next trial. Participants in the visual-auditory (VA) condition experienced the same stimuli but in the reverse order. For the VA participants, the line drawing was displayed for 2.5 seconds before being replaced by the phrase “Listen carefully to the melody” (placed at the same location in order to mask residual visual input). Concomitantly, the melody began playing.

All participants (AV and VA) rated the similarity of the contour between the picture and the melody on a scale of 1 to 7 (1 being not at all similar, 7 being very similar). Trials were presented in random order, with the restriction that no individual melody was heard twice in a row. Twenty-five possible (original) melodies combined with six possible visual displays (the match plus five mismatches) resulted in 150 trials in total. To clarify, because only the original melodies were presented, there were no additionally generated melodies. Instead, pairing generated visual sequences with the original melody constituted a mismatch. Participants were either run individually or in pairs (on different computers, separated by a divider). The entire experimental session lasted about one hour for both AV and VA conditions.

2.2 Results

To provide a baseline measure of maximal similarity, participants' ratings for the matching auditory-visual stimuli were first compared with the ratings for the mismatched stimuli by means of a one-way repeated-measures Analysis of Variance (ANOVA). The within-subjects factor was match (matching versus mismatching). In the initial analysis the different levels of auditory-visual mismatch were thus collapsed. For the AV condition, this analysis revealed that ratings of similarity were significantly higher for matches (M = 4.83, SD = .55) than for mismatches (M = 4.27, SD = .65), $F(1,18) = 15.69$, $MSE = .19$, $p < .001$, $\eta_p^2 = .47$. Interestingly, two participants failed to show this trend; this result indicates that they were not attending to the task, and therefore their data were removed from further analyses. Similar results were observed for the VA condition; matches (M = 4.86, SD = .72) were rated as being more similar to the melody than mismatches (M = 4.18, SD = .54), $F(1,22) = 72.58$, $MSE = .07$, $p < .001$, $\eta_p^2 = .77$. In this case, one participant did not show this pattern; the data of this participant were removed from further analyses. Overall, therefore, the average ratings of perceived similarity of melodies and matching sequences exceeded those of melodies and mismatching sequences.

The preceding analysis demonstrates that participants were sensitive to the similarity between auditory and visual melodic contours. However, the analysis does not determine whether listeners differentiated between visual contours having varying degrees of similarity with the auditory contour. To explore this issue, subsequent analyses focused on examining whether or not the various models of contour similarity described earlier could predict listeners' perceived contour similarity. Because this question is one of predicting perceived levels of mismatch between auditory and visual stimuli, these analyses focused on the mismatched sequences only and excluded the match trials.

Based on the various contour models described earlier, a host of contour similarity predictors were generated, including models based on those outlined by Schmuckler (1999). The Fourier analysis model produces two possible predictors (as already discussed): amplitude spectra and phase spectra similarity. As described in the stimulus section, amplitude and phase spectra information for all integer series were calculated, and absolute difference scores standardized to the length of the melody were computed between the auditory (original) and visual (mismatch) sequences[1]. Along with these Fourier analysis measures, Schmuckler (1999) also described an oscillation model, in which the interval information between consecutive pitches is quantified to produce both a summed and a mean interval measure (see Schmuckler, 1999, for detailed discussion of these measures). Accordingly, four measures were derived from this earlier work – amplitude and phase spectra difference scores, and summed and mean interval difference scores.

Along with these measures, three additional theoretical predictors were calculated. The first is based on the combinatorial model (Friedmann, 1985; Marvin & Laprade, 1987; Polansky & Bassein, 1992; Quinn, 1999) and involves the CSIM measure described earlier, which characterizes each contour as a matrix in terms of whether a subsequent tone is higher (coded as 1) or equal to/lower (coded as 0) than each of the other tones in the melody. Then, the mean number of shared elements between the matrices of each mismatch and its corresponding match was calculated and used as the CSIM predictor. Second, a surface correlation measure was calculated by simply correlating the integer codes for each melody. Third, a measure based on comparing the number of reversals in the match and mismatch was calculated. Dividing the number of reversals in the match by the number of reversals in the mismatch gave a ratio of reversals. This ratio was subtracted from 1 so that the absolute value of this difference indicated the percent difference in number of reversals between match and mismatch (a higher number would indicate greater difference, thus presumably less similarity).

Preliminary analyses revealed that the length of the melody was a strong predictor of perceived similarity, perhaps because two of the 25 melodies were longer than the rest (56 notes; beyond two standard deviations of the mean of 35 notes). Given that remembering the first contour and comparing it to the second was a challenging task, and only these two melodies were much longer than the others, listeners may have systematically rated longer melodies (and line drawings) as more similar than shorter stimuli. Therefore, the data of these two melodies were excluded, leaving 23 melodies (each with five mismatches); in addition melody length was included as a potential predictor of similarity.

Table 1 provides an intercorrelation matrix for these eight measures across all the mismatching stimuli in this study. This table reveals a few significant intercorrelations between variables. As expected, CSIM and surface correlation measures were essentially equivalent ($r = .96$, $p < .001$), corroborating Shmulevich's (2004) calculations. Melody length correlated significantly with amplitude spectra, summed interval and mean interval. These correlations are not surprising given that these three variables were all standardized to the length of the melody. Mean interval was significantly correlated with amplitude spectra and reversal ratio; reversal ratio was also related to summed interval. The interrelation of these variables most likely indicates the extent to which these measures mutually indicate some aspect of the cyclical ups and downs of contour.

Table 1: Intercorrelations of Theoretical Predictors of Contour Similarity for Experiment 1

Predictor	Phase Spectra	Summed Interval	Mean Interval	Surface Correlation	CSIM	Reversal Ratio	Melody Length
Amplitude Spectra	-.07	-.01	.27**	.14	.15	.07	-.62***
Phase Spectra		.03	-.01	-.18	-.12	-.03	.10
Summed Interval			-.18	-.09	-.11	.35***	.28**
Mean Interval				.12	.11	.30**	-.30**
Surface Correlation					.96***	.05	-.14
CSIM						.02	-.17
Reversal Ratio							.07

** *p* < .01. *** *p* < .001.

All eight of these predictors were correlated with the averaged similarity ratings for the AV and VA conditions. The results of these analyses appear in Table 2 and demonstrate that surface correlation, CSIM, and melody length all significantly correlated with listeners' cross-modal similarity ratings in both conditions. Amplitude spectra difference scores correlated negatively and significantly with the VA similarity ratings but not for the AV condition. The AV and VA ratings themselves were significantly related, (r = .39, p < .001).

As a follow-up to these analyses, two multiple regression analyses were performed to determine the unique contribution of these models to predicting perceived similarity, for the AV and VA conditions separately. Given the high correlation between the surface correlation and CSIM variables (leading to an unacceptably low tolerance value of .087 in the regression equation), and the fact that surface correlation had the larger unique contribution of explanatory variance in both AV and VA conditions, only surface correlation was retained in the final regression equations. Both AV and VA similarity ratings were thus predicted from the three variables of amplitude spectra differences, surface correlation, and melody length. For the AV condition these three variables significantly predicted similarity ratings, R (3,111) = .41, p < .001, with significant contributions by surface correlation, B = .61, β = .28, p < .01, and melody length, B = .03, β = .36, p < .01. In contrast, amplitude spectra failed to contribute significantly, B = 1.49, β = .03, ns. For the VA condition these three variables also significantly predicted similarity ratings, R (3,121) = .49, p < .001, with significant contributions from amplitude spectra, B = -3.06, β = -.24, p < .05, and surface correlation, B = .18, β = .37, p < .001. In this case, melody length failed to contribute significantly, B = .004, β = .19, ns.

Finally, a set of analyses looked at the impact of musical experience on contour similarity. For this analysis each participant's ratings were averaged across the 23 matching stimuli and compared with the average ratings from four different sets of mismatches. The first set consisted of the averaged ratings for the complete set of mismatches (N = 115); the second set consisted of averaged the ratings for the 23 mismatches with the largest amplitude spectra difference score; the third set consisted of averaged the ratings for the 23 mismatches with the largest phase spectra difference score; the fourth set consisted of averaged the ratings for the 23 mismatches with the lowest surface correlation with each melody. Each participant's data were transformed into z-scores (each participant as a separate population), and the differences between the z-scores of the matches and the four mismatched sets were calculated. Thus, each participant had four scores: an over-

Table 2: Correlations of Theoretical Predictors with Auditory-Visual (AV) Similarity Ratings and Visual-Auditory (VA) Similarity Ratings of Experiments 1 and 2

	Experiment 1		Experiment 2	
Predictors	AV similarity rating	VA similarity rating	AV similarity rating	VA similarity rating
Amplitude Spectra	-.15	-.30**	-.10	-.22*
Phase Spectra	-.12	.06	-.37***	-.30**
Summed Interval	.06	.04	-.08	-.04
Mean Interval	-.05	-.07	.10	-.01
Surface Correlation	.24*	.31***	.46***	.45***
CSIM	.19*	.27**	.41***	.39***
Reversal Ratio	.07	.02	.19*	.13
Melody Length	.30**	.28**	-.15	-.02

* *p* < .05. ** *p* < .01. *** *p* < .001.

Table 3: Correlations Between the Years of Musical Training and Difference Score Measures in Similarity Ratings for Experiment 1 and 2

	Experiment 1		Experiment 2	
Difference score	AV Condition	VA Condition	AV Condition	VA Condition
Overall	.54*	-.15	.57*	.05
Amplitude Spectra	.33	-.12	.59*	-.05
Phase Spectra	.49*	-.34	.49*	.08
Surface Correlation	.53*	.03	.63**	.01

* $p < .05$. ** $p < .01$.

all difference score, an amplitude spectra difference score, a phase spectra difference score and a surface correlation difference score. These difference scores were then correlated with participants' degree of musical training (for AV and VA conditions separately), as indexed by the number of years of formal instruction on an instrument or voice. Table 3 shows the results of these analyses.

Participants in the AV condition with more formal training differentiated more between matches and mismatches, and relied more on both phase spectra and surface correlation differences to form their ratings of perceived similarity. For the VA condition, however, musical training did not affect participants' difference scores. There were no overall differences between the AV and VA condition in the absolute size of the overall difference score, F(1,37) < 1, MSE = .08, ns, the amplitude spectra difference score, F(1,37) = 2.91, MSE = .11, ns, the phase spectra difference score, F(1,37) < 1, MSE = .12, ns, or the surface correlation difference score, F(1,37) < 1, MSE = .11, ns.

2.3 Discussion

There are three main findings of Experiment 1. First, listeners matched contours of long melodies cross-modally, as demonstrated by higher similarity ratings between the auditory melodies and matching visual representations of their contour, relative to ratings of similarity between melodies and mismatched visual representations. Second, established theoretical models of contour similarity can partly explain the perceived similarity of cross-modal melodic contours, however there were differences between the AV and VA conditions. Third, only in the AV condition did musical expertise aid listeners in rating the difference between match and mismatch; it also enabled them to be more sensitive to phase spectra and surface correlation in forming their ratings.

Our observation that listeners were able to recognize the similarity between contours presented cross-modally replicates previous findings on cross-modal contour perception (Balch, 1984; Balch & Muscatelli, 1986; Davies & Jennings, 1977; Messerli et al., 1995; Mikumo, 1997; Miyazaki & Rakowski, 2002; Morrongiello & Roes, 1990; Waters et al., 1998). Because only one contour (either auditory or visual) was presented at a time, listeners could not simply compare the auditory and visual contours element by element and check for differences. Accordingly, this task required listeners to extract and subsequently remember contour information for use in a later comparison.

What attributes of the contours contributed to listeners' perceived similarity? Both conditions showed strong effects of surface correlation, a finding that extends previous research on within-modal auditory contour similarity (Hermes, 1998a, 1998b; Quinn, 1999) to cross-modal applications. To the extent that surface correlation conveys both the local, note-to-note characteristics and overall global shape of a contour, this finding implies that listeners can use a combination of both local and global cues when converting contours between the auditory and visual domains, regardless of the modality in which the contour is initially presented.

The effect of Fourier components on perceived similarity was mixed, and varied for the AV and VA conditions. Phase did not contribute to the AV or VA condition regressions, a result that replicates and extends findings of the unreliable nature of phase in modeling melodic contour perception (Schmuckler, 1999, 2004; although see Schmuckler, 2008). In contrast, amplitude spectra differences were significant, but only in the VA condition. These results suggest that listeners can use the global cues of cyclic oscillation that Fourier analysis captures for evaluations of cross-modal melodic contour similarity, but only when comparing a visual contour to a subsequently occurring auditory contour. However, discussing this finding in detail requires reference to the results of the second experiment, therefore the general discussion considers the implications of this finding.

Largely because Experiment 2 replicates the findings of the variable role of musical expertise for AV and VA conditions, this result also is explored in greater detail in the general discussion. However, the fact that this finding emerges only in the AV condition suggests that converting contour information from the auditory to the visual domain exploits the skills that musical training confers. It is likely that the AV condition is more challenging than the VA condition due to differential memory demand. Specifically, because the melodies were presented in a gated (note-by-note) fashion, participants had to remember the melody in its entirety in the AV condition and subsequently compare it to a visual contour. Conversely, in the VA condition participants could compare their memory of the visual contour to the gated presentation of the melody as it progressed note-by-note rather than waiting until the melody finished. Thus the relatively higher memory demand of the AV condition may differentiate across levels of musical training more so than the VA condition.

Indeed, one potential concern with this study concerns the high memory demand of the task. In particular, this study employed melodies of considerable length, which may have strained listeners' memory capacities and made the evaluation of cross-modal contour similarity difficult. Accordingly, it is of interest to replicate the principal findings of this work with melodies that make lesser memory demands. Specifically, can listeners recognize cross-modal melodic contour similarity, and can current models of contour information such as surface correlation and Fourier analysis components explain perceived similarity when memory demands are less? Experiment 2 provided such a replication by testing the cross-modal similarity of shorter melodies than those employed here, thus also extending this work.

3 EXPERIMENT 2

The results of Experiment 1 suggest that surface correlation and Fourier analysis components both contribute to the perceived similarity of long melodies compared across auditory and visual modalities. However, listeners often hear shorter melodies, and furthermore most of the previous work on melodic contour (within and across modalities) uses much shorter melodies. It is also possible that the length of the stimulus melodies and the concomitant memory demands might have influenced the nature of listeners' cross-modal comparisons, along with how well these different approaches characterized cross-modal contour similarity.

Therefore it is of interest to replicate these results with shorter melodies, for two main reasons. First, these models of melodic contour may perform differently under conditions more similar to existing melodic contour research. Thus, repeating these tests with shorter melodies can investigate this possibility and potentially extend the validity of these models to melodies of various lengths. Second, listeners may or may not use similar contour information for short as well as long melodies. Consequently, testing shorter melodies provides the opportunity to ascertain if listeners use the same information to evaluate cross-modal melodic similarity regardless of contour length.

To test these possibilities, Experiment 2 employed the same task as the earlier study but used new, shorter melodies for cross-modal comparisons.

3.1 Method

Participants

Participants were undergraduate students in an introductory psychology course at the University of Toronto at Mississauga, and received course credit for their participation. There was no prerequisite or exclusion based on participants' level of musical training.

There were 17 participants in the AV condition, with an average age of 18.5 years (SD = .86), and an average of 1.5 years (SD = 2.6; range = 0 to 10 years) of formal musical instruction. For the VA condition, there were 17 participants, with an average age of 19.1 years (SD = 2.19), and an average of 1.3 years (SD = 2.9, range = 0 to 10 years) of formal musical instruction.

Stimuli, Apparatus, and Procedure

Twenty-five tonal melodies from a compilation of sight singing melodies (Smey, 2007) were used for this study. All of these melodies were between 14 and 18 notes long, and did not modulate to a new key. The average length of the melodies was 16.7 notes (SD = 1.3) and the average duration was 7.5 s (SD = .4). As in Experiment 1, the tempo of these melodies was 120 beats per minute, and the timbre was an acoustic grand piano MIDI patch. The melodies were coded as integer series in the same manner as Experiment 1, to form the "matching" visual contours. The non-matching mismatched visual contours were created in the same fashion as Experiment 1, using the same rules and theoretical similarity measures.

The apparatus and procedures were the same as in Experiment 1. There were 150 trials in total, and the experimental session lasted about 45 minutes.

3.2 Results

As in Experiment 1, an initial step in the data analysis was designed to establish the average similarity rating for conditions of maximal similarity (match). A one-way repeated measures ANOVA compared participants' ratings for the matching auditory-visual stimuli with the ratings for the mismatched stimuli, with the within-subjects factor of match (matching versus mismatching). For the AV condition, this analysis revealed that ratings of similarity were significantly higher for matches (M = 4.93, SD = .72) than for mismatches (M = 4.13, SD = .55), $F(1,16) = 18.86$, $MSE = .29$, $p < .001$, $\eta_p^2 = .54$. For the VA condition the results were similar, with matches (M = 5.18, SD = .61) rated as more similar to the melody than mismatches (M = 4.22, SD = .47), $F(1,16) = 40.93$, $MSE = .19$, $p < .001$, $\eta_p^2 = .72$. Again, therefore, listeners recognized the greater similarity of contours that matched the melodies to those that were mismatched.

Subsequent analyses determined the extent to which the various contour similarity models correlated with participants' perceived similarity ratings, again focusing only on the ratings of the mismatch trials. This analysis tested the same contour similarity predictors as Experiment 1, including the difference score measures of amplitude spectra, phase spectra, summed interval and mean interval, as well as the CSIM/surface correlation measure, reversal ratio and melody length measures. Table 4 shows the intercorrelations between the predictors for Experiment 2. The correlations between these predictors and the perceived similarity ratings for the AV and VA conditions appear in Table 2. For both the AV and VA conditions, phase spectra, surface correlation and CSIM measures were significantly related to participants' ratings. Counterintuitively, the reversal measure was significantly positively correlated with AV similarity ratings, a finding suggesting that a greater difference in reversals between a melody and its mismatch produced higher perceived similarity. As in Experiment 1 the amplitude spectra significantly

Table 4: Intercorrelations of Theoretical Predictors of Contour Similarity for Experiment 2

Predictor	Phase Spectra	Summed Interval	Mean Interval	Surface Correlation	CSIM	Reversal Ratio	Melody Length
Amplitude Spectra	.08	.24**	.37***	-.14	-.17	.07	-.23**
Phase Spectra		.12	.07	-.42***	-.37***	-.14	.01
Summed Interval			.20*	.08	.07	.01	-.27**
Mean Interval				.02	.00	.45***	-.42***
Surface Correlation					.97***	.11	-.22*
CSIM						.09	-.23**
Reversal Ratio							-.22*

* $p < .05$. ** $p < .01$. *** $p < .001$.

correlated with perceived similarity in the VA condition only. Finally, the AV and VA condition similarity ratings correlated significantly with each other (r = .58, p < .001).

Two multiple regression analyses examined the strength and unique contribution of each of the potential predictors to perceived similarity. As in Experiment 1, surface correlation was included instead of CSIM in both AV and VA conditions because of its stronger relation with similarity ratings. For both sets conditions, similarity ratings were predicted from amplitude spectra differences, phase spectra differences, surface correlations, and reversal scores.

For the AV condition, these variables significantly predicted similarity ratings, R(4,120) = .51, p < .001, with significant contributions of phase spectra differences, B = -.22, β = -.19, p < .05, and surface correlation, B = .44, β = .36, p < .001. In contrast, there was no significant effect of either amplitude spectra differences, B = -.79, β = -.04, ns, or of reversals, B = .23, β = .13, ns. For the VA condition, these variables also significantly predicted similarity ratings, R(4,120) = .49, p < .001, with significant contributions by amplitude spectra differences, B = -3.68, β = -.16, p < .05, and surface correlations, B = .47, β = .37, p < .001. In contrast, there was no significant effect of either phase spectra differences, B = -.15, β = -.11, ns, or of reversals, B = .18, β = .1, ns.

The last set of analyses tested the effect of musical experience on contour similarity ratings. Each participant's ratings for the 25 matching stimuli were averaged and compared with the same four sets of mismatches described in Experiment 1. As in this previous study, the differences between the z-scores of the matches and the four mismatch sets were calculated, and correlated with participants' degree of musical training, as indexed by the number of years of formal musical instruction. Table 3 presents these analyses, and indicates the same general pattern as Experiment 1. Participants in the AV condition with more formal training differentiated matches and mismatches more, were better able to use amplitude and phase spectra and surface correlation differences between matches and mismatches in forming a perceived similarity rating. But in the VA condition, musical training did not correlate with participants' difference scores. Also similar to Experiment 1, there were no differences in absolute size of the difference scores between the AV and VA condition. Neither the overall difference score F(1,32) < 1, MSE = .14, ns, nor the amplitude spectra difference score, F(1,32) < 1, MSE = .19, ns, nor the phase spectra difference score, F(1,32) = 1.2, MSE = .16, ns, nor the surface correlation difference score, F(1,32) = 1.8, MSE = .17, ns showed any difference in absolute size between the AV and VA condition.

3.3 Discussion

In Experiment 2, listeners again succeeded at recognizing matching cross-modal melodic contours. Furthermore, surface correlation and Fourier components predicted their ratings of perceived similarity between non-matching contours. Lastly, musical expertise allowed listeners to make better use of the available cues in evaluating contour similarity in the AV condition. Therefore the results of Experiment 2 are quite similar to those of Experiment 1, while ruling out the potentially confounding effects of melody length from Experiment 1.

The surface correlation measure was a good predictor of cross-modal contour similarity ratings for both the AV and VA condition, again demonstrating the importance of correlation coefficients in modeling contour similarity and generalizing its validity to cross-modal perception. The Fourier components, on the other hand, varied in their predictive value depending on the order of presentation of the contours. Specifically, listeners' ratings were related to phase spectra for the AV condition, and amplitude spectra in the VA condition. Neither Fourier component significantly predicted perceived similarity in both conditions. Other than the significant contribution of phase in the AV condition of Experiment 2 (that did not occur in Experiment 1), these results echo Experiment 1.

4 GENERAL DISCUSSION

4.1 Summary

Together, Experiments 1 and 2 provide a number of insights into contour processing. First, and most fundamentally, these studies demonstrate that listeners can recognize the similarity of melodic contours when presented cross-modally, regardless of melody length. Both studies revealed higher similarity ratings for matching auditory and visual contours relative to mismatching contours. Although this finding may seem relatively intuitive, this result is noteworthy in the sense that the majority of research on cross-modal melodic contour (Balch, 1984; Balch & Muscatelli, 1986; Cupchik, Phillips, & Hill, 2001; Lidji et al., 2007; Mikumo, 1994; Miyazaki & Rakowski, 2002; Morrongiello & Roes, 1990; Waters et al., 1998) has used relatively short melodies (five to seven notes) that were within the capacity of working memory. Because both studies in this work employed melodies well beyond the limitations of short term processes, recognition of cross-modal similarity in this case is not a foregone conclusion, particularly given that the sequential presentation of the contours exacerbated the difficulty of the task. Nevertheless, listeners were able to recognize the similarity of cross-modal melodic contours.

Second, these results provided an additional validation of the applicability of current models of contour structure and similarity to a previously untested domain. Specifically, theoretical similarity between cross-modal contours was predictable based on the combinatorial CSIM (or surface correlation) model proposed by Quinn (1999), as well the Fourier analysis model of Schmuckler (1999, 2004, 2008). In both experiments, these two models significantly predicted cross-modal contour similarity. This result suggests that at least some of the information that listeners use when constructing a mental representation of an auditory or visual contour is embodied by these quantitative contour descriptions.

4.2 Differences between experiments

One important distinction between these two models in these studies that merits deeper consideration is the variable success of the Fourier components (amplitude and phase) across modality presentation order and melody length. Whereas the surface correlation model was predictive across both presentation orders and melody lengths, amplitude and phase were not. Specifically, amplitude spectra differences were predictive of contour similarity for both short and long melodies, but only when the visual contour preceded the auditory contour (the VA condition), but not for the opposite order (the AV condition). In contrast, phase spectra differences were predictive only for the AV presentations with the short melodies.

Why might a VA, but not an AV, ordering of contours allow for the use of amplitude spectra information, whereas an AV ordering with short melodies enable the use of phase spectra information? One possibility is that listeners mentally convert what they remember of the contour presented first into the modality of the contour that occurs second to facilitate a direct comparison between the two. That is, listeners might attempt to create an auditory analogue of a visually presented contour for a VA ordering, or vice versa for an AV ordering. Such a recoding would make similarity judgements predictable based on the optimum way of characterizing the latter contour. Research on the applicability of Fourier analysis to visual scenes has revealed that in general, phase information is more important than amplitude information for visual perception (Bennett & Banks, 1987, 1991; Kleiner, 1987; Kleiner & Banks, 1987). Conversely, amplitude spectra information is more important than phase spectra information when perceiving auditory contours (Schmuckler, 1999, 2004).

There is good reason for the variable importance of amplitude and phase for audition and vision, respectively. In vision, variation in amplitude corresponds to stimulus energy (essentially degrees of light and dark), whereas phase corresponds to stimulus structure, or roughly the presence and placement of edge information. Clearly, of the two, edges and their locations are more fundamental for visual object recognition. For auditory contours, however, stimulus energy indexes the relative strength of the cyclic components (i.e., whether the signal repeats once, twice, and so on, over its length), whereas phase indexes the relative timing within the contour of ascending and descending patterns. Although both forms of information are potentially important in understanding the general shape and structure of a melody, the former intuitively seems to have a greater perceptual priority. In support of this idea, Schmuckler (1999) found that listeners can make use of phase information for perceived contour similarity when the melodies were constructed specifically to contain important phase relations. More recently, Schmuckler (2008) found a consistent correlation between phase spectra differences and perceived contour similarity when phase information was calculated based on a rhythmically weighted contour code (see Schmuckler, 1999, 2004, for discussions of this form of coding). However, in a multiple regression context phase spectra differences failed to add significantly to predictions of contour similarity.

The idea that listeners convert what they remember of the first contour into the modality of the second contour predicts well the observed pattern of results for the amplitude spectra differences. Specifically, because the VA condition would encourage listeners to recode the visual contour into an auditory one, amplitude spectra information would thus become maximally important for contour comparisons; this was what was observed in this study. This hypothesis, however, also predicts the opposite pattern for the AV condition. In this case, listeners would mentally convert the initial auditory contour into a visual analogue, with similarity judgements primarily predictable based on phase spectra differences. In partial support of this idea, similarity judgements in the AV condition were predictable based on phase information, at least for the shorter melodies of Experiment 2. However, phase played no role in the AV condition for the longer melodies of Experiment 1, implying a melody length effect on the use of phase information.

In short, the predictive value of phase changed across melody length, indicating that cross-modal contour similarity may be evaluated differently under varying musical conditions. But why should melody length have such an impact on listeners' use of phase, but not amplitude? Simply put, because phase information indexes the relative timing of the ups and downs in an auditory signal, shorter melodies enable the use of local ups and downs, and thus foster listeners' mental recoding of the melodic contour as a visual analogue. However, longer melodies (on average 35 notes in Experiment 1) vitiate the usefulness of local information, such as the timing and/or position of rises and falls in the contour, as measured by phase spectra. Accordingly, phase information will be of less use with such melodies. In contrast, because amplitude information captures global contour shape, such information is equally accessible in short and long melodies; in fact, global contour information is likely the most accessible information in longer melodies. Consequently, melody length should have less influence on the use of amplitude spectra information, provided that the melodies are long enough to contain sufficiently differentiated amplitude spectra (see Schmuckler, 2004, for a discussion of this point).

A final point about the differences between Experiments 1 and 2 concerns the relationship between AV and VA similarity ratings. In Experiment 1, the correlation between similarity ratings for the AV and VA conditions was relatively low (r = .39) compared to Experiment 2 (r = .58). This difference is likely a result of greater task difficulty of the first experiment due to longer melodies (and thus increased memory demand), thereby introducing more variability into the similarity ratings. However, the level of task difficulty did not only differ across experiments, but also within conditions; the latter variance reveals some interesting findings with regard to musical expertise, discussed below.

4.3 Role of musical experience

The third principal finding from these studies involves the role of musical experience in cross-modal melodic contour similarity. In both experiments, musical training aided participants' ability to differentiate between matches and mismatches, but only in the AV conditions. Further, in these conditions, musical training enabled listeners to make better use of amplitude spectra, phase spectra, and surface correlation information. These results give rise to two questions – how can musical training confer an advantage on listeners' cross-modal melodic contour perception generally, and why is this facilitation specific to the AV condition?

Musical training involves extensive practice with cross-modal contours. Specifically, musicians receive extensive practice in translating between written musical notation (essentially a system of horizontal lines with vertically arranged dots) and auditory sequences, experience that intuitively seems quite comparable to the tasks used in these studies. Accordingly, simple practice effects with comparably structured stimuli may account for the overall advantage conferred by musical training. In keeping with this argument, there are reports in the literature of processing advantages in cross-modal musical stimuli due to musical training. Brochard et al. (2004) similarly found that musicians possess a spatial advantage for processing dots placed above and below horizontal lines (similar to musical notation). Further, these authors also observed that musicians processed dots placed to the left and right of vertical lines faster than nonmusicians. Lidji et al. (2007) had similar findings, in that pitch height automatically activated congruent left-right spatial mappings for musicians but not nonmusicians. Specifically related to contour perception, Balch and Muscatelli (1986) found that musicians outperformed nonmusicians in all contour comparison tasks, including within-modal (AA and VV) and cross-modal (AV and VA) conditions. Furthermore, accuracy at recognizing transformations to melodic contours predicts the ability to judge spatial transformations of three-dimensional figures (Cupchik et al., 2001). Thus musical training may improve the perception and processing of cross-modal contour more generally.

However, musical experience was not helpful in the current studies for all conditions, but only the AV condition. The relative difficulty of the AV versus VA condition may explain why the facilitation effect of musical training only occurred in the AV condition, as evidenced by the difference score measures (Table 3). In both experiments, the similarity ratings between melodies and their matching visual contours was higher than for non-matching, mismatched contours, but the effect was always larger for the VA condition than for the AV condition. Inspecting the partial eta-squared values reveals that the effect size of differentiating between match and mismatch was higher for VA than AV conditions in Experiment 1 (AV $\eta_p^2 = .47$; VA $\eta_p^2 = .77$) and Experiment 2 (AV $\eta_p^2 = .54$; VA $\eta_p^2 = .72$). This difference makes sense intuitively, because the AV condition was more taxing on memory demands than the VA condition. Accordingly, the more difficult task of the AV condition accentuated the difference in abilities to compare melodic contours cross-modally as a result of musical training. Conversely, the VA condition was less difficult for participants, and so the contour processing advantage of musically-trained listeners was not as apparent. Thus if listeners encounter a situation that resembles a music-specific task, then musicians' experience will give them an advantage. However if the task changes (in this case, even just the order of presentation of stimuli), the domain-specific skills that musicians have developed may not confer the same benefits. Additionally, presenting the visual line drawing in a gated fashion may make the VA condition more difficult and consequently differentiate more between musically trained and untrained listeners.

4.4 Limitations

Along with the positive findings of these studies, there are a number of important limitations to this work that require consideration. Probably the most critical such concern involves the fact that although the various theoretical models of contour structure were predictive of cross-modal similarity, ultimately these models only explained part of the variance in such predictions. Such a finding raises the question of exactly how important such information is in participants' perception and processing of contour. As a partial answer to this concern, it is worth noting that the level of predictiveness of these variables is generally equivalent to what has been previously reported in the literature (Eerola, Järvinen, Louhivuori, & Toiviainen, 2001; Quinn, 1999; Schmuckler, 1999). Accordingly, although there are clearly many other factors that also enter into contour perception, the information captured by these predictors seems to be a consistently influential. Both Eerola and colleagues (Eerola & Bregman, 2007;

Eerola, Himberg, Toiviainen, & Louhivuori, 2006; Eerola et al., 2001) and Schmuckler (1999, 2004) have posited and investigated a variety of other factors, ranging from rhythmic components to structural factors (such as tonality and meter) to individual contour features, with varying degrees of success.

A second form of limitation with this work involves issues with the theoretic predictors themselves. Specifically, both Fourier analysis and surface correlations have inherent constraints that raise concerns when applying such procedures to models of contour structure and perceived similarity. For instance, both correlation techniques and Fourier analysis procedures are constrained by factors related to the length of the series being analyzed. Correlation measures are adversely affected by sequence length, such that the shorter the sequence the more susceptible the measure is to outlying values of the individual elements. Accordingly, shorter melodies limit the utility of correlation measures. Correlations are also limited in that they can only be applied to sequences containing the same number of elements. Given that contour comparisons rarely involve contours of the same length, this poses a methodological problem for applying surface correlations to models of melodic contour.

Fourier analysis techniques also present important methodological concerns. For one, as a mathematical procedure Fourier analysis makes a variety of assumptions about the signal that are generally not met in an application to melodic contour. Perhaps the most obvious is that Fourier analysis assumes that the signal is continuous and periodic (i.e., it has been on forever and will continue indefinitely). Needless to say, other than the occasional annoying tune that perversely gets stuck in one's head, melodies do not repeat ad infinitum. Yet in order to achieve continuity and function as a cohesive piece of music, repetition of some of the musical structure must occur; contour is one of the most important forms of pitch structure and as such could function as one of components that help to achieve this continuity. Another assumption of Fourier analysis concerns the length of the signal. When the signal is too short, Fourier analysis spectra are prone to distortions such as edge effects. The length of the melodies used in this research help to insulate the Fourier analysis from this phenomenon, but this is an issue in any application of this tool. Ultimately, the success of this approach in predicting contour similarity in this and other contexts provides support for the applicability of these procedures for the quantification of contour perception and processing, despite these potentially problematic issues.

5 Conclusion

In conclusion, this investigation into the cross-modal similarity of melodic contour has enabled insights into how listeners accomplish the transfer of contour information between the visual and auditory modalities. The multimodal nature of music highlights the importance of understanding how listeners convert musical information between modalities, and melodic contour is a prime example. For example, there are numerous musical skills that depend on the accurate conversion of melodic contour between the visual and auditory modalities, such as the ability to read music, record melodies in written form and monitor the accuracy of musical performances in real-time.

There are several potential implications of the current results. First, they validate the applicability of theoretical models of contour structure to cross-modal investigations. Second, these findings have the potential to inform models of music expertise and cross-modal music cognition. Third, this research may have relevance to practical applications such as remedial speech perception training and pedagogical approaches to musical instruction.

REFERENCES

't Hart, J., Collier, R., & Cohen, A. J. (1990). *A perceptual study of intonation: An experimental-phonetic approach to speech melody*. Cambridge, UK: Cambridge University Press.

Adams, C. R. (1976). Melodic contour typology. *Ethnomusicology, 20(2)*, 179-215.

Bachem, A. (1950). Tone height and tone chroma as two different pitch qualities. *Acta Psychologica, 7*, 80-88.

Balch, W. R. (1984). The effects of auditory and visual interference on the immediate recall of melody. *Memory and Cognition, 12(6)*, 581-589.

Balch, W. R., & Muscatelli, D. L. (1986). The interaction of modality condition and presentation rate in short-term contour recognition. *Perception and Psychophysics, 40(5)*, 351-358.

Bartlett, J. C., & Dowling, W. J. (1980). Recognition of transposed melodies: A key-distance effect in developmental perspective. *Journal of Experimental Psychology: Human Perception and Performance, 6(3)*, 501-515.

Bennett, P. J., & Banks, M. S. (1987). Sensitivity loss in odd-symmetric mechanisms and phase anomalies in peripheral vision. *Nature, 326(6116)*, 873-876.

Bennett, P. J., & Banks, M. S. (1991). The effects of contrast, spatial scale, and orientation on foveal and peripheral phase discrimination. *Vision Research, 31(10)*, 1759-1786.

Boltz, M. G., & Jones, M. R. (1986). Does rule recursion make melodies easier to reproduce? If not, what does? *Cognitive Psychology, 18(4)*, 389-431.

Boltz, M. G., Marshburn, E., Jones, M. R., & Johnson, W. W. (1985). Serial-pattern structure and temporal-order recognition. *Perception and Psychophysics, 37(3)*, 209-217.

Bregman, A. S., & Campbell, J. (1971). Primary auditory stream segregation and perception of order in rapid sequences of tones. *Journal of Experimental Psychology, 89(2)*, 244-249.

Brochard, R., Dufour, A., & Després, O. (2004). Effect of musical expertise on visuospatial abilities: Evidence from reaction times and mental imagery. *Brain and Cognition, 54(2)*, 103-109.

Chang, H.-W., & Trehub, S. E. (1977). Auditory processing of relational information by young infants. *Journal of Experimental Child Psychology, 24(2)*, 324-331.

Cuddy, L. L., Cohen, A. J., & Mewhort, D. J. (1981). Perception of structure in short melodic sequences. *Journal of Experimental Psychology: Human Perception and Performance, 7(4)*, 869-883.

Cupchik, G. C., Phillips, K., & Hill, D. S. (2001). Shared processes in spatial rotation and musical permutation. *Brain and Cogni-*

tion, 46(3), 373-382.

Davies, J. B., & Jennings, J. (1977). Reproduction of familiar melodies and perception of tonal sequences. *Journal of the Acoustical Society of America, 61(2)*, 534-541.

Démonet, J. F., Price, C. J., Wise, R., & Frackowiak, R. S. J. (1994). A pet study of cognitive strategies in normal subjects during language tasks: Influence of phonetic ambiguity and sequence processing on phoneme monitoring. **Brain, 117(4)**, 671-682.

Deutsch, D. (1972). Octave generalization and tune recognition. *Perception and Psychophysics, 11(6)*, 411-412.

Deutsch, D., & Feroe, J. (1981). The internal representation of pitch sequences in tonal music. *Psychological Review, 88(6)*, 503-522.

Dowling, W. J. (1978). Scale and contour: Two components of a theory of memory for melodies. *Psychological Review, 85(4)*, 341-354.

Dowling, W. J. (1991). Tonal strength and melody recognition after long and short delays. *Perception and Psychophysics, 50(4)*, 305-313.

Dowling, W. J., & Fujitani, D. S. (1971). Contour, interval, and pitch recognition in memory for melodies. *Journal of the Acoustical Society of America, 49(2, Pt. 2)*, 524-531.

Dowling, W. J., & Harwood, D. L. (1986). *Music cognition*. San Diego: Academic Press.

Dowling, W. J., & Hollombe, A. W. (1977). The perception of melodies distorted by splitting into several octaves: Effects of increasing proximity and melodic contour. *Perception and Psychophysics, 21(1)*, 60-64.

Dyson, M. C., & Watkins, A. J. (1984). A figural approach to the role of melodic contour in melody recognition. *Perception and Psychophysics, 35(5)*, 477-488.

Eerola, T., & Bregman, M. (2007). Melodic and contextual similarity of folk song phrases. *Musicae Scientiae, Discussion Forum 4A-2007*, 211-233.

Eerola, T., Himberg, T., Toiviainen, P., & Louhivuori, J. (2006). Perceived complexity of western and african folk melodies by western and african listeners. *Psychology of Music, 34*, 337-371.

Eerola, T., Järvinen, T., Louhivuori, J., & Toiviainen, P. (2001). Statistical features and perceived similarity of folk melodies. *Music Perception, 18*, 275-296.

Eerola, T., & Toiviainen, P. (2004). *Midi toolbox: Matlab tools for music research.* University of Jyväskylä: Kopijyvä, Jyväskylä, Finland. Available at http://www.jyu.fi/musica/miditoolbox/

Eiting, M. H. (1984). Perceptual similarities between musical motifs. *Music Perception, 2(1)*, 78-94.

Francès, R. (1988). *The perception of music*. Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.

Frankish, C. (1995). Intonation and auditory grouping in immediate serial recall. *Applied Cognitive Psychology, 9*, S5-S22.

Freedman, E. G. (1999). The role of diatonicism in the abstraction and representation of contour and interval information. *Music Perception, 16(3)*, 365-387.

Friedmann, M. L. (1985). A methodology for the discussion of contour, its application to schoenberg music. *Journal of Music Theory, 29(2)*, 223-248.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science, 7(2)*, 155-170.

Halpern, A. R., Bartlett, J. C., & Dowling, W. J. (1998). Perception of mode, rhythm and contour in unfamiliar melodies: Effects of age and experience. *Music Perception, 15(4)*, 335-355.

Hermes, D. J. (1998a). Auditory and visual similarity of pitch contours. *Journal of Speech, Language, and Hearing Research, 41(1)*, 63-72.

Hermes, D. J. (1998b). Measuring the perceptual similarity of pitch contours. *Journal of Speech, Language, and Hearing Research, 41(1)*, 73-82.

Idson, W. L., & Massaro, D. W. (1978). A bidimensional model of pitch in the recognition of melodies. *Perception and Psychophysics*, 24(6), 551-565.

Jones, M. R., & Boltz, M. G. (1989). Dynamic attending and responses to time. *Psychological Review, 96(3)*, 459-491.

Jones, M. R., & Ralston, J. T. (1991). Some influences of accent structure on melody recognition. *Memory and Cognition, 19(1)*, 8-20.

Kleiner, K. A. (1987). Amplitude and phase spectra as indices of infants' pattern preferences. *Infant Behavior & Development, 10(1)*, 49-59.

Kleiner, K. A., & Banks, M. S. (1987). Stimulus energy does not account for 2-month-olds' face preferences. *Journal of Experimental Psychology: Human Perception and Performance, 13(4)*, 594-600.

Ladd, D. R. (1996). *Intonational phonology*. Cambridge, England: Cambridge University Press.

Lamont, A., & Dibben, N. (2001). Motivic structure and the perception of similarity. *Music Perception, 18(3)*, 245-274.

Lidji, P., Kolinsky, R., Lochy, A., & Morais, J. (2007). Spatial associations for musical stimuli: A piano in the head? *Journal of Experimental Psychology: Human Perception and Performance, 33(5)*, 1189-1207.

Lieberman, P. (1967). *Intonation, perception, and language*. Cambridge, MA: M.I.T. Press.

Marvin, E. W., & Laprade, P. A. (1987). Relating musical contours - extensions of a theory for contour. *Journal of Music Theory, 31(2)*, 225-267.

McDermott, J. H., Lehr, A. J., & Oxenham, A. J. (2008). Is relative pitch specific to pitch? *Psychological Science, 19(12)*, 1263-1271.

Messerli, P., Pegna, A., & Sordet, N. (1995). Hemispheric dominance for melody recognition in musicians and non-musicians. *Neuropsychologia, 33(4)*, 395-405.

Mikumo, M. (1994). Motor encoding strategy for pitches of melodies. *Music Perception, 12(2)*, 175-197.

Mikumo, M. (1997). Multi-encoding for pitch information of tone sequences. *Japanese Psychological Research*, 39(4), 300-311.

Miyazaki, K., & Rakowski, A. (2002). Recognition of notated melodies by possessors and nonpossessors of absolute pitch. *Perception and Psychophysics, 64(8)*, 1337-1345.

Monahan, C. B., Kendall, R. A., & Carterette, E. C. (1987). The effect of melodic and temporal contour on recognition memory for pitch change. *Perception and Psychophysics, 41(6)*, 576-600.

Morris, R. D. (1993). New directions in the theory and analysis of musical contour. *Music Theory Spectrum, 15(2)*, 205-228.

Morrongiello, B. A., & Roes, C. L. (1990). Developmental changes in children's perception of musical sequences: Effects of musical training. *Developmental Psychology, 26(5)*, 814-820.

Morrongiello, B. A., Trehub, S. E., Thorpe, L. A., & Capodilupo, S. (1985). Children's perception of melodies: The role of contour, frequency, and rate of presentation. *Journal of Experimental Child Psychology, 40(2)*, 279-292.

Narmour, E. (1990). *The analysis and cognition of basic melodic structures: The implication-realization model*. Chicago, IL, US: University of Chicago Press.

Ottman, R. W. (1986). *Music for sight-singing* (3rd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Peretz, I., & Babaï, M. (1992). The role of contour and intervals in the recognition of melody parts: Evidence from cerebral asymmetries in musicians. *Neuropsychologia, 30(3)*, 277-292.

Perry, D. W., Zatorre, R. J., Petrides, M., Alivisatos, B., Meyer, E., & Evans, A. C. (1999). Localization of cerebral activity during simple singing. *Neuroreport, 10(18)*, 3979-3984.

Pick, A. D., Palmer, C. F., Hennessy, B. L., & Unze, M. G. (1988). Children's perception of certain musical properties: Scale and contour. *Journal of Experimental Child Psychology, 45(1)*, 28.

Pierrehumbert, J., & Beckman, M. (1988). *Japanese tone structure*. Cambridge, MA: The MIT Press.

Pierrehumbert, J., & Hirschberg, J. (1990). *The meaning of intonation contours in interpretation of discourse*. In P. R. Cohen, J. Morgan & M. E. Pollack (Eds.), Intentions in communication (pp. 271-311). Cambridge, MA: M.I.T. Press.

Polansky, L., & Bassein, R. (1992). Possible and impossible melody - some formal aspects of contour. *Journal of Music Theory, 36(2)*, 259-284.

Quinn, I. (1999). The combinatorial model of pitch contour. *Music Perception, 16(4)*, 439-456.

Ruckmick, C. C. (1929). A new classification of tonal qualities. *Psychological Review, 36*, 172-180.

Schmuckler, M. A. (1999). Testing models of melodic contour similarity. Music Perception, 16(3), 295-326.

Schmuckler, M. A. (2004). *Pitch and pitch structures*. In J. Neuhoff (Ed.), Ecological psychoacoustics (pp. 271-315). San Diego, CA: Elsevier Science.

Schmuckler, M. A. (2008). Melodic contour similarity using folk melodies. Submitted.

Schoenberg, A. (1967). *Fundamentals of musical composition*. New York: St. Martins.

Schwarzer, G. (1993). Development of analytical and holistic processes in the categorization of melodies/entwicklung analytischer und holistischer prozesse bei der kategorisierung von melodien. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 25(2)*, 89-103.

Shepard, R. N. (1982). Geometrical approximations to the structure of musical pitch. *Psychological Review, 89(4)*, 305-333.

Shmulevich, I. (2004). A note on the pitch contour similarity index. *Journal of New Music Research, 33(1)*, 17-18.

Smey, D. (2007). *Sight-singing bonanza*. from http://davesmey.com/eartraining/sightsing.pdf

Thomassen, J. M. (1982). Melodic accent - experiments and a tentative model. *Journal of the Acoustical Society of America, 71(6)*, 1596-1605.

Trehub, S. E., Bull, D., & Thorpe, L. A. (1984). Infants' perception of melodies: The role of melodic contour. *Child Development, 55(3)*, 821-830.

Waters, A. J., Townsend, E., & Underwood, G. (1998). Expertise in musical sight reading: A study of pianists. *British Journal of Psychology*, 89(1), 123-149.

Watkins, A. J. (1985). Scale, key, and contour in the discrimination of tuned and mistuned approximations to melody. *Perception and Psychophysics*, 37(4), 275-285.

White, B. W. (1960). Recognition of distorted melodies. *American Journal of Psychology, 73*, 100-107.

Xu, Y. (2005). Speech melody as articulartorily implemented communicative functions. *Speech Communication, 46*, 220-251.

Zatorre, R. J., Evans, A. C., & Meyer, E. (1994). Neural mechanisms underlying melodic perception and memory for pitch. *Journal of Neuroscience, 14(4)*, 1908-1919.

Zatorre, R. J., Perry, D. W., Beckett, C. A., Westbury, C. F., & Evans, A. C. (1998). Functional anatomy of musical processing in listeners with absolute pitch and relative pitch. *Proceedings of the National Academy of Sciences of the United States of America, 95(6)*, 3172-3177.

Author Notes

Grants from the Natural Sciences and Engineering Council of Canada to Mark A. Schmuckler and William F. Thompson supported this research. Please address correspondence concerning this article to Jon. B. Prince (jon.prince@utoronto.ca).

Notes

Schmuckler (1999) used both difference scores and correlational measures for computing perceived similarity. Interestingly, in that work as well as subsequent research (Schmuckler, 2004), difference scores have proven to be somewhat more sensitive than correlations to perceived contour similarity. One possible reason for this finding is that outliers can greatly influence correlation values. Such extreme values occasionally occur with Fourier analysis information, in terms of the relative strengths of high frequency information, which typically tends to be quite low. Such outliers, then, would have a more dramatic effect on correlations than average difference scores, and could thus lead to somewhat distorted similarity predictions.

Musicophilia: Tales of Music and the Brain
By Oliver Sacks

Alfred A. Knopf, 2007,
List price $19.95 CDN (Paperback) - 381 pp.,
ISBN 13: 9780676979794

Oliver Sacks is Professor of Clinical Neurology and Clinical Psychiatry at Columbia University Medical Center, author of 10 books including "The Man Who Mistook His Wife for a Hat", and clearly an avid music lover. In Musicophilia, Dr. Sacks shares stories of patients, composers, performers, and even some of his own personal experiences to illustrate musical eccentricities, disorders related to music, and the healing power of music.

The book contains 29 chapters, divided into 4 sub-sections: 1) Haunted by Music; 2) A Range of Musicality; 3) Memory, Movement, and Music; 4) Emotion, Identity, and Music.

In the first section, Sacks illustrates various ways in which music can appear in our heads, unbidden and sometimes intrusive. In Chapter 1, we learn about individuals who had not been especially interested in music, but suddenly became musicophiles, one after being struck by lightning. Chapter 2 tells the story of individuals who hear mysterious, unrecognizable but somehow familiar music before experiencing seizures. The stories in Chapter 3 also involve individuals experiencing seizures, but in this case triggered by external music, often specific types, songs, instruments or noises. In Chapter 4, Sacks shares research that shows how our auditory cortex is active even when we are just imagining music. Sometimes the music we imagine is outside of our control, and we are subject to repetitive "brainworms" (or "earworms"), which Sacks describes in Chapter 5. To close the section, Chapter 6 speaks of musical hallucinations. Sacks introduces individuals who started hearing music, only to discover that, despite its apparent exteriority, it was only in their heads.

The second section of the book speaks to the wide range of skills related to musicality. In Chapter 7, we are introduced to the idea that different aspects of musicality are dissociable; one person may be technically proficient, but lack understanding of tone and phrasing, while another may be passionate about music, but lack an ear for it. When an individual has selective impaired processing of music (i.e., speech is spared), they are said to have amusia or what is commonly referred to as tone deafness (Chapter 8). This encompasses a range of issues, including the inability to hear differences between tones or timbres, or to make sense of strings of notes. Sacks describes absolute (perfect) pitch in Chapter 9, and asks why some people have it and not others. In Chapter 10 we learn about the peculiar case of cochlear amusia, characterized by distortions of notes that seem to shift and run into each other. We come to appreciate the value of having two ears when we read Chapter 11, which describes how we lose the auditory equivalent of depth perception in vision, and emotional force as well, when we experience deafness in one ear. The remarkable musicality of savants is introduced in Chapter 12. These individuals are usually impaired in abstract thinking and linguistic skills, yet they may have astonishing memory for music after hearing it for the first time, as well as transposition and improvisation skills. Chapter 13 looks at the links between musicality and blindness; studies have shown that blind individuals are more interested in music than their sighted counterparts, more likely to have absolute pitch, and often show exceptional music ability. To conclude the section, Chapter 14 talks of music and synesthesia, a melding of the senses in which different keys, rhythms, timbres, tempos or instruments may be "seen" in different colours, or different intervals may have distinctive "tastes".

The third section of the book explores relationships between music and physical movement; not only do we use movement to produce music, but we tend to move in response to music. First, in Chapter 15, Sacks shares the captivating story of Clive Wearing, who experiences amnesia as a result of a brain infection. He is incapable of remembering anything that took place more than a few seconds ago, but incredibly he can still perform and conduct music with great sensitivity. In another story of musical sparing, Chapter 16 tells how individuals with aphasia, who cannot produce spoken language, may yet be able to sing, and thus learn to speak again through melodic intonation therapy. Chapter 17 briefly tells the tale of an individual with dyskinesia who has developed a musical cantillation (chant) to match his involuntary movements. In Chapter 18 we learn how the tics associated with Tourette's syndrome can be affected by the frequency and intensity of music, and how some individuals are free of tics when they perform or conduct music. General links between music and rhythm are discussed in Chapter 19: how humans (and no known animals) entrain to a rhythm, how we can use music to help remember sequences, how rhythm can bind us communally. In Chapter 20, we return to the Parkinsonian patients that Sacks famously wrote about in "Awakenings", which was later transformed into an Oscar-nominated movie. These patients are unable to initiate movement, but can respond to an event, such as catching a ball thrown at them, or dancing to music. A pianist with an amputated arm and "phantom" fingers is the subject of Chapter 21. The final chapter in this section, Chapter 22 exposes musician's dystonia, which is a neurological movement disorder akin to writer's cramp. The instinct to fight this disorder by working harder may paradoxically cause it to worsen.

One of the main reasons we listen to music is because of how it makes us feel. The final section of the book looks at music and emotion. Sacks shares his own musical dreams in Chapter 23, as well as relating how Wagner and Berlioz found musical inspiration in their dreams. In Chapter 24 we learn about how there seem to be different mechanisms for intellectual and emotional appreciation of music; patients with acquired amusia may still feel emotions although they can no longer comprehend the music, whereas individuals with Asperger's syndrome may not feel musical emotion, and yet still enjoy and appreciate music. In Chapter 25, Sacks shares stories of how music has shattered the frozen feelings and melancholia associated with bereavement. Chapter 26 briefly showcases one patient who showed flat affect when speaking, but seemed to express emotions when singing. Chapter 27 introduces individuals who show musical disinhibition, likely related to frontotemporal dementia. These individuals are alert, oriented and able to write and draw pictures, but they sing perpetually and irrepressibly and can therefore be difficult to communicate with. In Chapter 28, we learn about the love of music common to many individuals with Williams syndrome. These individuals are extremely sensitive to all sounds, and have a tendency to become completely absorbed by music. In a testament to the emotional power of music, Chapter 29 relates how music can elicit a response from individuals with even advanced forms of dementia. Some individuals can still play an instrument or sing, and even learn and memorize new songs. The positive moods engendered by the music can persist long after the individual remembers having heard the music.

Although the book is intended for a general audience, some knowledge of brain structures would be helpful; neurological terminology is often used without definition or context. While some disorders are described in detail, including history about the discovery and study of the disorder, rate of incidence and theorized etiology, others are named without establishing this background information and the reader is left wanting more.

This reader was intrigued by the recurring idea that we all might have abilities that are suppressed or inhibited. For example, the memory, transposition and improvisation skills of musical savants might be lying dormant in many of us, released in savants only by an accident in the normal course of development, a brain injury or stroke. Similarly, absolute (perfect) pitch might be available to many of us if only a stimulus could trigger its release.

Dr. Sacks is a renowned storyteller, and he provides enough details about his patients to bring their stories to life. He often shares his subject's own words, which is an extremely effective means of illuminating their personal experiences. However, on occasion the stories seem disjointed from one another and the reader is left to make their own connections between them and wonder at what conclusions may be drawn.

Overall, one is left marveling at the fragility of the brain and the variety of disorders that can arise, but equally at the plasticity of the brain and how we can adapt to changes that occur.

Gillian Sandstrom,
Department of Psychology, Ryerson University

CANADIAN NEWS

Retirement Biography of Dr. J. David Quirt

It was 1973, thirty-five years ago, when young J. David Quirt, with a freshly minted PhD in physics from UBC, joined the Noise and Vibration Group in the Department of Building Research (which became Institute for Research in Construction).

In the 1970's, David's research established the technical basis for federal guidelines for sound insulation near airports and highways, which he co-authored with CMHC colleagues. They were subsequently adapted for uses such as Part 10 of the Alberta Building Code and Ontario Ministry of Housing regulations for development near highways or railways, as well as policy for noise from residential A/C units.

To gain the thorough understanding of the science necessary to create the Guides, David completed extensive studies of sound transmission into buildings and through specific components. Using these data he developed empirical prediction methods for sound from railways and highway traffic. David then led multi-year validation studies that combined laboratory experiments with parametric field studies in a special test building on Armstrong Road near Ottawa Airport, using hundreds of low-level flypasts. Extensive laboratory studies that followed established the first clear delineation of how sound transmission through windows depends on glazing details. Resulting papers won best paper awards in the Journal of the Acoustical Society of America in 1982 and 1983.

In the nineties David led a number of multidisciplinary joint research projects with IRC's Fire Risk Management Program and industry partners. Most notable was his project on sound transmission of walls where he tested a sparse matrix of assemblies, to establish the dependence of sound transmission on generic construction parameters and to provide comparisons of competing products, such as insulation types. The experiment design supported both development of empirical models and calibration of analytic models for sound and fire resistance. David led the team to work closely with industry experts, ensuring that key issues for both fire and sound control were addressed. This enabled the findings to be adapted into "approved constructions" listed in the Part 9 Tables in the National Building Code.

Latterly, David has been collaborating with his NRC-IRC colleagues on consortium-funded projects on structure-borne sound (Flanking Transmission) that are changing the approach to noise control in wood-framed construction. Most recently, David completed a consortium project on Firestops/blocks with more than 21 partners to generate another award winning best practice Guide.

David was Group Leader of the NRC-IRC Acoustics Subprogram from 1987 to 2008, and Director of Building Envelope and Structure Program from 1993 to 1998. David is a Fellow of the Acoustical Society of America (since 1992); Associate Editor for Architectural Acoustics of the Journal of the Acoustical Society of America (1987-2001); Executive secretary of the Canadian Acoustical Association, Member of CSA technical committees, Chair of the Canadian Advisory Committee for ISO TC43/SC2 "building acoustics" (since 1992).

The Acoustics Group is very fortunate because David has decided to have a long transition between full-time acoustics and retirement, which means we will continue to benefit from the stimulating technical discussions drawn from his vast knowledge, and he will continue to work at NRC-IRC on selected projects and standards activities. (Should you wish to contact David, his NRC-IRC contact information remains unchanged.)

On behalf of your friends and colleagues at NRC-IRC we wish you all the very best for retirement.

Trevor Nightingale

Ramani Ramakrishnan, Editor-in-Chief of Canadian Acoustics Journal adds:

I have known David Quirt since 1984 and got to know him better after I started working at the Ontario Ministry of the Environment (MOE) as a Policy Analyst. The National Research Council provided assistance to MOE while the Ministry was formulating its land use and noise assessment guidelines. In particular, Dave Quirt was instrumental in providing information about sound transmission issues as well as performance data of building assemblies. David Quirt, as the head of the Noise Analysis Group at the Institute for Research in Construction, worked as a consultant to MOE in the assessment of noise from home air-conditioning units. Since I was the project leader at MOE, it was a pleasure for me to liaise with IRC and David Quirt.

My association with David became closer after I became the Editor-in-Chief of the Canadian Acoustics Journal. David, as the secretary of the Canadian Acoustical Association, had the responsibility to provide me, with various information such as minutes of meetings, membership directory, back page updates and membership forms, at least four times a year. David was very punctual and always supported me in my responsibility to bring out each issue of the journal on time to the CAA membership. I found working with David very easy and cordial. I want to express my appreciation and thanks to him. I wish him well during his retirement years and hope that his strong relationship with CAA will continue unabated.

What's new in acoustics in Canada?

Jérémie Voix <voix@caa-aca.ca>

I recently made a commitment to a revived "What's new in acoustics in Canada?" section for our journal. I have now all my keywords alerts, RSS feeds, web spider and other automated watching tools in place to harvest some potentially interesting news. Nevertheless, you could really contribute to this section by kindly pointing me to institutions or individuals that are involved in Acoustics -at large- in Canada.

For this journal issue, we will start with special “noisy” news... stay tuned!

14th Annual International Noise Awareness Day is scheduled April 29, 2009

On Wednesday, Aril 29, 2009, the 14th Annual International Noise Awareness Day will be observed throughout North America and in many other countries around the world, coordinated by the Noise Center of the New York League for the Hard of Hearing. Locally, members and supporters of the Right to Quiet Society will distribute leaflets in Vancouver and other places to raise awareness about the ever increasing level of noise pollution in our environment.

[Source: http://www.quiet.org/ Accessed on Feb 24th, 2009]

Le bruit – Gare au bruit : un avertissement à ébruiter !

Le bruit est susceptible d’affecter les capacités d’écoute et de concentration, d’accroître le degré de stress et, à la longue, de réduire la capacité auditive. Cette fiche qui veint de paraitre à la CSST (Commission de la santé et de la sécurité du travail (CSST) propose au personnel des centres de la petite enfance (CPE) et autres services de garde à l’enfance des solutions pour assurer la sécurité du travail. I donne un aperçu des situations génératrices de bruit dans les CPE et fournit des exemples de moyens de prévention, dont l’installation de tuiles acoustiques au plafond.

[Source: http://www.irsst.qc.ca/files/documents/fr/prev/v22_01/4.pdf Accédé le 24 février 2009]

FAQs: Whale and dolphin strandings and beachings

After the plight of three dolphins trapped in the ice off the northeast coast of Newfoundland last week, a very interesting article written by Emily Chung, CBC News to understand why dolphins would get trapped in the ice along Canada's coasts. It explains that human activity that causes a lot of underwater noise might also play a role when the animals are near the edge of the ice.

[Source: http://www.cbc.ca/technology/story/2009/02/24/f-tech-whale-dolphin-strandings.html
Accessed Feb. 27Th, 2009]

Lorraine reçoit le premier prix du concours Villes et villages paisibles

La ville de Lorraine (QC) a obtenu le premier prix du concours lancé en 2006 par l'Ordre des orthophonistes et audiologistes du Québec, Villes et villages paisibles, une reconnaissance des efforts réalisés pour conserver et améliorer la quiétude sur son territoire.

[Source: http://www.journallecourrier.com/article-263625-Lorraine-recoit-le-premier-prix-du-concours-Villes-et-villages-paisibles.html Accédé le 31 décembre 2008]

Waterfall Honours Canada with Niagara Speakers

Speakers made of glass? It's true. Each Waterfall speaker, made by the French company Waterfall Audio, is constructed almost completely of glass. Adding to the clever design is the fact that each one is named after a famous waterfall. And the latest to join the line up honours Canada (or at least the half of the waterfall that's on our side) with the name Niagara. Waterfall Audio's Niagara speakers aren't for the average buyer, though: a pair will run you a cool $53,000!

[Source: http://www.hereshow.ca/news_detail.asp?nid=1643 Last accessed Feb. 27Th, 2009]

Coûts sous-estimés du bruit en milieu de travail

L'Institut national de santé publique du Québec (INSPQ) a publié à la fin de l'automne 2007 une étude sur l'analyse des coûts associés au bruit en milieu de travail pour le régime d'indemnisation québécois. Les résultats montrent que les déboursés de la Commission de la santé et de la sécurité du travail (CSST) en matière de surdité ne représentent qu'environ le quart des coûts réels dubruit en milieu de travail. Ainsi, les chercheurs estiment que le coût annuel réel pour le régime d'indemnisation s'élève à 35,4 M$ au lieu des 8,4 M$ basés sur les seules données du régime d'indemnisation, notamment en considérant les accidents survenus attribuables au bruit ou à une perte d'audition.

[Source : http://www.inspq.qc.ca/publications/notice.asp?E=p&NumPublication=712 Accédé le 20 novembre 2008]

'Let's make some noise, Canada!' promo launched

Vancouver's 2010 Olympic committee is launching a "Let's make some noise, Canada!" campaign to mark the one-year countdown to the opening ceremonies. One of the easiest ways to participate was by simply making some noise on February 12 — one year from the start of the Olympic Winter Games Opening Ceremony. At exactly 6:00 pm local time, in time zones across the country, Canadians of all ages were invited to make some noise. You could honk car horns in downtown Toronto or sound foghorns in Halifax, ring sleigh bells in Prince George, chant folkloric songs, or dance to Aboriginal drumming in St. John's . . . anything to show Canadians' pride as the Games draw near. The noise will roll like a wave of sound across the land as the clock strikes 6:00 pm in each of the country's six time zones.

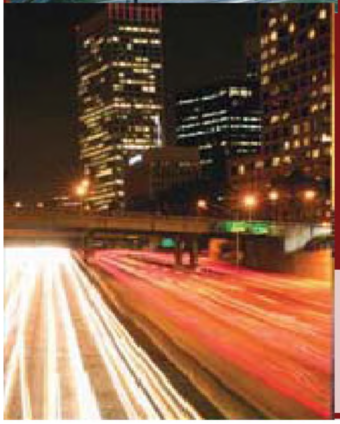
[Source: http://www.vancouver2010.com/en/news/news-releases/-/62786/32566/xi1wuv/one-year-to-go-to-the-vancouve.html Last Accessed Feb. 19Th, 2009]

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

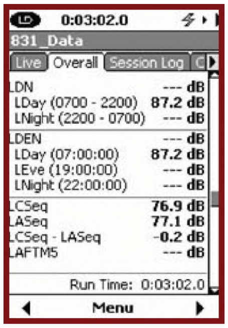


FEATURES

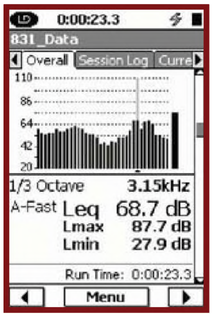
- Class 1/Type 1 sound level meter
- Small size with large display. Ergonomic
- User friendly operator interface
- 120MB standard memory expandable up to 2GB
- Single measurement range from 20 to 140 dB SPL
- Up to 16 hours of battery life
- Provided with utility software for instrument set-up and data download
- Field upgradeable
- AUX port for connection to USB mass storage & cellular modems

MEASUREMENT CAPABILITIES

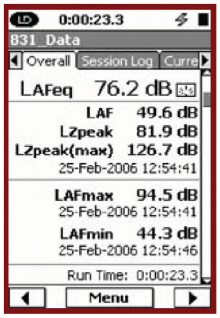
- Real time 1/1 & 1/3 octave frequency analysis
- Simultaneous display of several noise measurements—*ANY DATA* (Leq, Lmax, Spectra, etc
- Automatic logging of user selectable noise measurements (Leq, Lmax, Spectra, etc…)
- Exceedance logging with user selectable trigger levels
- Audio and voice recording with replay



```
0:03:02.0
831_Data
LDN                    --- dB
 LDay (0700 - 2200)   87.2 dB
 LNight (2200 - 0700)  --- dB
LDEN                   --- dB
 LDay (07:00:00)      87.2 dB
 LEve (19:00:00)       --- dB
 LNight (22:00:00)     --- dB
LCSeq                 76.9 dB
LASeq                 77.1 dB
LCSeq - LASeq         -0.2 dB
LAFTM5                 --- dB
Run Time: 0:03:02.0
```



```
0:00:23.3
831_Data
1/3 Octave    3.15kHz
A-Fast Leq    68.7 dB
       Lmax   87.7 dB
       Lmin   27.9 dB
Run Time: 0:00:23.3
```



```
0:00:23.3
831_Data
LAFeq  76.2 dB
        LAF   49.6 dB
     LZpeak   81.9 dB
LZpeak(max)  126.7 dB
  25-Feb-2006 12:54:41
     LAFmax   94.5 dB
  25-Feb-2006 12:54:41
     LAFmin   44.3 dB
  25-Feb-2006 12:54:46
Run Time: 0:00:23.3
```

Larson Davis
A PCB Group Co.

WWW.DALIMAR.CA INFO@DALIMAR.CA
QUEBEC 450-424-0033 ONTARIO 905-707-9000 ALBERTA 403-288-4416

inter·noise 2009
innovations in practical noise control
2009 August 23-26 — Ottawa, Canada

Invitation and call for papers

Dear Colleagues,

The Organizing Committee of the 38th International Congress and Exposition on Noise Control Engineering (INTER-NOISE 2009) extends a warm welcome and invitation to participate fully in what promises to be the premier noise control engineering conference of 2009. The INTER-NOISE 2009 Congress, sponsored by the International Institute of Noise Control Engineering and co-organized by the Canadian Acoustical Association and the Institute of Noise Control Engineering-USA, will be held in Ottawa, from 23–26 August 2009.

The Congress will feature a broad range of high-level research papers from around the world, as well as an extensive exhibition of noise and vibration control and measurement equipment and systems. Distinguished speakers will provide additional stimulation for our technical sessions and discussions with a focus on our theme of "Innovations in Practical Noise Control."

The 2009 International Symposium on Active Control of Sound and Vibration (ACTIVE 2009) will be held 20-22 August, immediately before the INTER-NOISE 2009 congress.

The INTER-NOISE 2009, and ACTIVE2009, will be held at the Westin Ottawa Hotel, which is located in the heart of Canada's Nation's capital close to all major attractions, Parliament Buildings, National Gallery, Royal Mint, and many museums.

Both INTER-NOISE and ACTIVE symposium will have the same schedule for abstract and paper submission:

Abstracts Due: 23 January 2009
Notification of Acceptance: 20 March 2009
Papers Due: 22 May 2009

The congress website provides complete information on the congress including, instructions on paper and abstract submission, planned technical sessions, distinguished lectures, exposition, registration, and social events, so please visit internoise2009.com often.

It is our pleasure to welcome you to INTER-NOISE 2009 and ACTIVE 2009 in Ottawa.

Trevor Nightingale and Joe Cuschieri,
Co-Presidents
&
Brad Gover and Stuart Bolton,
Technical Co-Chair

In a Class *of its Own*

The unmistakable look of Hand-held Analyzer Type 2270 can overshadow a number of discrete yet significant distinctions which make this powerful instrument the complete toolbox for sound and vibration professionals. These include:

- Integrated digital camera
- Two-channel measurement capability
- Integrated LAN and USB interfaces for fast data transfer to PC and remote control and monitoring of Type 2270
- Environmental protection IP 44

Versatile in the Extreme

Type 2270 also boasts a wide range of application software modules that can be licensed separately so you get what you need when you need it.

Currently available measurement software includes:

- Sound Level Meter application
- Real-time frequency analysis
- Logging (noise level profiling)
- Sound and vibration recording
- Building acoustics
- Tonal assessment

Type 2270 meets the demands of today's wide-ranging sound and vibration measurement tasks with the accuracy and reliability associated with Brüel & Kjær instrumentation.

To experience the ease-of-use of Type 2270, just go to www.bksv.com and view the on-line video demonstrations.

For more information please contact your local Brüel & Kjær representative

Hand-held Analyzer *Type 2270*

HEADQUARTERS: DK-2850 Nærum · Denmark · Telephone: +4545800500
Fax: +4545801405 · www.bksv.com · info@bksv.com

Australia (+61)29889-8888 · Austria (+43)18657400 · Brazil (+55)115188-8166
Canada (+1)514695-8225 · China (+86)1068029906 · Czech Republic (+420)267021100
Finland (+358)9-755950 · France (+33)169907100 · Germany(+49)42117870
Hong Kong (+852)25487486 · Hungary (+36)12158305 · Ireland (+353)18037600
Italy (+39)025768061 · Japan (+81)337798671 · Republic of Korea (+82)234730605
Netherlands (+31)318 55 9290 · Norway (+47)66771155 · Poland (+48)228167556
Portugal (+351)214711453 · Singapore (+65)3774512 · Slovak Republic (+421)254430701
Spain (+34)916590820 · Sweden (+46)84498600 · Switzerland (+41)18807035
Taiwan (+886)227139303 · United Kingdom (+44)1438739000 · USA (+1)8003322040

Local representatives and service organisations worldwide

Brüel & Kjær

Canadian Acoustical Association
Association Canadienne d'Acoustique

PRIZE ANNOUNCEMENT • ANNONCE DE PRIX

Prize

Edgar and Millicent Shaw Postdoctoral Prize in Acoustics
Alexander G. Bell Graduate Student Prize in Speech Communication and Hearing
Eckel Graduate Student Prize in Noise Control
Fessenden Graduate Student Prize in Underwater Acoustics
Raymond Hetu Undergraduate Student Prize in Acoustics

Prix

Prix Post-doctoral Edgar et Millicent Shaw en Acoustique
Prix etudiant Alexander G. Bell en Communication orale et audition (2e ou 3e cycle)
Prix etudiant Eckel en Controle du bruit (2e ou 3e cycle)
Prix etudiant Fessenden en Acoustique sous-marine (2e ou 3e cycle)
Prix etudiant Raymond Hetu en Acoustique (1er cycle)

Deadline for Applications:
April 30th 2009

Date limite de soumission des demandes:
30 Avril 2009

Consult CAA website for more information
Consultez le site Internet de l'ACA pour de plus amples renseignements
(http://www.caa-aca.ca)

PDA & Laptop
Displays
Wired
Wireless
SLARM
94.0 dBA
98.3 dB
102.5 dB
RT History
Stored History
Excedence Events
94.1 dB
Adjustable Thresholds
ACO Pacific, Inc.
SLARM™
NORMAL
ALERT
ALARM
CONNECT

--- SECOND ANNOUNCEMENT ---

ACOUSTICS WEEK IN CANADA

Niagara-on-the-Lake, 14 - 16 October 2009

Acoustics Week in Canada 2009, the annual conference of the Canadian Acoustical Association, will be held in Niagara-on-the-Lake, Ontario from 14 to 16 October 2009. This is the premier Canadian acoustical event of the year, and is being held in beautiful, quaint Niagara-on-the-Lake village, making it an event that you do not want to miss. The conference will include three days of plenary lectures, technical sessions on a wide range of areas of acoustics, the CAA Annual General Meeting, an equipment exhibition, and the conference banquet and other social events.

Venue and Accommodation – The conference will be held at the Pillar & Post Inn, Niagara-on-the-Lake, Ontario [http://www.vintage-hotels.com/niagara-on-the-lake/hotels/pillar-and-post.php]. Pillar and Post is located in a quiet residential area, surrounded by a wonder of gardens and only a five minute walk from Niagara-on-the-Lake's main street. Originally built in the late 1800s, it was used as a canning factory in the midst of Niagara's wine and fruit region. Since 1970 it has been gradually transformed into a luxurious country inn with 122 newly redesigned guestrooms, plus ample meeting space with the latest business amenities. Participants registering with the hotel before 5 September 2009 will receive the reduced room rate of $179/night (single or double). Stay at the conference hotel to be near all activities and your colleagues, and to help make the conference a financial success, to the benefit of all CAA members.

Plenary Lectures – Plenary lectures will be presented in the areas of Architectural Acoustics/Ecology (Dr. Bradley & Prof. Westerkamp), Psychological Acoustics (Prof. Neuhoff) and Bio Acoustics (Prof. Kolios).

Special Sessions – Special sessions (about 23) consisting of invited and contributed papers are currently being organized on the following topics:

- Architectural Acoustics
- Acoustic Ecology and Soundscape
- Sound Absorbing Materials
- Biomedical Acoustics
- Speech Production, Speech Perception and Speech Disorders
- Noise Control
- Aeroacoustics
- Wavelet Acoustics
- Occupational Noise Standards
- Psychological Acoustics
- Vibroacoustics

If you would like to propose and/or organize a special session in your technical area, please contact the Conference Chair or the Technical Chair as soon as possible.

Equipment Exhibition – The conference will include one one-day exhibition of acoustical equipment and products on Thursday 15 October 2009. If you are an equipment supplier interested in participating in the exhibition, please contact the Exhibition Coordinator as soon as possible.

Social Events – The conference will begin on Wednesday morning with an opening ceremony and welcome by Prof. Kendra Schank Smith, Chair, Architectural Science Department, Ryerson University. On Wednesday evening, a reception will be held for all delegates, followed by a potential Soundscape Walk along the village's main street. Banquet will be held on Thursday with a musical listening experience, "EMOTICHAIR."

Spousal Program – If the number warrants, Mrs. Suman Ramakrishnan will organize a spousal program with a walking tour of Niagara-on-the-Lake as well as a trip to Niagara Falls with stops along the Niagara Parkway for all attending spouses and children. Details will be made available with registration forms.

Student Participation – The participation of students is strongly encouraged. Travel subsidies and reduced registration fees will be available. A hotel room-sharing program will be available to reduce costs. Student presenters are eligible to win prizes for the best presentations at the conference.

Paper Submission – Following are the deadlines for submission of abstracts, and of two-page summaries for publication in the proceedings issue of *Canadian Acoustics*: Abstracts submission: 1 June 2009; submission of two-page summaries: 15 July 2009.

Registration – details of registration fees and the registration form will be made available on the conference website. Early registration at a reduced fee is available until 1 September 2009.

Local Organizing Committee

- *Conference Chair: Ramani Ramakrishnan [rramakri@ryerson.ca]*
- *Technical Chair: Frank Russo [russo@ryerson.ca]*
- *Technical Committee: Ben Dyson [bdyson@ryerson.ca]*
- *Local Co-Chair: Moustafa Osman [moustafa.osman@sympatico.ca]*
- *Treasurer: Dalila Guisti [dalila@jadeacoustics.com]*
- *Equipment Exhibition: Rich Peppin [RPeppin@aol.com]*
- *Registration: Mandy Chan [machan@hgcengineering.com]*
- *Registration: Megan Munro [mmunro@hgcengineering.com]*
- *Registration: Payam Ezzatian [payam.ezzatian@gmail.com]*
- *Website: Payam Ashtiani [pashtiani@aercoustics.com]*
- *Translations: Inna Petrennic [inna@echologics.com]*

Conference Website at http://www.caa-aca.ca/

--- SECONDE ANNONCE---

SEMAINE CANADIENNE D'ACOUSTIQUE

Niagara-on-the-Lake, 14-16 octobre 2009

La conférence annuelle de l'Association Canadienne d'Acoustique se tiendra à Niagara-sur-le-Lac en Ontario du 14 au 16 octobre 2009. Il s'agit de plus important événement canadien de l'acoustique de l'année que vous ne voulez pas manquer, car il aura lieu à Niagara-sur-le-Lac, un bel et pittoresque village. Trois jours de sessions plénières, ainsi que des sessions techniques seront présentées, couvrant un large éventail du domaine de l'acoustique. La conférence comprendra aussi la réunion annuelle générale de l'ACA, l'exposition de divers équipements acoustique, un banquet et autres événements sociaux.

Lieu du congrès et hébergement – La conférence se tiendra au Pillar & Post Inn à Niagara-sur-le-Lac en Ontario [http://www.vintage-hotels.com/niagara-on-the-lake/hotels/pillar-and-post.php] dans la zone résidentielle tranquille, entourée des jardins étonnants à seulement cinque minutes de marche à pied de la rue principale de Niagara-sur-le-Lac. Construit au départ à la fin des années 1800, ce bâtiment a été conçu comme une conserverie au millieu de la région de vin et fruit de Niagara. Depuis 1970 on le transforme graduellement dans une auberge de luxe aux 122 chambres récemment renouvellées, ainsi qu'au ample espace pour des réunions avec les derniers outils de bureatique. Les délégués qui résevront leur chambre avant le cinque septembre 2009 bénéficieront d'un tarif préférentiel de 179 $/nuit (occupation simple et double). Choisissez cet hôtel pour participer pleinement au congrès, à proximité de toutes les activités et de vos collegues, et pour assurer le succès de la confèrence pour le bénéfice de tous les membres de l'ACA.

Sessions plénières – Les sessions plénières séront présentées dans les domaines de l'acoustique architectural/écologie (le Dr. Bradley et le Pr. Westerkamp), la psychoacoustique (le Pr. Neuhoff) et la bioacoustique (le Pr. Kolios).

Sessions spéciales – Environ 23 sessions spéciales présentées par des conférenciers invités ou par des communications soumises par les délégués sont actuellement organisées autour de divers sujets, tels que:

- Acoustique architecturale
- Ecoacoustique et Soundscape
- Matériaux absorbants
- Acoustique biomédicale
- Production, perception et troubles de langage
- Contrôle de bruit
- Aéroacoustique
- Acoustique de vaguelettes
- Normes des bruits au travail
- Psychoacoustique
- Vibroacoustique

Si vous désirez suggérer un sujet de session spéciale et/ou organiser une de ces sessions, veuillez communiquer avec le président du congrès ou le directeur scientifique.

Exposition technique – La Conférence comprendra une exposition d'équipement et de produits de l'acoustique, qui aura lieu jeudi, le 15 octobre 2009. Si vous êtes un fournisseur d'équipement intéressé de participer, veuillez contacter la personne en charge de la coordination de l'exhibition le plus vite possible.

Activités – La conférence débutera le mercredi matin avec une cérémonie d'ouverture et un discours de bienvenue par la Pr. Kendra Schank Smith, Directrice du Département d'architecture à l'Université Ryerson. Mercredi soir, une réception est prévue pour tous les délégués, suivie par une marche Soundscape le long de la rue principale du village, ainsi qu'un concert musical "EMOTICHAIR".

Programme de famille – Si le nombre de personnes intéressées permet, Mme. Suman Ramakrishnan organisera un programme de famille prévu pour tous les époux et les enfants. Ce programme comprendra une promenade à Niagara-on-the-Lake, ainsi qu'une excursion aux chutes Niagara avec arrêts le long de Niagara Parkway. Renseignements seront disponibles avec formulaires d'inscription.

Participation étudiante – La participation d'étudiants au congrès est vivement encouragée. Des aides financières pour le déplacement et une réduction pour l'inscription seront mises à disposition. Un programme pour faciliter le partage des chambres sera mis sur pied pour réduire les dépenses. Les étudiants présentant leurs travaux seront éligibles pour les prix des meilleurs présentations au congrès.

Soumission des présentations – Les dates limites pour soumission pour la publication dans l'issue en cours de "L'Acoustique Canadienne" sont le 1 juin 2009 pour les résumés at le 15 juillet 2009 pour les sommaires de deux pages.

Inscription – Les détails ainsi que le formulaire d'inscription seront mis en ligne sur le site Web de la conférence. Une réduction sera effective pour tout inscription avant le cinque septembre 2009.

Comité d'organisation

- *Président: Ramani Ramkrishnan [rramakr@ryerson.ca]*
- *Directeur scientifique: Frank Russo [russo@ryerson.ca]*
- *Comité scientifique: Ben Dyson [bdyson@ryerson.ca]*
- *Directeur: Moustafa Osman [moustafa.osma@sympatico.ca]*
- *Trésorier: Dalila Guisti [dalila@jadeacoustics.com]*
- *Exposition technique: Rich Peppin [Rpeppin@aol.com]*
- *Inscription: Mandy Chan [machan@hgcengineering.com]*
- *Inscription: Megan Munro [mmunro@hgcengineering.com]*
- *Inscription: Payam Ezzatian [payam.ezzatian@gmail.com]*
- *Site Web: Payam Ashtiani [pashtiani@aercoustics.com]*
- *Traductrice: Inna Petrennic [inna@echologics.com]*

Site Web de la conférence à http://www.caa-aca.ca/

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 $70.00 for individual members and $30.00 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 $70.00. 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 $ 70.00
[ ] CAA Student Membership $ 30.00

[ ] Institutional Subscription $ 70.00
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:

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 70.00$ pour les membres individuels, et de 30.00$ 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 70.00$. Des compagnies et établissements préfèrent souvent la cotisation de membre bienfaiteur, de 300.00$ 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 70.00 $
[ ] Membre étudiant(e) 30.00 $
[ ] Abonnement institutionnel 70.00 $
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

The Canadian Acoustical Association
l'Association Canadienne d'Acoustique

PRESIDENT
PRÉSIDENT

Christian Giguère
Université d'Ottawa
Ottawa, Ontario
V8W 3P6
(613) 562-5800 x4649
cgiguere@uottawa.ca

PAST PRESIDENT
PRÉSIDENT SORTANT

Stan Dosso
University of Victoria
Victoria, British Columbia
V8W 3P6
(250) 472-4341
sdosso@uvic.ca

SECRETARY
SECRÉTAIRE

David Quirt
P. O. Box 74068
Ottawa, Ontario
K1M 2H9
(613) 993-9746
dave.quirt@nrc-cnrc.gc.ca

TREASURER
TRÉSORIER

Dalila Giusti
Jade Acoustics
411 Confederation Parkway, Unit 19
Concord, Ontario
L4K 0A8
(905) 660-2444
dalila@jadeacoustics.com

EDITOR-IN-CHIEF
RÉDACTEUR EN CHEF

Ramani Ramakrishnan
Dept. of Architectural Science
Ryerson University
350 Victoria Street
Toronto, Ontario
M5B 2K3
(416) 979-5000 #6508
rramakri@ryerson.ca
ramani@aiolos.com

WORLD WIDE WEB HOME
PAGE: http://www.caa-aca.ca

Dave Stredulinsky
(902) 426-3100

ASSISTANT EDITOR
RÉDACTEUR ADJOINT

Ralph Baddour
Department of Medical Biophysics
University of Toronto
rbaddour@uhnres.utoronto.ca

DIRECTORS
DIRECTEURS

Tim Kelsall
(905) 403-3932
tkelsall@hatch.ca

Vijay Parsa
(519) 661-2111 Ex. 88947
parsa@nca.uwo.ca

Sean Pecknold
(902) 426-3100
sean.pecknold@drdc-rddc.gc.ca

Richard Peppin
(410) 290-7726
peppinr@scantekinc.com

Roberto Racca
(250) 483-3300
rob@jasco.com

Frank Russo
(416) 979-5000 ext. 2647
russo@caa-aca.ca

Jérémie Voix
(514) 932-2674
voix@caa-aca.ca

Clair Wakefield
(250) 370-9302
nonoise@shaw.ca

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

ACOUSTIKALAB Inc.
Jean Laporte - (514) 692-1147
jlaporte@acoustikalab.com - Montral, QC

Bruel & Kjaer North America Inc.
Mr. Andrew Khoury - (514) 695-8225
andrew.khoury@bksv.com - Pointe-Claire, QC

ECORE International
Mr. Paul Downey - (416) 440-1094
pcd@ecoreintl.com - Toronto, ON

HGC Engineering Ltd.
Mr. Bill Gastmeier - (905) 826-4044
info@hgcengineering.com - Mississauga, ON

J.E. Coulter Associates Ltd.
Mr. John Coulter - (416) 502-8598
jcoulter@on.aibn.com - Toronto, ON

JASCO Research Ltd.
Mr. Scott Carr - (902) 405-3336
scott@jasco.com - Halifax, NS

Novel Dynamics Test Inc.
Mr. Andy Metelka - (519) 853-4495
ametelka@cogeco.ca - Acton, ON

Peutz & Associés
M. Marc Asselineau +33 1 45230500
m.asselineau@peutz.fr
Paris, FRANCE

Pyrok Inc.
(914) 777-7770; FAX: (914) 777-7103
info@pyrokinc.com - Mamaroneck, NY

Sensor Technology Limited
Niru Somayajula, (705) 444-1440
techsupport@sensortech.ca - Collingwood, ON

Soft dB Inc.
M. André L'Espérance - (418) 686-0993
contact@softdb.com - Sillery, QC

Swallow Acoustic Consultants Ltd.
Mr. John Swallow - (905) 271-7888,
jswallow@jsal.ca - Mississauga, ON

Vibro-Acoustics
Mr. Tim Charlton - (800) 565-8401
tcharlton@vibro-acoustics.com
Scarborough, ON

Wilrep Ltd.
Mr. Don Wilkinson - (905) 625-8944
info@wilrep.com - Mississauga, ON

ACO Pacific Inc.
Mr. Noland Lewis - (650) 595-8588
acopac@acopacific.com - Belmont, CA

AECOM
Frank Babic - (905) 747-7411
frank.babic@aecom.com - Markham, ON

Dalimar Instruments Inc.
Mr. Daniel Larose - (514) 424-0033
daniel@dalimar.ca - Vaudreuil-Dorion, QC

H.L. Blachford Ltd.
Mr. Dalton Prince - (905) 823-3200
amsales@blachford.ca - Mississauga, ON

Hydro-Quebec
M. Blaise Gosselin - (514) 840-3000x5134
gosselin.blaise@hydro.qc.ca - Montréal, QC

J.L.Richards & Assoc. Ltd.
Mr. Terry Vivyurka, P.Eng. - (613) 728-3571
mail@jlrichards.ca - Ottawa, ON

Mc SQUARED System Design Group
Mr. Wade McGregor - (604) 986-8181
info@mcsquared.com - North Vancouver, BC

Owens-Corning Canada Inc.
Mr. Salvatore Ciarlo - (800) 988-5269
salvatore.ciarlo@owenscorning.com -
St.Leonard, QC

Michel Picard
(514) 343-7617; FAX: (514) 343-2115
michel.picard@umontreal.ca - Brossard, QC

RWDI AIR Inc.
(519) 823-1311; FAX: (519) 823-1316
peter.vandelden@rwdi.com - Guelph, ON

SILEX Innovations Inc.
Mr. Mehmood Ahmed - (905) 612-4000
mehmooda@silex.com, Mississauga, ON

Spaarg Engineering Ltd.
Dr. Robert Gaspar - (519) 972-0677
gasparr@kelcom.igs.net - Windsor, ON

Tacet Engineering Ltd.
Dr. M.P. Sacks - (416) 782-0298,
mal.sacks@tacet.ca - Toronto, ON

Wakefield Acoustics Ltd.
Mr. Clair Wakefield - (250) 370-9302
clair@wakefieldacoustics.com - Victoria, BC

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

Eckel Industries of Canada Ltd.
- (613) 543-2967
eckel@eckel.ca - Morrisburg, ON

Hatch Associates Ltd.
Mr. Tim Kelsall - (905) 403-3932
tkelsall@hatch.ca - Mississauga, ON

Integral DX Engineering Ltd.
Mr. Greg Clunis - (613) 761-1565
greg@integraldxengineering.ca - Ottawa, ON

Jade Acoustics Inc.
Ms. Dalila Giusti - (905) 660-2444
dalila@jadeacoustics.com - Concord, ON

MJM Conseillers en Acoustique Inc.
MJM Acoustical Consultants Inc.
M. Michel Morin- (514) 737-9811
mmorin@mjm.qc.ca - Montréal, QC

OZA Inspections Ltd.
Mr. David Williams - (800) 664-8263x25
oza@ozagroup.com - Grimsby, ON

Pinchin Environmental Ltd.
Ms. Robin Connley - (905) 363-0678
info@pinchin.com - Mississauga, ON

Scantek Inc.
Mr. Richard J. Peppin - (410)-290-7726
peppinr@scantekinc.com - Columbia, MD

SNC/Lavalin Environment Inc.
M. Jean-Luc Allard - (450) 651-6710
jeanluc.allard@snclavalin.com - Longueuil, QC

State of the Art Acoustik Inc.
Dr. C. Fortier - (613) 745-2003
sota@sota.ca - Ottawa, ON

Valcoustics Canada Ltd.
Dr. Al Lightstone - (905) 764-5223
solutions@valcoustics.com
Richmond Hill, ON

West Caldwell Calibration Labs
Mr. Stanley Christopher - (905) 595-1107
info@wccl.com - Brampton, ON