Canadian Acoustics / Acoustique Canadienne 21(1) 3-8 (1993)

COMPARISON OF SPECTRAL AND VOICE QUALITY PARAMETERS OF VOICE TRAINED ANGLOPHONES AND FRANCOPHONES IN CONTINUOUS SPEECH

William Weiss
Department of Theatre
University of Ottawa
Ottawa, Ontario, Canada K1N 6N5.

ABSTRACT

Continuous speech of 23 subjects was recorded. The group was composed of Voice Trained (n=11) and Untrained (n=12) Male and Female Anglophone subjects. The objective of the investigation was to find out how are *spectral levels* and *voice quality* compared to an equivalent group of Francophones that were the object of a previous study. Long term spectral analysis was applied to all recorded samples and *Spectral Levels* were determined for regions of $F0$, $F1$, <1000Hz (B_{1K}), >1000Hz (A_{1K}); and the following ratios were computed: 1. $\theta = F1 - F0$, 2. α_{AB}= >1000Hz - <1000Hz. Analyses of variance were carried out in order to ascertain differences between Trained and Untrained subjects and between the linguistic groups. Results show: 1. α_{AB} greater for Trained Anglophones than for Untrained Anglophones. 2. α_{AB} not greater for Trained Francophones than for Untrained Francophones. 3. α_{AB} greater for Trained Anglophones than for Trained Francophones. 4. α_{AB} the same for Untrained Anglophones and Untrained Francophones. 5. Trained Anglophones have higher Spectral Levels than Untrained Anglophones. 6. Trained Francophones have smaller Spectral Levels than Untrained Francophones. 7. Trained Anglophones have significantly higher Spectral Levels than Trained Francophones in the higher frequencies. 8. Untrained Francophones have significantly higher Spectral Levels than Untrained Anglophones in the lower frequencies. 9. There were no significant differences for θ. Key Words: Spectral analysis- L.T.A.S.- Voice Quality - Voice Training - Anglophones - Francophones.

SOMMAIRE

La parole continue de 23 sujets a été enrégistrée. Le groupe était composé de sujets anglophone masculins et féminins avec formation (n=11) et sans formation (n=12). Le but de l'étude était de comparer leurs *hauteurs spectraux* et leurs *timbres vocaliques* avec un groupe semblable de francophones qui avaient fait l'objet d'une recherche antérieure. Le spectre moyenné à long terme a été calculé pour tous les échantillons et des *hauteurs spectraux* ont été détérminés pour les zones $F0$, $F1$, <1000Hz (B_{1k}), >1000Hz (A_{1k}); ainsi que les proportions suivantes: 1. $\theta = F1 - F0$, 2. α_{AB}= >1000Hz - <1000Hz. Des analyses de variance ont été réalisées pour vérifier des différences entre sujets avec et sans formation ainsi qu'entre les deux groupes linguistiques. Résultats: 1. α_{AB} plus grand pour anglophones avec formation que pour anglophones sans formation. 2. α_{AB} n'est pas plus grand pour francophones avec formation que pour francophones sans formation. 3. α_{AB} plus grand pour anglophones avec formation que pour francophones avec formation. 4. α_{AB} semblable pour anglophones et francophones sans formation. 5. Anglophones avec formation ont des niveaux spectraux plus élevés que les anglophones sans formation. 6. Francophones sans formation ont des niveaux spectraux plus élevés que les francophones avec formation. 7. Anglophones avec formation ont des niveaux spectraux plus élevés que francophones avec formation d'une façon significative dans les fréquences aigües. 8. Francophones sans formation ont des niveaux spectraux plus élevés que anglophones sans formation d'une façon significative dans les fréquences basses. 9. Il n'y avait pas des différences significatives pour θ. Mots clés: analyse spectrale - s.m.l.t. - timbre vocalique - formation vocale - anglophones - francophones.

1. INTRODUCTION

Long Term Average Spectra as an acoustic measure of voice quality has been used in order to ascertain progress in voice therapy (Frokjaer-Jensen and Prytz, 1976) and voice training (Wedin *et al.* 1978), to distinguish trained from untrained singers (Sundberg and Gauffin, 1978) and across language groups (Zalewski and Majewski, 1971, Banuls-Terol, 1971, Tarnóczy and Fant, 1964, Boysson-Bardies *et al.* 1984, Harmegnies and Landercy, 1985, Esling, 1983). Given the very wide intra-speakers' variations (Tarnoczy, 1956, Bordone-Sacerdote and Sacerdote, 1969, Doddington, 1985, Harmegnies and Landercy, 1985) spectra at conversational levels have not been used widely in differentiating language-groups.

An acoustic measure of voice quality was proposed by Frokjaer-Jensen and Prytz (1976) as the ratio α = intensity above 1kHz/ intensity below 1kHz. In that study patients with a history of unilateral laryngeal paralysis, after undergoing voice therapy increased the energy in the harmonics above 1000Hz. Wedin *et al.*, (1978) seemed to confirm the usefulness of this measure in normal speech with a group of professional singing teachers who had undergone one week of intense voice training. Their normal speech voice contained a significant increase of energy in the spectra above 1000Hz. The higher intensity in the upper spectra seems to be associated with a "clearer" speaking voice (Kiukaanniemi *et al.*, 1982). Sundberg and Gauffin, 1978 suggested, on the other hand, that in singing, judging the higher spectra as a measure of good quality is misleading because it could be obtained with an increased vocal effort ("pressed" phonation) which is not characteristic of trained male singers. They proposed that a measure of good quality is θ, a higher increase of energy in the $F0$ area relative to the $F1$ area, which is a characteristic of trained subjects ("flow" phonation). In a recent study (Weiss, 1991), which was also the first part of the present investigation, both of these voice quality measures showed no significant differences between voice trained and untrained Francophone subjects. It was concluded that possibly the voice quality measurement α is linguistically related, given that those studies were conducted with English, Danish Swedish and Finnish speakers, and were, perhaps, not appropriate for French, whereas the other voice quality measurement, θ might be adequate for singing but not for speech.

Some authors have expressed the need to give special consideration to normative acoustical data of voice trained speakers (Peppard *et al.*, 1988). If acoustical measurements of desired voice quality parameters as represented by voice trained speakers are different across languages, it would be important to compare the relevant acoustical targets of both trained and untrained speakers of different language groups. These normative parameters could be of interest in the fields of voice training, speaker's recognition and synthetic speech. In this study the above mentioned four frequency intervals LTAS's ($F0$, $F1$, <1000Hz (B_{1K}), >1000Hz (A_{1K})) and the ratios: 1. $\theta = F1 - F0$, 2. α_{AB}= >1000Hz - <1000Hz, were used to compare continuous read speech, between **voice trained** and **untrained** subjects of Canadian English and French, both within and across languages. The data for the French speaking group, which was extracted at the same time as the English speaking group, was used again from the previous paper (Weiss, 1991).

2. SUBJECTS

The experiment for both groups was carried out at the same time and under the same conditions. The group of 23 Anglophone subjects was composed of Voice Trained (n=11) and Untrained subjects (n=12). The group of 23 Francophone subjects was composed of Voice Trained (n=12) and Untrained subjects (n=11). All subjects were given a detailed audiogram and those with abnormal hearing were screened out. Those with mother tongues other than Canadian English (Anglophones) or Canadian French (Francophones) were also excluded. An experienced voice trainer listened to the subjects in order to preclude any voice abnormalities. There were 10 male and 13 female Anglophones, and 12 male and 11 female Francophones. The average age of the Anglophone group was 36.83 years and that of the Francophone group 29.16 years. The average years of training of trained Anglophones group was 6.59 years, and that of trained Francophones 4.96 years. The trained subjects were either members of a well known choir or professional actors and radio announcers. The subjects donated their time without pay. The problem pertaining to the fact that different types of voice training for singers and speakers were involved in this experiment was considered. Given the large array of voice training methods given to singers and speakers, and the fact that speakers get training of the sort given to singers, it is impossible to control the pedagogical input in such a large group. It was deemed that the selection of trained subjects for investigations of the normative type is probably quite adequate when based on their careers or artistic activities.

3. MATERIALS

The English and French texts, of phonetically balanced contents lasting approximately one minute of reading time, was edited from existing literary materials. Given the length of the reading materials and the minor importance of their contents (Benson and Hirsh, 1953) they will not be reported.

4. PROCEDURE

The subjects were recorded while reading the same one minute text. The subjects were given a few minutes to become acquainted with the text, but beyond that they were not given instruction or practice. All the recordings, and the audiometric screening, were conducted in a soundproof cabin (I.A.C.). During the recording the distance from the microphone was one foot. A 1000Hz tone was recorded on the tapes for calibration purposes. The microphone was a Sennheiser MD441-U (filtration switch on 'M'), the tape recorder a full track Revox 77A (tape speed 15 ips), and the tapes Ampex 406.

5. ANALYSES

The recorded samples were analyzed with an Ono Sokki CF300 spectral analyzer for Long Term Average Spectra at 1/3 octave intervals, 16-kHz range, for 128 spectra used to compute the long term average, during one minute. The data were transferred and digitized in an IBM micro-computer through a software package designed for the project and then transferred to the mainframe (Amdhal) computer where Spectral levels were determined.

Long Term Average Spectra were computed for the following intervals:

***F*0e**: Log energy at interval 80-160Hz for men, 160-250 Hz for women
***F*1e**: Log energy at interval 315-600Hz
B_{1K}: Log energy below 1kHz (80-800Hz)
A_{1K}: Log energy above 1kHz (1000-5000Hz)
θ_{F1F0}: ***F*1e** minus ***F*0e**
$\alpha_{AB} = A_{1K}$ minus B_{1K}

These intervals were chosen in order to calculate the ratios θ_{F1F0} and α_{AB} which were the voice quality acoustical parameters reported in previous investigations and tested in the present study. Given that the recording and the analyses conditions were identical for all subjects, and that the interest lied in the relative energy levels, the computed intervals served to compare spectral levels and it was not necessary to SPL normalize them. The log energy measures also allowed for a simple subtraction for the calculation of the ratios. SAS (SAS Institute Inc.) analyses of variance (t-tests) were computed in order to ascertain differences between the Voice Trained and the Untrained, as well as between Anglophones and Francophones.

6. RESULTS AND DISCUSSION

Unlike Francophones (Table 2, Fig. 2) the Anglophone data (Table 1, Fig. 1) confirms previous findings (Wedin *et al.* 1978) that voice training provides significantly more spectral energy in the higher frequencies relative to the lower frequencies (α_{AB}: $p<.004$; the smaller log difference indicates more relative energy in the higher spectra). However, the spectral levels of the trained Anglophone subjects are higher in all intervals (i.e. have smaller negative numbers) and significantly at A_{1k} ($p<.02$). That is, Trained Anglophones speak louder that their Untrained counterparts, and those higher vocal levels produce more energy in the higher frequencies.

Unlike Anglophones, Untrained Francophones have higher spectral levels than the Voice Trained yet α is not significantly different.

θ was not significantly different for either language group.Trained Anglophones (Table 3, Fig.3) have significantly higher spectral levels in the higher frequencies (A_{1K} ($p<.05$)) than Trained Francophones and there are significant differences for α_{AB} ($p<.002$). This confirms that voice training increases the energy in the higher frequencies for English speaking subjects as it did for Swedish speaking subjects (Wedin et al. 1978). Untrained Francophones (Table 4, Fig.4), on the other hand, have higher vocal levels than Untrained Anglophones mainly in the lower frequencies (F0e: $p<.03$; B_{1K}: $p<.03$) but there are no significant differences for α_{AB}.

Voice training increases or decreases vocal levels differentially, depending on the language group. These spectral differences are probably linked to language dependent articulatory settings (Laver, 1980). However there were no significant spectral level differences in the higher frequencies between the untrained Anglophones and Francophones. Trained Anglophones, on the other hand, have higher energy in the upper spectra than untrained Anglophones and trained Francophones. Clearly, voice training produced a language specific acoustical target that was not operant in French.

There are no clear perceptual evaluations of acoustical determinants of voice quality. One can only infer tentatively some possibilities from these results. The tables show that the spectral levels of voice trained Anglophones were higher than those of their untrained counterparts. It is well known that higher vocal levels produce a shift of energy towards the higher spectra (Van Summers *et al.* 1988; Stanton, 1988) thereby producing more 'sonority' (Wedin *et al.*,1978) and perhaps more 'tension' (Laver, 1980). Trained Francophones have on average lower spectral levels than their untrained counterparts, indicating perhaps that for Francophones the vocal target is one of relaxation. This is found in Table 3 where vocal levels for trained Anglophones and Francophones is similar except in the higher frequencies, whereas in Table 4 untrained Francophones show significantly higher levels in the frequencies below 1000Hz (B1K) where most of the vocal energy resides. The settings for English and French are different in lip, jaw and tongue tip positioning creating a 'laxer' English and a 'tenser' French (Laver, 1980), the adjustments resulting from voice training seem to equalize vocal levels but accentuate the difference in the higher frequencies that result from the settings.

A comparison by gender did confirm the result of the combined groups and was not reported in order not to unduly burden this paper.

The fact that α_{AB} distinguished the trained Anglophones from all the other groups, but that it did not distinguish the untrained Anglophones from the untrained Francophones made it clearly a linguistic *voice training* parameter, an acoustic measure that characterises good English voices. It is noteworthy that a similar measure for French was not found, and that the English target should not be applied to French speech.

7. SUMMARY

1. α_{AB} greater for Trained Anglophones than for Untrained Anglophones. 2. α_{AB} not greater for Trained Francophones than for Untrained Francophones. 3. α_{AB} greater for Trained Anglophones than for Trained Francophones. 4. α_{AB} the same for Untrained Anglophones and Untrained Francophones. 5. Trained Anglophones have higher Spectral Levels than Untrained Anglophones. 6. Trained Francophones have smaller Spectral Levels than Untrained Francophones. 7. Trained Anglophones have significantly higher Spectral Levels than Trained Francophones in the higher frequencies. 8. Untrained Francophones have significantly higher Spectral Levels than Untrained Anglophones in the lower frequencies. 9. There were no significant differences for θ.

8. REFERENCES

BANULS-TEROL, V. (1971). "Weighted average spectrum of human speech: an approach," Proc.7th I.C.A., 253-256.

BENSON, R., HIRSH, I. (1953). "Some variables in audiospectrometry," J. Acoust. Soc. Am. 25, 499-505.

BORDONE-SACERDOTE, C., SACERDOTE, G. (1969). "Some spectral properties of individual voices," Acustica 21, 199-210.

BOYSSON-BARDIES, B. DE, HALLE, P., SAGART, L. (1984). "Acoustic investigations of cross-linguistic variability in babbling," Interdisciplinary Sympos. on Precursors of Early Speech.

DODDINGTON, GEORGE R.(1985). "Speaker Recognition - Identifying People by their Voices," Proc. IEEE, 73, 11, 1651-1664.

ESLING, J. H. (1983). "Quantitative analysis of acoustic correlates of supralaryngeal voice quality features in the long-time spectrum," Abstr. 10th Inter. Congr. Phonet. Sc., Dordrecht, Cinnaminson, Foris Publ., 363.

FROKJAER-JENSEN B, PRYTZ S. (1976). " Registration of voice quality," Bruel & Kjaer,Technical Revue, 3, 3-17.

HARMEGNIES,B., LANDERCY, A. (1985). "Language features in the long-term average spectrum," Revue de Phonétique Appliquée, 73-74-75.

KIUKAANNIEMI, H., SIPONEN, P., MATTILA, P. (1982). "Individual Differences In The Long-term Speech Spectrum," Folia Phoniat. 34, 21-28.

LAVER, J. (1980). The Phonetic Description of Voice Quality. Cambridge, Cambridge University Press.

PEPPARD,R.C.,BLESS, D.M., MILENKOVIC, P. (1988). "Comparison of Young Adult Singers and Nonsingers with Vocal Nodules," J. of Voice, 2, 3, 250-260.

STANTON, B. J.(1988). "Acoustic phonetic analysis of normal, loud, and Lombard speech in simulated cockpit conditions,"J. Acoust. Soc. Am.84, Suppl.1, 115.

SUNDBERG, J., GAUFFIN J. (1978). "Waveforms and spectrum of the glottal voice source," Speech Transmis. Lab. Q. Prog. Stat. Rep., STL-QPSR 2-3, Royal Institute of Technology, Stockholm, Sweden, 35-50.

TARNOCZY, T. (1956). "Determination of the speech spectrum through measurements of superimposed samples," J. Acoust. Soc. Am.28, 1270-1275.

TARNOCZY,T., FANT, G. (1964). "Some remarks of the average speech spectrum," Speech Transmis. Lab. Q. Prog. Stat. Rep. STL-QPSR,4, Royal Institute of Technology, Stockholm, Sweden, 13-14.

VAN SUMMERS, W.,PISONI, D.B., BERNACKI, R. H., PEDLOW, R.I., STOKES, M.A. (1988). "Effects of noise on speech production: Acoustic and perceptual analyses," J. Acoust. Soc. Am. 84, 3, 917-928.

WEDIN, S., LEANDERSON, R., WEDIN, L. (1978). "Evaluation of Voice Training. Spectral Analysis Compared with Listeners' Judgements," Folia Phoniat., 30, 103-112.

WEISS, W. (1991). "Some Acoustic-phonetic Parameters Of The Lombard Effect For The Voice Trained," Proceedings of the 12th International Congress of Phonetic Sciences, Aix-en-Provence.

ZALEWSKI, J., MAJEWSKI, W. (1971). "Polish speech spectrum obtained from superposed samples and its comparison with spectra of other languages," Proc. 7th I.C.A., 249-252.

Table 1: Mean energy levels (dB) of voice TRAINED **Anglophones** (N=11) and UNTRAINED **Anglophones** (N=12) measured over selected (1/3 octave) intervals.

Interval		F0e	F1e	θF1F0	B1K	A1K	αAB
						*	**
dB	Tr. Ang.	-21.73	-21.99	-0.13	-17.48	-25.93	-8.45
	Untr. Ang.	-23.50	-23.34	0.15	-19.42	-30.94	-11.51

Table 2: Mean energy levels (dB) of voice TRAINED **Francophones** (N=12) and UNTRAINED **Francophones** (N=11) measured over selected (1/3 octave) intervals.

Interval		F0e	F1e	θF1F0	B1K	A1K	αAB
dB	Tr. Fra.	-21.73	-22.14	-0.41	-17.93	-29.61	-11.67
	Untr. Fra.	-20.96	-20.48	0.47	-16.53	-28.04	-11.51

Table 3: Mean energy levels (dB) of **Trained** ANGLOPHONES (N=11) and **Trained** FRANCOPHONES (N=12) measured over selected (1/3 octave) intervals.

Interval		F0e	F1e	θF1F0	B1K	A1K	αAB
						*	**
dB	Tr. Ang.	-21.73	-20.99	0.74	-17.48	-25.93	-8.45
	Tr. Fra.	-21.73	-22.14	-0.41	-17.93	-29.61	-11.67

Table 4: Mean energy levels (dB) of UNTRAINED ANGLOPHONES (N=12) UNTRAINED FRANCOPHONES (N=11) measured over selected (1/3 octave) intervals.

Interval		F0e	F1e	θF1F0	B1K	A1K	αAB
		*			*		
dB	Untr. Ang.	-23.50	-23.34	0.15	-19.42	-30.94	-11.51
	Untr. Fra.	-20.96	-20.48	0.47	-16.26	-28.04	-11.51

* significant at the 0.05 level
** significant at the 0.01 level
*F*0e:Energy at interval 80-160Hz for men, 160-250Hz for women
*F*1e:Energy at interval 315-600Hz
B_{1K}:Energy below 800Hz (80-800Hz in 1/3 octaves)
A_{1K}:Energy above 1000Hz (1000-5000Hz in 1/3 octaves)
θ_{F1F0}: F1e minus F0e
$\alpha_{AB} = A_{1K}$ minus B_{1K}

Normal Speech for Anglophone

Voice Trained VS Untrained

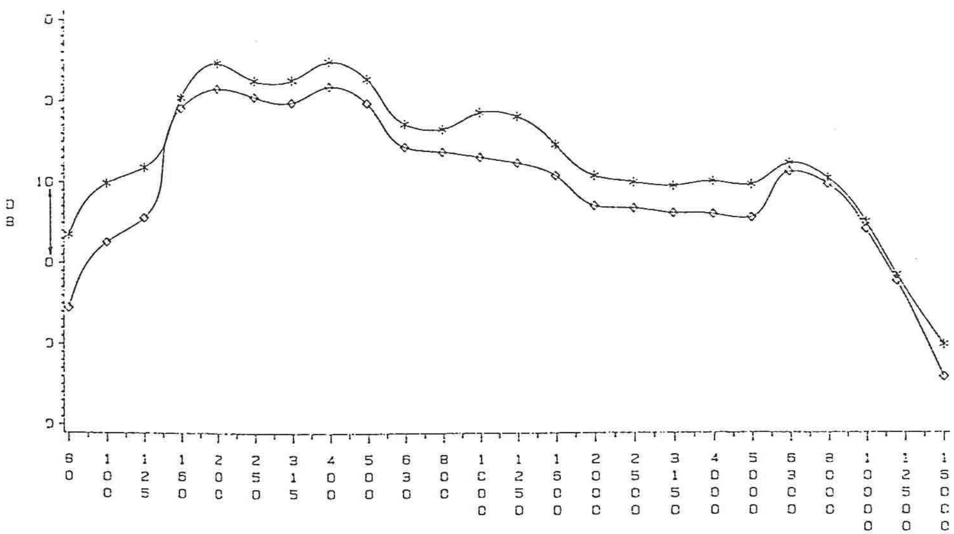

Fig. 1

Normal Speech for Francophone

Voice Trained VS Untrained

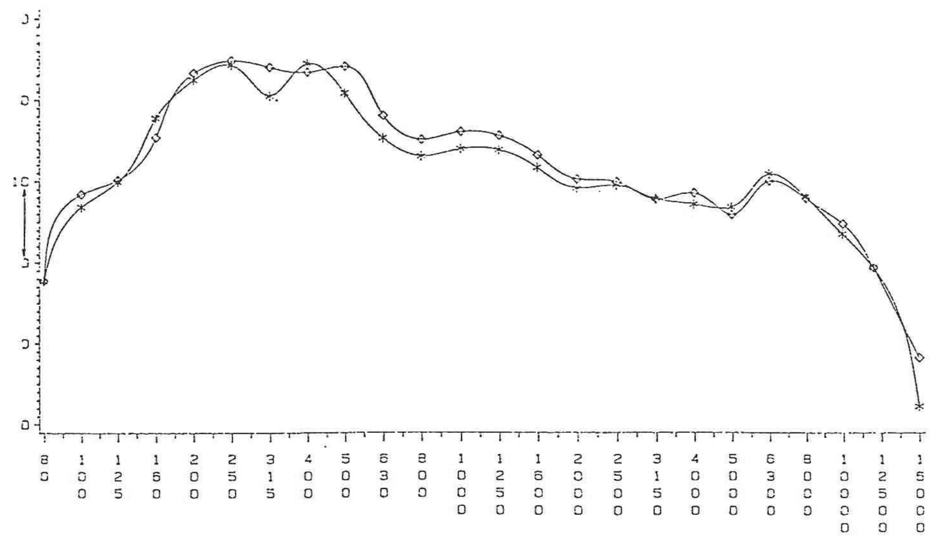

Fig. 2

Normal Speech for Trained Subjects

Anglophones vs Francophones

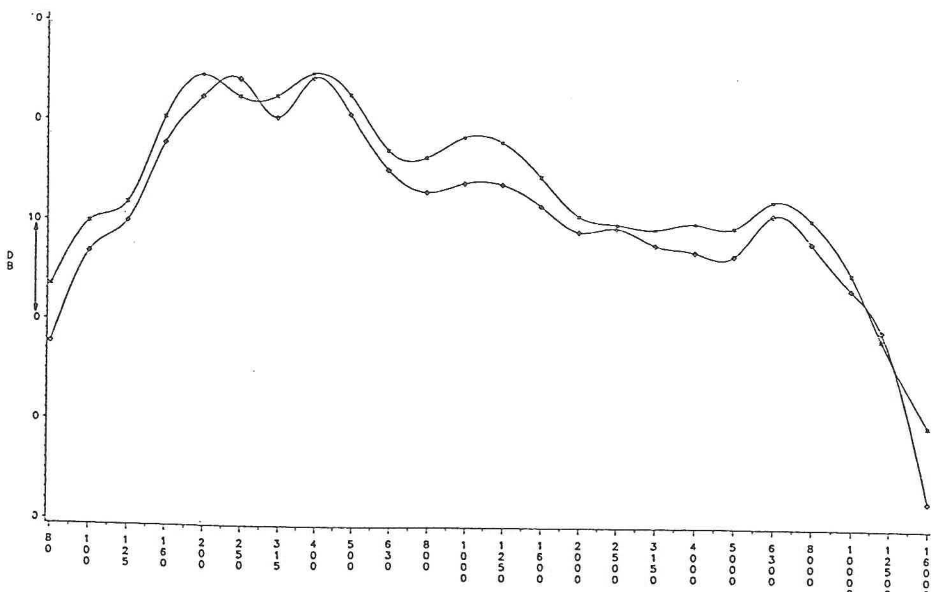

Fig. 3

Normal Speech for Untrained Subjects

Anglophones vs Francophones

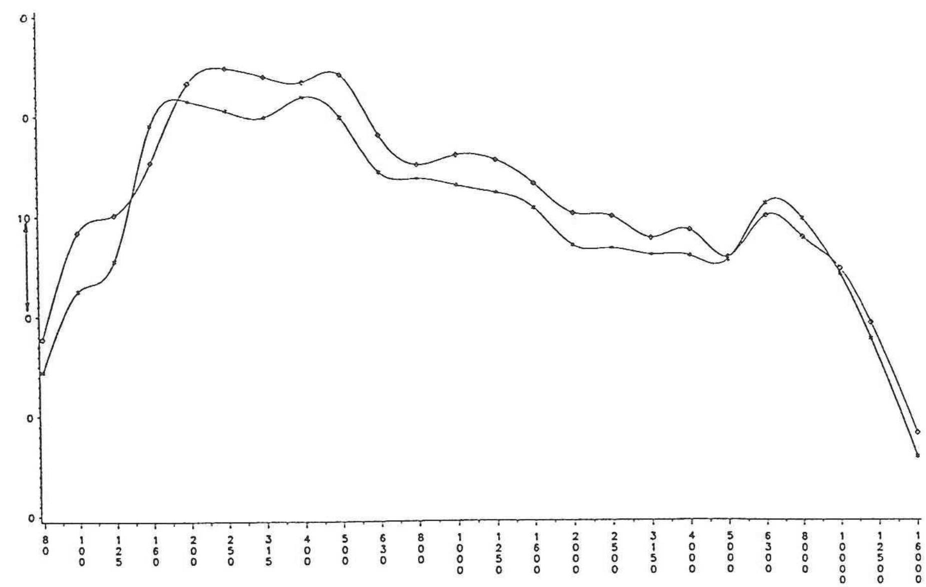

Fig. 4