AN EXPERIMENT IN INTER-LANGUAGES SPEAKER RECOGNITION USING THE SDDD INDEX

Bernard HARMEGNIES Albert LANDERCY Marielle BRUYNINCKX

Université de l'Etat à Mons - Département de Phonétique et Psycho-acoustique - Avenue du Champ de Mars, B-7000 MONS - BELGIQUE

ABSTRACT

The voices of 10 Belgian bilingual (Dutch-French) subjects were analysed by means of a high resolution frequency analyser (400 channels FFT). Long Term Average Spectra (LTAS) of the subjects'voices were computed both on the basis of French and of Dutch utterances (balanced texts). The SDDD index was used in crder to compare these LTAS. Its discriminating ability in an inter-language speaker recognition task was evaluated by means of the Receiver Operating Characteristics (ROC) curves for all the comparison conditions under investigation and revealed to be greater than the one of the cross-correlation coefficient.

INTRODUCTION

Although Long Term Average Spectra (LTAS) have been used in various contexts and are usually considered as good acoustical cues to voice quality, several of their properties are not yet well known. Among others, the question of the LTAS resistance to changes in the languages used by speakers is still controversial.

On the one hand, several experiments suggest that languages exert strong effects on LTAS. KIUKAANNIEMI and MATTILA report differences between Finnish and English data [1]. HALLE, de BOYSSON - BARDIES and SAGARD suggest that even LTAS from 8 and 10 month old babies can be influenced by the language of the social group they belong to [2]. MAJEWSKI and HOLLIEN [3] and ZALEWSKI, MAJEWSKI and HOLLIEN [4] obtain recognition rates different for Americans and Poles; this seems to suggest language-related LTAS differences. On the other hand, some authors consider that LTAS are language-independant to some extent. BYRNE [5] notices that LTAS he drew from English texts uttered by Australian look much like those from ANIANSSON [6] drawn from Swedish speech. On the basis of an experiment involving Piamontes, Italian and French, TOSI [7] concludes that each speaker possesses "relative" LTAS invariance, irrespective of the language spoken. HARMEGNIES and LANDERCY [8] report few differences between LTAS drawn from Dutch and French utterances produced by bilingual subjects. As NOLAN remarks [9], there is a conflict between these research trends and it is unclear whether LTAS can be considered as language-independent cues to voices quality.

In this paper, which constitutes a contribution to this problem, we will study to what extent inter-language speaker recognition based on LTAS is possible. Because its discriminatory ability is supposed to overcome those of classical indices, a new dissimilarity index, SDDD [10, 11] will be used for the purpose of comparing spectra; its power will be assessed by comparison with the correlation coefficient.

EXPERIMENT

Experimental setting

The speakers were 10 bilingual Belgian subjects, between 18 and 21 years old. Each of them uttered two texts ten times in succession : a phonetically balanced French text and a phonetically balanced Dutch text. Both texts were about 18 seconds long. The recording sessions took place in a sound-proof room. The subjects were sitting in front of the microphone, placed at a constant 40 cm distance from their lips. All texts were recorded on a NAGRA IV S recorder, by means of a KM 84 NEUMANN microphone.

Acoustical analysis

The acoustical analyses were performed later by means of a 400-channels 2033 Brüel Kjaer FFT analyser (BK 2033). Its sampling frequency was set to 12.8 kHz, in order to obtain a 0-5 kHz frequency span. With this setting, the spectra presented a 12.5 Hz resolution over the whole frequency band under investigation. The BK 2033 built-in linear averaging process was used in order to compute LTAS. The 200 (10 subjects x 2 languages x 10 utterances) so-obtained LTAS were then transmitted from the analyser to a 4341

IBM computer via a personal computer, for storage and further computations.

Comparison procedure

Inter- and intra-language comparisons were performed. For intra-language comparisons, the same procedure was used both for the Dutch and the French LTAS : 1. (intra-speaker comparisons) for each of the 10 speakers, one comparison was performed for each possible non-redundant pair of his 10 LTAS (i.e. 45 comparisons); 2. (inter-speaker comparisons) for each possible non-redundant pair of different speakers (i.e. 45 pairs), all possible comparisons of their respective 10 LTAS were performed (i.e. 100 comparisons for one pair). For each language, 450 intra-speaker and 4500 inter-speaker comparisons were therefore performed.

For inter-language comparisons, all the French LTAS were compared with all the Dutch LTAS; 1000 intra-speaker and 9000 inter-speaker comparisons were therefore performed.

For each comparison, both a similarity (R) and a dissimilarity (SDDD) index were computed.

Indices for the comparison of LTAS

In order to define the indices, it is convenient to consider each LTAS as a K-dimensional vector, with k being the total number of frequency channels taken into account in the spectrum. Therefore, spectrum S may be defined as :

$$S = (S_1, \ldots, S_i, \ldots, S_k) \qquad (1)$$

with S_i, the level of the i^{th} frequency component. In this paper as well as in most previous ones [1,4,8,10,12], the S_i values will be expressed in decibels.

The Bravais-Pearson cross-correlation coefficient (R) can be used as a similarity index for the comparison of LTAS. It expresses the tendency of the s_i values to covary with the S_i values and it ranges, in absolute values from 0 (complete independance of the S_i and S'_i variabilities) to 1 (perfect correlation of the S_i and S_i values). R can be defined as :

$$R_{SS'} = \frac{1}{K} \frac{\sum_{i=1}^{K} (S_i - M_S)(S'_i - M'_S)}{\sigma_S \sigma'_S} \qquad (2)$$

Where M_S and M'_S are the means for all S_i and S'_i values, respectively, and σ_S and σ'_S are the corresponding standard deviations. If the spectra beeing compared are identical, the correlation between the S_i and S_i values is 1. On the contrary, a weak correlation indicates a lack of similarity of the spectral shapes. R is usually considered as one of the best indices because : 1. it exhibits a discriminating ability in the same range than the one of other classical indices (e.g. the euclidean distance) [4]; 2. unlike other classical indices, R is insensitive to changes in the overall levels of the spectra and, therefore, does not require any intensity normalization.

The Standard Deviation of the interspectral Differences Distribution (SDDD) has been recently introduced [10, 11]. SDDD measures the variability of the $S_i - S'_i$ differences. It is defined as :

$$SDDD_{SS} = \sqrt{\frac{1}{K} \sum_{i=1}^{K} \{ S_i - S'_i - MD \}^2} \qquad (3)$$

where MD is the average of the $S_i - S'_i$ differences. If the shapes of the spectra compared are highly similar, the differences values are almost invariant and tend to concentrate around a given central tendency influenced only by the between-spectra overall level difference. If the shapes are different, one can find large level differences in certain frequency channels and small ones in others; the standard deviation of the differences increases. SDDD can therefore be used as a dissimilarity index for LTAS. Like R, SDDD is insensitive to changes in the levels of the spectra; moreover, in recent intra-language speaker recognition experiments [10, 11], SDDD has revealed to be more discriminative than R.

RESULTS

The distributions characteristics of the SDDD values drawn from all kinds of comparisons are presented in table 1. This table shows that, in the case of intra-language comparisons, the intra-speaker distributions are

	Mean	Standard deviation	Extreme Values
French/French			
Intra spk.	2.912	.421	1.9-4.7
Inter spk.	4.966	.805	3.0-8.3
Dutch/Dutch			
Intra spk.	3.162	.510	2.0-4.9
Inter spk.	5.138	.789	3.0-7.8
French/Dutch			
Intra spk.	3.941	.514	2.6-5.6
Inter spk.	5.221	.866	2.9-8.7

Table 1 : Characteristics of the inter- and intra-speaker distributions of the SDDD values drawn from intra-language and inter-language comparisons.

well separated from the inter-speaker distributions (French : mean SDDD intra = 2.912 against mean SDDD inter = 4.66; Dutch : mean SDDD intra = 3.162 against mean SDDD inter = 5.138). Nevertheless the separation of inter- and intra-speaker distributions is less important in the case of inter-language comparisons (mean SDDD intra = 3.941 against mean SDDD inter = 5.221). Similar observations can be drawn from table 2, about the distributions of the correlation index.

In order to study more accurately the relationships between these distributions, we decided to plot the corresponding Relative Operating Characteristic (ROC) curves. For each

	Mean	Standard deviation	Extreme Values
French/French			
Intra spk.	.932	.002	.84-.97
Inter spk.	.805	.007	.52-.94
Dutch/Dutch			
Intra spk.	.928	.002	.81-.97
Inter spk.	.817	.006	.57-.94
French/Dutch			
Intra spk.	.886	.003	.78-.96
Inter spk.	.798	.007	.43-.94

Table 2 : Characteristics of the inter- and intra-speaker distributions of the R values drawn from intra-language and inter-language comparisons.

comparison condition (French/French, Dutch/Dutch and French/Dutch), a series of values across the entire range of variation of each index were successively considered as rejection thresholds for a recognition task. The corresponding false alarm- and correct recognition rates were drawn from the observed distributions and considered as couples of coordinates in the ROC space. Six (2 indices x 3 comparison conditions) ROC curves were plotted this way (see fig. 1).

It is well known that the area enclosed in the entire ROC space beneath a ROC curve is a distribution-free measure of sensitivity [13]. It is therefore very easy, even from simple direct examination of figure 1, to perform a ranking of the six curves on the basis of the corresponding discriminative powers. In order of decreasing discriminating ability, this ranking is : 1. SDDD, intra-language comparisons (French/French); 2. SDDD, intra-language comparisons (Dutch/Dutch); 3. R, intra-language comparisons (French/French); 4. R, intra-language comparisons (Dutch/Dutch); 5. SDDD, inter-language comparisons; 6. R, inter-language comparisons.

In order to obtain more informative figures than this ranking, we also evaluated the surfaces beneath the curves. For this purpose, each curve was fitted by a polynomial function thanks to polynomial regression techniques. Polynomes of the sixth and seventh order were used and a good fitting was achieved in every case (residual sum of squares was in the .01 - .001 range). The polynomial functions were thereafter integrated. Table 3 gives the values of the surfaces thus obtained (taking into consideration that a unit-surface would mean perfect discrimination and a .5 surface would mean random recognition).

DISCUSSION

Both indices under investigation lead to the same main conclusion, i.e., LTAS-based speaker recognition is language dependant : the ROC curves clearly show that the inter-language comparisons are less speaker-discrimi-

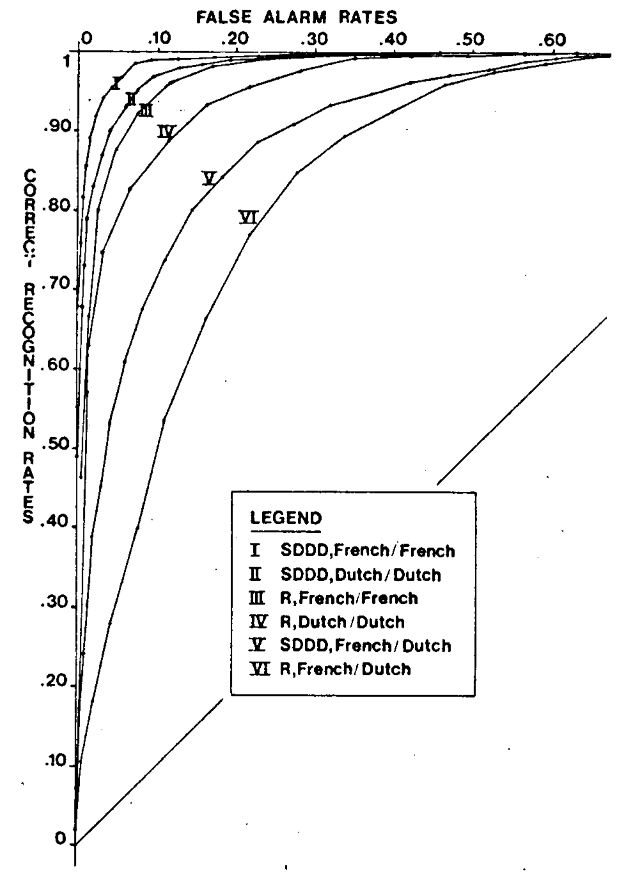

Figure 1 : Receiver Operating Characteristics curves for both indices in each comparison condition.

nant than the intra-language comparisons. This finding seems to plead for the idea that languages exert some effects on LTAS. Moreover, both indices reveal better performances with LTAS drawn from French utterances than with those drawn from Dutch texts. This is in agreement with Majewski and Hollien's suggestion that the power of long term spectrum as an identification tool might be somewhat language dependant [3]. It should be noticed, however, that although LTAS turn out to be language-dependant, it still convey enough cues to the speaker's personnality to make inter language speaker recognition possible : our ROC curves demonstrate that the power of the inter-language recognition process is still far better than chance. In this sense, we can agree with Tosi's conclusion [7] that LTAS possess "relative" invariances, irrespective of the language they come from.

In other words, our general conclusion as to the LTAS resistance to changes in languages could be : long term spectra are influenced by the language spoken (at least when bilingual Dutch/French subjects read 18 seconds long texts), but the speaker influence is greater; LTAS-based inter-language recognition is therefore less safe than intra-language recognition, but quite possible.

This conclusion leads to the question of the relative power of inter-language recognition. Firstly, it is quite obvious, from figure 1 and table 3, that SDDD is more powerfull

	SDDD	R
French/French	.994	.983
Dutch/Dutch	.987	.963
French/Dutch	.910	.856

Table 3 : Areas of the entire ROC space beneath each ROC curve.

in all comparison situations. In a case where one suspects that the comparison situation could lower the discriminative ability of the comparison procedure (e.g. in the case of inter-language recognition), SDDD should therefore be prefered. Furthermore, if the ROC surfaces listed in table 3 are measures of the corresponding discriminative abilities, their ratios can inform about the relative powers of the indices in each situation. One can compute, this way, that the power of SDDD in inter-language recognition is about 92 % of its own power for intra-language recognition, although the power of R in inter-language recognition is only about 88 % of its own power for intra-language recognition. Thus, not only SDDD is more speaker-discriminant than R, it is moreover less sensitive to changes in the comparison conditions (at least language variations).

As a final remark, we must emphasize that our data were collected on a restricted number of subjects only : any overgeneralization would therefore be hazardous. We nevertheless think that they convey structures strong enough to consider our findings at least as firm working hypotheses for our future research.

REFERENCES

[1] KIUKAANNIEMI, H.J. and P. MATTILA, "Long-term speech spectra : a computerized method of measurement and a comparative study of Finnish and English data", Scandinavian Audiology, 9, 1980, 67-72.

[2] HALLE, P., de BOYSSON-BARDIES and L. SAGARD, "Utilisation des spectres à long terme pour dégager des propriétés acoustiques des langues : étude comparative et développementale", Actes des 13èmes J.E.P., GALF, Bruxelles, 1984, 147.

[3] W. MAJEWSKI, and H. HOLLIEN, "Euclidean distances between long-term speech spectra as a criterion for speaker recognition" Speech perception and automatic recognition, vol. 3 (Communication seminar), Stockholm, Almqvist and Wiksell, 1974, 303-310.

[4] J. ZALEWSKI, W. MAJEWSKI, and H. HOLLIEN, "Cross correlation of long-term speech spectra as a speaker identification technique", Acustica, 34, 1975, 20-24.

[5] BYRNE, D., "The speech spectrum - some aspects of its significance for hearing aid selection and evaluation", British Journal of Audiology, 11, 1977, 40-46.

[6] ANIANNSON, G., "Speech discrimination predicted from tone audiometry and articulation index", Acta Otolaryngogica, Suppl. 320, 1974, 36-43.

[7] TOSI, O., Voice Identification, University Park Press, Baltimore, 1979.

[8] HARMEGNIES, B., and A. LANDERCY, "Language features in the long-term average spectrum", Revue de Phonétique Appliquée, 73-74-75, 1985, 51-68.

[9] NOLAN, F. The phonetic bases of speaker recognition, Cambridge University Press, Cambridge, 1983.

[10] HARMEGNIES, B. and A. LANDERCY, "Comparison of spectral similarity indices for speaker recognition", Proceedings of the 12th International Congress on Acoustics, vol. I, A1-4, Toronto, 1986.

[11] HARMEGNIES, B., "SDDD : a new dissimilarity index for the comparison of speech spectra" (submitted).

[12] HOLLIEN H. and W. MAJEWSKI, Speaker identification by long-term spectra under normal and distorted speech conditions, J. Acoust. Soc. Am., 62, 4, (1977), 975-80.

[13] SWETS, J.A., "The relative operating characteristic in psychology", Science, 182, 1973, 990-1000.