

Evaluation comparative du pouvoir discriminant d'indices de (dis)similarité spectrale

P. PAIROUX et B. HARMEGNIES

Département de la Communication Parlée, Université de Mons-Hainaut, 18 Place du Parc, 7000 Mons, Belgium

Abstract: In this paper, the discriminatory abilities of 16 similarity and dissimilarity indices are tested by means of a speaker-recognition experiment involving high-dimensional long-term speech spectra. The results give the ranking of the 16 indices on the basis of their discriminatory abilities.

1. INTRODUCTION

Harmegnies [1], dans une large synthèse de la littérature consacrée au SMLT (Spectre Moyen à Long Terme), a mis l'accent sur les carences en matière d'approches métrologiques abordant la quantification des variations des spectres. Il a ainsi développé un indice original (SDDD) qu'il a confronté aux deux indices les plus fréquemment utilisés. L'objectif de notre étude consiste à élargir cette comparaison à un nombre plus important d'indices de (dis)similarité de façon à obtenir leur rangement en fonction de leur pouvoir discriminant. Dans cette optique, nous allons recourir à la technique des courbes ROC (Receiver Operating Characteristics) [2] inspirée de la théorie de la détection du signal [3].

1.0. Les indices dans la littérature

Pour statuer objectivement sur la ressemblance caractérisant 2 spectres, il est nécessaire de réaliser une mesure de leur degré de (dis)similarité. La littérature [1-4-5-6-7-8] révèle l'existence d'une multitude de moyens visant ce but. Aucun d'entre eux ne semble cependant pouvoir être considéré à titre de panacée. Il paraît pourtant logique que le choix de l'un plutôt que de l'autre influe de manière plus ou moins sensible sur les résultats obtenus.

De ce foisonnement de procédures et d'une analyse de la littérature générale sur la reconnaissance de formes, nous avons extrait 16 indices appartenant aux deux grandes familles que constituent d'une part, les indices de similarité et d'autre part, les indices de dissimilarité. ils sont présentés dans l'annexe 1.

1.1. L'évaluation du pouvoir discriminant des indices

Importées de la théorie de la détection du signal, les courbes ROC ont

pour but de quantifier les relations existant entre deux distributions caractérisant une procédure de détection et de fournir ainsi une mesure du pouvoir discriminant. Les distributions sont celles des valeurs de l'indice respectivement sous conditions de détection erronée (bruit) et de détection correcte (signal).

Le plan ROC est caractérisé par le carré-unité dont l'axe des ordonnées représente la proportion de reconnaissance correcte tandis que l'axe des abscisses correspond à la probabilité de fausse alerte. Chaque valeur de la variable de mesure est, tour à tour, considérée comme un seuil et située dans le plan ROC, les valeurs des distributions correspondant au seuil étant utilisées à titre de coordonnées dans le plan ROC.

L'ensemble des points ainsi obtenus forme la courbe. L'aire incluse sous celle-ci est une mesure non paramétrique du pouvoir discriminant [2]. Plus cette valeur est grande (à la limite, unitaire) et plus la technique employée est performante pour séparer le signal du bruit.

2. LE PLAN EXPERIMENTAL

2.0. L'objectif

Chacun des indices sera tour à tour utilisé dans le cadre d'expériences simulées de reconnaissance de locuteur. Une seule base de données sera utilisée à cet effet, et les différences de succès de la tâche de reconnaissance seront dès lors utilisées pour statuer sur les différences de pouvoir discriminant des indices.

2.1. Les spectres

Les analyses portent sur 400 SMLT issus des productions de 40 locuteurs francophones belges (20 de chaque sexe). Chacun d'entre eux a émis 10 réalisations d'un texte phonétiquement équilibré d'une durée approximative de trente secondes. L'analyse acoustique, réalisée au moyen d'un analyseur Brüel Kjaer 2033, a fourni pour chaque production, un SMLT défini sur 400 canaux dans la bande DC-5 kHz, avec une résolution constante de 12.5 Hz.

2.2. La normalisation des spectres

Certains indices étant sensibles aux différences de niveaux globaux, les spectres ont fait l'objet d'une normalisation préalable à leur utilisation.

Parmi les techniques décrites dans la littérature pour trouver une norme commune de niveau ou d'intensité, on a opté pour celle qui consiste à ramener le niveau global à une constante arbitraire.

2.3. La procédure de comparaison

La tâche de reconnaissance simulée a été réitérée 48 fois : pour chacun des 16 indices, elle a été appliquée, à l'échantillon des hommes, à celui des femmes, et enfin au groupe total. Toutes les comparaisons intra locuteur et inter locuteurs non redondantes possibles ont été calculées, soit :

- pour chaque groupe sexuel, 900 valeurs intra locuteur et 19000 valeurs inter locuteurs.
- pour le groupe total, 1800 valeurs intra sujet et 78000 valeurs inter sujets.

2.4. Les résultats

Pour chaque indice étudié, nous avons, au sein de chacun des six groupes, regroupé toutes ses valeurs sous forme d'une distribution intra et inter

locuteurs, laquelle nous a permis d'établir sa courbe ROC. Toutes les valeurs situées dans l'aire de recouvrement des distributions inter et intra locuteurs ont été localisées dans l'espace ROC et l'aire sous la courbe a ensuite été calculée au moyen de la méthode d'intégration des trapèzes. Le tableau 1 résume toutes ces données et indique le pouvoir discriminant de tous les indices appliqués dans chacun des 3 groupes.

3. CONCLUSIONS

Sur base du tableau 1, divers points peuvent être mis en évidence. Premièrement, la majorité des indices utilisés confirment par leur valeur élevée que le SMLT est un excellent indicateur acoustique d'identité vocale.

Deuxièmement, en accord avec la littérature, dans la majorité des cas, le SMLT discrimine mieux les voix masculines que féminines.

Troisièmement, les indices sensibles aux différences de niveaux apparaissent plus discriminants que les autres. Cette constatation est sans doute liée au fait que la normalisation consiste, par essence, en un ajout d'information dans la procédure. Comme, d'une part, l'information sur le niveau intrinsèque n'est pas toujours accessible et que d'autre part, le choix d'une procédure spécifique de normalisation est, in fine, arbitraire, il importe de pouvoir disposer d'indices non affectables par les différences de niveaux.

Parmi ces derniers, l'indice SDDD apparaît le meilleur à la fois dans le groupe féminin et dans le groupe total. Dans le groupe masculin, il est rangé en deuxième position derrière le coefficient de corrélation de Bravais-Pearson.

Si on se place dans la perspective de tâches de reconnaissance in situ où la mesure du niveau et la connaissance du sexe du locuteur ne sont pas accessibles (contrôle d'accès, téléphone, etc.), l'indice SDDD apparaît dès lors, comme le meilleur choix.

On notera que sa légère infériorité de pouvoir discriminant par rapport au coefficient de corrélation de Bravais-Pearson dans le groupe masculin, est en contradiction avec le raisonnement théorique [1] sur les propriétés mathématiques des 2 indices. Cette constatation pourrait s'expliquer par une tendance des spectres masculins à présenter des macrostructures plus similaires entre sujets, par opposition aux spectres féminins. Cette hypothèse fera l'objet de développements ultérieurs.

4. BIBLIOGRAPHIE

[1] Harmegnies B., SDDD : a new dissimilarity index for the comparison of speech spectra, *Pattern Recognition Letters* 8 (1988) 153-158.

[2] SWETS T.A., The relative operating characteristic in psychology, *Science* 182 (1973) 990-1000.

[3] Coombs C.H., Dawes R.M., Tversky A., *Psychologie mathématique* (P.U.F., Paris, 1975).

[4] Achard F., Choix d'un indice de similarité en classification automatique, *Quality & Quantity* 6 (1972) 289-302.

[5] Chandon J.L., Pinson S., *Les méthodes d'analyse typologique* (Encyclopédie du Marketing II, Editions Techniques, Paris, 1977).

[6] Chandon J.L., Pinson S., *Analyse typologique* (Masson, Paris, 1981).

[7] Majewski W., Basztura C., "Speaker recognition in open sets", *Proc. of the 10th Congress of the Phonetic Sciences, Utrecht* (1983) pp. 322-325.

[8] Bhattacharya C.G., A simple method of resolution of a distribution into its Gaussian components, *Biometrics* 23 (1967) 115-137.

INDICES	FEMMES	HOMMES	ENSEMBLE
1)D. euclidienne	.9912634 (2)	.9994374 (1)	.9968417 (1)
2)D. de Bhattacharya	.9911153 (4)	.9991964 (7)	.9963234 (2)
3)SDDD (I)	.9887673 (7)	.9994123 (4)	.9962488 (3)
4)Coef. de corrélation de Bravais-Pearson (I)	.9860203 (9)	.9994369 (2)	.9962137 (4)
5)D. du Chi Deux	.9885125 (8)	.9992840 (6)	.9958913 (5)
6)D. de Bray et Curtis	.9921086 (1)	.9993774 (5)	.9958249 (6)
7)Coef. de similarité angulaire	.9896483 (6)	.9994319 (3)	.9957873 (7)
8)D. de Clark	.9906379 (5)	.9988132 (9)	.9955281 (8)
9)D. de Canberra	.9912297 (3)	.9991727 (8)	.9953869 (9)
10)D. pondérée par l'inverse de l'écart type	.9801566 (10)	.9905694 (10)	.9822119 (10)
11)Coef. de Kolmogorov-Smirnov	.9474944 (11)	.9511292 (11)	.9602561 (11)
12)D. de Chebychev	.9403828 (12)	.9252719 (12)	.9567777 (12)
13)Covariance (I)	.7771032 (13)	.7846393 (13)	.8221137 (13)
14)Chi Deux de Basztura	.6254754 (16)	.5597416 (15)	.6340068 (14)
15)C. aplatissement (I)	.6910466 (14)	.6282366 (14)	.4811614 (15)
16)C. de dissymétrie	.6367286 (15)	.5271626 (16)	.4756967 (16)

Tableau 1. Surface incluse sous la courbe ROC des divers indices. Les indices insensibles aux variations de niveaux sont signalés par (I). Le numéro entre parenthèses indique le rang de l'indice par ordre de pouvoir discriminant décroissant.

(01) $T = \sqrt{\frac{\sum_{j=1}^f W_j (X_j - Y_j)^2}{\sum_{j=1}^f W_j}}$ avec $W_j = 1$

(02) $T = \sqrt{\frac{\sum_{j=1}^f W_j (\sqrt{X_j} - \sqrt{Y_j})^2}{\sum_{j=1}^f W_j}}$ avec $W_j = 1$

(03) $T = \sqrt{\frac{1}{f} \sum_{j=1}^f (X_j - Y_j)^2 - \left(\frac{1}{f} \sum_{j=1}^f (X_j - Y_j)\right)^2}$

(04) $T = \frac{1}{f} \frac{\sum_{j=1}^f (X_j - \bar{X})(Y_j - \bar{Y})}{\sigma_x \sigma_y}$

(05) $T = \frac{\sum_{j=1}^f W_j |X_j - Y_j|}{\sum_{j=1}^f W_j}$ avec $W_j = \frac{1}{\sum_{j=1}^f (X_j + Y_j)}$

(06) $T = \frac{\sum_{j=1}^f \left[\frac{(X_j - A)^2}{A+B} + \frac{(Y_j - B)^2}{A+B} \right]}{\sum_{j=1}^f \left[\frac{X_j}{A+B} + \frac{Y_j}{A+B} \right]}$

avec : $A = \sum_{j=1}^f X_j$ et $B = \sum_{j=1}^f Y_j$

(07) $T = \frac{\sum_{j=1}^f X_j * Y_j}{\sqrt{\sum_{j=1}^f X_j^2 * \sum_{j=1}^f Y_j^2}}$

(08) $T = \sqrt{\frac{1}{f} \sum_{j=1}^f W_j (X_j - Y_j)^2}$ avec $W_j = \frac{1}{(X_j + Y_j)^2}$

(09) $T = \sum_{j=1}^f W_j |X_j - Y_j|$ avec $W_j = \frac{1}{(X_j + Y_j)}$

(10) $T = \sqrt{\sum_{j=1}^f W_j (X_j - Y_j)^2}$ avec $W_j = \frac{1}{\max_{1 \leq j \leq f} (X_j - Y_j)^2}$

(11) $X_{original} - X_1 \ X_{i+1} \ X_{i+2} \ \dots \ X_f$

$Y_{original} - Y_1 \ Y_{i+1} \ Y_{i+2} \ \dots \ Y_f$

$X_{cumulé} - X_{i+1} = X_i \ X_{i+1} = X_i + X_{i+1} \ \dots$

$Y_{cumulé} - Y_{i+1} = Y_i \ Y_{i+1} = Y_i + Y_{i+1} \ \dots$

Diff. → $|\Delta_i| \quad |\Delta_{i+1}| \quad \dots$

$T = AU |\Delta| \text{ MAXIMUM}$

(12) $T = \max_{1 \leq j \leq f} |X_j - Y_j|$

(13) $T = \frac{1}{f} \sum_{j=1}^f (X_j - \bar{X})(Y_j - \bar{Y})$

(14) $T = \sqrt{\frac{\sum_{j=1}^f \frac{1}{(X_j + Y_j)}}{\sum_{j=1}^f X_j} \left[\frac{X_j}{\sum_{j=1}^f X_j} - \frac{Y_j}{\sum_{j=1}^f Y_j} \right]^2}$

(15) $T = \frac{1}{f \sigma^2} \left[\sum_{j=1}^f D_j^4 - \frac{4}{f} \sum_{j=1}^f D_j \sum_{j=1}^f D_j^3 + \frac{6}{f^2} \left(\sum_{j=1}^f D_j \right)^2 \sum_{j=1}^f D_j^2 - \frac{3}{f^3} \left(\sum_{j=1}^f D_j \right)^4 \right]$

avec : $\sigma^2 = \text{écart type des } X_i - Y_i$
 $D_i = X_i - Y_i$ avec $j \leq i \leq f$

(16) $T = \frac{1}{f \sigma^3} \left[\sum_{j=1}^f D_j^3 - \frac{3}{f} \sum_{j=1}^f D_j \sum_{j=1}^f D_j^2 + \frac{2}{f^2} \sum_{j=1}^f D_j^3 \right]$

avec $\sigma^2 = \text{écart type des } X_i - Y_i$
et $D_i = X_i - Y_i$ avec $j \leq i \leq f$

Annexe 1. Indices étudiés