

# Evolution temporelle de la variabilité du spectre moyen à long terme

M. BRUYNINCKX et B. HARMEGNIES

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

**Abstract:** Most research concerning the Long Term Average Spectrum (LTAS) variability has consisted in analyses of contemporary vocal utterances but others studies have nevertheless revealed the presence in LTAS of a specific variability depending on time. Furui et al ([1], [2]) have shown a good Long Term Cepstrum stability for periods that do not exceed 3 weeks but a significant variability is progressively detected for longer periods. This communication must be inserted in the research work aiming at systematization of all the studies centered on the comparisons of recordings collected at 5-year intervals in similar experimental conditions.

## 1. INTRODUCTION

Dans leur grande majorité, les recherches relatives à la variabilité du spectre à long terme (SMLT) ont recouru à l'analyse de productions vocales contemporaines. Quelques travaux ont cependant porté sur l'étude de la variabilité spécifique liée au temps. Ainsi, Furui et coll. ont, sur base d'enregistrements en japonais, mis en évidence une bonne stabilité du *cepstre* à long terme pour des périodes allant de quelques jours à trois semaines mais montrent que, passé ce délai, une variabilité substantielle s'installe graduellement ([1], [2], [3], [4], [5], [6]). On ignore cependant les caractéristiques de la variabilité temporelle des spectres à long terme, en particulier provenant d'autres langues.

Cette communication s'inscrit dans le cadre de recherches visant à systématiser ces études. Elle se centre sur les comparaisons d'enregistrements de sujets francophones, enregistrés lors de deux sessions à 5 ans d'intervalle, à la faveur de conditions expérimentales invariantes.

## 2. DISPOSITIF EXPERIMENTAL

### 2.1. Sujets

Les sujets, tous de sexe masculin et âgés d'une vingtaine d'années au moment des premiers enregistrements, étaient au nombre de cinq. Tous étaient francophones natifs, originaires de Belgique et y résidant. Ils n'étaient atteints d'aucune pathologie ORL.

### 2.2. Conditions d'enregistrement

La première [7] et la seconde session d'enregistrement se sont déroulées dans la chambre anéchoïque du Service de Physique Générale de la Faculté polytechnique de Mons avec un matériel inchangé : un microphone NEUMANN U87i, connecté via le circuit de préamplification d'un magnétophone NAGRA IV-S, à un digitaliseur SONY PCM-501 ES utilisant comme mémoire de masse le canal image d'un magnétoscope SONY SL-HF 950; les sujets étaient maintenus à une distance constante (40cm) du micro.

### 2.3. Corpus

Lors des deux sessions, les sujets ont été invités à répéter cinq fois un même texte phonétiquement équilibré comportant 277 phonèmes et d'une durée approximative de production de vingt secondes [7]. La présente expérience porte donc sur cinquante productions (5 sujets x 5 productions x 2 sessions).

## 3. TRAITEMENT DES DONNEES ET RESULTATS

### 3.1. Analyses acoustiques

Les enregistrements recueillis furent traités grâce à un analyseur spectral Brüel Kjaer 2033 dont la fréquence d'échantillonnage du signal d'entrée était fixée à 12.8 kHz. Les spectres produits étaient ainsi définis sur 400 canaux avec une résolution constante de 12.5 Hz dans la bande 0-5 kHz. Un algorithme de calcul de moyenne à pondération linéaire a fourni un SMLT pour chaque production. Les SMLT obtenus ont été ensuite transmis à un ordinateur de type IBM PC via une interface GPIB, en vue du traitement statistique.

### 3.2. Quantification des dissimilarités spectrales

Les spectres ont fait l'objet de comparaisons paire par paire, réalisées au moyen de l'indice de dissimilarité SDDD [8]. Soient  $S$  et  $S'$ , deux spectres à comparer,  $s$  et  $s'$ , leurs niveaux respectifs pour la  $i$ ème bande de fréquence, on a :

$$SDDD_{ss'} = \sqrt{\frac{1}{K} \sum_{i=1}^k (s_i - s'_i - M)^2}$$

avec  $K$ , le nombre de bandes de fréquences définissant le spectre, et  $M$ , la moyenne des différences  $s_i - s'_i$ .

La valeur de l'indice est d'autant plus grande que les formes des spectres comparés sont différentes. Elle est nulle lorsque les formes sont identiques. En outre, au contraire d'indices tels la distance euclidienne, SDDD ne requiert pas de normalisation des SMLT en intensité.

### 3.3. Traitement des spectres dans la bande DC-5kHz

Seules des comparaisons intra sujet ont été effectuées : d'une part, deux types de comparaisons intra période (session 1 vs session 1 et session 2 vs session 2) et, d'autre part, des comparaisons inter périodes (session1 vs session2).

Pour chacune des deux sessions, nous avons ainsi procédé à la comparaison des productions de chaque locuteur et ce, pour chacune des dix paires non redondantes de SMLT, soit un total de 100 comparaisons (2 sessions x 5 sujets x 10 comparaisons de spectres contemporains). La même procédure a été utilisée pour les comparaisons inter périodes, soit 125 comparaisons (5 sujets X 25 comparaisons de spectres non contemporains).

Le tableau 1 résume les valeurs des indices individuels moyens de dissimilarité SDDD obtenus à partir des différentes comparaisons intra et inter périodes effectuées entre les spectres moyens à long terme issus des productions de chacun des deux sujets pour les sessions 1 et 2. Chaque nombre  $y$  figurant est une moyenne résultant de 10 comparaisons dans le cas des comparaisons intra période et 25, dans celui des comparaisons inter périodes.

	1	2	3	4	5
INTRA 1	2.24	2.67	2.41	2.24	2.43
INTRA 2	2.46	2.38	2.42	2.32	1.94
INTER 1/2	6.06	5.19	6.22	4.52	3.76

Tableau 1 : indices individuels moyens de dissimilarité pour les comparaisons intra et inter périodes.

On constate que les valeurs résultant des comparaisons inter périodes sont considérablement plus élevées que celles issues des comparaisons intra période, alors

que les différences entre les deux types de comparaisons intra période sont minimes. Cette observation confirme, sur base de ces enregistrements francophones, l'existence d'une variabilité naturelle du SMLT dans le temps signalée par Furui (1977) sur base de productions en japonais.

### 3.4. Traitement des spectres en basses fréquences

Dans le but de contribuer à une investigation des sources de variation temporelle du SMLT, nous avons réitéré la procédure de comparaison exposée en 3.3. à trois reprises. Seule une partie de l'information spectrale a cependant été utilisée à chaque occasion. Le premier traitement a ainsi été limité aux quarante premiers points du spectre (0-500 Hz), le deuxième aux cinquante premiers (0-625 Hz) et le troisième aux soixante premiers (0-750 Hz). Nous visons, par ce biais, à évaluer la validité de l'hypothèse selon laquelle l'information portée par la partie basses fréquences du spectre offre une meilleure stabilité dans le temps que celle caractérisant les fréquences hautes. Les 300 valeurs d'indices SDDD ainsi calculées, ne pouvant être présentées ici, sont résumées sous forme graphique à la figure 1.

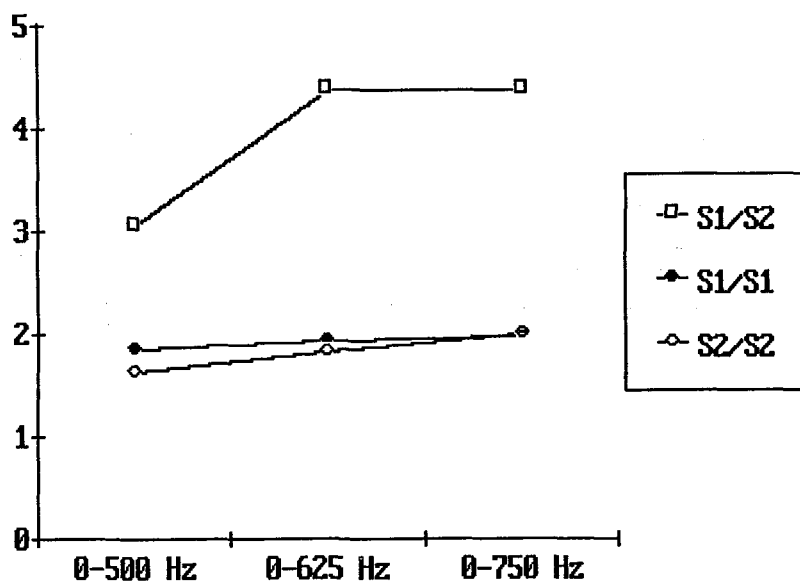


Figure 1 : Moyennes des indices SDDD tenus à la faveur des comparaisons intra sujet intra session (S1/S1 et S2/S2) et inter sessions (S1/S2) pour chacun des trois empanns fréquentiels

On constate d'abord, que les dissimilarités interspectrales caractérisant les comparaisons inter sessions sont supérieures à celles que révèlent les comparaisons intra session. Il se confirme donc que le délai séparant les deux sessions d'enregistrement contribue à modifier la forme du spectre à long terme.

Il est clair, cependant, que le rapport entre les dissimilarités intra session et la dissimilarité inter session varie en fonction de la bande de fréquences considérée. Ce sont les spectres définis dans la zone des fréquences les plus basses qui résistent le mieux à l'introduction du délai inter sessions.

### 4. CONCLUSION

Les différentes comparaisons effectuées au moyen de procédures de reconnaissance de forme sur les SMLT obtenus à travers plusieurs bandes d'analyse fréquentielle, ont permis de mettre en évidence une meilleure stabilité dans le temps des fréquences basses. Cette conclusion obtenue à partir d'enregistrements francophones est en accord avec les observations de Furui et coll., dérivées de corpus en japonais. Elle apporte ainsi un début de confirmation à l'idée que le phénomène présente quelque généralité.

Il appartiendra aux recherches futures de procéder à une investigation systématique des zones fréquentielles résistantes. Il conviendrait également que la recherche soit ouverte à d'autres langues.

En tout état de cause, on peut noter que dans le cas présent, la prise en compte graduellement plus importante d'une zone exploitée par le premier formant vocalique a amoindri la fidélité du SMLT. Le poids relatif plus important de la zone correspondant au fondamental conforte au contraire sa stabilité. Bien que cette observation résulte d'une analyse trop superficielle, on peut noter qu'elle est liée à la problématique de l'étiologie physique des variations du spectre à long terme. En effet, si ce dernier est, comme le suggèrent certaines recherches ([9], [10]), plus influencé par la source que par le conduit, il y a effectivement lieu d'investiguer dans quelle mesure le fonctionnement laryngien est plus invariant que celui des articulateurs supraglottiques.

## 5. REFERENCES

- [1] Furui S., Itakura F., Time variation of long-time average spectrum, Effects on talker recognition, *Conv. Rec. Acoust. Soc. of Japan*, 3 (1972[a]) pp. 2-5.
- [2] Furui S., Itakura F., Time variation by long-term averaged speech spectrum, *Trans. IECE Japan*, 55A (1972[b]) p. 549.
- [3] Furui I., Analysis of temporal variation of talker dependant features, *Review of the electrical communication laboratories*, 25(3-4) (1977) pp. 231-244.
- [4] Furui S., Effects of long-term spectral variability on speaker recognition, *Journal of the Acoustical Society of America*, 64(suppl.1), paper NN28 (1978), p. 183.
- [5] Furui S., Itakura F., Saito S., Personal information in the long time average speech spectrum, *The electrical communication laboratories*, 23 (1985), pp. 9-10.
- [6] Furui S., Research on individuality features in speech waves and automatic speaker recognition techniques, *Speech Communication*, 5 (1986) pp. 183-197.
- [7] Harmegnies B., Contribution à la caractérisation de la qualité vocale. Analyses plurielles de spectres moyens à long terme de parole, Dissertation doctorale (Université de Mons-Hainaut, 1988[a]).
- [8] Harmegnies B., SDDD, A new dissimilarity index for the comparison of speech spectra, *Pattern Recognition Letters*, 8 (1988[b]) pp. 153-158.
- [9] Harmegnies B., Esling J., Delplancq V., Quantitative study of the effects of settings changes on the LTAS, Acts of the European Conference on Speech Communication and Technology, Paris 1989 (1989), pp. 132-149.
- [10] Nolan F., The phonetic bases of speaker recognition (Cambridge University Press, Cambridge 1983).