

## Etude acoustique du réflexe Lombard en vue de la reconnaissance de la parole produite en milieu bruité

R. HAJISLAM, Y. ANGLADE, J.-C. JUNQUA\* et J.-M. PIERREL

CRIN-CNRS & INRIA-Lorraine, Campus Scientifique, BP. 239, 54506 Vandœuvre-lès-Nancy cedex, France

\* et Speech Technology Laboratory, Santa Barbara, California, U.S.A.

### Résumé

Le but de cette étude est de déterminer pour le français, au niveau phonétique, les différences acoustiques entre la parole normale et la parole prononcée en milieu bruité (soumise au réflexe Lombard). Ce travail fait suite à une première étude que nous avons effectué sur l'anglais-américain. La même méthodologie a été adoptée pour les deux langues, et les mêmes paramètres acoustiques ont été extraits, ce qui se justifie par leur utilisation similaire dans les différents systèmes de reconnaissance de la parole étudiés pour les deux langues.

Les résultats obtenus au cours de la première étude sur l'anglais-américain [1] ont montré des tendances d'évolution de nombreux paramètres acoustiques. Dans ce papier, nous présentons les résultats obtenus pour le français. Ils confirment les tendances observées dans [1], et montrent ainsi qu'il existe une forte corrélation entre les modifications acoustiques engendrées par le réflexe Lombard sur l'anglais-américain et le français.

## 1 INTRODUCTION

Il est bien connu qu'en présence de bruit, l'être humain a tendance à accentuer son effort vocal. Ceci est appelé le réflexe Lombard [2]. Plus récemment, Schulman [3] a montré que l'augmentation de l'effort vocal modifie les mouvements articulatoires et que cette modification n'est pas une amplification linéaire de l'articulation normale, mais plutôt une réorganisation complexe des mouvements articulatoires. Par ailleurs, Junqua [4] a montré que la complexité du problème est telle que les changements acoustiques liés au réflexe Lombard dépendent du type du bruit.

La grande majorité des systèmes actuels de reconnaissance automatique de la parole ont été conçus pour travailler sur la parole normale sans tenir compte du réflexe Lombard. Or plusieurs études [5] ont montré que dans certains cas la variation de la production de la parole due à la présence du bruit dégrade beaucoup plus les performances des systèmes que le bruit ambiant lui-même.

Dans la perspective d'améliorer la robustesse des systèmes de reconnaissance, pour qu'ils puissent tenir compte du réflexe Lombard, plusieurs études ont été consacrées à l'étude des différences acoustiques entre la parole Lombard et la parole normale [6, 7, 8].

Notre travail s'insère dans le cadre du projet européen ROARS<sup>1</sup> dont l'objectif est de construire un système analytique robuste de reconnaissance automatique de la parole continue. Notre tâche consiste à étudier les différences acoustiques entre la parole Lombard et la parole normale. Dans un premier temps, nous avons mené une étude sur une base de données anglo-américaine déjà existante [1]

<sup>1</sup>ROARS, RObust Analytical Recognition System, projet ESPRIT 2 réunissant Thomson Sintra DASM (F), le CRIN-CNRS & INRIA-Lorraine (F), ENA Telecomunicaciones (SP) et l'université de Valencia (SP).

où nous avons relevé des modifications dues à l'effet Lombard pour certains paramètres acoustiques. L'étude que nous présentons sur le français dans cet article fait suite à ce travail : il vise à vérifier si les résultats et les tendances d'évolution des paramètres acoustiques observées pour l'anglais-américain sont les mêmes pour le français.

## 2 PROCEDURE EXPERIMENTALE

Pour réaliser notre étude, nous avons enregistré une base de données comportant 36 mots : les chiffres et les lettres du français. Ces mots ont été prononcés de façon isolée par quatre locuteurs : 2 hommes et 2 femmes originaires de différentes régions de France. Chaque locuteur a prononcé quatre fois la liste de mots dans une chambre sourde, deux fois dans une ambiance calme et deux fois dans une ambiance bruitée où un bruit blanc Gaussien de 85 dB SPL était injecté dans un casque calibré. Le corpus a été digitalisé à 20 kHz, puis étiqueté à l'aide de *Snorri* [9], un logiciel d'analyse et de traitement de la parole développé au CRIN-CNRS & INRIA Lorraine. Pour analyser l'influence du réflexe Lombard et valider les résultats obtenus pour l'anglais-américain, nous avons étudié les paramètres suivants, choisis du fait de leur importance dans un système de reconnaissance de la parole :

- l'énergie dans sept bandes de fréquences : 0-250 Hz, 250-500, 500-1000, 1-3 kHz, 3-5 kHz, 5-7 kHz, 7-10 kHz. Ces bandes sont plus précises dans les basses fréquences, où se trouve le maximum de l'information utile pour la reconnaissance.
- Le centre de gravité énergétique : il permet de synthétiser en un seul paramètre le déplacement fréquentiel de l'énergie.
- La position fréquentielle des trois premiers formants calculés par la méthode LPC (Linear Predictive Coding).
- La limite inférieure de bruit fricatif : c'est la fréquence à partir de laquelle commence la concentration d'énergie d'une fricative.
- La durée de chaque phonème : obtenue grâce à l'étiquetage manuel.

Tous ces paramètres (à l'exception de la durée) ont été calculés en trois points de chaque phonème, répartis de façon à négliger au maximum les influences des contextes phonétiques gauche et droit (au premier quart du phonème, à la moitié et au dernier quart). Nous avons calculé une moyenne de ces trois points pour obtenir la valeur finale qui a été utilisée pour les tests de comparaison.

Pour analyser les résultats obtenus pour chacun des paramètres calculés, nous avons utilisé des tests statistiques. L'objectif de ces tests est de fournir les tendances générales concernant les différences entre les populations *parole normale* et *parole Lombard*. Lors de la comparaison, nous avons pris la parole normale comme référence. Pour analyser les différences, nous avons utilisé deux tests classiques de comparaison : le Tp-test et l'analyse de variance à deux dimensions. Nous avons relevé très peu de différences de résultats entre les deux tests. Ils ont été appliqués pour chaque paramètre calculé sur l'ensemble des occurrences d'un phonème présent dans la base de données, tous locuteurs et contextes phonétiques confondus.

## 3 LES RESULTATS

Pour présenter les résultats liés à chacun des paramètres, nous allons utiliser les notations suivantes : la tendance d'évolution du paramètre est stable ( $\Leftrightarrow$ ), en légère diminution ( $\Downarrow$ ), en importante diminution ( $\Downarrow\Downarrow$ ), en légère augmentation ( $\Uparrow$ ) ou en importante augmentation ( $\Uparrow\Uparrow$ ). L'importance d'une variation est déterminée en fonction de deux valeurs de seuil qui sont calculées à partir des variations minimale et maximale des paramètres. La table 1 contient les résultats obtenus pour tous les phonèmes. Nous pouvons les résumer par les conclusions suivantes :

- l'énergie (E1 à E7) : concernant les voyelles, nous notons une baisse d'énergie dans les basses fréquences et une augmentation dans les hautes fréquences. Il y a une stabilité pour les fréquences intermédiaires. Les semi-voyelles, les liquides et les nasales sont affectées par l'effet Lombard dans les basses fréquences seulement. En effet, nous pouvons voir que l'énergie

diminue entre 0 et 500 Hz, et reste stable dans les autres bandes de fréquences. Quant aux fricatives, on remarque que l'énergie baisse dans toutes les bandes de fréquences sauf pour les phonèmes /f/ et /v/ qui ne suivent pas cette tendance. Pour les occlusives, nous notons aussi une diminution de l'énergie dans toutes les bandes de fréquence avec une certaine stabilité pour les fréquences intermédiaires;

- **le centre de gravité énergétique (CGE)** : le centre de gravité énergétique augmente pour les voyelles. Concernant les autres phonèmes, nous avons noté une stabilité dans sa position fréquentielle avec une légère augmentation pour le /s/, /ʃ/ et /k/;
- **la durée phonétique** : l'analyse de ce paramètre montre qu'il y a une légère augmentation pour les voyelles et une stabilité pour les semi-voyelles et les liquides. Concernant les fricatives, nous avons noté une stabilité dans la durée, à l'exception du /s/ et du /v/, où une légère diminution a été notée. Comme dans notre base de données, la plupart des occlusives sont situées au début des mots, leur durée totale (silence compris) est difficilement mesurable, particulièrement pour les occlusives non-voisées. En conséquence, nous n'avons pas étudié la variation de la durée phonétique pour cette catégorie de phonème;

Table 1 : Synthèse des résultats obtenus pour chaque phonème

phonème	E 1	E 2	E 3	E 4	E 5	E 6	E 7	CGE	F 1	F 2	F 3	SF
i	↓	↓	↔	↔	↑	↑	↔	↑	↑	↔	↔	
e	↓	↓	↔	↔	↑	↑	↔	↑	↑	↔	↓	
ɛ	↓	↓	↔	↔	↑	↑	↑	↑	↑	↔	↔	
a	↓	↓	↔	↔	↑	↑	↑	↑	↑	↔	↓	
o	↓	↓	↑	↔	↔	↑	↑	↑	↑	↑	↔	
u	↓	↓	↔	↔	↔	↔	↑	↔	↑	↔	↔	
y	↓	↓	↔	↔	↑	↔	↑	↑	↑	↔	↔	
ø	↓	↓	↔	↔	↔	↑	↑	↑	↑	↔	↔	
œ	↓	↓	↔	↔	↑	↑	↑	↑	↑	↔	↔	
ë	↓	↓	↔	↔	↔	↔	↑	↑	↔	↔	↔	
ö	↓	↓	↔	↔	↔	↔	↑	↑	↑	↔	↔	
R	↓	↔	↔	↔	↔	↓	↓	↔				
l	↓	↓	↔	↔	↔	↔	↓	↔				
w	↓	↓	↓	↔	↔	↔	↓	↔				
q	↓	↓	↔	↔	↔	↔	↔	↔				
m	↓	↓	↔	↔	↔	↔	↔	↔				
n	↓	↓	↔	↔	↔	↔	↓	↔				
p	↓	↓	↔	↓	↓	↓	↓	↔				
t	↓	↑	↑	↑	↑	↔	↔	↔				
k	↓	↓	↔	↔	↔	↔	↓	↑				
b	↓	↓	↔	↔	↓	↓	↓	↔				
d	↓	↓	↔	↔	↔	↔	↓	↔				
g	↓	↓	↔	↔	↓	↓	↓	↓				
f	↔	↔	↔	↔	↔	↔	↓	↔				↑
s	↓	↓	↓	↓	↓	↓	↓	↑				↑
ʃ	↓	↓	↓	↓	↓	↓	↓	↑				↑
v	↓	↔	↔	↔	↔	↔	↓	↔				↑
z	↓	↓	↓	↔	↓	↓	↓	↔				↑
ʒ	↓	↓	↓	↓	↓	↓	↓	↔				↑

- **les formants (F1, F2 et F3)** : comme nous l'avons également constaté pour l'anglais-américain, la tendance principale d'évolution des formants concerne le premier formant. En

effet, on observe une augmentation de la fréquence de ce formant pour toutes les voyelles. Concernant les deuxième et troisième formants, nous avons noté une stabilité de leurs positions fréquentielles;

- **le seuil de friction (SF)** : une augmentation nette de la limite fréquentielle inférieure de l'énergie a été noté pour toutes les fricatives.

Globalement, ces résultats sont très similaires à ceux obtenus pour l'anglais-américain [1].

## 4 CONCLUSION

Ce travail présente une étude acoustique comparative de différents paramètres acoustiques calculés sur de la parole normale et sur de la parole produite dans un milieu bruité (parole Lombard). Il s'insère dans le cadre du projet ESPRIT européen ROARS et est un complément à une précédente étude effectuée avec la même méthodologie et les mêmes algorithmes sur l'anglais-américain. Cette étude avait montré l'existence de modifications acoustiques dues au réflexe Lombard de nombreux paramètres. En particulier, ces évolutions concernaient l'énergie dans les basses fréquences, le centre de gravité énergétique, le premier formant, les passages par zéro, le seuil de friction et la durée phonétique. Les résultats obtenus dans notre étude montrent que ces tendances d'évolution sont semblables pour le français. Ils témoignent de l'importance de l'influence du réflexe Lombard sur les paramètres acoustiques utilisés dans les systèmes analytiques de reconnaissance automatique de parole, tel celui proposé dans le système ROARS, et de la nécessité d'en tenir compte pour améliorer leur robustesse au bruit. L'intégration de ces résultats dans ces systèmes peut se faire de deux manières : soit en réalisant, au vu du rapport *Signal/Bruit* de la parole à reconnaître, une adaptation automatique des valeurs de référence des principaux paramètres de reconnaissance au sein du système. Soit en fondant d'avantage la reconnaissance sur les indices les moins affectés par le réflexe Lombard afin de s'affranchir le plus possible de la variabilité non linéaire de certains paramètres acoustico-phonétiques.

## Références

- [1] J.C. Junqua and Y. Anglade. Acoustic and perceptual studies of lombard speech : Application to isolated-word automatic speech recognition. In *ICASSP*, pages 841–844, 1990.
- [2] E. Lombard. Le signe de l'élévation de la voix. *Ann. Maladies Oreille, Larynx, Nez, Pharynx*, 37:101–119, 1911.
- [3] R. Schulman. Articulatory dynamics of loud and normal speech. *JASA*, 85(1):295–312, 1989.
- [4] J.C. Junqua. Acoustic and production pilot studies of speech vowels produced in noise. In *ICSLP*, 1992.
- [5] J.C. Junqua and H. Wakita. A comparative study of cepstral lifters and distance mesures for all pole models of speech in noise. In *ICASSP*, pages 476–479, 1989.
- [6] R. Hajislam. Technical Report on Changes in Speaker Articulation Due to Ambient Noise (complement report). Rapport de fin de contrat, ESPRIT II project 5516 ROARS, 1992.
- [7] B.J. Stanton, L.H. Jamieson, and G.D. Allen. Acoustic-phonetic analysis of loud and lombard speech in simulated cockpit conditions. In *ICASSP*, 1988.
- [8] W.V. Summers, D.B. Pisoni, R.H. Bernacki, R.I. Pedlow, and M.A. Stockes. Effects of noise on speech production : Acoustic and perceptual analyses. *JASA*, 84(3):917–927, 1988.
- [9] Y. Laprie. SNORRI : Un système d'étude interactif de la parole. In *17e JEP*, Nancy, 1988.