

Étude acoustique et articulatoire de la parole Lombard

Maëva Garnier, Lucie Bailly, Marion Dohen, Pauline Welby, Hélène Loevenbruck

▶ **To cite this version:**

Maëva Garnier, Lucie Bailly, Marion Dohen, Pauline Welby, Hélène Loevenbruck. Étude acoustique et articulatoire de la parole Lombard : Effets globaux sur l'énoncé entier. XXVIemes Journees d'Etudes sur la Parole, Jun 2006, Dinard, France. pp.077. hal-00214304

HAL Id: hal-00214304
https://hal.science/hal-00214304

Submitted on 29 Jan 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Étude acoustique et articulatoire de la parole Lombard : Effets globaux sur l'énoncé entier

Maëva Garnier[1], Lucie Bailly[1,2], Marion Dohen[2], Pauline Welby[2], Hélène Lœvenbruck[2]

[1]Laboratoire d'acoustique musicale, CNRS UMR 7604, UMPC, Ministère de la culture, 11 rue de Lourmel, 75015 Paris
[2]Institut de la communication parlée, CNRS UMR 5009, INPG, Univ. Stendhal, 46 av. F. Viallet, 38031 Grenoble
Mél : garnier@lam.jussieu.fr, {lucie.bailly, pauline.welby, helene.loevenbruck}@icp.inpg.fr, marion.dohen@gmail.com

ABSTRACT

This study aims at characterizing the articulatory modifications that occur in speech in noisy environments, and at examining them as compensatory strategies. Audio, EGG and video signals were recorded for a female native speaker of French. The corpus consisted of 33 short sentences with a Subject-Verb-Object (SVO) structure. The sentences were recorded in three conditions : silence, 85dB white noise, 85dB "cocktail party" noise. Labial parameters were extracted from the video data. The analyses enabled us to examine the effect of the type of noise and to show that hyper-articulation concerns lip aperture and spreading rather than lip pinching. The analysis of the relationship between acoustic and articulatory parameters show that this speaker especially adapts to noise not only by talking louder or increasing vowel recognition cues but also by increasing spectral emergence.

1. INTRODUCTION

Les environnements bruyants affectent les situations de communication car ils perturbent à la fois la perception que le locuteur a de sa propre voix et son intelligibilité vis-à-vis de son interlocuteur. Pour compenser ces perturbations, le locuteur modifie sa façon de parler. Ce phénomène est connu sous le nom d'effet Lombard et a été principalement décrit des points de vue acoustique et phonétique [1-4]. Dans cette étude, nous nous intéressons aux modifications articulatoires concomitantes de l'effet Lombard. Nous cherchons à caractériser ces modifications articulatoires par rapport à la parole produite dans le silence. Existe-t-il une hyper-articulation? Est-elle conjointe à la modification de certains paramètres acoustiques ? Quels sont les paramètres articulatoires concernés par ces modifications ? Enfin, l'articulation de la parole Lombard dépend-elle du type de bruit ? À notre connaissance, quelques études se sont intéressées à l'articulation de certains types de parole "hyper" [5-7] mais très peu d'études concernent spécifiquement l'articulation de la parole Lombard [8-9].

2. MATÉRIEL ET MÉTHODE

2.1. Corpus

Les 33 phrases du corpus ont été construites sur une même structure Sujet-Verbe-Objet et selon un enchaînement de syllabes de structure CV, afin de simplifier la segmentation acoustique [10] (cf. (1) pour exemple). Seules des consonnes sonores ont été choisies afin de minimiser les perturbations de la courbe de F0.

(1) Le monument et les moulins rallient la vallée.

2.2. Enregistrement audiovisuel

Nous avons enregistré simultanément les signaux audio, électroglottographiques (EGG) et articulatoires d'une locutrice française, n'ayant aucune connaissance du protocole ou de l'effet Lombard . Elle avait pour consigne de lire les phrases en s'adressant à l'expérimentatrice située à 2m en face d'elle. Les données articulatoires ont été extraites d'enregistrements vidéo (25 images/s) de face des lèvres de la locutrice, en utilisant un dispositif labiométrique développé à l'ICP [11]. Dans le cadre de cette étude, nous nous sommes concentrées sur l'analyse de **l'étirement (A)** et de **l'ouverture (B)** des lèvres, ainsi que de **l'aire intéro-labiale (S)** (cf. Figure 1). Pour ces trois paramètres articulatoires, nous avons examiné à la fois les maxima d'amplitude des mouvements (**max**) ainsi que leur variation plus globale (**glob**) bien représentée par l'aire sous la courbe d'évolution de A, B, ou S en fonction du temps [12, p113-114].

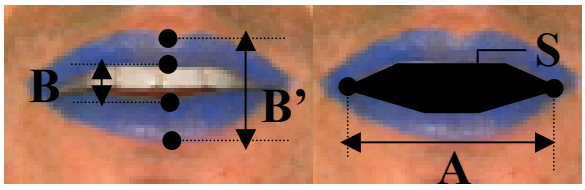


Figure 1 : Représentation des paramètres articulatoires

Nous avons également examiné un quatrième paramètre : le pincement, défini comme la compression des lèvres au moment de leur fermeture sur les segments [m], et donc équivalent au paramètre B' lorsque B est égal à 0 (cf. Figure 1). On observe deux types de pincement : **le pincement protrus** pour lequel la surface labiale visible augmente au moment de la compression, et **le pincement avalé** correspondant à un écrasement des lèvres vers l'intérieur de la bouche avec rétrécissement de la surface labiale visible.

Le signal audio a été enregistré à l'aide d'un microphone AKG placé à 20cm des lèvres, et le signal EGG à l'aide d'un électroglottographe EG2. Ils ont été numérisés à 44.1kHz sur 16bits. Deux environnements bruyants : un bruit blanc **(bb)** et un bruit de "cocktail party" **(cktl)**, tirés de la base BD_Bruit [13], ont été diffusés à 85dB (intensité mesurée au niveau des oreilles de la locutrice) par deux haut-parleurs placés à 2m de la locutrice et espacés de 2m. Leur spectre moyenné est représenté Figure 2. La locutrice a d'abord été enregistrée en situation silencieuse de référence, puis dans chacun des deux environnements bruyants.

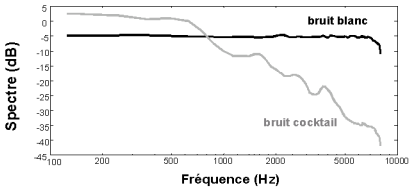


Figure 2: Spectre moyenné des deux types de bruit étudiés.

Les données audio ont été traitées pour supprimer le bruit (algorithme de [14]) et étiquetées sous Praat en déterminant les frontières des phrases, des groupes nominaux du sujet et de l'objet des phrases, et de leurs segments. La fréquence fondamentale **(F0)** et le quotient ouvert **(Oq)** ont été mesurés à partir du signal EGG par une méthode d'autocorrélation.

3. RÉSULTATS

Nous avons adopté dans la suite de l'article la notation suivante pour exprimer les résultats statistiques : * pour p<0,05, ** pour p<0,01, *** pour p<0,001, et ns pour p>0,05 (non significatif).

3.1. Articulation des énoncés dans le bruit

Amplitude des mouvements articulatoires

On observe une augmentation significative de A, B et S dans le bruit, pour les deux types de bruit, et une diminution significative du pincement protrus dans le bruit blanc (cf. Table 1). On remarque un effet significatif du type de bruit sur l'augmentation de tous les paramètres articulatoires, excepté le pincement avalé. Cette augmentation de l'amplitude des paramètres articulatoires que nous nommerons par la suite "hyper-articulation" est globalement plus marquée dans le bruit cocktail que dans le bruit blanc.

Table 1 : Évolution de l'amplitude des mouvements articulatoires dans les deux environnements bruyants.

		Δ dans bb	Δ dans cktl	Δ cktl-Δbb
A	max	+3,4% ***	+6,5% ***	+3% **
	glob	+8,9% ***	+14,2% ***	+5,3% ***
B	max	+19,5% ***	+27,5% ***	+7,9% ***
	glob	+16,0% ***	+27,1% ***	+11,2% ***
S	max	+28,4% ***	+42,3% ***	+13,9% ***
	glob	+25,3% ***	+ 43,6% ***	+18,3% ***
Pincement protrus	max	-16,3% ***	-11,0% *	+27,3% ***
	glob	-36,7% ***	-10,2% *	+26,5% ***
Pincement avalé	max	+83,0% **	+95,5% **	+12,5% -
	glob	+149,2% **	+141,1% **	-8,1% ns

Pics de vitesse des mouvements articulatoires

Les pics de vitesse (Vmax) augmentent significativement dans le bruit pour A, B et S, quel que soit le type de bruit, et diminue significativement pour le pincement protrus dans le bruit blanc (cf. Table 2). On observe une différence significative entre les deux types de bruit au niveau de l'augmentation des pics de vitesse de B, S et du pincement protrus. Les pics de vitesse augmentent davantage dans le bruit cocktail que dans le bruit blanc. Par contre, on n'observe pas d'effet du type de bruit sur les pics de vitesse de A ni sur ceux du pincement avalé. Les pics de vitesse du mouvement peuvent être reliés à la notion d'effort articulatoire [10].

Table 2 : Évolution des pics de vitesse des mouvements articulatoires dans les deux environnements bruyants.

		Δ dans bb	Δ dans cktl	Δ cktl-Δbb
Vmax de A		+12,7% **	+13,9% ***	+1,2% ns
Vmax de B		+14,9% ***	+29,9% ***	+15,0% ***
Vmax de S		+17,8% ***	+38,5% ***	+20,7% ***
Vmax du Pincement	protrus	-16,5% ***	+3,3% ns	+19,8% **
	avalé	+120,1% **	+127,9% **	+7,8% ns

3.2. Évolution des paramètres acoustiques dans le bruit

Sur les énoncés

La table 3 présente les valeurs moyennes des paramètres acoustiques sur la globalité des énoncés. Les résultats acoustiques obtenus sont cohérents avec les études antérieures sur la parole Lombard [1-3] : l'intensité et la fréquence fondamentale augmentent significativement quel que soit le type de bruit, de même que la durée des énoncés (cf. Table 3). On observe une diminution conséquente de Oq. La différence d'énergie entre le maximum spectral de la zone 80-1500Hz (1ers harmoniques de la voix) et le maximum spectral de la zone 1500-3500Hz augmente significativement dans le bruit. Il en va de même pour la différence d'énergie spectrale entre la zone 80-1500Hz et la zone 3500-5500Hz.

Table 3 : Évolution des paramètres acoustiques dans les deux environnements bruyants.

		Δ dans bb	Δ dans cktl	Δ cktl-Δbb
Intensité		+12,9dB ***	+8,6dB ***	-4,3dB ***
F0		+55,5Hz ***	+65Hz ***	+9,5Hz ***
Énergie spectrale	1500-3500Hz	+23dB ***	+19dB ***	-4dB ***
	3500-5500Hz	+27dB ***	+19dB ***	-8dB ***
Oq		-0,123 ***	-0,118 ***	+0,005 ns
Durée		+11% ***	+8.2% ***	-2,8% *

Le type de bruit a un effet significatif sur l'évolution des paramètres acoustiques, à l'exception de Oq. L'augmentation de l'intensité, de l'énergie spectrale et de la durée des mots est plus importante dans le bruit blanc que dans le bruit cocktail. Par contre, l'élévation de la fréquence fondamentale est plus importante dans le bruit cocktail que dans le bruit blanc.

Sur les voyelles et les consonnes

L'intensité des voyelles augmente davantage dans le bruit que celle des consonnes (cf. Table 4). Par contre, le renforcement de l'énergie spectrale n'est pas significativement différent dans le bruit pour les voyelles et les consonnes sonores du corpus. Les voyelles sont significativement allongées dans le bruit tandis que les consonnes ont tendance à être raccourcies, conformément aux observations de [1-3]. L'augmentation de l'intensité et de l'énergie spectrale dans la zone 3500-5500Hz est plus importante dans le bruit blanc que dans le cocktail, que ce soit sur les voyelles ou les consonnes. Les voyelles sont également plus allongées dans le bruit blanc. Par contre, l'effet du type de bruit sur le renforcement de l'énergie spectrale dans la zone 1500-3500Hz n'est significatif que pour les consonnes, avec un renforcement plus important dans le bruit blanc que dans le bruit cocktail.

Table 4 : Évolution des paramètres acoustiques dans le bruit pour les voyelles et les consonnes.

	Intensité	Énergie spectrale		Durée
		1500-3500Hz	3500-5500Hz	
ΔVoyelles dans bb	+12,9dB ***	+17,7dB ***	+24,1dB ***	25,2% ***
ΔVoyelles dans cktl	+8,4dB ***	+17,4dB ***	+13,3dB ***	16% ***
Différence cktl-bb pour ΔVoyelles	-4,5dB ***	-0,3dB ns	-10,8dB ***	-9,2% ***
ΔConsonnes dans bb	+8,5dB ***	+21,6dB ***	+28,2dB ***	-5,7% ns
ΔConsonnes dans cktl	+3,5dB ***	+19,6dB ***	+11,8dB ***	-3,6% ns
Différence cktl-bb pour ΔConsonnes	-5,0dB ***	-2,0% ***	-16,4dB ***	+2,1% ns
Différence ΔVoyelles - ΔCons	+4,6dB ***	+ 2,9dB ns	+1,3dB ns	+25,2% ***

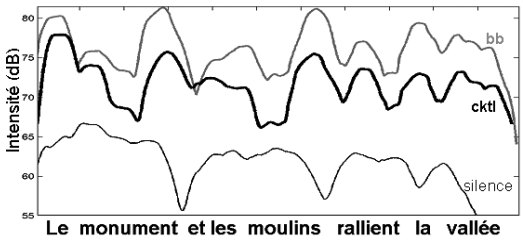


Figure 3 : Intensité d'un énoncé dans le silence, le bruit blanc (bb) et le bruit cocktail (cktl).

Comme attendu, les voyelles sont plus fortes que les consonnes dans le silence (effet de 4,5dB, p<0,001), mais l'intensité suit globalement les constituants SVO, avec une chute marquée au niveau de leur frontière. Dans le bruit, on constate un renforcement de la dynamique des "lobes" d'intensité par syllabe (+4dB, p<0,001), comme illustré Figure 3, pouvant contribuer à une perception plus hâchée de la parole.

Sur les trois voyelles /u/, /i/, et /a/

Nous avons ensuite examiné l'évolution dans le bruit des trois voyelles /u/, /i/, et /a/, formant les extrema du triangle vocalique. Seules l'intensité et l'énergie spectrale évoluent de façon différente dans le bruit pour ces trois voyelles (cf. Table 5). L'évolution de la durée et de Oq ne dépend pas significativement de la catégorie phonétique.

Table 5 : Évolution des paramètres acoustiques dans le bruit pour les 3 voyelles /a/, /i/ et /u/. La dernière colonne du tableau représente la moyenne quadratique des différences entre /u/ et /i/, entre /u/ et /a/ et entre /i/ et /a/.

		Δ / u / dans le bruit	Δ / i / dans le bruit	Δ / a / dans le bruit	Effet du type de voyelle
Intensité		+9,6dB ***	+8,9dB ***	+14,6dB ***	4,5dB ***
Énergie spectrale	1500-3500Hz	+18,4dB ***	+20,5dB ***	+11,4dB ***	7,0dB ***
	3500-5500Hz	+21,5dB ***	+19,2dB ***	+13,8dB ***	5,7dB **
Oq		-0,111 ***	-0,115 ***	-0,092 ***	0,017 ns
Durée		26,7% ***	23,2% ***	14,3% **	9,2% ns

L'observation du triangle vocalique montre une tendance globale d'augmentation du premier formant dans le bruit, en particulier pour la voyelle ouverte /a/ et la voyelle protruse /u/ (cf. Figure4). Ce résultat est en accord avec les observations de Van Summers et al., 1988 [3] et de Rostolland [17]. Le deuxième formant est sensiblement modifié pour le /i/ et le /a/ et augmente considérablement dans le bruit pour le /u/ . L'augmentation de F1 est indépendante de l'augmentation de la F0 (r=0,22 dans le bruit blanc et r=0,014 dans le bruit cocktail).

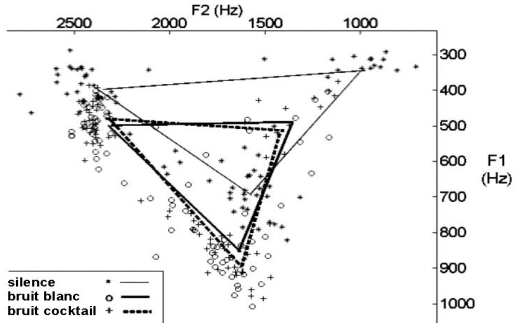


Figure 4 : Évolution du triangle vocalique dans le bruit blanc et le bruit cocktail. (Les sommets des triangles correspondent aux barycentres des /a/, /i/ et /u/ de chaque condition, dans le plan F1/F2).

3.3. Corrélations entre paramètres acoustiques et articulatoires

Sur la globalité des énoncés, on observe une très forte corrélation entre l'amplitude maximale de A et B (r=0,91) ainsi qu'entre les pics de vitesse de A et B (r=0,70). La fréquence fondamentale et l'énergie spectrale sont corrélées avec les maxima de B (r=0,66 et r=0,70) davantage que l'intensité avec les maxima de B (r=0,59). La durée est très peu corrélée avec les paramètres acoustiques et articulatoires (|r|<0,15).

4. DISCUSSION ET CONCLUSION

On observe une hyper-articulation en environnement bruyant, portant davantage sur l'ouverture et l'étirement des lèvres que sur leur pincement. Elle concerne autant l'amplitude des mouvements articulatoires que leurs pics de vitesse, pouvant être reliés à la force articulatoire. On observe un renforcement conjoint des paramètres acoustiques, qui confirment les observations d'études antérieures [1-3].

Le type de bruit a un effet significatif sur les paramètres acoustiques (augmentation plus importante dans le bruit blanc que dans le bruit cocktail, excepté pour la F0). Cette influence du contenu spectral a déjà été évoqué ou mise en évidence dans d'autres études [2,14]. Il a également un effet sur les paramètres articulatoires (augmentation plus importante dans le bruit cocktail que dans le bruit blanc). Par ailleurs, l'énergie spectrale est le paramètre de plus fort taux d'augmentation dans le bruit. L'ouverture des lèvres est davantage corrélée à l'énergie spectrale qu'à l'intensité de la voix. Ces différents points argumentent plutôt en faveur d'une adaptation acoustique et articulatoire dans le bruit cherchant à augmenter l'émergence spectrale de la voix plutôt qu'à augmenter prioritairement le rapport voix sur bruit. On observe également que l'adaptation acoustique favorise les voyelles dans le bruit par rapport aux consonnes. Cela pourrait sembler logique pour des consonnes non sonores, risquant d'être masquées par le bruit environnant. Le fait que l'on observe cette adaptation également pour des consonnes sonores pourrait corroborer l'hypothèse d'une intelligibilité dans le bruit reposant sur des patrons vocaliques, proposée par Dohalska [16]. Il est à noter que des observations inverses ont été faites en parole claire [5]. Enfin, les premières observations articulatoires et formantiques ne semblent pas aller dans le sens d'une spécialisation des voyelles en leur lieu d'articulation, à part pour les voyelles ouvertes. Cela pourrait expliquer l'intelligibilité dégradée de la voix criée observée par Rostolland [17]. Dans le but d'explorer plus rigoureusement l'hyper-articulation des différents types de voyelles, un deuxième corpus contrôlant la coarticulation vient d'être enregistré. Il permettra d'analyser la protrusion en plus des paramètres d'étirement et d'ouverture. En outre, nous avons enregistré plusieurs locuteurs pour vérifier si les résultats de cette présente étude peuvent être généralisés. Il serait également intéressant de prolonger cette exploration à l'analyse des mouvements linguaux pour déterminer s'il existe une compensation au niveau de la langue, notamment pour les voyelles protruses.

5. REMERCIEMENTS

Nous tenons à remercier Jacques Poitevineau, Nathalie Henrich, Christophe Savariaux, Alain Arnal, Aude Noiray, Claire Lalevée et Coriandre Vilain pour leur contribution à cette étude. La participation de Pauline Welby à cette recherche a été soutenue par une bourse internationale Marie Curie du 6th European Community Framework Programme.

6. BIBLIOGRAPHIE

[1] A. Castellanos, J. M. Benedi et F. Casacuberta. An analysis of general acoustic-phonetic features for Spanish speech produced with the Lombard effect. *Speech Comm.,* 20 : 23–35, 1996.

[2] J. C. Junqua. The Lombard reflex and it role on human listener and automatic speech recognisers. *JASA,* 93 : 510–524, 1993.

[3] W. Van Summers, D. B. Pisoni, R. H. Bernacki et R. I. Pedlow, M. A. Stokes. Effect of noise on speech production: acoustic and perceptual analyses. *JASA,* 84 : 912–928, 1988.

[4] M. Garnier, N. Henrich, D. Dubois et J. D. Polack. Est-il valide de considérer l'effet Lombard comme un phénomène linéaire en fonction du niveau de bruit ? *Actes du 8ème Congrès Français d'Acoustique, Tours,* 2006.

[5] B. Lindblom. Speech transforms. *Speech Comm*., 11 : 357–368, 1992.

[6] H. Lœvenbruck. Effets articulatoires de l'emphase contrastive sur la Phrase Accentuelle en français. *Actes des XXIIIèmes JEP*, pages 165–168, 2000.

[7] R. Schulman. Articulatory dynamics of loud and normal speech. *JASA,* 85 : 295–312, 1988.

[8] J. C. Junqua. Acoustic and production pilot studies of speech vowels produced in noise. *Actes de ICSLP*, pages 811–814, 1992.

[9] J. Kim, C. Davis, G. Vignali et H. Hill. A visual concomitant of the Lombard reflex. *Actes de Auditory-Visual Speech Processing 2005*, pages 17–21, 2005.

[10] L. Bailly. Étude articulatoire de la parole produite en environnement bruyant. *Mémoire de DEA ATIAM*, *université Paris 6,* 2005.

[11] M. T. Lallouache, Un poste "Visage-Parole" couleur. Acquisition et traitement automatique des contours des lèvres. *Thèse de doctorat, INPG*, 1991.

[12] M. Dohen. Deixis prosodique multisensorielle : production et perception audiovisuelle de la focalisation contrastive en français. Thèse de doctorat, INPG, 2005.

[13] J. Zeiliger. BD_Bruit, une base de données de parole de locuteurs soumis à du bruit. *Actes des Xèmes JEP,* 287–290, 1994.

[14] S. Ternström, M. Sodersten et M. Bohman, Cancellation of simulated environmental noise as a tool. *J. Voice,* 16 : 195–206, 2002.

[15] W. L. Nelson, Physical principles for economies of skilled movements. *Biol. Cybern.,* 46 : 135–147, 1983.

[16] M. Dohalska et J. Mejvaldova. Rôle de la prosodie dans la communication en milieu bruité. *Actes des Xèmes JEP,* pages 265–268, 2000.

[17] D. Rostolland, Phonetic structure of shouted voice. *Acustica,* 51 : 80–89, 1982.