An Acoustic and Articulatory Study of Lombard Speech: Global Effects on the Utterance

Maëva Garnier[1], Lucie Bailly[1,2], Marion Dohen[2,3], Pauline Welby[2], Hélène Lœvenbruck[2]

[1]Laboratoire d'Acoustique Musicale, CNRS UMR 7604, UMPC, Ministère de la culture, Paris, France
[2]Institut de la Communication Parlée, CNRS UMR 5009, INPG, Univ. Stendhal, Grenoble, France
[3]ATR – Cognitive Information Science Laboratories, Kyoto, Japan
garnier@lam.jussieu.fr, {lucie.bailly, pauline.welby, helene.loevenbruck}@icp.inpg.fr, marion.dohen@gmail.com

Abstract

This study aims at characterizing the acoustic and 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 short sentences with a subject-verb-object (SVO) structure. The sentences were recorded in three conditions: silence, 85dB white noise, and 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 shows that this speaker especially adapts to noise not only by talking louder or increasing vowel recognition cues but also by increasing spectral emergence.

Index Terms: Lombard speech, noise, acoustic, articulation.

1. Introduction

Communication in noisy environments is disturbed for two main reasons: speakers get attenuated feedback of their own voices, and their intelligibility is decreased for the listener. The speech compensatory adaptation to these perturbations is called the Lombard effect, and has been mainly described from acoustic and phonetic points of view [1-4]. In this study, we aim at characterizing the articulatory gestures that accompany these acoustic modifications in noise and at comparing them to speech produced in silence. Is there a hyper-articulation? Which articulatory parameters are concerned by this speech adaptation? Are the articulatory and acoustic modifications correlated? Lastly, does the articulatory adaptation depend on the type of noise? Some studies have already described articulation of "hyper" types of speech [5-6] but few specifically examine Lombard speech [7-8].

2. Methods

2.1. Corpus

The corpus consisted of 33 short sentences with a subject-verb-object (SVO) structure. They consisted of CV syllables (see (1)) and voiced consonants were chosen in order to minimize the segmental perturbations of the intonation contour.

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

'The monument and the mills are rallying the valley.'

2.2. Audiovisual recordings

We simultaneously recorded audio, electroglottographic (EGG) and articulatory signals from a female native speaker of French. She read sentences to a person standing two meters in front of her. Articulatory data were extracted from video recordings (25 images/s) of the speaker's lips, using a labiometric device developed at the Institut de la Communication Parlée [9]. In this study, we focused on the analysis of lip **spreading (A),** lip **aperture (B)**, and **inter-lip area (S)** (see Figure 1). For these three parameters, we examined the amplitude maxima of the articulatory movements (**max**), as well as their more global evolution (**glob**), which corresponds to the integral over a normalized time period [9, pp. 113- 114].

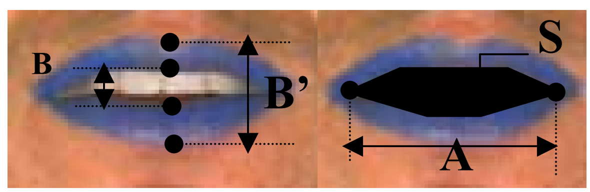

Figure 1. *Articulatory parameters.*

We also analyzed **lip pinching**, defined as lip compression when the mouth is closed for [m] segments. It is equivalent to B' when B is equal to zero (see Figure 1). Two types of lip pinching could be observed: **protruded lip pinching**, i.e. when the visible labial surface increases with lip compression, and **swallowed lip pinching**, i.e. when the lips flatten and move towards the inside of the mouth, resulting in a decrease of the visible labial surface.

The audio signal was recorded with an AKG microphone placed 20cm away from the lips. The EGG signal was recorded with an electroglottograph EG2. Both signals were digitized at a rate of 44.1kHz, over 16bits. Two noisy environments were used, both extracted from the BD_Bruit database [11]: white noise **(wn)** and cocktail party noise **(cktl)**. They were played over two loudspeakers located 2m away from the speaker and 2m away from each other. The noise level was calibrated to 85dB at the participant's ears. The speaker was first recorded in a silent reference condition, and then in both noisy environments. Noise was removed from the acoustic signal using the method designed by [12]. The utterance boundaries were then labelled using Praat [13], as well as word and segment boundaries. Fundamental frequency **(F0)** and open quotient **(Oq)** were extracted from the EGG signal using an autocorrelation method.

3. Results

The following notation was adopted for indicating statistical significance: * $p < .05$, ** $p < .01$, *** $p < .001$, and *ns* (not significant) $p > .05$.

3.1. Articulation of the utterances in noise

3.1.1. Amplitude of the articulatory movements

A significant increase of A, B and S was observed in noisy conditions for both types of noises. Protruded lip pinching decreased significantly in white noise (see Table 1). The type of noise had a significant effect on the increase of all articulatory parameters, except swallowed lip pinching. This increase of the articulatory parameters (henceforth called "hyper-articulation") was globally greater in cocktail party noise than in white noise.

Table 1. *Evolution of the amplitude of the articulatory movements in the two noisy environments.*

		Δin wn	Δin cktl	Effect of the type of noise. Δcktl-Δwn
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%** ***
Protruded pinching	**max**	**-16.3%** ***	**-11.0%** *	**+27.3%** ***
	glob	**-36.7%** ***	**-10.2%** *	**+26.5%** ***
Swallowed pinching	**max**	**+83.0%** **	**+95.5%** **	+12.5% ns
	glob	**+149.2%** **	**+141.1%** **	-8.1% ns

3.1.2. Velocity peaks of the articulatory movements

We observed a significant increase in the velocity peaks **(Vmax)** of A, B and S for both types of noise. The protruded lip pinching velocity peaks decreased significantly in white noise (see Table 2). The velocity peaks of B, S and protruded lip pinching increased significantly more in cocktail party noise than in white noise. However, the type of noise had no significant effect on the increase in the velocity peaks of A and swallowed lip pinching. Peak velocity has been related to the notion of articulatory effort [14].

Table 2. *Evolution of the articulatory velocity peaks in the two noisy environments.*

		Δ in wn	Δ in cktl	Effect of the type of noise. Δcktl-Δwn
Vmax of A		**+12.7%** **	**+13.9%** ***	+1.2% ns
Vmax of B		**+14.9%** ***	**+29.9%** ***	**+15.0%** ***
Vmax of S		**+17.8%** ***	**+38.5%** ***	**+20.7%** ***
Vmax of	**protruded pinching**	**-16.5%** ***	**+3.3%** ns	**+19.8%** **
	swallowed pinching	**+120.1%** **	**+127.9%** **	+7.8% ns

3.2. Evolution of acoustic parameters in noise

3.2.1. Over the whole utterance

Table 3 presents the mean values of the acoustic parameters over the whole utterance for all utterances. The acoustic results confirmed those obtained in previous studies of Lombard speech [1-3]: voice intensity, F0 and word duration increased significantly for both types of noise (see Table 3). We also observed a large decrease of Oq in noise, that can be related to a more pressed voice quality, and an increase in the difference between the spectrum maximum in the 80-1500Hz frequency band (first harmonics) and the spectrum maximum in the 1500-3500Hz frequency band. A comparable result was obtained for the energy difference between the 80-1500Hz frequency band and the 3500-5500Hz frequency band.

Table 3. *Evolution of the acoustic parameters in the two noisy environments.*

		Δ in wn	Δ in cktl	Effect of the type of noise. Δcktl-Δwn
Intensity		**+12.9dB** ***	**+8.6dB** ***	**-4.3dB** ***
F0		**+55.5Hz** ***	**+65Hz** ***	**+9.5Hz** ***
Spectral energy	**1500-3500Hz**	**+23dB** ***	**+19dB** ***	**-4dB** ***
	3500-5500Hz	**+27dB** ***	**+19dB** ***	**-8dB** ***
Oq		**-0.123** ***	**-0.118** ***	+0.005 ns
Word duration		**+11%** ***	**+8.2%** ***	**-2.8%** *

The type of noise had a significant effect on the evolution of the acoustic parameters in noisy conditions, except for Oq. The increase of voice intensity, spectral energy, and word duration was greater in white noise than in cocktail party noise. On the contrary, mean F0 increased more in cocktail party noise than in white noise.

3.2.2. For vowels and consonants

Vowel intensity increased more in noisy conditions than did consonant intensity (see Table 4). Vowel duration increased significantly in noise whereas consonant duration tended to decrease, as it has already been observed ([1-3]). On the contrary, spectral energy did not significantly increase more in noise for vowels than for consonants.

Vowel duration increased more in white noise than in cocktail party noise. The type of noise had a significant effect on the increase of spectral energy in the 3500-5500Hz frequency band, and for consonants, on the increase of spectral energy in the 1500-3500Hz frequency band. The reinforcement is greater in white noise than in cocktail party noise.

Table 4. *Evolution of the acoustic parameters in noisy conditions for vowels and consonants.*

	Intensity	Spectral energy		Word duration
		1500-3500Hz	3500-5500Hz	
ΔVowels in wn	+12.9dB ***	+17.7dB ***	+24.1dB ***	25.2% ***
ΔVowels in cktl	+8.4dB ***	+17.4dB ***	+13.3dB ***	16% ***
Difference cktl-wn for ΔVowels	-4.5dB ***	-0.3dB ns	-10.8dB ***	-9.2% ***
ΔConsonants in wn	+8.5dB ***	+21.6dB ***	+28.2dB ***	-5.7% ns
ΔConsonants in cktl	+3.5dB ***	+19.6dB ***	+11.8dB ***	-3.6% ns
Difference cktl-wn for ΔConsonants	-5.0dB ***	-2.0% ***	-16.4dB ***	+2.1% ns
Difference ΔVowels - ΔConsonants	+4.6dB ***	+ 2.9dB ns	+1.3dB ns	+25.2% ***

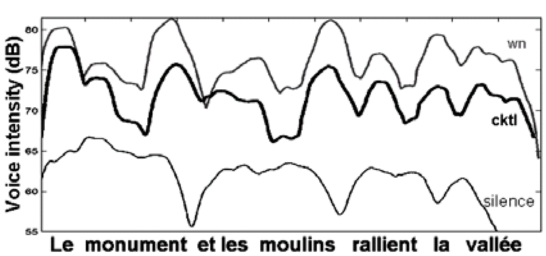

Figure 2. *Voice intensity of an utterance in silence, white noise (wn) and cocktail party noise (cktl).*

As expected, vowels were louder than consonants in silence (with an average increase of 4.5dB, $p < .001$). We especially observed that voice intensity followed the prosodic phrasing of the utterance, falling at prosodic phrase boundaries. In noisy conditions, we noticed that the dynamic of the voice intensity "lobes" per syllable increased significantly (+4dB, $p<.001$). This could contribute to the sometimes "choppy" nature of Lombard speech (see Figure 2).

3.2.3. For the vowels /u/, /i/ and /a/

Table 5. *Evolution of the acoustic parameters in noisy conditions for the vowels* /a/, /i/ *and* /u/.

		Δ / u / in noise	Δ / i / in noise	Δ / a / in noise	Effect of the vowel type
Intensity		+9.6dB ***	+8.9dB ***	+14.6dB ***	4.5dB ***
Spectral energy	**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
Word duration		26.7% ***	23.2% ***	14.3% **	9.2% ns

Table 5 presents the mean values of the acoustic parameters over all the /a/, /i/ and /u/ segments of the corpus. These vowels correspond to the extrema of the vowel triangle in the F1/F2 plane. Voice intensity and spectral energy varied differently for these three vowels (see Table 5). On the contrary, the evolution of word duration and Oq did not significantly depend on the vowel considered.

Figure 3 shows that F1 tended to increase in noisy conditions, particularly for the open vowel /a/ and the rounded vowel /u/. This result confirms the observations made by Van Summers et al., 1988 for one of their speakers [3], and by Rostolland for shouted speech [16]. The second formant is slightly modified for /i/ and /a/ and increases considerably in noisy conditions for /u/.

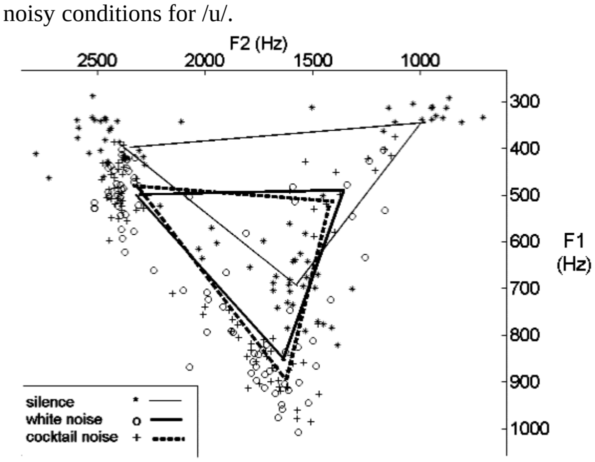

Figure 3. *Modification of the vowel system in white noise and in cocktail party noise. (The corners of the triangles correspond to the barycenters of* /a/, /i/ *and* /u/ *for each condition, in the F1/F2 plane).*

3.3. Correlations between acoustic and articulatory parameters

We observed a strong correlation between the maximum amplitudes of A and B over the whole utterance (r = .91), and also a strong correlation between their velocity peaks (r = .70). F0 and the spectral energy are correlated with maximal amplitude of B (r = .66 and r = .70 respectively). This correlation is stronger than the one between voice intensity and maximum amplitude of B (r = .59). Word duration is weakly correlated with other acoustic and articulatory parameters (|r| < .15).

4. Discussion and Conclusions

Hyper-articulation was observed when the speaker spoke in noisy environments. This amplification of the articulatory movements concerned lip aperture and spreading rather than lip pinching, not only for the movement amplitude, but also for the velocity peaks, which can be related to articulatory effort. We observed an associated evolution of the acoustic parameters that confirmed the observations of earlier studies [1-3]. Type of noise had a significant effect on the evolution of the acoustic parameters (the increase was greater in white noise than in cocktail party noise, except for F0). An influence of the spectral characteristics of noise has already been noticed in previous studies [2,12]. We observed here that it also had a significant effect on the evolution of the articulatory parameters (the increase was more important in cocktail party noise than in white noise). In addition, spectral energy was the parameter which increased the most. Lip aperture was more correlated with spectral energy than with voice intensity. Therefore it seems that acoustic and articulatory adaptations in noise include not only talking louder but also specifically increasing spectral emergence.

As it has already been noticed in previous studies, acoustic adaptation in noisy conditions favors vowels rather than consonants in general. This result is understandable for voiceless consonants, which are expected to be masked by the ambient noise. However, it is also observed for voiced consonants. This could be in line with the hypothesis of Dohalska [15], who suggested that vowels are used as “patterns” to help the listener reconstitute the intelligibility of the message. However, we note that the opposite observation has been made for clear speech, for which consonants are reinforced in comparison to vowels [5].

The observation of both articulatory and acoustic (formant values) data show that hyper-articulation does not seem to reinforce vowel recognition cues, except for the open vowel /a/. This could explain the decrease in intelligibility noticed for shouted speech [16]. A new corpus controlling for coarticulation has recently been recorded in order to rigorously explore the hyper-articulation corresponding to different vowels. This corpus will also enable us to additionally analyze lip protrusion. Several speakers have been recorded in order to analyze whether the results of the present study can be generalized. It would also be interesting to go further and explore whether there exists compensatory tongue movements for the increased aperture of the lips, especially for rounded vowels. Compensation strategies have indeed been observed for the vowel [u] produced with a lip-tube (increased lip aperture) [17].

5. Acknowledgments

We would like to thank N. Henrich, C. Savariaux, A. Arnal, A. Noiray, J. Poitevineau, C. Lalevée, and C. Vilain for their contribution to the project. Pauline Welby’s participation was supported by a Marie Curie International Fellowship within the 6th European Community Framework Programme.

6. References

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

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

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

[4] Garnier, M., Henrich, N., Dubois, D., and J. D. Polack. "Est-il valide de considérer l’effect Lombard comme un phénomène linéaire en fonction du niveau de bruit ?", Proc. Congrès Français d'Acoustique, 2006.

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

[6] Schulman, R. "Articulatory dynamics of loud and normal speech", J. Acoust. Soc. Amer., 85: 295–312, 1988.

[7] Junqua, J. C. "Acoustic and production pilot studies of speech vowels produced in noise", Proc. ICSLP: 811–814, 1992.

[8] Kim, J., Davis, C., Vignali, G., and Hill, H. "A visual concomitant of the Lombard reflex", Proc. Listenery-Visual Speech Processing 2005: 17–21, 2005.

[9] Lallouache, M. T. "Un poste "Visage-Parole" couleur. Acquisition et traitement automatique des contours des lèvres", Doctoral dissertation, Institut National Polytechnique de Grenoble, 1991.

[10] Dohen, M. "Deixis prosodique multisensorielle: production et perception audiovisuelle de la focalisation contrastive en français", Doctoral dissertation, Institut National Polytechnique de Grenoble, 2005.

[11] Zeiliger, J. "BD_Bruit, une base de données de parole de locuteurs soumis à du bruit", Proc. Xèmes Journées d’Étude sur la Parole: 287–290, 1994.

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

[13] Boersma, P. and Weenink, D. "Praat: doing phonetics by computer (Version 4.2.28) [Computer program]". Retrieved from http://www.praat.org/, 2004.

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

[15] Dohalska, M. and Mejvaldova, J. "Rôle de la prosodie dans la communication en milieu bruité", Proc. Xèmes Journées d’Étude sur la Parole: 265–268, 2000.

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

[17] Savariaux, C., Perrier, P. and Orliaguet, J.P., "Compensation strategies for the perturbation of the rounded vowel [u] using a lip tube: A study of the control space in speech production", J. Acoust. Soc. Amer., 98 (5): 2428-2442, 1995.