The Lombard Effect: a physiological reflex or a controlled intelligibility enhancement?

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

[1]Laboratoire d'Acoustique Musicale (CNRS UMR 7604, UMPC, Ministère de la culture)
[2]Institut de la Communication Parlée (CNRS UMR 5009, INPG, Université Stendhal)

garnier@lam.jussieu.fr,{Marion.Dohen, Helene.Loevenbruck,welby, Lucie.Bailly}@icp.inpg.fr

Abstract. *The purpose of this study was to examine whether speech in noisy environments consists of global acoustic and articulatory modifications or if there are some changes specific to units within the utterance. Changes on a more local level could be interpreted as a controlled intelligibility enhancement of specific speech cues such as cues to word segmentation or prosodic phrasing. Audio and video signals were recorded for a female native speaker of French in three conditions: silence, 85dB white noise, and 85dB "cocktail party" noise. The corpus consisted of 33 short sentences with a subject-verb-object (SVO) structure. Labial parameters were extracted from the video data. A controlled intelligibility enhancement in noise was observed for some cues to word segmentation and utterance structure.*

1. Introduction

1.1. Background

Speech production is influenced by the immersion of the speaker in a noisy environment. Speech changes in noise are collectively called the Lombard effect and include an increase in vocal intensity, fundamental frequency (F_0), and word duration (Lombard, 1911; Junqua, 1993), as well as an increase in amplitude of articulatory movements (Garnier et al., 2006; Davis, 2006). A number of perception studies have found that speech produced in noisy conditions is more intelligible than speech produced in silence (see, for example, Junqua, 1993; there are however, limits: past a certain point, shouting decreases intelligibility, Rostolland, 1982).

This increased intelligibility raises the question of whether speech changes in noisy environments are motivated by increased intelligibility. This question is part of a larger question on the physiological and cognitive mechanisms underlying the Lombard effect. Communication in noisy environments might be disturbed for at least two reasons: speakers get attenuated feedback of their own voices, and their intelligibility is decreased for listeners. Two main interpretations of the Lombard effect have been proposed. The first argues that the effect is a physiological audio-phonatory reflex (Lombard, 1911), the second that Lombard changes are motivated by compensation on the part of the speaker for decreased intelligibility (Lane and Tranel, 1971). Some authors have also argued that both mechanisms may contribute to the changes made by the speaker in noisy environments (Junqua, 1993).

A number of arguments have been put forth in favor of these different interpretations: On the one hand, speakers cannot totally inhibit speech changes in noise (Pick et al., 1989); and similar changes occur in animals (for example in monkeys, Sinnott et al., 1975). On the other hand, Lombard changes are greater in adults than in children and in spontaneous speech than in reading tasks (Lane and Tranel, 1971; Amazzi and Garber, 1982).

1.2. Aims

One aspect has not yet been examined that could bring arguments to this debate. If the Lombard effect is motivated by a search for intelligibility, Lombard changes might enhance some cues rather than others, and thus might vary over the utterance as a whole or within a smaller unit within the utterance. In this study we therefore examine some specific aspects of

speech production in French, across the utterance as a whole and within smaller units, in both quiet and noisy conditions (Condition 1: NOISE).

First, in French, as in other languages, there are prosodic differences between function words (determiners, conjunctions, etc.) and content words (nouns, verbs, etc.). (Delais-Roussarie, 1995; Welby, 2006; *inter alia*). For example, content word syllables tend to be longer, *ceteris paribus*. The evidence suggests that these differences between function words and content words help listeners in the task of word segmentation (for example, Christophe, 1993). A greater difference between the duration of function word syllables and content word syllables in noisy conditions could suggest an effort to enhance this linguistic category distinction. We therefore examined differences between syllables like [le] in the article *les* and the first and the last syllables of content words like [mu], and [nE] in *les moulinets* 'the reels', in quiet and in noise. (Condition 2: WORD TYPE).

In addition, utterance-initial syllables have been shown to be longer and over-articulated compared to other syllables (Fougeron and Keating, 1997). Similarly, syllables at the end of prosodic units are lengthened (Beckman and Edwards, 1994) and over-articulated (Tabain, 2003; Lœvenbruck, 1999), with greater lengthening associated with higher prosodic levels. This relative lengthening is considered to be a cue to word and higher-level boundaries (e.g. Rietveld, 1980; Christophe, 1993; Bagou et al., 2002). We therefore compared initial, intermediate, and final position in the utterance (Condition 3: POSITION). Duration differences at unit boundaries may be enhanced in noise, that is, the difference between the duration of initial syllables in phrase-initial position (like [mi] in *Le minet leva le nez* 'The kitty raised his nose') and the other positions and the difference between the duration of final syllables in phrase-final position (like [nE] in *Loulou nourrit le minet* 'Loulou is feeding the kitty') and the other positions.

Finally, in French, as in other languages, an F_0 declination is observed across the utterance, as is a final lowering at the very end of an utterance (in declaratives, for example). If we consider this declination as a cue to an utterance boundary, we might expect it to be enhanced in noise.

2. Methods

2.1. Background

The corpus consisted of 33 short sentences with a subject-verb-object (SVO) structure. Only CV syllables were used in order to simplify the acoustic labeling of the corpus. Four groups of 4 targets were selected, each consisting of a 2- or 3- syllable content word, with or without a preceding determiner. The 33 sentences combined these 16 targets, so that each target appeared in three positions in the utterance[1]:

Initial	(1) Loulou nourrit le minet. 'Loulou is feeding the kitty.'
Intermediate	(2) Nina et Loulou mimaient les lamas. 'Nina and Loulou were imitating the llamas.'
Final	(3) Maman ramena Loulou. 'Mom brought back Loulou.'

2.2. Audiovisual recordings

Audio and articulatory signals were simultaneously recorded for a female native speaker of Hexagonal French. The speaker read the 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 (Lallouache, 1991). In this study, we focused on the analysis of lip spreading (A), lip aperture (B), and inter-lip area (S) (see Figure 1). We then examined the mean amplitude of the articulatory movements, corresponding to the integral over a normalized time period (Dohen, 2005). We also analyzed lip pinching, defined as lip compression when the mouth is closed for [m] segments (i.e. B' when B=0, see Figure 1). The audio signal was recorded with an AKG

[1] To reduce recording time, some sentences contain more than one target word

microphone placed 20cm away from the lips and digitized at a rate of 44.1kHz, over 16bits. Two noisy environments (white noise and cocktail party noise) were used, both extracted from the BD_Bruit database (Zeiliger, 1994). 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 a method especially designed for that problem (Ternström et al., 2002). The utterance, target, and syllable boundaries were then labeled using Praat (Boersma and Weenink, 2004). The results presented here report the two types of noise considered together.

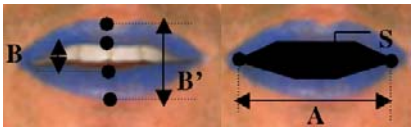

Figure 1. Articulatory parameters.

The parameter values are highly dependent on the segmental composition of the syllable studied (e.g. [mu] vs. [li]; smaller lip area (S) for rounded [u] compared to spread [i]). The following normalization procedure was applied to allow inter-syllable comparisons: for each item: parameter values in the noisy conditions were divided by the corresponding values in the silent condition. After normalization, a value of 1 corresponds to no variation between the noisy and silent conditions; a value above 1 corresponds to an increase in noise compared to the silent condition; and a value below 1 to a decrease in noise compared to silence.

3. Results

Global articulatory and acoustic effects across the utterance as a whole from the quiet to the noisy condition are described in an earlier study (Garnier et al. 2006). We focus here on effects within smaller units. Figure 2 gives an example of the variation across several parameters from (a) silence to (b) noise. Note, for example, that F_0, intensity, as well as articulatory parameters show a general increase in noise.

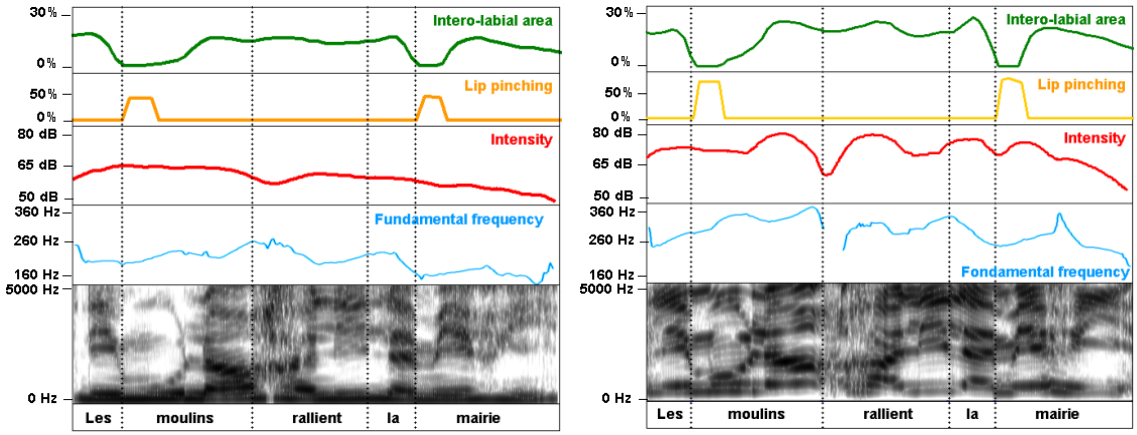

Figure 2. Several parameters for the same item produced in (a) silence and in (b) noise. From top to bottom: inter-lip area, lip pinching, intensity, fundamental frequency, time aligned with the spectrogram and segmentation into words.

As detailed in the following, analyses of variance (ANOVA) were carried out with Matlab. The following notation is adopted in reporting statistical significance:
* $p < .05$, ** $p < .01$, *** $p < .001$, and *ns* (not significant) $p > .05$.

3.1. Comparison of function and content words

The first research question we examined was the following: Is there a difference in enhancement between content word syllables and function word syllables? In order to address

this question, we considered the eight targets containing a determiner (*le*, *la* or *les*).[2] Of these, four consisted of 2-syllable content words and the other four of 3-syllable content words. We measured the values for the function word syllable, and the two syllables at the edges of the content word, i.e. the first syllable and the last syllable. Figure 3 presents the normalized values for duration and several articulatory parameters in noise relative to silence. For each normalized parameter, we conducted a one-way ANOVA (factor: SYLLABLE TYPE, three levels: function word syllable, content word initial syllable, content word final syllable).

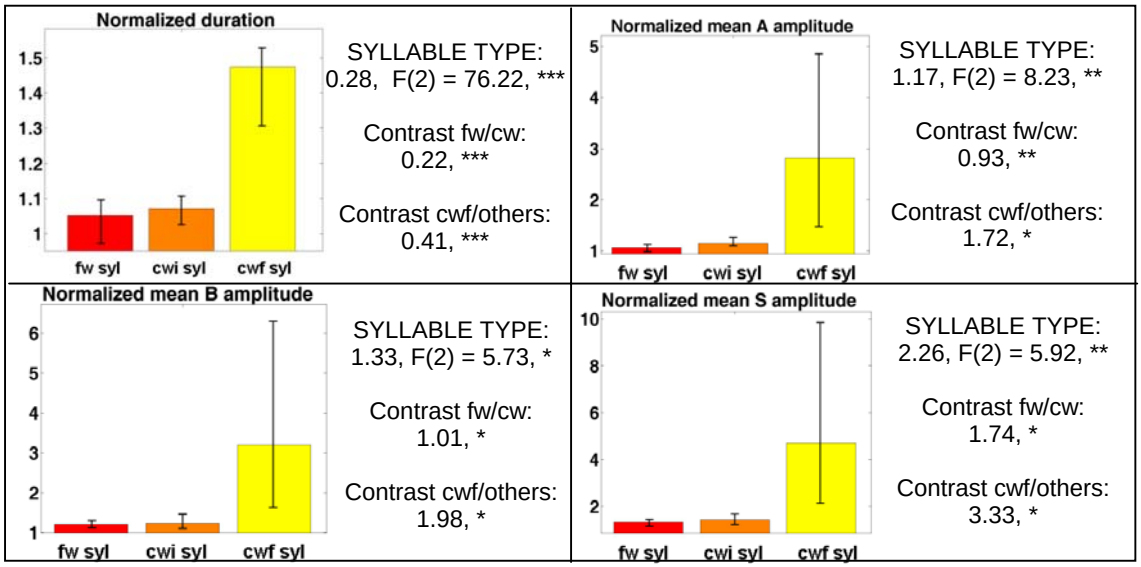

Figure 3. Normalized values for duration and several articulatory parameters in noise relative to silence. Bars represent standard deviation. *fw syl* represents the function word syllable, *cwi syl* the content word-initial syllable and *cwf syl* the content word-final syllable.

First, it is interesting to note that all normalized values are greater than 1, which means that all parameters increased in noise relative to silence, regardless of the word type (function word or content word).

Secondly, a significant effect of word type on the variation from silence to noise was found for duration and mean A, B, and S amplitude. Specifically, the durational and articulatory amplitude parameters are more enhanced in noise for content word-final syllables than for other syllables. It therefore seems, for this speaker, that content word-final syllables are more enhanced in noise than are function word syllables or content word-initial syllables. This could suggest that the speaker tried to reinforce cues to the ends of content words.

3.2. Analysis of the initial syllable of a content word depending on word position within the utterance

The second research question we examined was the following: Is there a difference in initial syllable enhancement when the content word to which the syllable belongs appears in initial, intermediate or final position? For example, is the first syllable in *Loulou* more enhanced in noise in (1), (2) or (3)? (see 2.1). In order to address this question, we considered the initial syllable of all 16 content words, each of which appears in three positions in the utterance (initial, intermediate and final). Figure 4 presents the normalized values for duration and several articulatory parameters in noise relative to silence for the word-initial syllables in the 3 positions. The results of the one-way ANOVA (factor: position, three levels: initial, intermediate, final) are reported beside each graph (label POSITION). The results labeled *contrast init/others* correspond to a comparison between the initial level and the other two levels. These results were obtained using multiple comparison tests derived from the ANOVA.

[2] Note that throughout the paper, the values reported for function words were measured for these determiners; they do not include, for example, conjunctions.

First, it is interesting to note that all normalized values, except lip pinching in non-initial positions, are greater than 1, which means that on initial syllables, all parameters increased in noise relative to silence, regardless of the position of the word to which the initial syllables belong (initial, intermediate or final).

Secondly, a significant effect of position on the variation from silence to noise was found only for the mean B amplitude. More specifically, for the content word initial syllables, only the B mean amplitude showed significantly greater increases in noise in the initial position than in the other positions. However, syllable duration, mean A amplitude, as well as mean amplitude of lip pinching, all tended to show a greater increase in noise for initial syllables of content words in utterance-initial position relative to content words in other positions.

It therefore seems that for this speaker, content word-initial syllables tend to be longer, produced with a more open articulation in noise than in silence, and especially so when the content word is in initial position in the utterance.

As mentioned in section 1.2, it has been suggested in the literature that initial syllable lengthening could be a cue to word and higher-level boundaries. The fact that the initial position in the utterance is accompanied by greater articulatory and durational increases in noise (relative to silence) suggests that the speaker might have been trying to enhance articulatory and durational cues to word segmentation and prosodic hierarchy.

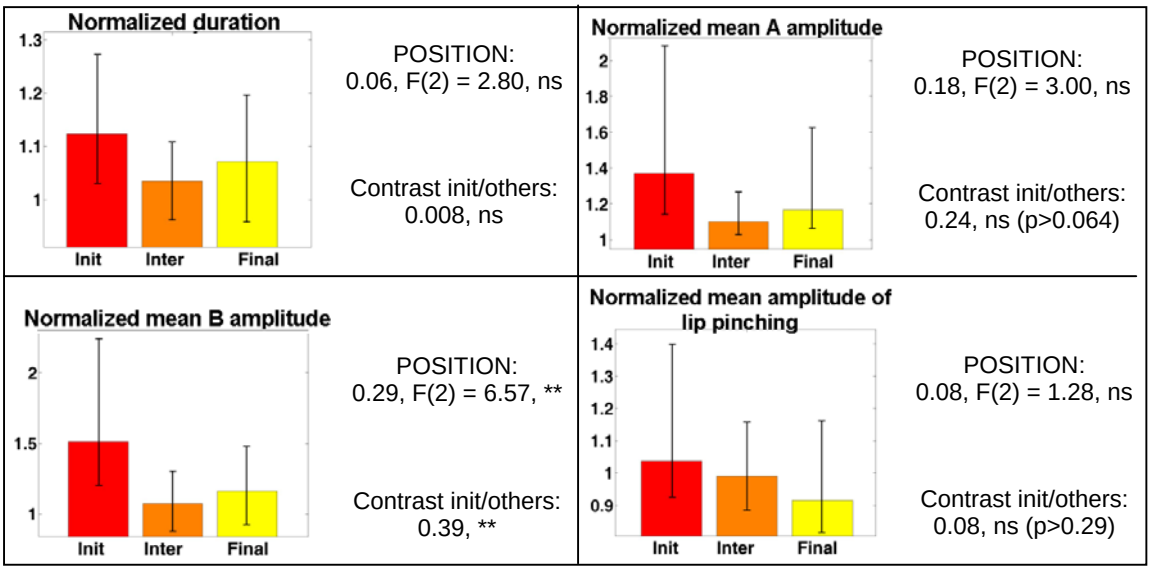

Figure 4. Normalized values for duration and several articulatory parameters in noise relative to silence for initial syllables in content words in initial, intermediate and final position in the utterance. Bars represent standard deviation.

3.3. Analysis of the final syllable of a content word depending on word position within the utterance

Next we examined the question: Is there a difference in final syllable enhancement when the content word to which the syllable belongs appears in initial, intermediate or final position? For example, is the last syllable in *Loulou* more enhanced in noise in (1), (2) or (3) (see 2.1)?

In order to address this question, we considered the initial syllable of all 16 content words, each of which appears in three positions in the utterance (initial, intermediate and final). For eight of the content words, the final syllable was the second one; for the other eight, it was the third one. In order to address this question, we considered the initial syllable of all 16 content words, each of which appears in three positions in the utterance (initial, intermediate and final). Figure 5 presents the normalized values for duration and several articulatory parameters in noise relative to silence for the word-final syllables in the three positions. The results of the one-way ANOVA (factor: position, three levels: initial, intermediate, final) are reported beside each graph(label POSITION). The results labeled *contrast init/others* correspond to a comparison between the initial level and the other two levels. These results were obtained using multiple comparison tests derived from the ANOVA.

First, it is interesting to note that all normalized values are greater than 1, which means that on final syllables, all parameters increased in noise relative to silence, regardless of the position of the word to which the final syllables belong (initial, intermediate or final).

Secondly, a significant effect of position on the variation from silence to noise was found for duration. More specifically, for the content word-final syllables, the syllable duration showed a significantly higher increase in noise in the utterance-final position than in the other positions. In addition, all articulatory parameters tend to show a greater increase in noise for final syllables of content words in utterance-final position relative to content words in other positions.

It therefore seems that for this speaker, final content word syllables tend to be longer, produced with a more open articulation in noise than in silence, particularly when the content word is in final position in the utterance. As mentioned in section 1.2, it has been suggested in the literature that final syllable lengthening could be a cue to word and higher-level boundaries. The fact that the final position in the utterance is accompanied by articulatory and durational increases in noise (relative to silence) suggests that the speaker might have been trying to enhance articulatory and durational cues to word segmentation and prosodic hierarchy.

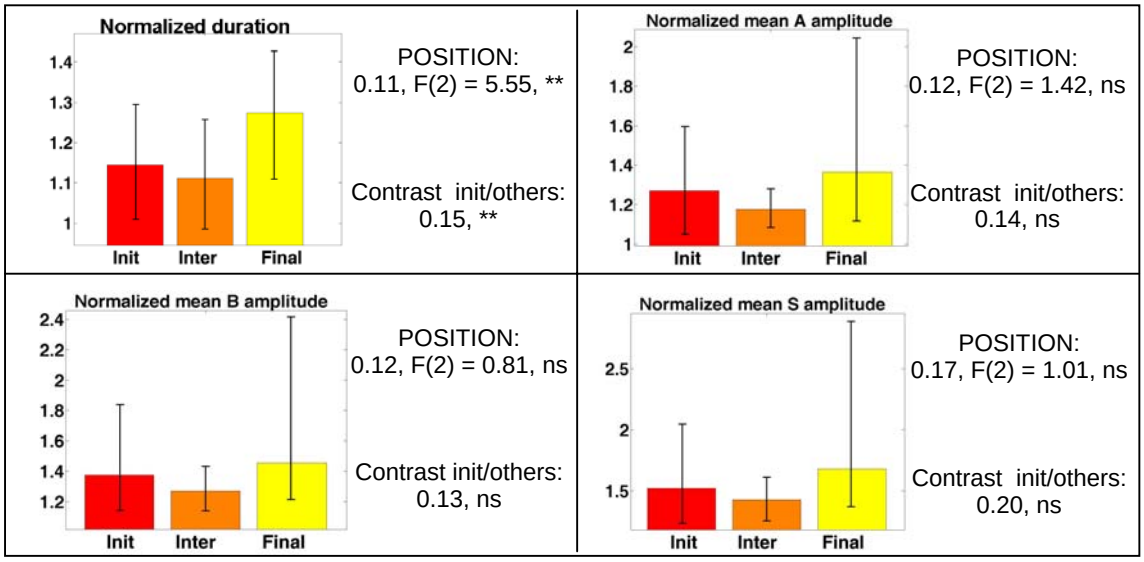

Figure 5: Normalized values for duration and several articulatory parameters in noise relative to silence for final syllables in content words in initial, intermediate and final position in the utterance. Bars represent standard deviation.

3.4. Analysis of F_0 and intensity declination over the utterance

Finally, we examined the question: Are F_0 and intensity declination across the utterance enhanced in noise? To answer this question, we considered mean F_0 and intensity of all 16 content words, each of which appears in three positions in the utterance (initial, intermediate and final). Figure 6 shows the results of the F_0 and intensity analyses.

In silent conditions, as expected, an F_0 declination utterance is observed across the (POSITION effect: *23 Hz, F(2) = 64.14, ****). This declination is still observed in noisy conditions (POSITION effect: *30 Hz, F(2) = 96.64, ****) but is not significantly enhanced. It has been previously demonstrated (Lombard, 1911; Junqua, 1993) that speech in noise is produced with a globally higher F_0 than in silence. We might suspect that the lesser declination observed here in noise could be due to the fact that F_0 values for initial and intermediate content words have reached a ceiling (cf. Rostolland, 1992). However, the standard deviation of mean F_0 does not decrease in noise, even for the initial and intermediate content words, which are produced with higher mean F_0. Thus the declination attenuation does not seem to be due to a ceiling effect. Similar observations can be made about the intensity declination across the utterance (POSITION effect*: 3.3 dB, F(2) = 90.78, ***, in silence; 2.3 dB, F(2) = 21.66, ***, in noise*).As mentioned in 1.2, F_0 declination can be considered to be a cue to an utterance boundary. An enhancement of F_0 declination or intensity declination would thus reinforce cues to utterance boundary. It seems that this speaker, however, did not particularly reinforce these

F_0 and intensity cues to utterance boundaries in noise. It may be that other cues playing the same role (such as final syllable lengthening and over-articulating) were sufficiently enhanced. F_0 value could thus be maintained to a sufficiently high level, even at the utterance end.

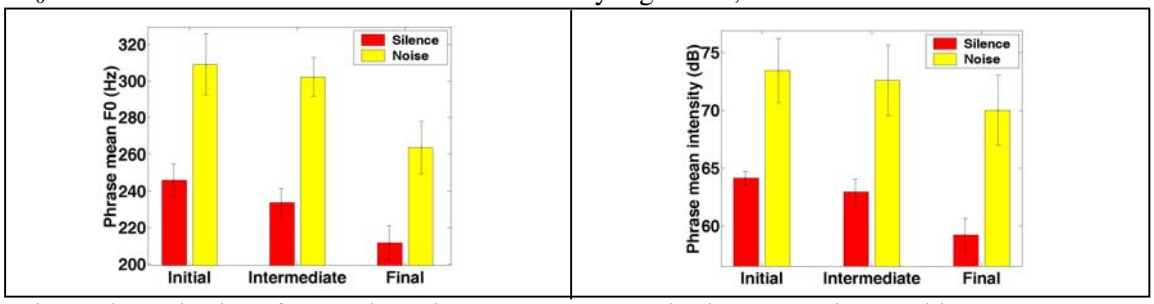

Figure 6. Declination of F_0 and intensity over the utterance in silent and noisy conditions. Bars represent standard deviation.

4. Discussion and Conclusions

In this study, we examined the variation from silence to noise of specific aspects of speech that could enhance intelligibility in noise. We conducted these analyses in order to explore if the Lombard effect affects the whole utterance in the same way or if cues to intelligibility are more enhanced for certain units of the utterance. This could shed light on the question of whether the Lombard effect is simply a physiological reflex or a controlled intelligibility enhancement.

Our previous research has shown global lengthening and over-articulation in noise (Garnier et al. 2006). We sought here to examine whether more localized effects would be found, in particular ones which might contribute to easing word segmentation or phrasing decisions. The results show that content word-final syllables are more enhanced in noise than are other syllables (increased lengthening and increased over-articulation). These changes are therefore not purely physiological in nature, since the changes from silence to noise depend on linguistic category (function word *vs.* content word). The fact that our speaker enhances content word-final syllables more than content word-initial syllables is consistent with a reinforced marking of the end of content words. We note that in French, the ends of content words (or of accentual phrases) tend to be more consistently and redundantly marked than the beginnings (see Welby 2006 and references therein). We also observed a general tendency for initial content word syllables in utterance-initial position to be accompanied by greater articulatory and durational increases in noise. This is consistent with the enhancement of articulatory and durational cues to word segmentation and prosodic phrasing. We observe similar results for absolute utterance-final syllables.

In addition, F_0 and intensity declinations across the utterance were attenuated rather than enhanced in noise. Rostolland, 1992 observed a plateau for shouted voice at the top of the intonational curve, which he interpreted as being a limitation to voice frequency modulation. Our data, however, do not show a significantly lower standard deviation corresponding to the increase of the strongest units and the weakest ones. Thus, it does not seem that the speaker reached her limit.

This study therefore shows some variability in acoustic and articulatory changes from silence to noise depending on unit within an utterance or a word, changes which seem to be consistent with enhanced cues to word segmentation and prosodic phrasing. These results are also in line with those obtained in previous studies on clear speech (Cutler and Butterfield, 1990) and emphatic speech (Lœvenbruck, 2000). The lack of statistical significance for some measures may (or may not) be due the fact that this was a single speaker study with a small corpus.

Our results call for follow-up studies in a number of areas, including: the extent to which the results found are dependent on the speaker, on the language studied, and on the corpus used. In addition, if the changes observed do indeed enhance cues to word segmentation and prosodic phrasing, perception studies should demonstrate this. Finally, other intonational parameters are clearly worth examining. For example, Welby, 2006 suggested that the intonational cues to content word beginning in French might be enhanced in Lombard speech.

5. Acknowledgments

The authors would like to thank Christophe Savariaux, Alain Arnal, Aude Noiray, Claire Lalevée and Coriandre Vilain for their contribution to this study. Pauline Welby's participation in this project was supported by a Marie Curie International Fellowship of the 6th European Community Framework Programme.

References

Amazi, D. K. and Garber, S. R. The Lombard sign as a function of age and task. *The Journal of Speech and Hearing Research*, 25(4): 581–585, 1982.

Bagou, O. and Frauenfelder, U. H. Stratégie de segmentation prosodique: rôle des proéminences initiales et finales dans l'acquisition d'une langue artificielle. In *Proceedings of the XXVIèmes Journées d'Etude sur la Parole*, pages 571–574, 2006.

Beckman, M.E. and Edwards, J. Articulatory evidence for differentiating stress categories. In Keating, P.A, editor, *Phonological structure and phonetic form: Papers in laboratory phonology III,* pages 7–33, Cambridge University Press, 1994.

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

Christophe, A. Role de la prosodie dans la segmentation en mots. Doctoral thesis, École des Hautes Études en Sciences Sociales, Paris, 1993.

Cutler, A. and Butterfield, S. Durational cues to word boundaries cues in clear speech. *Speech Communication*, 9: 485–495, 1990.

Davis, C., Kim, J., Grauwinkel, K. and Mixdorff, H. Lombard speech: Auditory(A), Visual(V) and AV effects. In *Proceedings of Speech prosody*, pages 361–365, 2006.

Delais-Roussarie, E. Pour une approche parallèle de la structure prosodique : étude de l'organisation prosodique et rythmique de la phrase française. *Doctoral dissertation, University of Toulouse*, 1995.

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

Fougeron, C. and Keating, P. A. Articulatory strengthening at edges of prosodic domains. *The Journal of the Acoustical Society of America*, 101(6): 3728–3740, 1997.

Garnier, M., Bailly, L., Dohen, M., Welby, P. and Lœvenbruck, H. An acoustic and articulatory study of Lombard speech : global effects on the utterance. In *Proceedings of ICSLP*, pages 2246–2249, 2006.

Junqua, J. C. The Lombard reflex and its role on human listener and automatic speech recognizers. *The Journal of the Acoustical Society of America*, 93: 510–524, 1993.

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

Lane, H. and Tranel, B. The Lombard sign and the role of hearing in speech. *The Journal of Speech and Hearing Research*, 14: 677–709, 1971.

Lœvenbruck, H. An investigation of articulatory correlates of the accentual phrase in French. In *Proceeding of ICPhS 99*, vol. 1, pages 667–670, 1999.

Lombard, E. Le signe de l'élévation de la voix. *Annales des maladies de l'oreille et du larynx*, 37: 101–119, 1911.

Pick, H. L., Siegel, G. M., Fox, P. W., Garber, S. R. and Kearney, J. K. Inhibiting the Lombard effect. *The Journal of the Acoustical Society of America*, 85(2): 894–900, 1989.

Rietveld, A. C. M. French word bounderies. *Language and Speech*, 23(3): 289–296, 1980.

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

Sinnott, J. M., Stebbins, W. C. and Moody, D. B. Regulation of voice amplitude by the monkey. *The Journal of the Acoustical Society of America*, 58: 412–414, 1975.

Tabain, M. Effects of prosodic boundary on /aC/ sequences: articulatory results. *The Journal of the Acoustical Society of America*, 113(5): 2834–2849, 2003.

Ternström, S., Sodersten, M. and Bohman, M. Cancellation of simulated environmental noise as a tool. *The Journal of Voice*, 16: 195–206, 2002.

Welby, P. Intonational differences in Lombard speech: looking beyond F_0 range. In *Proceedings of Speech Prosody*, pages 763–766, 2006.

Welby, P. French intonational structure: Evidence from tonal alignment. *The Journal of Phonetics*, 34(3): 343-371, 2006.

Zeiliger, J. BD_Bruit, une base de données de parole de locuteurs soumis à du bruit. In *Proceedings of the Xèmes Journées d'Étude sur la Parole*, pages 287–290, 1994.