May speech modifications in noise contribute to enhance audio-visible cues to segment perception?

Maëva Garnier

IJLRA-LAM (Université Paris 6, CNRS, Ministère de la Culture)
Music Acoustics Group-School of Physics (University of New South Wales, Sydney)
garnier@lam.jussieu.fr

Abstract

In this study we explore how acoustic and lip articulatory characteristics of bilabial consonants and three extreme French vowels vary in Lombard speech. In the light of several theories of segments perception we have shown that formant modifications should decrease the audio intelligibility of vowels in noise. On the contrary, modification in lip articulation should improve the visual intelligibility of vowels and bilabial consonants. This is not in agreement with previous studies which reported a global increased intelligibility of Lombard speech especially in the audio domain and not a lot in the visual one [1-3]. Thus, more detailed research is needed about the segmental and prosodic contribution to the increased intelligibility of Lombard speech

Index Terms: Lombard speech, production, audiovisual cues.

1. Introduction

Speech produced in noisy environment, also called Lombard speech, demonstrates several acoustic and articulatory modifications as compared to speech produced in quiet condition [1,4-6]. Are these modifications communicative strategies [7] or just a consequence of the increased voice intensity [8]? In any case, they globally contribute to enhance speech intelligibility in the audio domain [1], but not so much in the visual one [2,3]. However, we do not know very precisely yet in which extent all the different speech modifications contribute (or not) to this intelligibility improvement. Some previous studies have shown that some of the acoustic modifications could contribute to enhance the emergence of speech over a background noise [9-11]. It has also been shown that some prosodic cues to discourse structure are enhanced in noise [9,12,13] as well as words carrying the most information in the utterance [9]. In this article, we aim at exploring whether articulatory changes observed in Lombard speech could potentially contribute to increase segmental intelligibility, by prototyping audible or/and visible cues to segments recognition or distinction.

The reasons for hyper-articulating in noise are indeed not very clear. First, some studies support the idea that increased jaw aperture observed in shouted or Lombard speech may only be related to the increased voice intensity [14,15]. However, hyperarticulation is observed not only in loud speech but also in clear speech or infant addressed speech [16,17], so that we can assume that hyper-articulation aims at increasing audiovisual intelligibility in these specific situations, and that it may be the case for Lombard speech too. On the other hand, articulatory transform in loud and Lombard speech has been shown to consist of a reorganization rather than an homothetic amplification of the gestures [14] or a simple translation of the vowel system in the first formants space [15,18]. Consequently, it is far from obvious that these modifications improve segmental intelligibility. In fact, the increased jaw aperture can even be considered as a perturbation to intelligible speech production instead [19,20].

In that study, we explore acoustic and articulatory modifications of French vowels and bilabial consonants in noise. We envisage how these modifications may contribute to ‘prototype” some audible and/or visible cues to segment recognition [21-24] or to enhance the audiovisual contrast between phonological categories [17,25,26].

2. Material and methods

2.1. Protocol

3 female native speakers of French (L1 to L3) have successively been recorded when playing an interactive game with the experimenter, standing 2.5m in front of each of them (see [9,27] for more details). The game required the production of 15 highly confusable logatoms, standing as river names ([l**a**l**a**], [l**e**la], [lal**e**], [lal**i**], [l**i**la], [l**y**la], [lal**y**], [lal**u**], [l**u**la], [**p**ala], [la**p**a], [**b**ala], [la**b**a], [**m**ala], [la**m**a]). Analysed segments of these logatoms are indicated in bold. A carrying sentence : “L**a** __ **l**onge l**a** __” (The *river1* is passing by the *river2*) was imposed so that the analysed vowels were always in the same segmental context /l_(l)/, and the bilabial consonants in the same context /a_a/.

Speakers were first recorded in silence and then in a cocktail party noise, extracted from the BD_Bruit database [28]. Noise was played over two loudspeakers located 2m away from the speaker and 2m away from each other. Its level was calibrated to 85dB at the participant's ears. Noise was removed from the acoustic signal using the method designed by Ternstrom et al. [29].

2.2. Audiovisual measurements

The Audio signal was recorded with an AKG microphone placed 20cm away from the lips, then digitized at a rate of 44.1kHz. Segments boundaries were labelled using Praat [30]. The frequencies of the vowels’ first three formants were estimated with Matlab from the audio signal by using a conventional autocorrelation-based LPC method [31]. The number of poles has been chosen as a function of the sampling frequency.

Articulatory data were extracted from video recordings (25 images/s) of the speaker's lips, using a labiometric device [32]. We have measured the maximum amplitude of lip aperture, lip spreading (both from inner lip contour), and upper lip protrusion movements on every analysed vowels (cf. **Figure 1**). All these articulatory parameters were normalized by their speaker-dependant maximum, measured from extreme articulatory gestures recorded at the end of the experiment. Therefore they are presented in percents instead of millimeters. Lastly, we also estimated lip compression on bilabial consonants by the parameter (B'$_{neutral}$-B')/B'$_{neutral}$, only defined when the mouth is closed. B' stands for lip aperture measured from the external lip contour and its neutral value is measured from an instant of the recording when the speaker shows a closed and resting position of the mouth, without any lip contraction (see Figure 1). Lip compression is minimum (0%) for resting closure of the mouth, and maximum (100%) when the lips are so compressed that they are no longer visible (i.e. when B'=0).

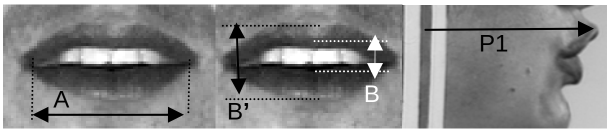

Figure 1. Articulatory parameters : Lip spreading (A), Lip aperture from external and inner contour (B and B'), Protrusion of the upper lip (P1).

3. Audible cues to vowel perception

There are different theories of segments perception, and thus, different approaches to the problem of understanding how acoustic and articulatory modifications of speech segments may affect their intelligibility.
First, the theory of acoustic invariance [21] or the « direct realist » theory [22] defend the idea that segment recognition is determined by invariant acoustic properties. In particular, it is admitted that the frequency of the 3 first formants play a determinant role in vowel perception [33]. Other theories rather support the idea that human perception is not parametric and sensitive to invariant properties, but sensitive to ratios or to integrated properties instead [34,35]. Lindblom even suggests that the interpretation of these cues in terms of phonological categories may depend on the communicative context and may involve high level cognitive processes [36].
Then, some studies brought arguments in favor of the existence of prototypic values for the recognition of phonological categories [23,24], which means that a vowel may be better recognized when its acoustic properties are close to those of its prototype. On the other hand, other studies support the idea that languages tend to maximize the distance in the space of acoustic properties between phonological categories [25], and that speakers also tend to expand their vowel system when speaking clearer or slower [17,26].
Therefore we are going to explore how the acoustic modification of vowels produced in noise can be interpreted at the light of these different theories of vowel perception.

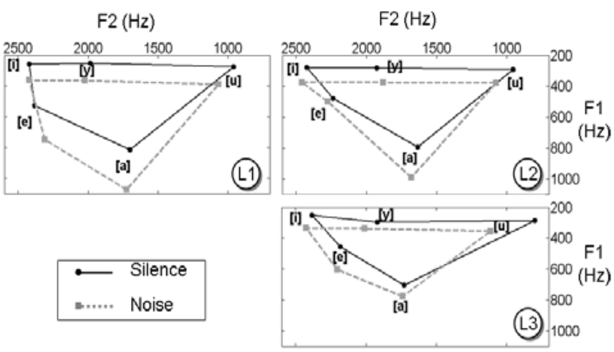

Figure 2. Modification in the F1/F2 plane of the vowel system from silence to cocktail-party noise for the three female speakers L1, L2 and L3.

Figure 2 presents how the first two formants (F1 and F2) of 5 French vowels vary in average from silent to noisy conditions for the 3 recorded speakers. We have left the third formant (F3) aside as its variations from silence to noise do not exceed 3% for all the considered vowels except [u] for which F3 increases by 5% in average.

We can see that vowel production mainly changes in noise along the F1 dimension (from + 22 to 36%). Most of the vowels do not vary much along F2 (less than + 2%), except [u] for which F2 increases by 21% in average from silence to noise for these 3 considered speakers.
In a conception of acoustic invariance or prototypic perception of vowels, these acoustic modifications may contribute to increase the intelligibility of [a] vowels produced in noise. On the contrary, all the other vowels move away from their prototypes and might be less intelligible: out of context, [i] vowels produced in noise might be perceived as [e], and [u] vowels as [o] (cf. Figure 2).
Then, in a contrastive conception of vowel perception, we can see that formant modifications in noise tend to enhance the distance along the F1 dimension, and thus the potential distinction between high and close vowels, whereas the contrast is instead reduced between front and back vowels along the F2 dimension. These results do not confirm those of Rostolland on French shouted speech, who reported that F2 varies with vocal effort in a different way for every vowel category but in a way which maintains their distinction [18]. On the contrary, he observed a significant modification of F3 with vocal effort, which decreases the vowel contrast.

Now, formants do not vary alone in noise but in conjunction with fundamental frequency (F0). Traunmüller has shown that the difference between F1 and F0, in barks, is more relevant than F1 alone to account for the perception of vowel height [37]. On Figure 3, we can see that (F1-F0)$_{barks}$ remains constant in noise (for L2) or even decreases (for L1 and L3) for close vowels, whereas it remains constant (for L3) or increases (for L1 and L2) for [a]. From a prototypic as well as from a contrastive conception of vowel perception, this may increase the intelligibility and the distinction between close and open vowels. On the other hand, the semi-open vowel [e] shows different variations among speakers. (F1-F0)$_{barks}$ increases for L1 and L3, with a similar trend to open vowels, whereas it decreases for L2.

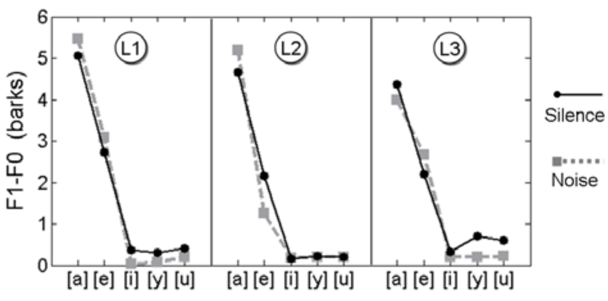

Figure 3. Modification in noise of the vowel system along the (F1-F0)$_{barks}$ dimension, for three female speakers L1, L2 and L3.

In addition, Traunmüller suggested that (F2-F1)$_{barks}$ well accounts for the distinction of back and front rounded vowels [38], being important in the case of [i] and weak for [u]. On **Figure 4**, we can see that (F2-F1)$_{barks}$ tends to decrease in noise for close vowels (except for [u] produced by L3), which is contrary to what Lienard et al. observed with an increased vocal effort [39]. We also notice that the contrast between front rounded and back vowels decreases in noise along this $(F2\text{-}F1)_{barks}$ dimension, which should rather lead to a decreased intelligibility.

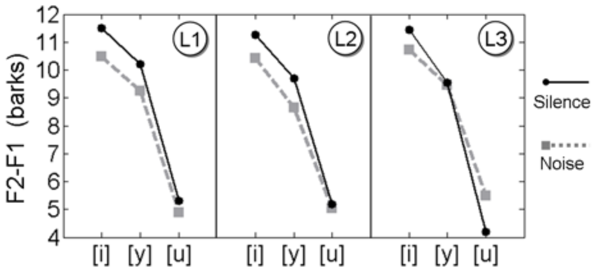

Figure 4. Modification in noise of close vowels [i], [y] and [u] along the $(F2\text{-}F1)_{barks}$ dimension, for three female speakers L1, L2 and L3.

Syrdal and Gopal also suggested $(F3\text{-}F2)_{barks}$ as a good correlate of the front–back contrast [40], being high in the case of [u] and weak for [i] or [y]. For speech produced in noise, we have seen that F2 and F3 vary few except for vowel [u]. On Figure 5, we can see that this contributes to reduce the contrast between [u] and [y] along that $(F3\text{-}F2)_{barks}$ dimension and thus, to potentially reduce their perceptual distinction.

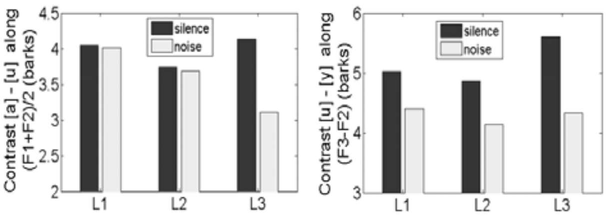

Figure 5. Modification in noise of the acoustic distance between [a] and [u] along the $(F1+F2)/2_{barks}$ dimension, and between [u] and [y] along the $(F3\text{-}F2)_{barks}$ dimension, for three female speakers L1, L2 and L3.

Then, Chistovich *et al.* have introduced the notion of ''center of gravity effect'' to account for the integrated perception of F1 and F2 in the region of the vowel space where these first two formants are close together [35]. Thus, an integrated value such as (F1+F2)/2 in barks could be relevant to characterise [u], which has a very low spectrum barycentre. This tonotopic parameter could also well account for the distinction between [u] and [o] or [a], which demonstrate higher centers of gravity. As F1 and F2 increase in noise for [u], its spectrum barycentre consequently raises, which may affect its intelligibility. However, we can see on Figure 5 that the contrast between [a] and [u] is still preserved in noise along the $(F1+F2)/2_{barks}$ dimension for speakers L1 and L2. This is not the case of L3 for whom this contrast strongly decreases in Lombard speech.

Last, we have computed the area of the vowel triangle formed by the three extrema [a], [i] and [u] of the French vowel system, in the (F1-F0)*F2*F3 space (in barks). Figure 6 shows how this area is similar for speech produced in silence and in noise for speakers L1 and L2, but decreases in noise for L3. This means that contrary to clear speech, Lombard speech does not demonstrate any expansion of the vowel system but a conservation or even a reduction instead, confirming previous observations of Lombard speech [5] and shouted speech [18].

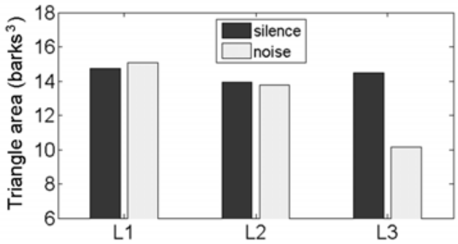

Figure 6. Modification in noise of the area of the triangle constituted by [a], [u] and [i], the three extrema of the French vowel system, in the $(F1\text{-}F0)*F2*F3_{barks}$ space.

4. Visible cues to segments perception

We have seen previously how acoustic modifications of the vowel production in noise may affect segmental intelligibility. Now, speech is not only audible but also visible. Visual modality helps not only hard of hearing people but every listener to better understand speech [41]. Indeed, for some theories of speech perception, like the “direct realist” theory [22], the speech motor theory [42], or the theory of perception for action control [43], phonological categories are defined at the level of the production gesture and the acoustic signal is only a cue, as much as the visual one, to get back to articulatory gestures.

Thus, French vowels can be categorised and perceptually distinguished from 4 articulatory features:

- The degree of vocal tract aperture, related to jaw lowering and lip opening.
- The position of the vocal tract constriction by the tongue, at the front or back or the oral cavity.
- The protrusion and the rounding of the lips, or on the contrary their spreading.
- The degree of coupling between oral and nasal cavities, thanks to the velum.

Among these different features, jaw and lip movements are highly visible. Tongue position can be guessed but more hardly. Velum movements cannot be visually perceived. Therefore we have focused here on the modification of lip articulatory movements and examined how these modifications may affect visual intelligibility of Lombard speech.

First, we can see on **Figure 7** that lip aperture increases in noise for every vowel and speaker, which is consistent with the global increase of F1 [44]. Consequently, this may improve the intelligibility of open vowels but reduce it for close ones. However, from a contrastive point of view, the articulatory distance between the open [a] vowel and close rounded vowels ([y], [u]) is preserved for speakers L1 and L3 along the lip aperture dimension, and even slightly increases for L2. This confirms results of Schulman who observed that vowel distinction is preserved along the lip opening/closing dimension in loud speech [14].

Spread vowels show more inter-speaker variability. Indeed, speakers L1 and L2 do not really open more the lips in noise for [i] vowels whereas the greatest amplification of lip aperture in noise is observed for [i] produced by L3. This results in an improved contrast in lip aperture between [i], [e] and [a] produced in noise by speakers L1 and L2, whereas [i] and [e] show similar lip aperture in noise for L3.

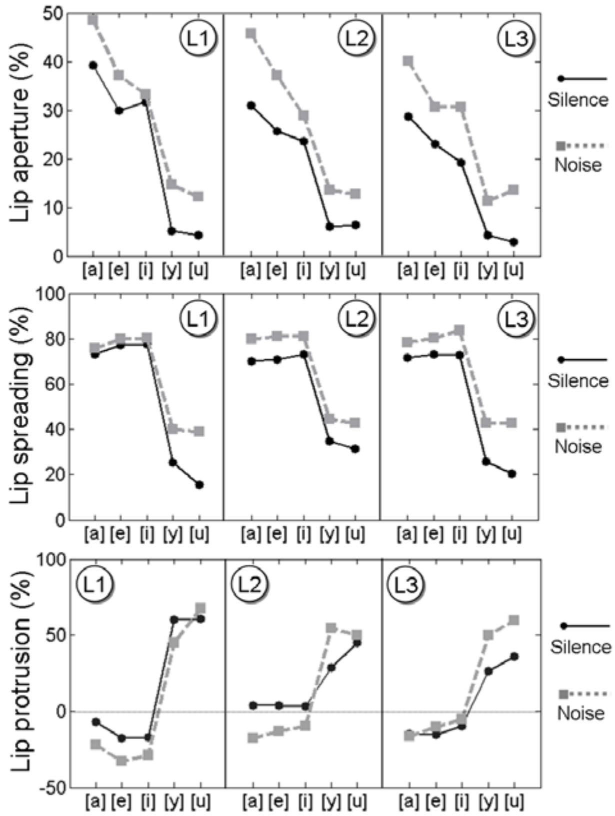

Figure 7. Modification in noise of the vowel system along lip articulatory dimensions, for three female speakers L1, L2 and L3. Articulatory parameters are normalised by their maximum value measured for each speaker.

Lip spreading increases in noise for rounded vowels (cf. **Figure 7**), which may reduce their visual intelligibility. On the other hand, spread vowels are also more spread in noise by speakers L2 and L3, but not by L1. This results in a conservation of the spreading/rounding contrast for L2, and a reduction for L1 et L3. Thus, this does not completely confirm the observations of Schulman who observed a preservation of this contrast in loud speech [14].

Then, we can see on **Figure 7** that vowels are more contrasted in noise along the protrusion dimension. This results from different strategies for the 3 speakers: L3 enhances the protrusion of [u] and [y] as compared to the other vowels. On the contrary, L1 conserves the protrusion of protruded vowels but enhances back movements of the lips on spread and open vowels. L2 enhances the protrusion of protruded vowels as well as back movement of open and spread vowels.

Last, we noticed that speakers amplify not only lip opening movements in noise, but also closing ones. Thus, **Figure 8** shows how lip compression of bilabial consonants is enhanced in noise by every speaker. Such an increased lip compression for bilabial stops has already been observed in loud speech [14]. This could be related to the increase of intra-oral pressure with an increased vocal effort. However, we observe an increase of lip compression in noise not only for bilabial stops but also for nasal segments [m]. Whatever causes these articulatory modifications, they may result in an improved visual recognition of bilabial consonants produced in noise. In addition, we notice on **Figure 8** a larger increase of lip compression for [p] segments, for speakers L2 and L3, which is consistent with previous studies [45,46].

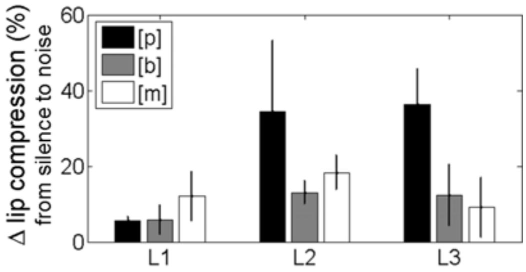

Figure 8. Increased lip compression in noise for 3 bilabial consonants [p], b] and [m], and three female speakers L1, L2 and L3.

5. Discussion

Previous studies in perception reported a global increased intelligibility of Lombard speech, especially in the audio domain and not a lot in the visual one [13].

From a production point of view, we have shown in this article that acoustic modification of vowels in noise may preserve or enhance their specific features or their contrast along the $(F1\text{-}F0)_{barks}$ dimension, on the contrary to all the other examined dimensions such as $(F1+F2)/2_{barks}$, $(F1\text{-}F2)_{barks}$ or $(F3\text{-}F2)_{barks}$. In the space of the first 3 formants, we do not either observe in Lombard speech any similar expansion of the vowel system to what Lindblom reported in clear speech [17]. Thus, whatever the theory of speech perception we may consider to interpret acoustic modifications in noise, it does not seem that the global increase of audio intelligibility observed for Lombard speech comes from the frequency variation of vowel formants. On the contrary, we may instead expect a degradation of segmental intelligibility in the audio domain for speech produced in noise.

In the visual domain, we have shown that lip movements in aperture and protrusion contribute to prototype vowels visible features or to increase the contrast between vowel categories. However, this is not the case for lip spreading movements. Lip compression is also significantly enhanced in noise for every bilabial consonant. Thus, we may expect an increased segmental intelligibility of Lombard speech in the visual domain.

The main conclusion of this study is that observations in segment production cannot explain the results of perceptual studies of Lombard speech [13].

A first explanation is that we may have focused on production descriptors which are not relevant. Thus, we have especially focused here on vowels whereas consonants may be more crucial. Indeed, clear speech has been characterised not only by an expansion of the vowel system but also by an enhancement of obstruent sounds, especially of stop consonants [47]. Likewise, we have focused here on static cues whereas dynamic features may be more important for segment perception in noise [48,49].

A second main possibility is that the increased intelligibility globally observed over whole utterances produced in noise do not come from the segmental level but more from the enhancement of prosodic cues or from the highlighting of important information in the utterance [9,12,13]. Therefore it seems necessary at this point to investigate with more details the different segmental and suprasegmental contributions to the global increased intelligibility of Lombard speech. To that goal, we are currently conducting some audiovisual tests to explore the perception of vowels produced in noise.

Then, if the increased intelligibility of Lombard speech does not come from the segmental level, there still remains the question of interpreting the use or the cause of hyper-articulation and formants modification in noise.

The first hypothesis would be that they are all related to the increase of vocal intensity. However, we have precisely reported throughout the article some difference between our observations on Lombard speech and previous studies on loud or shouted speech [14,18]. Formant modifications observed here can be pretty well explained by the increase in vocal effort and mouth aperture which go together. On the other hand, we observe some articulatory modifications in noise, such as an increased protrusion for protruded vowels (for 2 speakers over 3) or an enhancement of lip compression for [m] segments, which cannot be directly explained by the increase in vocal intensity.

However, we have shown in the article that isolated vowels are distorted rather than prototyped in noise, but that the contrast between them can be conserved along some audible and visible dimensions, though. We have seen for example that speaker L1 does not enhance the contrast in protrusion by enhancing this feature for protruded vowels, but on the contrary by enhancing the antagonistic movement for non protruded vowels. This lets us think that acoustic and articulatory modifications may compensation strategies more than strategies to increase segmental intelligibility. Indeed, the speaker may adopt a main communicative strategy to cope with the noise (maybe “speaking louder” ?), which may induce some perturbation at the segmental level (by increasing the mouth aperture and raising formants [19,20]). Some of the acoustic and articulatory modifications observed for Lombard speech may thus be related to this first main strategy. But on the contrary, some other ones may be done to compensate for the first ones and achieve a compromise between the goal of the first main strategy and the preservation of an acceptable segmental intelligibility. A previous study showed that loud speech is more intelligible that “normal” speech until a threshold of vocal intensity from which intelligibility decreases a lot [50]. This phenomenon could argue in favour of our hypothesis and indicate that above a given threshold of vocal intensity, speakers do not achieve any longer to reach a compromise between the increase of vocal intensity and the preservation of vowel articulation.

The enhancement of visible cues could usefully contribute to this compromise by compensating for the acoustic degradation of audible features, thanks to the complementary nature of both modalities [51]. Indeed, rounded and protruded vowels are the most intelligible vowels in the visual domain [52]. This means that they are still well recognized even if their articulation is a few distorted. In addition, maintaining this visible cue can be enough for the speech partner to identify this vowel category, even if formants are modified a lot.

Moreover, there are some motor equivalences so that movements of an articulator can be compensated by another articulator to achieve a similar acoustic result [53]. In the case of [u] in particular, speakers can compensate for an increased mouth opening by moving their tongue back [19]. Therefore we have recently recorded a new database of Lombard speech to explore tongue movements in addition to lip articulation, in order to determine whether there are some compensation strategies at the tongue level and whether hyper-articulation observed in Lombard speech concerns only visible articulators or all articulators.

6. Acknowledgements

Many thanks to the 3 speakers of that study. I am also grateful to Lucie Bailly, Marion Dohen, Helene Loevenbruck and Pauline Welby with whom I conducted a preliminary experiment to that study [6,12]. I also want to thank Nathalie Henrich, Chistophe Savariaux and Alain Arnal for their precious help in the acquisition of these data.

7. References

[1] Junqua, J., "The lombard reflex and it role on human listener and automatic speech recognizers", Journal of the Acoustic Society of America, 93(1): 510-524, 1993.

[2] Davis, C., Kim, J., Grauwinkel, K. et al., "Lombard speech: Auditory(A), Visual(V) and AV effects", in Proceedings of Speech prosody, Dresden, Germany2006.

[3] Chung, V., Mirante, N., Otten, J. et al., "Audiovisual processing of Lombard speech", in Proceedings of AVSP-2005, 55-56, 2005.

[4] Castellanos, A., JM.Benedi and Casacuberta, F., "An analysis of general acoustic-phonetic features for spanish speech produced with the lombard effect", Speech Communication, 20: 23-35, 1996.

[5] Junqua, J. C., "Acoustic and production pilot studies of speech vowels produced in noise", in Proceedings of ICSLP, Banff, Canada1992.

[6] Garnier, M., Bailly, L., Dohen, M. et al., "An Acoustic and Articulatory Study of Lombard Speech: Global Effects on the Utterance", in Proceedings of ICSLP, Pittsburgh, PA, USA, 17-22, 2006.

[7] Lane, H. and Tranel, B., "The Lombard sign and the role of hearing in speech", Journal of Speech and Hearing Research, 14: 677-709, 1971.

[8] Lombard, E., "Le signe de l'elevation de la voix", Annales des maladies de l'oreille et du larynx, 37: 101-119, 1911.

[9] Garnier, M., "Communiquer en environnement bruyant : de l'adaptation jusqu'au forcage vocal", Ph.D thesis of the University Paris 6, Paris, 2007.

[10] Mokbel, C., "Reconnaissance de la parole dans le bruit: bruitage / débruitage", Ph.D thesis of the Ecole nationale supérieure des télécommunications, Paris, 1992.

[11] Egan, J. J., "Psychoacoustics of the Lombard voice response", Journal of Auditory Research, 12: 318-324, 1972.

[12] Garnier, M., Dohen, M., Lœvenbruck, H. et al., "The Lombard Effect: a physiological reflex or a controlled intelligibility enhancement?" in Proceedings of 7th International Seminar on Speech Production Ubatuba, Brazil, 255–262, 2006.

[13] Welby, P., "Intonational Differences in Lombard Speech: Looking beyond F0 Range." in Proceedings of Speech Prosody, Dresden, Germany, 763-766, 2006.

[14] Schulman, R., "Articulatory dynamics of loud and normal speech", Journal of the Acoustic Society of America, 85(1): 295-312, 1989.

[15] Garnier, M., Wolfe, J., Henrich, N. et al., "Interrelationship between vocal effort and vocal tract acoustics: a pilot study", in Proceedings of ICSLP, Brisbane, 2008.

[16] Picheny, M. A., Durlach, N. I. and Braida, L. D., "Speaking Clearly for the Hard of Hearing II : Acoustic Characteristics of Clear and Conversational Speech", Journal of Speech and Hearing Research, 29: 434-446, 1986.
[17] Lindblom, B., Brownlee, S., Davis, B. et al., "Speech transforms", Speech Communication, 11: 357-368, 1992.
[18] Rostolland, D., "Phonetic structure of shouted voice", Acta Acustica, 51: 80-89, 1982.
[19] Savariaux, C. and Perrier, P., "Compensation strategies for the perturbation of the rounded vowel [u] using a lip tube: A study of the control space in speech production", Journal of the Acoustic Society of America, 98(5): 2428-2442, 1995.
[20] Geumann, A., Kroos, C. and Tillmann, H. G., "Are there compensatory effects in natural speech ?" in Proceedings of ICPhS, San Francisco, US, 399-402, 1999.
[21] Blumstein, S. and Stevens, K., "Acoustic invariance in speech production", Journal of the Acoustic Society of America, 66: 1001-1017, 1979.
[22] Fowler, C. A., "An event approach to the study of speech perception from a direct-realist perspective", Journal of Phonetics, 14: 3-28, 1986.
[23] Grieser, D. and Kuhl, P. K., "Categorization of Speech by Infants: Support for Speech-Sound Prototypes", Developmental Psychology, 25(4): 577-588, 1989.
[24] Sussman, J. E., "A preliminary test of prototype theory for a [ba]-to-[da] continuum (A)", Journal of the Acoustic Society of America, 93(4): 2392, 1993.
[25] Manuel, S. Y. and Krakow, R. A., "Universal and Language particular aspects of vowel-to-vowel coarticulation", Haskins Lab. Stat. Rep. Speech Res: 69-78, 1984.
[26] Johnson, K., Flemming, E. and Wright, R., "The Hyperspace Effect: Phonetic Targets Are Hyperarticulated", Language 69(3): 505-528, 1993.
[27] Garnier, M., Henrich, N. and Dubois, D., "Methodologies for the investigation of speech produced in noise: A comparative study", (submitted)
[28] Zeiliger, J., Serignat, J. F., Autresserre, D. et al., "BD_Bruit, une base de données de parole de locuteurs soumis à du bruit", in Proceedings of Xèmes Journées d'Etude de la Parole, Grenoble, France, 287-290, 1994.
[29] Ternström, S., Sodersten, M. and Bohman, M., "Cancellation of simulated environmental noise as a tool for measuring vocal performance during noise exposure", Journal of Voice, 16(2): 195-206, 2002.
[30] Boersma, P. and Weenink, D., "Praat: doing phonetics by computer (Version 4.2.28) [Computer program]", (Issue): retrieved from http://www.praat.org/ in 2005,
[31] Markel, J. E. and Gray, A. H., "Linear Prediction of Speech", Springer-Verlag New York, Inc. Secaucus, NJ, USA.1982.
[32] Lallouache, T., "Un poste "Visage-Parole" couleur. Acquisition et traitement automatique des contours des lèvres", Ph.D thesis of the INPG, 1991.
[33] Carlson, R., Granström, B. and Fant, G., "Some studies concerning perception of isolated vowels", STL-QPSR, 2-3: 19-35, 1970.
[34] Schwartz, J. L. and Escudier, P., "A strong evidence for existence of a large scale integrated spectral representation in vowel perception", Speech Communication, 8: 235-259, 1989.
[35] Chistovich, L. A. and Lublinskaya, V. V., "The ‘center of gravity’effect in vowel spectra and critical distance between the formants: " Hearing Research 1 185-195, 1979.
[36] Lindblom, B., "Role of articulation in speech perception: clues from production", Journal of the Acoustic Society of America, 99(3): 1683-1694, 1996.
[37] Traunmüller, H., "Perceptual dimension of openness in vowels", Journal of the Acoustic Society of America, 69(5): 1465-1475, 1981.
[38] Traunmüller, H., "The role of the fundamental and the higher formants in the perception of speaker size, vocal effort, and vowel openness", in Proceedings of Franco-Swedish Seminar on Speech, Grenoble, France1985.
[39] Lienard, J. S. and Di Benedetto, M. G., "Effect of vocal effort on spectral properties of vowels", Journal of the Acoustic Society of America, 106(1): 411-422, 1999.
[40] Syrdal, A. K. and Gopal, H. S., "A perceptual model of vowel recognition based on the auditory representation of American English vowels", The Journal of the Acoustical Society of America, 79: 1086-1100, 1986.
[41] Sumby, H. and Pollack, I. W., "Visual Contribution toSpeech Intelligibility in Noise", Journal of the Acoustic Society of America, 26: 212-215, 1954.
[42] Liberman, G. and Mattingly, A. M., "The motor theory of speech perception revised", Cognition, 21: 1-36, 1985.
[43] Schwartz, J. L., Abry, C., Boë, L. J. et al., "Phonology in a theory of perception-for-action-control " in, J. Durand and B. Laks [Ed], Phonology: from Phonetics to Cognition, 255-280, Oxford University Press, Oxford, 2002.
[44] Lindblom, B. E. and Sundberg, J. E., "Acoustical consequences of lip, tongue, jaw, and larynx movement", Journal of the Acoustic Society of America, 50(4): 1166-1179, 1971.
[45] Gracco, V. L., "Some organizational characteristics of speech movement control", Journal of Speech and Hearing Research, 37: 4-27, 1994.
[46] Malecot, A., "Mechanical pressure as an index of force of articulation", Phonetica, 14: 169-180, 1966.
[47] Picheny, M. A., Durlach, N. I. and Braida, L. D., "Speaking clearly for the hard of hearing I: intelligibility differences between clear and conversationnal speech", Journal of Speech an Hearing Research, 28: 96-103, 1985.
[48] Browman, C. P. and Goldstein, L., "Articulatory gestures as phonological units", Phonology, 6: 201-252, 1989.
[49] Carre, R. and Divenyi, P. L., "Modeling and perception of gesture reduction", Phonetica, 57(2-4): 152-169, 2000.
[50] Rostolland, D. and Parant, C., "Distorsion and intelligibility of shouted voice", in Proceedings of Symposium Speech Intelligibility, Liège, 293-304, 1973.
[51] Robert-Ribes, J., Schwartz, J. L., Lallouache, T. et al., "Complementarity and synergy in bimodal speech: Auditory, visual, and audio-visual identification of French oral vowels in noise", The Journal of the Acoustical Society of America, 103: 3677, 1998.
[52] Benoit, C., Mohamadi, T. and Kandel, S., "Effects of phonetic context on audio-visual intelligibility of French", Journal of Speech an Hearing Research, 37(5): 1195-1203, 1994.
[53] Perkell, J. S., Matthies, M. and Svirsky, M., "Trading relations between tongue-body raising and lip rounding in production of the vowel /u/: A pilot ‘Motor Equivalence’ study", Journal of the Acoustical Society of America, 93: 2948–2961, 1993.