Hyperarticulation in Lombard speech: A preliminary study

Juraj Šimko[1], *Štefan Beňuš*[2,3], *Martti Vainio*[1]

[1]Insitute of Behavioural Sciences, University of Helsinki, Finland
[2]Constantine the Philosopher University, Nitra, Slovakia
[3]Institute of Informatics, Slovak Academy of Sciences, Bratislava, Slovakia
juraj.simko@helsinki.fi, sbenus@ukf.sk, martti.vainio@helsinki.fi

Abstract

Over the last century researchers collected a considerable amount of data reflecting the properties of Lombard speech, i.e., speech in a loud environment. The documented phenomena include effects on intensity, fundamental frequency, spectral tilt, speech rate and articulation. Relatively little attention has been paid to the effects on relative extent of movement of individual articulators. In an attempt to fill in this gap we present a preliminary analysis of EMA data collected in increasing levels of babble noise. We introduce HH-index as a measure of overall relative activity of articulators. Our results indicate a non-linearity of the effect of noise on articulatory movement and quantitatively different effects on the movement extent for different groups of articulators. The effects of noise are compared with those brought out by other techniques for eliciting articulatory variation. We also discuss possible application of Lombard speech as an elicitation paradigm for studies of hyperarticulation.

Index Terms: Lombard speech, hyperarticulation, articulatory variation, Slovak, EMA recordings

1. Introduction

Speakers raise their voice when they speak in environmental noise. This adaptation of speech to noise in order to increase the signal-to-noise ratio is called the Lombard effect [1] and is realized by physiological means that have different consequences on speech acoustics. Typically, the speakers increase intensity and f_0, adjust intonational contours [2, 3, 4], and their mode of vocal fold vibration is more pressed decreasing the slope of the glottal voice-source spectrum. The loud environment also affects the duration and spectral characteristics of vowels, shifting the positions of the formants [5, 6].

Compared to the vast body of research focused on acoustic and perceptual correlates of the Lombard effect, experimental articulatory investigations are relatively sparse. Studies primarily focusing on the jaw and lip movements have confirmed that Lombard speech is realized with amplification of articulatory patterns identified in "normal" speech in silent environment, and that the extent of amplification is linked to the type of noise, its loudness and the degree to which speaker's self-monitoring feedback is compromised [7]. The amplification has been shown to involve complex, non-linear reorganization of articulatory movement patterns [8].

In this preliminary study we report the results of analysis of articulatory recording including the lip, jaw and tongue movement. We focus on relative expansion of movement trajectories induced by increasing volume of babble noise as well as on global durational correlates of the Lombard effect. Furthermore, we evaluate the relative sensitivity of individual articulators on the increasing level of noise.

In articulatory terms, the Lombard effect can be interpreted as hyperarticulation. Hyperarticulation – an increase of the extent of articulatory movement – and its hypoarticulation counterpart have been widely studied as the general source of phonetic variation (H&H variation) [9] and as an important factor underlying prosodic phenomena [10]. One of the aims of this work is to assess a possibility of eliciting hyperarticulation using the Lombard effect in a controlled, quantifiable fashion. We also include two additional conditions in our recordings – a hypospeech and a non-native speaker targeting speech – and compare the effects elicited by these paradigms with those brought up by environmental noise.

In order to evaluate the effects of elicitation conditions quantitatively, we introduce a measure of relative articulatory variation, *HH-index*, capturing a proportional increase/decrease of articulatory movement in terms of articulator trajectory.

2. Methods

This study is a part of a larger investigation of the effects of prosody on articulation and inter-articulator timing. Four Slovak stimuli sentences, each containing 17 syllables, were created for a three-way manipulation of utterance-internal boundary strength, resulting in a total of 12 stimuli. With a minimum 5 intended repetitions, each block thus contained at least 60 tokens with the order of stimuli randomised within each block. Non-linguistic manipulation of prosody included the elicitation of hyper- and hypo-articulation in the following way. A reference stimulus block (subsequently referred to as condition *0dB*) was recorded in silence and the speaker was instructed to speak naturally. Three stimuli blocks were produced with three levels of babble noise at the level of 60, 70, and 80 dB – conditions *60dB*, *70dB* and *80dB*, respectively – played over the subject's headphones. Additionally, another block was elicited with no noise, where the speaker was explicitly instructed to use relaxed, hypo-articulated speech (*0dB-r* condition). Finally, the assumed highest level of hyper-articulation was elicited with 80 dB babble noise simulating a communication with a non-native interlocutor (cf. [11]) who was present and visually interacted with the subject (condition *80dB-nn*). The blocks were recorded in the following order: *70dB*, *80dB*, *0dB*, *0dB-r*, *60dB*, *80dB-nn*. Overall, we obtained 365 sentences from one subject, a native Slovak speaker with no speech or hearing impairment.

The articulatory data come from kinematic trajectories of sensors attached to 6 active articulators – lower and upper lip (LL, UL), jaw, tongue tip (TT), tongue body (TB) and tongue dorsum (TD) – obtained using electro-magnetic articulography

10.21437/SpeechProsody.2014-162

(EMA, AG500, Carstens Medizinelektronik, IBS, University of Helsinki). The EMA data were post-processed using TAPAD routines [12]. Audio signal was used for automatic forced alignment using the SPHINX toolkit adjusted for Slovak [13] and the time points of speech initiation and cessation were subsequently manually corrected based on the initial amplitude increase at the beginning, and the cessation of formant structure (for vowels) or voicing (for sonorants), at the end of each sentence.

To assess the quantitative articulatory characteristics for each token we calculated the approximate length of trajectory of each sensor during the utterance. That is, for each time step (determined by sampling rate of the articulatograph, 200 Hz) we calculated the Euclidean distance between subsequent positions of the given sensor in midsagittal plane. The entire trajectory of the sensor during the utterance was then calculated as a sum of these small transitions. This measure of the articulatory activity closely corresponds to the definition of Bounded Variation norm used in [4] for investigating effects of noise on fundamental frequency in Lombard speech.

Naturally, even when no global articulatory variation is elicited, the distance covered by individual articulators varies with their anatomical properties and the segmental characteristics of the stimulus. In our data, for example, the TB sensor moved on average over approx. 340 mm per token, while the UL sensor covered on average less than 70 mm. Consequently, absolute hyperarticulation effects on the distance travelled are considerably greater for “livelier” articulators than for the more restricted ones. As we are primarily interested in relative and stimulus-independent effects on articulation we have normalized the measure defined above in the following way.

The *0dB* condition – the stimuli uttered with no background noise in a natural fashion – was used as a reference. First, for each sensor we computed the mean trajectory length of the recordings of the same stimulus in this condition. Then, in order to factor out the influence of segmental and prosodic structure of different stimuli, we divided the trajectory length of every sensor for every recording in the data-set by this *0dB*-mean value of the corresponding sensor-stimulus combination. We will call the resulting measure of relative hypo-/hyper-articutation the *HH-index* for the given articulator and token. For each token, the mean of *HH-indeces* for 5 out of 6 recorded sensors, excluding UL sensor[1], is referred to as the *overall HH-index* for the token.

Naturally, the mean values of all *HH-indeces* for *0dB* condition are all equal to 1. Greater values indicate proportionally greater articulatory movement; *HH-index* value 2 means that the given articulator (or the group thereof) covered twice the distance than in the reference condition – speaker hyperarticulated. Similarly, values less than 1 correspond to hypoarticulation.

To eliminate the influence of different segmental and prosodic structure of the 12 stimuli, analogous normalization was performed for durations: again, the duration of each token was divided by the mean duration of *0dB* tokens of the same stimulus, yielding the *normalized duration* measure.

As we only have one subject and the normalizations described above removed the influence of speech material, we used a simple ANOVA-based Tukey multiple comparisons of means (with corrected p-values) implemented in R for evaluation of the effects of noise levels as well as the other two manipulations.

3. Results

3.1. Durations

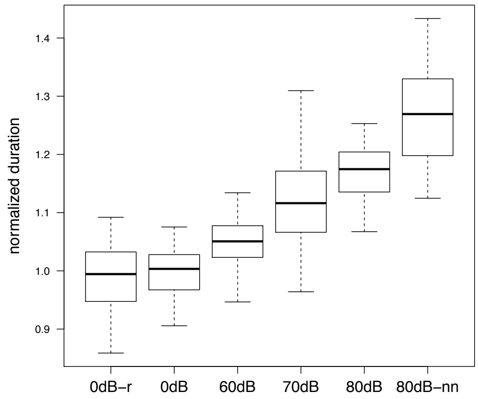

Figure 1: *Normalized durations* per condition.

Fig. 1 summarizes the effects of various conditions on duration of utterances as rendered by the subject. Tukey multiple comparisons of means showed that all differences among mean values for individual conditions are significant ($p < 0.001$), except the *0dB-r–0dB* pair ($p = 0.77$). Moreover, the pattern shown in the boxplot suggests approximately linear dependence of (non-linearly scaled) noise on the normalized duration. The differences between subsequent means, in the order captured in Fig. 1, are: 0.014, 0.050, 0.070, 0.051 and 0.096.

3.2. Overall articulatory variation

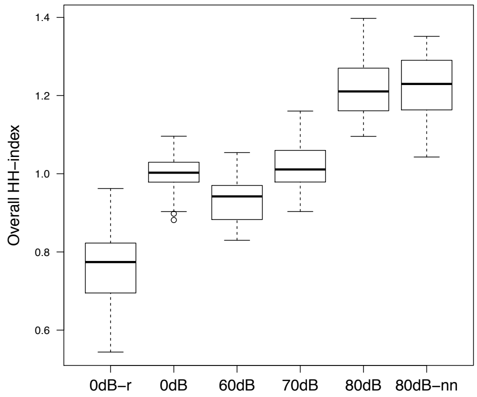

Figure 2: Overall *HH-indeces* per condition.

Fig. 2 shows the distributions of *overall HH-indeces* for individual tokens grouped by conditions. Again, most of the depicted differences are robustly significant ($p < 0.001$). The very strong effect of explicitly induced hypospeech indicates that the articulator trajectory measure as adopted in this work is a viable estimate of the magnitude of HH-variance. The difference of means between *0dB* and *0dB-r* is 0.238, the greatest among all the neighboring pairs; it means that during the “relaxed” speech the selected articulators covered almost one-quarter shorter distance than in the normal reference condition.

[1] We had to skip this sensor as we failed to recover its correct movement for multiple tokens in *70dB*-condition during the post-processing.

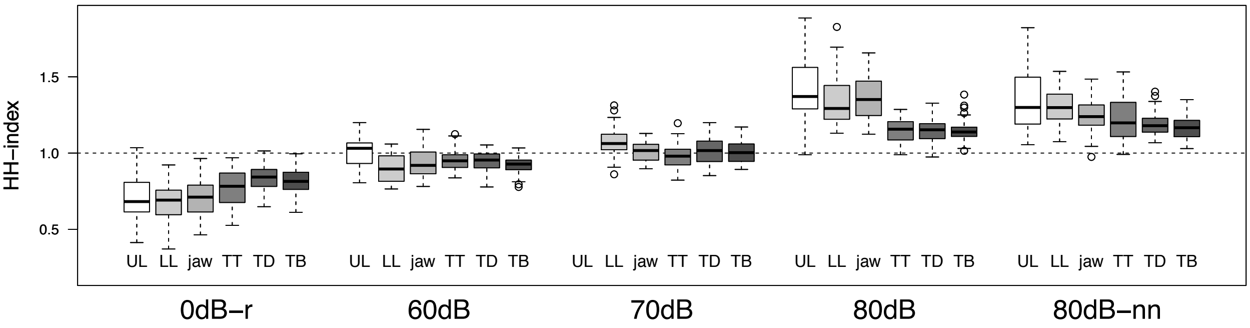

Figure 3: Sensor specific *HH-indeces* per condition.

The difference between means is not significant only in two cases. First is the difference between conditions *80dB* and *80dB-nn* ($p = 0.999$): the non-native speaker directed speech failed to elicit significantly greater articulatory movement, at least in our relatively small data-set.

Second case is the pair *0dB–70dB* ($p = 0.84$). This fact seems to be a consequence of a slightly surprising patterns present in our data. While for the three conditions with noise the mean value of *overall HH-index* significantly increases with the noise intensity, the mean value is actually significantly *lower* for *60dB* condition than for the reference quiet condition.

3.3. Sensitivity of individual articulators

We assessed the behavior of individual articulators – depicted by sensors UL, LL, jaw, TT, TD and TB – with respect to different recording conditions in two steps. First, we analyzed the effect of the conditions on each sensor separately, in a manner analogous to that in Section 3.2. Then, we compared the extents to which different sensors reacted to the influence of noise and the other two recording conditions.

Fig. 3 depicts the distributions of *HH-indeces* for individual sensors (shown in different shades of grey) and different conditions. Due to space restrictions, data for condition *0dB* are not plotted (the means for all sensors were by definition equal to 1 indicated by the dashed line). Also, the values for sensor UL are missing for *70dB*-condition for technical reasons, see Section 2.

The distributions of sensor-specific *HH-indeces* per individual conditions show, by and large, patterns very similar to that identified for the overall *HH-index* (Fig. 3.2). In general, the mean values significantly ($p < 0.001$) increase with increasing level of noise, and the comparison of *60dB* and *70dB* conditions with the quiet condition *0dB* in many cases fail to reveal a significant difference. For all sensors, the mean of *HH-indeces* for *0dB-r*-condition are significantly lower than for every other condition, and for *80dB* and *80dB-nn* conditions the means are significantly higher than for every condition with lower (or no) noise level ($p < 0.001$ in all cases).

Here we list all condition pairs for each sensor for which the difference of means of *HH-indeces* are not significantly different from 0 at $p < 0.001$ in our data set (Tukey multiple comparisons of means test).

For UL, the difference between the means for *60dB* and *0dB* is not significant ($p = 0.99$). For LL, the mean for *70dB* is significantly greater than that for *0dB* at $p < 0.01$. For both lip sensors the difference between *80dB* and *80dB-nn* is not significant ($p = 0.99, 0.86$, for UL and LL respectively).

For the jaw, the difference between *0dB* and *70dB* is not significant ($p = 0.99$), while the means for both *60dB* and *70dB* are both significantly smaller than that for *0dB* ($p < 0.05, 0.01$, respectively). The mean for *80dB* is significantly greater than that for *80dB-nn* ($p < 0.001$).

Interestingly, for TT and TD the relationship between *80dB* and *80dB-nn* is reversed compared to the jaw, the former being significantly lower that the latter ($p < 0.001, 0.05$, respectively). For TB sensor, the *80dB–80dB-nn* difference is also not significant ($p = 0.47$). For all three tongue sensors, the difference between *0dB* and *70dB* is not significant ($p = 0.63, 0.84, 0.84$, for TT, TD, TB, respectively). For TT sensor, neither are the differences between the means for *0dB* and *60dB* ($p = 0.06$) and *60dB* and *70dB* ($p = 0.83$).

Next we looked at differences in articulatory variability between individual articulators in different conditions. Table 1 summarizes the differences between the means of *HH-indeces* for all pairs of articulators (rows) organized by conditions (columns; the p-values are corrected for individual conditions). Figure 3 depicts the general trends.

First, note the very small differences between relative trajectory expansion/shrinking among tongue sensors TT, TB and

Table 1: Differences between mean values of *HH-indeces* for pairs of articulators for different conditions. Asterisks indicate significance of the difference being different from 0: ***:$p < 0.001$, **:$p < 0.01$,*:$p < 0.05$.

	0dB-r	*60dB*	*70dB*	*80dB*	*80dB-nn*
LL-UL	-0.03	-0.11^{***}		-0.10^{***}	-0.11
jaw-UL	-0.01	-0.07^{***}		-0.07^{*}	-0.17^{**}
jaw-LL	0.02	0.03	-0.07^{***}	0.03	-0.06
TT-UL	0.06	-0.05^{**}		-0.28^{***}	-0.19^{**}
TD-UL	0.13^{***}	-0.06^{***}		-0.28^{***}	-0.23^{***}
TB-UL	0.10^{***}	-0.09^{***}		-0.29^{***}	-0.25^{***}
TT-LL	0.09^{***}	0.05^{**}	-0.09^{***}	-0.18^{***}	-0.07
TD-LL	0.16^{***}	0.04^{*}	-0.06^{***}	-0.18^{***}	-0.12
TB-LL	0.13^{***}	0.02	-0.07^{***}	-0.19^{***}	-0.14^{*}
TT-jaw	0.07^{*}	0.02	-0.03	-0.21^{***}	-0.02
TD-jaw	0.13^{***}	0.01	0.01	-0.21^{***}	-0.06
TB-jaw	0.11^{***}	-0.02	0.00	-0.21^{***}	-0.09
TD-TT	0.06^{*}	-0.01	0.04	0.00	-0.04
TB-TT	0.04	-0.03	0.03	-0.01	-0.07
TB-TD	-0.03	-0.03	-0.01	0.00	-0.02

TT. The maximal difference is approximately 7 % between TT and TB for *80dB-nn* condition. Only one difference is significant (TD–TT for *0dB-r*, $p < 0.05$) in our data set.

On the other hand, there are relatively robust and in many cases significant differences between reaction of the tongue and lip-jaw articulatory systems on most conditions. The greatest differences between articulators from these groups can be seen in condition *80dB*. In this case, the lips and the jaw expanded their trajectory compared to the reference condition (*0dB*) by some 20–30 % more than the tongue articulators; all increases were significant. Smaller effect is present for *60dB* and *70dB* conditions. Interestingly, while in condition *70dB* LL trajectory expanded significantly more than tongue sensor's ones, it expanded significantly *less* than TT and TD trajectories in *60dB* condition. In the latter condition, however, UL trajectory shows significantly greater expansion of UL sensor relative to the tongue (unfortunately, data for UL in *70dB* are missing in our analysis). The jaw has not shown any significant differences compared to the tongue articulators in these two conditions.

The trend reversal suggested by LL sensor continues for the hypoarticulated condition *0dB-r*: with one exception (TT–LL), the effect is significantly greater for the tongue than for the lip-jaw articulators.

Within the lips-jaw system the results generally support the above observation of the greatest sensitivity of UL sensor compared to the LL and the jaw. The mean *HH-index* for UL is greater than both for LL (significantly so for *60dB* and *80dB*) and for the jaw (significantly for *60dB*, *80dB* and *80dB-nn*). Comparison between LL and the jaw reveals significant difference only in *70dB* condition.

Finally, for *80dB-nn* condition, the relative articulator sensitivity shows similar patterns to that of *80dB*, however, (with an exception of the UL sensor) the observed differences are generally not significant.

4. Discussion and conclusions

In the presence of a loud background noise, the utterances expand in duration and, at least in the case of 80 dB babble noise, also in the extent of articulatory movement. While temporal expansion seems to be approximately linear with logarithmic increase of the noise level, our results suggest a non-linear effect on articulatory trajectories. The overall effect on articulation seems to be negligible (or even, counterintuitively, negative) for lower noise levels, however, at the level of 80 dB the lengthening of trajectories is robust for all articulators.

Admittedly, the reported non-linearity can arise from various sources, predominantly the relatively small size of our data set limited to a single speaker. A possible source can also be the order in which the analyzed tokens were recorded: the reference, *0dB* stimuli were recorded right after the loudest *80dB* block, while *60dB* block followed the hypoarticulated *0dB-r* condition; clearly, some carryover effect could have influenced our measurements. At the same time, the lack of effect for the lower noise levels is intriguing as the subject was immersed in the noise through headphones with no self-monitoring feedback while he was wearing no headphones in quiet conditions: attenuating the external auditory feedback is to be expected to elicit a Lombard effect on its own even without the noise [14]. In any case, these initial findings warrant further investigation with additional speakers, randomized elicitation order and different ways of presenting/blocking self-monitoring feedback.

The robust effect of explicitly induced hypoarticulation (*0dB-r*) on both overall and articulatory specific *HH-indeces* in an expected direction justifies this measure as a way of quantitatively evaluating articulatory variation along HH dimension. As *HH-index* evaluates purely spatial extent of articulation and not articulatory velocity and/or duration, it can be used in a complementary fashion to other measures like the (normalized) duration used here. (Note the apparent "orthogonality" of the two measures for *0dB–0dB*-r and *80dB–80dB-nn* condition pairs.)

Furthermore, our results show that the Lombard effect is a viable methodology for eliciting global articulatory variation (more precisely, hyperarticulation) in a controlled manner: at least the loudest noise condition resulted in significant global and sensor-specific hyperarticulation patterns. In our limited data set, the other method intended to produce extra hyperarticulation – addressing non-native speaker (*80dB-nn*) – resulted in considerably longer durations but failed to elicit more overall articulatory movement compared to its nearest counterpart (*80dB*). To further evaluate and compare these two methods of triggering hyperarticulation, a condition when the subject speaks to non-native listener in quiet condition will be included in the follow up experiments.

The data analysis revealed interesting – albeit not altogether unexpected – patterns regarding relative sensitivity of articulators to conditions. Behavior of the articulators follows a plausible division to three anatomically meaningful groups: the tongue articulators, the lower lip-jaw system and the upper lip. In general, the sensors placed on the tongue exhibited mutually similar behavior as did the LL-jaw articulators, although the latter were more sensitive to noise-induced variations (at least for louder noise levels). The upper lip was still more sensitive. The greater sensitivity of the lips and the jaw is shown also for hypospeech, where the movement extent attenuated more for these articulators than for those of the tongue.

Two slightly different explanations can account for this phenomenon. It is possible that the greater extent of hyperarticulation for the lips and the jaw is specific to Lombard speech, in line with the other known correlates of the Lombard effect. Greater opening of the mouth simply contributes to better "audibility", salience in a loud environment, alongside increased loudness, pitch and spectral adjustments. The increase in motion of the visible articulators can also assist the interlocutor in parsing what has been said [15]. The observed effects on the tongue can be just a straightforward consequence of the more extensive movement of the anatomically connected jaw. Alternatively, the greater effect on the lips and the jaw compared to the tongue can be a consequence of greater freedom of the former in terms of physiological constraints and acoustic consequences of increased variation. In both aspects the tongue is more restricted than the jaw and the lips. In this case, the different sensitivity could be a hallmark of H&H variation in general, and has to be taken in consideration in research involving articulatory variation, for example, coarticulatory effects of bilabials and vowels in stressed vs. unstressed syllables. Our results provide a tentative support to both these interpretations: the differences in articulator sensitivity between *80dB* and *80db-nn* conditions for the former and the consistency of the *0dB-r* patterns with the general trend for the latter. More data and further research are required to shed more light on this issue.

5. Acknowledgements

The research leading to the results presented in this paper has received funding from the Finnish Academy, from the European Union FP7 under grant agreement n. 312382 and ERDF project RPKOM (ITMS 26240220064).

6. References

[1] E. Lombard, "Le signe de l'elevation de la voix," *Ann. Maladies Oreille, Larynx, Nez, Pharynx*, vol. 37, no. 101-119, p. 25, 1911.

[2] Y. Lu and M. Cooke, "Speech production modifications produced by competing talkers, babble, and stationary noise," *The Journal of the Acoustical Society of America*, vol. 124, p. 3261, 2008.

[3] ——, "The contribution of changes in f0 and spectral tilt to increased intelligibility of speech produced in noise," *Speech Communication*, vol. 51, no. 12, pp. 1253–1262, 2009.

[4] M. Vainio, D. Aalto, A. Suni, A. Arnhold, T. Raitio, H. Seijo, J. Järvikivi, and P. Alku, "Effect of noise type and level on focus related fundamental frequency changes," in *Proceedings of Interspeech 2012*, 2012.

[5] W. Van Summers, D. B. Pisoni, R. H. Bernacki, R. I. Pedlow, and M. A. Stokes, "Effects of noise on speech production: Acoustic and perceptual analyses," *The Journal of the Acoustical Society of America*, vol. 84, no. 3, p. 917, 1988.

[6] J.-C. Junqua, "The influence of acoustics on speech production: A noise-induced stress phenomenon known as the lombard reflex," *Speech Communication*, vol. 20, no. 1, pp. 13–22, 1996.

[7] M. Garnier, L. Bailly, M. Dohen, P. Welby, and H. Loevenbruck, "An acoustic and articulatory study of Lombard speech: Global effects on the utterance," in *Ninth International Conference on Spoken Language Processing*, 2006.

[8] R. Schulman, "Articulatory dynamics of loud and normal speech," *The Journal of the Acoustical Society of America*, vol. 85, p. 295, 1989.

[9] B. Lindblom, "Explaining Phonetic Variation: A Sketch of the H&H Theory," in *Speech Production and Speech Modelling*, W. J. Hardcastle and A. Marchal, Eds. Kluwer Academic Publishers, 1990, pp. 403–439.

[10] K. J. de Jong, "The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation," *Journal of the Acoustical Society of America*, vol. 97, no. 1, pp. 491–504, 1995.

[11] T. Cho, Y. Lee, and S. Kim, "Communicatively driven versus prosodically driven hyper-articulation in Korean," *Journal of Phonetics*, vol. 39, no. 3, pp. 344–361, 2011.

[12] P. Hoole and A. Zierdt, "Five-dimensional articulography," *Speech motor control*, pp. 331–349, 2010.

[13] S. Darjaa, M. Černák, M. Trnka, M. Rusko, and R. Sabo, "Effective triphone mapping for acoustic modeling in speech recognition." in *INTERSPEECH*, 2011, pp. 1717–1720.

[14] M. Garnier, N. Henrich, and D. Dubois, "Influence of sound immersion and communicative interaction on the Lombard effect," *Journal of Speech, Language and Hearing Research*, vol. 53, no. 3, p. 588, 2010.

[15] J. Kim, A. Sironic, and C. Davis, "Hearing speech in noise: Seeing a loud talker is better," *Perception*, vol. 40, no. 7, pp. 853–862, 2011.