PROSODIC BOUNDARIES IN LOMBARD SPEECH

Štefan Beňuš[1], Juraj Šimko[2]
[1]Constantine the Philosopher University in Nitra, Slovakia & Institute of Informatics, SAS, Bratislava, Slovakia
[2]Institute of Behavioural Sciences, University of Helsinki, Finland
sbenus@ukf.sk, Juraj.simko@helsinki.fi

ABSTRACT

Communicative intentions in realizing prosodic boundaries and in making speech more intelligible to the listener in ambient noise both utilize variation in F0 and duration. This paper asks how these cues relate when boundary type and the level of noise is varied. Two durational and two F0 measures of boundary strength extracted in the vicinity of boundaries are analyzed. Data suggest relatively weak local hyper-articulation, both cumulative and compensatory relationships between the cues, and subject-specific complementary strategies for cue selection in signalling communicative intentions.

Keywords: Lombard, prosodic breaks, Slovak

1. INTRODUCTION

Lombard speech is an umbrella term for the phonetic adjustments speakers make in response to the communicative requirements imposed by ambient noise [8]. It is typically associated with an increase in overall intensity, greater F0 range, temporal lengthening, flattening of spectral slope, greater center of gravity; in general those adjustments that facilitate speech intelligibility by increasing signal-to-noise ratio [14]. Studies also suggest that the Lombard effect is both an automatic process that results from attenuated feedback from the speaker's own voice, as well as a listener-oriented cognitive process under the control of a speaker who strives to increase intelligibility for the interlocutor [9,4].

It has been observed that these adjustments do not target speech material uniformly but selectively. For example, [12] examined average F0 increase in 90dB noise compared to quiet and found greater increases and variability in the in stressed (accented) than unstressed (unaccented) words. [11] observed that in an interactive communicative task in English, while all words were affected by Lombard rather uniformly at moderate noise levels (60dB), the agents of actions, presumably marked with pitch accents, were disproportionally hyper-articulated in terms of duration and F0 increase in high noise levels (90dB). [18] examining the Lombard effect (binary: no noise vs. 75dB) on the realization of Cantonese lexical tones in low and high frequency words observed that the dispersion of F0 curves was greater in tones of low frequency words compared to high-frequency ones.

These findings support the idea that speakers alter their speech to convey their intention to their interlocutors. Prosody signals speakers'intentions mostly through the distribution of prominences (pitch accents) and boundaries. Several studies have reported, albeit indirectly in some cases, on the Lombard effect on the realization of the prominent words. Little is known about the effect of ambient noise on the realization of prosodic boundaries.

The four most widely used features for marking prosodic boundaries and their strength is the lengthening of the pre-boundary material, F0 incursions, the duration of the silent pause, and the degree of pitch reset across the boundary [6,15,16,17]. These features are interconnected so that, for example, the longer the pause, the more final lengthening speakers produce and listeners expect [10]. Slowing down that lengthens pre-boundary material and the pauses is presumably also linked to greater F0 adjustments in that stronger breaks are realized with expanded pre-boundary F0 excursions and cross-boundary resets. In other words, lengthening is assumed to negatively correlate with undershoot of F0 targets [7].

However, these cues seem to be weighted in language-specific ways (e.g., pause is not needed for German infants while pre-boundary lengthening and pitch re-set are combined [16]. Also, positive correlation between lengthening and boundary strength has been questioned in [5].

The primary goal of this paper is to examine the effect of ambient noise on the realization prosodic boundaries. Specifically, we ask if the temporal and F0 adjustments work in compensatory or additive fashion. Also, we explore the possibility that no additional strategies are employed for local marking of the boundary strength. In other words, the overall hyper-articulation due to noisy environment has an effect of increasing the boundary strength but does not in any particular way targets the prosodic boundary marking locally. Additionally, we wish to expand on understanding communicatively induced hyper-articulation by employing several noise levels to explore finer dynamics of the Lombard effect (typically 2-3 levels in other studies), and a less researched Slavic language Slovak (typically Germanic or Romance languages).

2. METHODS

Three native speakers of Slovak (2F, 1M) read multiple repetitions of 12 Slovak prompt sentences under various noise conditions.

The stimuli sentences contained either *aby* [abi] 'so that' or *iba* [iba] 'only'. In Slovak, both words create syntactic and pragmatic affordance for a prosodic boundary to precede them [1]. The strength and type of this boundary was controlled. Break-0 (B0) corresponds to the weakest disjuncture typical for boundaries between words, Break-1 (B1), was marked in the stimuli with a comma and is typically realized with a continuation rise, and Break-2 (B2), marked with a full stop, was realized with a final fall. The target prosodic boundary appears twice in each sentence in a syntactically coordinative construction. The rhymes of the syllable preceding the boundary were controlled and contained a phonemically long nucleus ([a:] for *iba* sentences and [i:] for *aby* sentences) followed by either [m] or [n]. Hence, the target sequences analyzed in this study are [a:{m,n}(#)iba], or [i:{m,n}(#)abi].

Ambient noise was administered in blocks with 5 repetitions of each sentence in each block and 2 repetitions of each block. In the reference block with no noise, referred to as "0" the subject was instructed to speak naturally. In the "0r" block with intended hypo-articulation, the subject was asked to speak in a relaxed way (this block is not analyzed here). For all other blocks, subjects heard babble noise the headphones in three dB(A) levels: 60, 70, and 80. Finally, the assumed most hyper-articulated speech was elicited with 80 dB(A) noise simulating the communication of the subject with a non-native speaker [3].

Both the blocks as well as the sentences within the blocks were semi-randomized. The design thus included intended 2 repetitions of 6 blocks with 60 sentences each and 2 positions, a total of 1440 tokens per subject.

Data processing was done in three steps. First, automatic forced alignment of the words and phones to the acoustic signal was employed, followed by manual correction of the intervals in the vicinity of the boundaries. Second, a pitch tracking procedure implemented in Praat [2] was used to extract F0 in the vicinity of the boundaries and spurious values removed manually. Finally, labeled interval durations (e.g. pre-boundary rhyme, silent pause) and two F0 features were extracted: 1) for rhymes, a measure of F0 movement, Bounded Variation Norm (BVN) as a sum of absolute values of differences between subsequent F0 samples [13], and 2) F0 reset as absolute difference between mean F0 for the pre-boundary nasal and post-boundary vowel.

3. RESULTS

Fig. 1 shows speakers' response to noise manipulation in the pre-boundary rhyme and pause duration for separate boundary types, Table 1 summarizes post-hoc results.

Figure 1: Rhyme and pause durations normalized to the 1st word per speaker and condition.

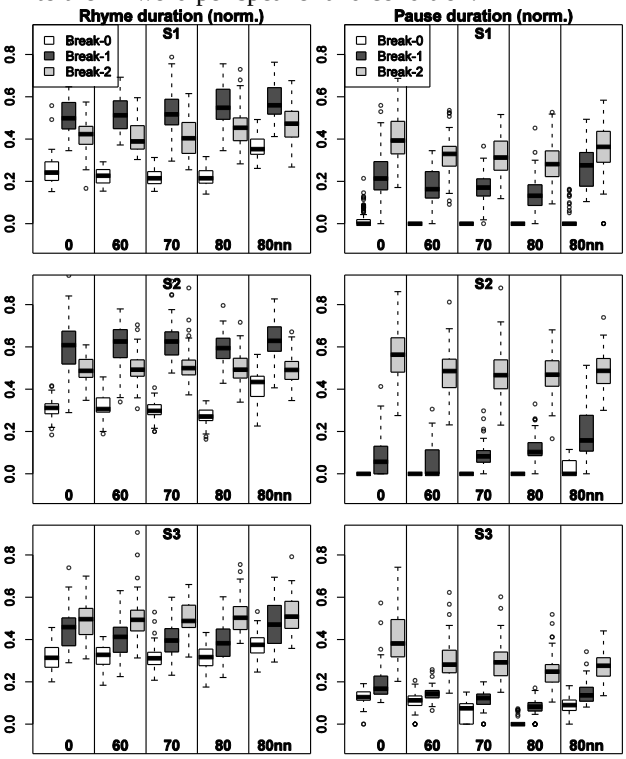


Table 1: Summary of differences between noise conditions (TukeyHSD post-hoc tests, p<0.05:"<")

	Rhyme	Pause
	Break1 (B1)	
S1	60,70,80 < 0 < 80nn	60,70,80 < 0,80nn
S2	ns	0,60,70 < 80,80nn
S3	60,70,80 < 0,80nn	80 < 70 < 60,80nn < 0
	Break2 (B2)	
S1	0,60,70 < 80,80nn	60,70,80,80nn < 0
S2	ns	60,70,80,80nn < 0
S3	ns	80 < 60,70,80nn < 0

In the left columns of Fig. 1 and Table 1, first consider the 'true' intonational boundaries, i.e. major breaks B1 and B2. The normalized rhyme durations are relatively stable. Only weak trends for lengthening with increasing noise can be observed, statistically supported only in S1's 0-70 vs. 80-80nn in B2. Interestingly, pre-boundary rhymes in 0 condition are either not different, or in two cases (S1 and S3 in B1) longer, than in 60-80. These results suggest that the tendency to locally hyper-articulate prosodic boundaries in response to increasing

ambient noise is weak. The rhymes in B0 (not shown in the table) display similar features: they are stable (or even slightly shortened in S1 and S2) for successive 0 → 80, and lengthened for 80nn.

Second, in the right columns of Fig. 1 and Table 1, increasing the noise level does not, in general, result in local lengthening of the pause associated with major prosodic boundaries. In five out of six cases, 0 has actually significantly longer pauses than 60-80 conditions, and S3 shows this trend rather clearly for both break types. The sole exception is S2's B1 with pauses lengthening with increasing noise with a significant 70-80 separation.

Third, subjects also responded differently to stimuli eliciting B1 and B2. While S1 and S2 produced longer pre-boundary rhymes in B2 than B1, S3 followed the opposite pattern. All three subjects had longer pauses in B2 than in B1.

Finally, extreme hyper-articulation in 80nn induces more local lengthening only in some cases.

We now move to examining F0 marking of prosodic boundary strength (PBS) employing BVN and F0 reset, depicted in Fig. 2 and Table 2. They filter to some extent the overall shifts of F0 due to the noise condition and are assumed to target localized adjustments to PBS.

Considering BVN first, we see the greatest effect of noise on marking PBS in S2 with much greater range of F0 movement than S1 and S3 (cf. y-axes), and statistically significant differences among the Lombard conditions for both major break types. S1 shows significant separation of 0-70 and 80-80nn conditions in B1 and, in addition to that, also 80 vs. 80nn in B2. Finally, S3 has a qualitatively different pattern consistent for both boundary types with F0 movement expanding less with increasing noise. In B0 condition, the response of BVN to noise is weaker, with only significant separation of 80nn from the rest for S1 and S2, and significantly greater F0 movement in 0 than the rest in S3.

Finally, consider data in the right columns of Fig. 2 and Tab. 2. S1 and S2 behave similarly but the response to noise is qualitatively different between the two breaks: little separation among the noise conditions in B1, and robust significant differences (greater reset with increasing noise) in B2. S3 again behaves differently with a trend of decreasing reset with noise in B1 and a binary distinction between 0 and all other noise conditions in B2. In sum, S1 and S2 show similar patterns although the amount of variability is smaller in S1 compared to S2. S3 shows qualitatively different behaviour with steady or decreasing F0 movement with noise increase.

We now ask if the F0 and temporal domains work cumulatively or in complementary fashion when responding to the communicative task of marking prosodic boundaries in ambient babble noise. We take our 4 measures of boundary strength and examine 2 pair-wise relationships with linear models for the three subjects separately. 80nn is excluded as it presents a slightly different communicative task from the remaining four conditions. Cumulative and complementary behaviours correspond to positive and negative correlation, respectively.

Figure 2: Bounded Variation Norm (BVN) and F0 reset (in bark) separately for noise conditions.

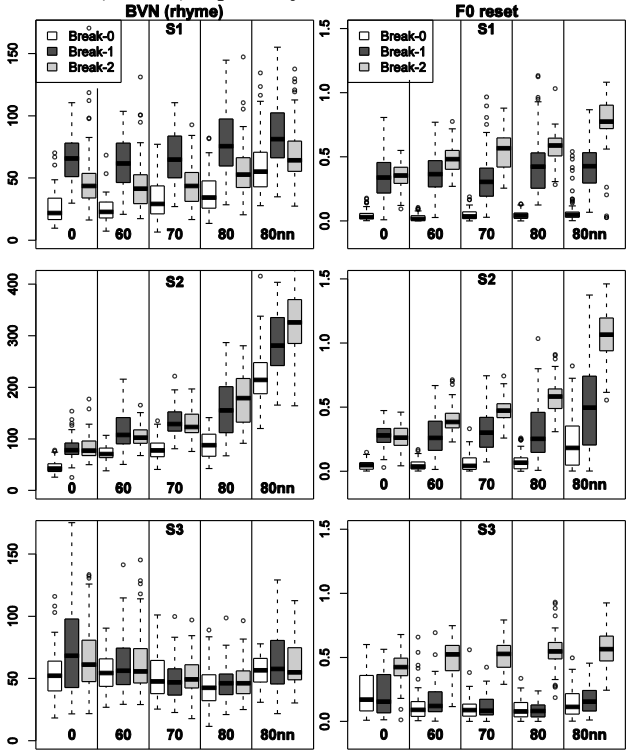


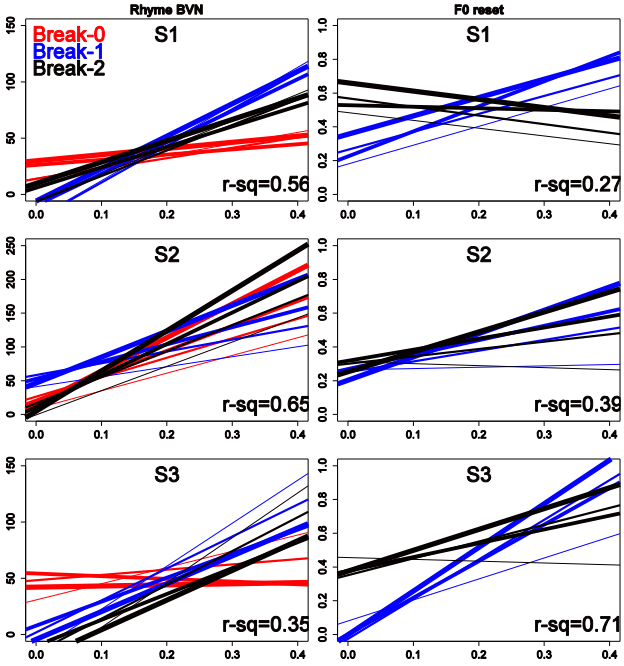
Table 2: Summary of differences between noise conditions (TukeyHSD post-hoc tests, p<0.05: <)

	BVN	F0 reset
	Break1	
S1	0,60,70 < 80,80nn	0,60,70 < 80,80nn
S2	0<60<70 < 80<80nn	0,60,70,80 < 80nn
S3	70,80 < 0,60,80nn	70,80 < 0,60,80nn
	Break2	
S1	0,60,70 < 80 < 80nn	0 < 60 < 70,80 < 80nn
S2	0<60<70 < 80<80nn	0 < 60 < 70 < 80< 80nn
S3	70,80 < 0,60,80nn	0 < 60,70,80,80nn

Consider first the relationship between rhyme duration and the extent of F0 movement over this rhyme (BVN) in the left of Fig. 3. The plots show that the communicative task of marking prosodic boundaries in babble noise is resolved differently for subjects and break types. S1 shows clear separation of the three breaks with increasingly cumulative behaviour of duration and F0 movement from B0 to B2 and B1, while no effect of noise on this

cumulative behaviour can be observed. Differences among the breaks are manifested for S2 in a different order: B1 shows the weakest positive relationship followed by B0, and B2 with the steepest positive slopes. B1, moreover, is marked with a clear increase in intercepts compared to the other two break types. This implies more F0 movement with little lengthening, which correspond to the compensatory behaviour. Finally, S3 displays a mixed behaviour. First, there is a clear difference between B0 and the other two breaks; the former shows no positive relationship irrespective of the noise level, similarly to S1, and the latter exhibit positive cumulative relationship. Yet, the effect of noise conditions can be observed at least when comparing 0 and 80 conditions in that the former has clearly greater slopes than the latter irrespective of the break type, which suggests a move away from the cumulative behaviour with increasing noise

Figure 3: Bounded Variation Norm (BVN) in the rhyme as a function of rhyme duration (left), F0 reset as a function of pause duration (right); thicker line => greater noise.



Finally, consider the relationship for the boundary itself between the amount of F0 reset and pause duration illustrated in the right column of Fig. 3. S1 again shows a clear difference between the 2 break types in that B1 is produced with a cumulative relationship and B2 with negative slopes, i.e., a compensatory relationship between the two measures. The noise is manifested only in increasing intercept, hence greater resets without a corresponding change in pause durations. S2 is different with a very clear effect of noise conditions on the slopes (greater slopes with increasing noise), and a small, but consistent, difference between the break types (B1 have slightly greater slopes than B2). Finally, S3 shows mostly cumulative relationship with positive slopes, similar to S2 in terms of the noise effect but with greater separation between the break types.

4. DISCUSSION AND CONCLUSIONS

We examined the effect of several levels of ambient babble noise on the durational and F0 marking in the vicinity of three types of prosodic boundaries. We found relatively weak local response of increasing noise in the durations of pre-boundary rhymes and silent pauses. Additionally, despite the prevalence of the cumulative relationship between the durational and F0 cues to boundary strength, we have also observed compensatory relationships suggesting trade-offs between the two types of PBS cues.

Our three boundary types were communicatively different, but phonologically, B1 and B2 shared a strong disjuncture, commonly realized with a pause. We found a clear separation between B0 and B1-B2 in both duration and F0 marking of the pre-boundary rhyme but also that ‘phonologically weak’ B0 in extreme hyper-articulation might be signalled similarly to ‘phonologicaly strong’ B1 or B2 in no or low noise. Regarding the relationship between B1 (rise) and B2 (fall), B1 has stronger cumulative relationship between F0 and duration in responding to noise than B2, possibly due to greater scaling of F0 maxima compared to minima or median.

Finally, we have uncovered a complex interaction between subjects and boundary types. Subjects behave differently, most clearly S1-S2 tend to mark PBS associated with noise on pre-boundary rhyme while S3 on the cross-boundary interval. S1 and S3 also clearly distinguish the boundary type by relationship between durational and intonational properties of the rhyme and cross-boundary interval. S2, on the other hand, keeps the relationships relatively stable but exhibits a consistent gradual response to the noise level in Lombard speech. This suggests possible complementary strategies for negotiating the dual task of marking the boundary and keeping the distinction between different communicative boundary types clear in adverse external conditions.

ACKNOWLEDGMENT

This material is based upon work supported by the Air Force Office of Scientific Research, Air Force Material Command, USAF under Award No. FA9550-15-1-0055, and was also supported in part by grant 2/0197/15 by Scientific Granting Agency.

5. REFERENCES

[1] Beňuš, Š., Šimko, J. 2014. Emergence of prosodic boundary: continuous effects of temporal affordance on inter-gestural timing. *Journal of Phonetics*, 44, 110–129.

[2] Boersma, P., Weenink, D. 2005. Praat: Doing phonetics by computer. (Version 5.3.12) [Computer program]. Retrieved from http://www.praat.org/.

[3] Cho,T., Lee,Y., Kim,S. 2011. Communicatively driven versus prosodically driven hyper-articulation in Korean. *Journal of Phonetics*, 39(3), 344–361.

[4] Garnier, M., Bailly, L., Dohen, M., Welby, P., Loevenbruck, H. 2006. An Acoustic and Articulatory Study of Lombard Speech: Global Effects on the Utterance. In *Proceedings of Interspeech*, 17–21.

[5] Heldner, M., Megyesi, B., 2003. Exploring the Prosody-Syntax Interface in Conversations. *Proc. 15th ICPhS* Barcelona, 2501–2504.

[6] Krivokapic, J., Byrd. D. 2012. Prosodic boundary strength: An articulatory and perceptual study. *Journal of phonetics*, 40, 430–442.

[7] Lindblom, B. 1999. Emergent phonology. Proc. 25th annual meeting of the Berkeley Linguistics Society. Berkeley: University of California.

[8] Lombard, E. 1911. Le Signe de l'Elevation de la Voix. *Ann. Malad. l'Oreille Larynx* 37, 101–19.

[9] Lu, Yan-Chen, and Martin Cooke. 2010. Spectral and Temporal Changes to Speech Produced in the Presence of Energetic and Informational Maskers. *Journal of the Acoustical Society of America* 128(4),: 2059–70.

[10] Noteboom, S. 1997. The prosody of speech: Melody and Rhythm. In: Hardcastle, W., Laver, J. (eds), *The Handbook of Phonetic Science*. Oxford: Blackwell, 640–673

[11] Patel, R., Shell, K. W. 2008. The influence of linguistic content on the Lombard effect. *Journal of Speech, Language, and Hearing Research* 51, 209–220.

[12] Rivers, C., Rastatter, M.P. 1985. The effects of multitalker and masker noise on fundamental frequency variability during spontaneous speech for children and adults. *Journal of Auditory Research* 25(1), 37–45.

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

[14] Van Summers, W, Pisoni, D., Bernacki, R., Pedlow, R., Stokes, M. 1988. Effects of Noise on Speech Production: Acoustic and Perceptual Analyses. *Journal of the Acoustical Society of America*. 84, 917–28.

[15] Wagner, M., Watson, D. 2010. Experimental and theoretical advances in prosody: A review. *Language and Cognitive Processes* 25, 905–945.

[16] Wellmann, C., Holzgrefe, J., Truckenbrodt, H., Wartenburger, I., Hohle, B. 2012. How each prosodic boundary cue matters: Evidence from German infants. *Frontiers in psychology* 3: 580.

[17] Wightman, C., Shattuck-Hufnagel, S., Ostendorf, M., Price, P. 1992 Segmental durations in the vicinity of prosodic phrase boundaries. *Journal of the Acoustical Society of America* 91: 1707–1717.

[18] Zhao, Y., Jurafsky, D. 2009. The effect of lexical frequency and Lombard reflex on tone hyperarticulation. *Journal of Phonetics* 37, 231–247.