THE EFFECT OF VOWEL LENGTHENING ON THE INTELLIGIBILITY OF OCCLUDED LOMBARD SPEECH

Xinyi Zhang[*1], **Meghan Clayards**[†2,3], **and Rachel E. Bouserhal**[‡1]
[1]Department of Electrical Engineering, École de technologie supérieure, Montréal, Québec, Canada
[2]School of Communication Sciences and Disorders, McGill University, Montréal, Québec, Canada
[3]Department of Linguistics, McGill University, Montréal, Québec, Canada

1 Introduction

When speaking in noise, people's vocal effort is automatically adjusted in order to improve the intelligibility of their speech; this phenomenon is called the Lombard effect [1]. However, when the ear canals are blocked by, for example, hearing protection devices or hearables, the occlusion of the ears causes significant changes in speech production, particularly in noise [2]. The result is a Lombard speech that is less intelligible than its typical form; we refer to this as occluded Lombard speech (*OLS*). The change in speech production is due to the lowered noise levels at the ear canals caused by the passive attenuation of the in-ear devices, as well as a change in the talkers' perception of their own voice. One main characteristic of the open-ear Lombard speech is elongated vowels. However, by analyzing the SpEAR database [3], an open-access bilingual in-ear speech database that includes OLS, this elongation has been found to be much shorter in OLS. This discrepancy leads to our research question: Can elongating the vowels in the OLS improve its intelligibility?

2 Method

Thirteen people from the Great Montreal Area volunteered to participate in this research. All participants self-reported to be native speakers of French with no history of speech or hearing disorders.

The sentence recordings (stimuli) that the participants listened to were derived from the French portion of SpEAR. Hearing In Noise Test sentences were used as the speech materials in this database. Seventy-eight OLS sentences with relatively similar speech levels were selected.

These sentences were further modified to have different signal-to-noise ratios (SNRs) and different vowel lengthening conditions. Four SNRs were chosen: -3, 0, 5, and 10 dB. The SNR manipulation was achieved by modifying the sound level of the sentence recordings relative to a noise (same for all conditions). Additionally, in each SNR condition, half of the sentences had their vowels elongated to 1.5 times of the original length (an addition of 48 ms on average).

The experiment was conducted in a sound-attenuated booth. On each trial, participants heard one of the stimuli, repeated and recorded what they heard, then rated how easy it was to understand using a rating scale. Participants completed each task at their own pace. The accuracy was calculated using the number of words repeated correctly out of the total number of words in a given sentence. The reaction time for each stimulus was measured by the difference between the time that a stimulus finished playing and the time that the participant clicked to start the repetition recording. After each sentence repetition, the participants were instructed to rate the ease of understanding of the stimulus on a 1000-pixel long slider with the leftmost being the most difficult and the rightmost being the easiest.

[*]xinyi.zhang.1@ens.etsmtl.ca
[†]meghan.clayards@mcgill.ca
[‡]rachel.bouserhal@etsmtl.ca

3 Results

Three measurements of the intelligibility of the stimuli are presented: the accuracy of the sentence repetition task, the participant-rated stimuli ease of understanding (EoU), and the reaction time (RT) for the sentence repetition task in milliseconds, where a shorter RT indicates a higher intelligibility.

Sentence repetition task word accuracy

Table 1: Descriptive statistics for sentence repetition accuracy.

	Unlengthened		Lengthened	
SNR	*Mean*	*SD*	*Mean*	*SD*
-3	0.61	0.38	0.50	0.37
0	0.82	0.30	0.81	0.27
5	0.96	0.12	0.90	0.20
10	0.98	0.08	0.89	0.20

Table 1 shows that with the increase of SNR, the sentence repetition accuracy increases and a ceiling effect can be observed. The group mean accuracy of the unlengthened conditions are always higher than in the lengthened condition except at SNR = 0, where they coincide. At SNR = 5 and SNR = 10, the group means of unlengthened condition are almost perfect with a small SD of approximately 0.1. However, group means plateaued at around 0.9 with a larger SD of 0.2 for the lengthened condition.

Participant-rated ease of understanding

As can be seen from Figure 1, on average, as the SNR goes up, so does the EoU. However, a large SD is observed at all SNRs for both of the vowel-lengthening conditions. In addition, the average EoU in the vowel-lengthened conditions is consistently lower than in the unlengthened conditions with the exception of SNR = 0.

Reaction time for sentence repetition task

As can be seen from Figure 2, for SNR from -3 to 5, RT negatively correlates with the SNR level as expected. However,

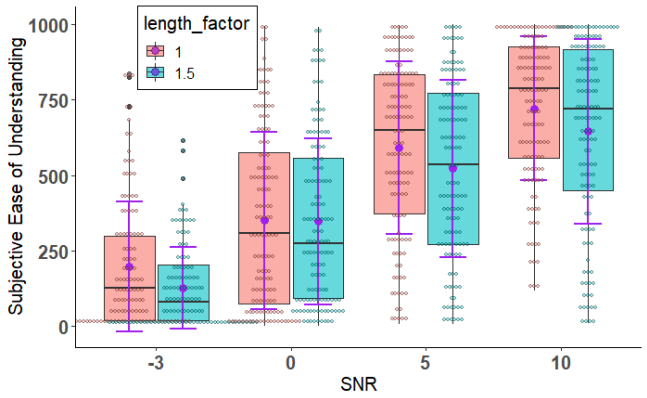

Figure 1: A box-plot overlaid with a dot-plot that compares the EoU of the stimuli, grouped by vowel lengthening condition and SNR levels. Group means and standard deviations are indicated by purple.

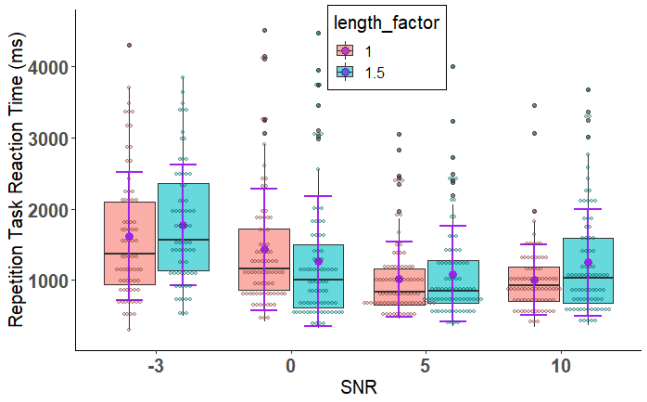

Figure 2: A box-plot overlaid with a dot-plot that compares the reaction time for the sentence repetition task, grouped by vowel lengthening condition and SNR levels. Group means and standard deviations are indicated by purple.

at SNR = 10, RT seems to have stagnated or even increased. The average RT of the unlengthened conditions are lower than of the lengthened ones at all SNRs but SNR = 0, where the lengthened condition has a lower RT.

4 Discussion and Conclusions

As shown in the results section, increasing SNR correlates with an increase in intelligibility. Comparing the two vowel-lengthening conditions, in general, vowel-lengthening seemed to correlate with a worsened intelligibility instead of an improvement. This may have happened because of the unnaturalness from the elongation manipulation (and/or its artefacts), even though the current degree of lengthening was chosen to have natural-sounding stimuli.

Another limitation related to the quality of naturalness of the stimuli was their non-communicative nature. Since Lombard speech is an instinctual adaptation for improving *communication* in noise, the elongation without the context of communication may have made the sentences sound less natural.

Interestingly, the participants seem to have behaved differently at SNR = 0 than at the other SNRs. To elaborate, the intelligibility of the stimuli at SNR = 0 did not seem to be worsened by vowel lengthening; in fact, looking at the reaction time results, it even seemed to have improved intelligibility. However, a large SD is seen for the unlengthened SNR = 0 condition compared to the other SNRs. Thus, it is possible that the unlengthened SNR = 0 stimuli were perceived to be less intelligible than they should have been, instead of the lengthening here being a (hypothetical) anomaly. Further inferential statistics with mixed-effects modelling needs to be performed for this apparent finding. Multivariate models that take into account all three measurements simultaneously may be preferred to have a higher power of analysis. In the mixed-effects modelling, in addition to treating participants and sentences as random effects, the speakers of the stimuli, eight in total, should be considered, too. It has been noticed that the speaker composition is not the same for every condition, and the intelligibility of the stimuli may vary due to the speakers' accents (given that all participants spoke Quebec French), especially in noise.

In case this finding should be statistically significant, why SNR = 0 is a special case here may be investigated with follow-up studies. Could it be that the effect of lengthening depends both on SNR and the *degree* of elongation, and the currently selected degree (i.e., 1.5 times) happened to be closer to the (hypothetical) optimal elongation at SNR = 0, compared to the other SNR - elongation combinations? Or could it be that SNR = 0 is in some way perceptually *categorically* different from the other SNR levels?

To conclude, the effects of vowel-lengthening on the intelligibility of occluded Lombard speech are still unclear, but it is unlikely that the current manipulation should improve intelligibility, especially independent of the SNRs. What cannot be ruled out, though, is the possibility that if the stimuli were perceived to be more natural-sounding, vowel-lengthening could still be a valid approach for intelligibility improvement.

Acknowledgments

The authors thank the previous work done for this research project by Colin Jones in Summer 2019, the funding from Natural Sciences and Engineering Research Council, and Fonds de recherche du Québec- Nature et technologies.

References

[1] Henrik Brumm and Sue Anne Zollinger. The evolution of the lombard effect: 100 years of psychoacoustic research. *Behaviour*, 148(11-13):1173–1198, 2011.

[2] Jennifer B Tufts and Tom Frank. Speech production in noise with and without hearing protection. *The Journal of the Acoustical Society of America*, 114(2):1069–1080, 2003.

[3] Rachel E. Bouserhal, Antoine Bernier, and Jérémie Voix. An in-ear speech database in varying conditions of the audiophonation loop. *The Journal of the Acoustical Society of America*, 145(2):1069–1077, February 2019.