THE LOMBARD EFFECT ON THE VOWEL SPACE OF NORTHERN VIETNAMESE

Giang Le[1], Yan Tang[1,2]

[1]Department of Linguistics, University of Illinois Urbana-Champaign, USA

[2]Beckman Institute for Advanced Science and Technology, University of Illinois Urbana-Champaign, USA

gianghl2@illinois.edu, yty@illinois.edu

ABSTRACT

This study investigates the Lombard effect triggered at different noise levels in Northern Vietnamese, as measured by formant, vowel space area, and vowel dispersion changes. As Lombard speech is generally found more intelligible than speech produced in quiet, we hypothesize that the vowel space area and between-group vowel dispersion would increase in general, in order to maximize vowel-tone contrasts. Acoustic analyses on speech produced in quiet and two noise levels show that contrary to expectation, the vowel space area decreased significantly with the increase of background noise level, as a consequence of the increase of the first formant for most vowels and the decrease of the second formant for non-back vowels. However, the variability within vowel groups decreased, showing better clustering of the same vowel sounds within the vowel space. These findings present a complex picture of the intelligibility of Lombard speech in a tonal language.

Keywords: Lombard speech, Northern Vietnamese, tone, vowel space

1. INTRODUCTION

The Lombard effect is talkers' involuntary, physiological, and vocal responses when speech communication takes place in a noisy environment. It has been extensively studied in the literature. While conclusive findings have been identified and reported for certain features of Lombard speech (e.g. increased duration of sonorous segments [1, 2, 3], increased F0 [1, 4, 3, 5], flattened spectral tilt [3], among others), findings regarding the vowel space have not always been consistent. Some studies reported changes in F1 and F2 for particular vowels of Lombard speech; others found that the vowel space area could either increase or decrease.

Often considered to be a type of clear speech, Lombard speech exhibits acoustic properties that other clear speech registers also do. For example, infant-directed speech (IDS) was found to have increased vowel duration and F0. Similarly, Ping et al. (2017) observed vowel space expansion in both IDS and Lombard speech. The authors explained that in an expanded vowel space, the vowel categories have less overlap and their contrasts are maximized [6]. This is conducive to the infants' perceptual development of phonemic contrasts that IDS helps to facilitate, and better intelligibility to overcome energetic masking in non-ideal listening environments. Smiljanic et al. (2017) also reported that noise-adapted speech and clear speech share similar features, such as a decreased speaking rate and increased energy in 1-3 kHz, sound pressure level (SPL), vowel space area and harmonics-to-noise ratio. Studying the acoustic-phonetic characteristics of speech produced in noise while wearing an oxygen mask [7], Bond et al. (1989) also reported a somewhat increased vowel area.

Other studies, however, reported either no evidence of vowel space expansion [8, 9, 10, 11] or vowel dispersion reduction [12]. Scobbie et al. (2012) remarked that there was a shift in the F1-F2 space due to an F2 increase for the back vowels, but overall the vowel space did not increase in area. Kim et al. (2013) concluded that there was a tendency for the formant values to be produced with less variation, but overall the vowel space was not expanded for speech produced in noise. In a study about speech produced in MRI noise, Gully et al. (2019) observed that the vowel space dispersion decreased, whereby the vowel space dispersion was measured by the mean Euclidean distance of vowel space vertices from the centroid.

In light of these mixed findings, this study analyzed the first (F1) and second (F2) formants, vowel space area formed by four vowels (/a/, /e/, /o/ and /u/), and a vowel space dispersion in Northern Vietnamese Lombard speech. This was to understand better how the Lombard effect impacts the articulation of vowels in a tonal language.

2. METHOD

2.1. Participants

Six native speakers of the Northern Vietnamese variety (three males, three females) between the ages of 19 and 34 were recruited to participate in this study. A hearing screening showed that all the participants had a normal hearing level.

2.2. Materials

Seventy-eight tokens in vowel, consonant-vowel and consonant-vowel-consonant formants were created for this study. These stimuli were embedded in a carrier sentence *"Toi noi cho ban nghe X bay gio"* (I say X to you now), where X was one of the target tokens. All the stimuli contained one of the three corner vowels /e/, /a/ or /u/. The vowel /i/ was not chosen because of the higher incidences of vulgar words when combined with certain consonants in the stimuli, which could trigger the participants to react in an unexpected manner. /o/ was also elicited as a backup for /u/ for one syllable template due to the same concern. All the syllables were combined with all possible tones. Note that not all the tokens are possible meaningful words in Vietnamese, but the combination of the tones on them is phonotactically possible.

2.3. Procedure

Three recording sessions took place in a sound-treated audio booth. Apart from the control condition, where participants uttered the stimuli *in quiet* ("quiet"), speech-shaped noise was presented to participants at 78 dB SPL ("78-dB SPL") and 90 dB SPL ("90-dB SPL") over a pair of open-back headphones in the other two sessions to elicit Lombard effect of different degrees. The speech-shaped noise was generated to have a representative long-term average spectrum of the Northern Vietnamese variety; it resulted in similar energetic masking on the target speech across frequencies. The participants were instructed to speak the stimuli displayed on a computer screen aloud twice each. While the stimuli were displayed to the participants in groups of the same syllable base, the tone ordering was randomized and the order by which the syllable bases were presented was also randomized. The participants could control how fast they moved through the stimuli via a mouse click. The recordings were saved as WAV files sampled at a rate of 44.1 kHz. In total, 5,136 vowel tokens were collected and analyzed.

2.4. Northern Vietnamese tones

Vietnamese is a tone language where a syllable can carry different pitch patterns, signifying semantic contrasts. Six phonemic tones in Northern Vietnamese can be described as follows.

Tone A1 is a level tone spoken with a modal voice. **Tone A2** is a low to mid-falling tone usually spoken in a modal voice but could also be spoken with a lax or breathy voice [13]. **Tone B1** is a mid-rising tone spoken with a modal voice. **Tone B2** is a mid-falling tone with strong glottalization at the end, or mid-falling *with creakiness*. **Tone C1** is also a falling tone, with a similar F0 contour as tone A2, but with slight laryngealization at the end [13]. Some speakers realize this tone with a mid-falling-rising contour, *similar to the contour of C2*. **Tone C2** is a rising tone with a glottal interrupt in its first half, also known as mid-rising *with creakiness*.

3. RESULTS

3.1. F1

A three-way ANOVA with repeated measures suggested that the effects of noise, vowel, and tone were found significant on the first formant in Hz across the examined sonorous segments and so were the two-way interactions [$\forall p < 0.001$]. Post-hoc comparisons looking at the vowel-noise interaction confirmed a significant increase in F1 for all the vowels from "quiet" to "78-dB SPL" and to "90-dB SPL" [$\forall p < 0.001$], except for the low vowel /a/ (Fig 1). Post-hoc comparisons looking at the tone-noise interaction showed that F1 increased for all tones from "quiet" to "78-dB SPL" [$\forall p < 0.001$] but there was no significant difference between the two noise conditions. The only exception was tone C2, where F1 increased significantly in "90-dB SPL" compared to "78-dB SPL" (Fig 2).

3.2. F2

For F2, a three-way ANOVA with repeated measures also found that noise, vowel, and tone as the main effects, and bi-factor interactions (i.e. noise-vowel, vowel-tone) were significant [$\forall p < 0.001$]. Post-hoc comparisons looking at the noise-vowel interaction showed that F2 increased significantly from "quiet" to "78-dB SPL" and to "90-dB SPL" [$p < 0.001$] for the back vowel /u/ but decreased significantly for

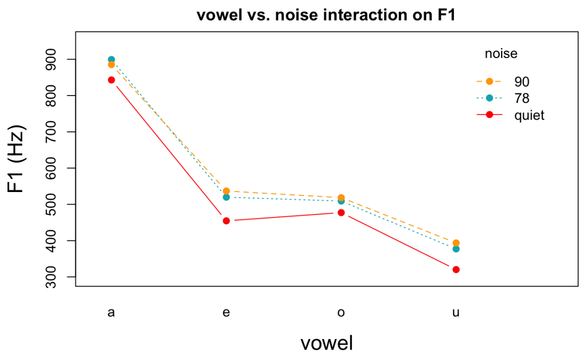

Figure 1: Interaction between vowel and noise on F1

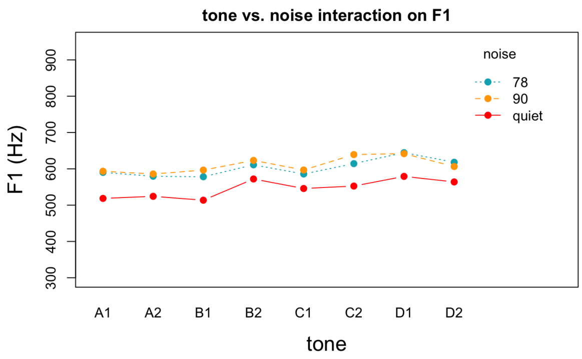

Figure 2: Interaction between tone and noise on F1

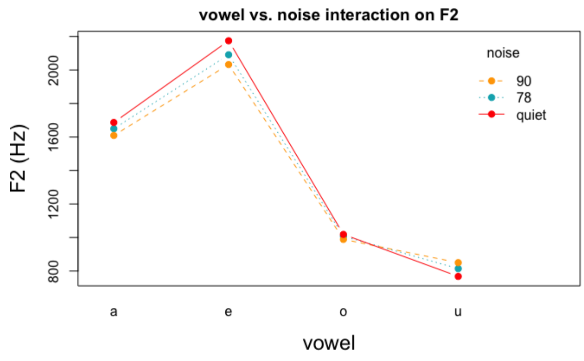

Figure 3: Interaction between vowel and noise on F2

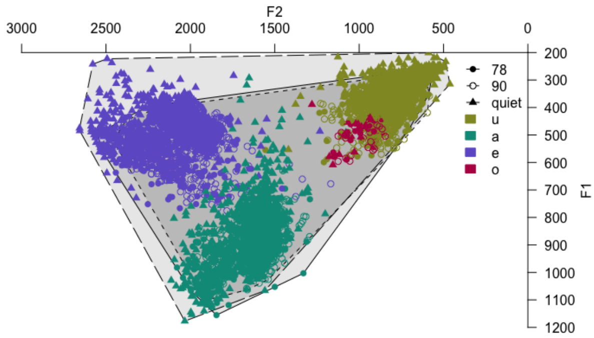

Figure 4: Vowel space area for all vowel tokens

vowels /a/ and /e/ (Fig 3). The increase in F2 for a back vowel result agreed with a previous finding by Scobbie et al. (2012).

3.3. Vowel space area and dispersion

The areas of the vowel space by tones and vowels were calculated as the area of the convex hull enveloping all the vowel tokens of each group, using the *phonR* package in R [14]. A two-way ANOVA showed that having vowel and noise as predictors and controlling for tones, noise has a significant main effect on the convex hull area [$p <$ 0.001, F=34.257] but vowel has no significant effect; specifically, the area decreased by 116444.3 Hz^2 in "78-dB SPL" compared to "quiet", and 120248.2 Hz^2 in "90-dB SPL" compared to "quiet". There was no significant difference in the vowel space area between the two noise conditions. The same result was found when controlling for vowels and having tone, noise as the main effects (See Fig 4). Post-hoc comparisons looking at the tone-noise interaction showed that while all phonemic tones followed the general pattern, tone C2 only had a significant difference in the vowel space area between "78-dB SPL" compared to "quiet" but not between the other conditions (See Fig 5).

We further examined the within- and between-group dispersions of the F1-F2 (Hz) vowel space area for different vowel groups. In the first measure, three centroids were identified for the three vowel groups, and the within and between-group sums of squares (SS) were calculated (See Table 1). It is evident that both the within and between SS measures decreased going from "quiet" to noise conditions, as an empirical observation. The second measure of dispersion is based on Whitfield and Goberman (2014)'s articulatory-acoustic vowel space's standardized general variance (SGV) [15], which is calculated as the positive p-th root of the determinant of a variance-covariance matrix of F1 and F2 vectors. The SGV can be considered a bivariate standard deviation and thus reflects the degree to which the data spreads out in an F1-F2 space. Our calculated values showed that the SVG decreased from "quiet" (133990.1) to "78-dB SPL" (115028.3), and further to "90-dB SPL" (101029.6).

4. DISCUSSION AND CONCLUSION

As F1 is inversely correlated to vowel height in general, producing a further higher F1 in Lombard

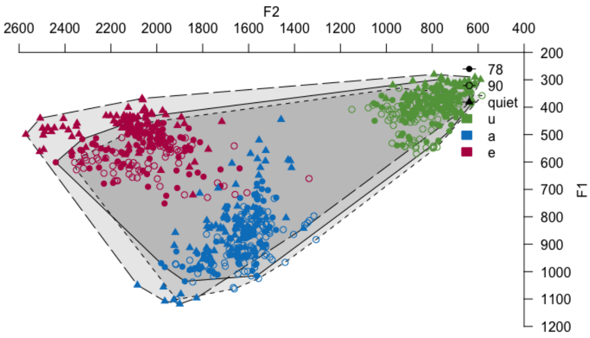

Figure 5: Vowel space area for vowels with tone C2

Table 1: Measures of dispersion within and between clusters

	total within SS	between SS
quiet	63,860,706.00	6.61E+08
78 dB SPL	44,502,705.00	5.66E+08
90 dB SPL	39,729,236.00	4.82E+08

speech could be difficult as a speaker's capacity for hyper-articulation could have reached the ceiling. This may explain why this study observed a significant increase in F1 for all the vowels but the low vowel /a/. Gully et al. (2019) also noted that the increase in F1 in noise is more evident for more closed vowels than for open vowels. As an increase in F1 is associated with a greater degree of jaw-lowering, the results in this study imply varying degrees of hyper-articulation for vowels in Lombard speech.

While F1 for most vowel sounds appeared to increase consistently, F2 seemed less predictable with the pattern changed over vowel backness as exhibited in Fig. 1 and 3. F1 is understood to be correlated with the volume of the back cavity in the vocal tract; hyperarticulation triggered by the Lombard effect could require the jaw to lower further in order to form a larger front cavity, leading to a further contraction of the back cavity, hence the increased F1. Because of the expansion of the front cavity, F2 decreased as a result, especially for non-back vowels such as /a/ and /e/. However, the F2 of back vowels such as /o/ and /u/ appear to be less affected by the change of the vocal configuration in noise, or even be affected conversely, possibly due to the complex acoustic coupling of the two cavities in the vocal tract when different sounds are produced, as shown in Fig. 3. Further investigation is thus warranted. Contrary to some previous findings, we found a decrease in the vowel space area going from quiet to higher noise levels (i.e. "78-dB SPL" and "90-dB SPL") in general. The observed increase in F1 for /e/, /o/ and /u/ but insignificant change for /a/ and a decrease in F2 for /a/ and /e/ but an increase for /u/ together seems adequate enough to shift the centroid of the vowel groups towards the convergence, hence the contraction of the overall vowel space formed by the four vowels as illustrated in Fig. 4. In the meantime, clearer clustering between vowel groups might interfere with the vowel space in the sense that this reduction in variability also leads to a reduced vowel space area produced by speakers in noise. How these two factors interact and what implications they have for the perception of vowels in noise are not yet clear; nonetheless, that reduced vowel space but tighter, more evident clustering might suggest a different type of intelligibility than a broader vowel space might. Additionally, while an expanded vowel space tends to be observed from "clear" speech, such as infant-directed speech or hearing impaired-directed speech, a reduced vowel space can happen in Lombard speech due to a change in speaking rate. Cowley (2020) also reported similar findings obtained here: the vowel space area decreased in Lombard speech. The author speculated that distracting noise could induce a faster speaking rate, leading to reduced articulatory movements and a smaller vowel space area [16]. Our findings also lent indirect support to the dubious correlation between speech's loudness and an expanding vowel space, as observed in [17], where elicited loud speech did not necessarily lead to formant variation and changes in the vowel space.

Our previous research [18] identified limitations to hyper-articulation as noise conditions were manipulated. For example, a reduced Lombard effect following increased noise levels, as measured by the range of F0 values, was detected in tone C2. In this study, we found that different vowels induced different degrees of formant changes in noise: F1 increase was not observed for the low vowel /a/ and F2 changes differed between back vowel and non-back vowels. Given that formants are resonances intensified by filtering effects in the supralaryngeal vocal tract, our findings suggested that filtering effects were impacted by hyper-articulation in varying degrees in Lombard speech. A reduced vowel space suggested that like loudness, an expanded vowel space was not always an automatic response to noise for better intelligibility. The high number of phonemic tonal contrasts in the language calls for greater intelligibility of Lombard speech, which is achieved by tighter clustering rather than increased vowel area.

5. REFERENCES

[1] J.-C. Junqua, "The lombard reflex and its role on human listeners and automatic speech recognizers." *Journal of the Acoustical Society of America*, vol. 93, no. 1, pp. 510 – 524, 1993.

[2] T. D. Hanley and M. D. Steer, "Effect of level of distracting noise upon speaking rate, duration and intensity." *Journal of Speech & Hearing Disorders*, vol. 14, pp. 363 – 368, 1949.

[3] 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 *INTERSPEECH*, 2006.

[4] W. Van Summers, D. Pisoni, R. Bernacki, R. Pedlow, and M. Stokes, "Effects of noise on speech production: Acoustic and perceptual analyses." *Journal of the Acoustical Society of America*, vol. 84, no. 3, pp. 917–928, 1988.

[5] Y. Lu and M. Cooke, "Speech production modifications produced in the presence of low-pass and high-pass filtered noise," *The Journal of the Acoustical Society of America*, vol. 126, no. 3, pp. 1495–1499, 2009. [Online]. Available: https://doi.org/10.1121/1.3179668

[6] T. Ping, R. Nan Xu, I. Yuen, and K. Demuth, "Phonetic enhancement of mandarin vowels and tones: Infant-directed speech and lombard speech." *Journal of the Acoustical Society of America*, vol. 142, no. 2, pp. 493 – 503, 2017. [Online]. Available: http://www.library.illinois.edu.proxy2.library.illinois.edu/proxy/go.php?url=https://search-ebscohost-com.proxy2.library.illinois.edu/login.aspx?direct=true&db=mah&AN=124939671&site=eds-live&scope=site

[7] Z. D. Bond, T. Moore, and B. Gable, "Acousticâphonetic characteristics of speech produced in noise and while wearing an oxygen mask," *The Journal of the Acoustical Society of America*, vol. 85, pp. 907–12, 03 1989.

[8] C. Davis and J. Kim, "Is speech produced in noise more distinct and/or consistent?" in *Australasian Speech Science and Technology Association*, 2010.

[9] E. Godoy, M. Koutsogiannaki, and Y. Stylianou, "Approaching speech intelligibility enhancement with inspiration from lombard and clear speaking styles," *Computer Speech Language*, vol. 28, no. 2, pp. 629–647, 2014. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0885230813000788

[10] J. Scobbie, J. Ma, and J. White, "The tongue and lips in lombard speech: A pilot study of vowel-space expansion," 09 2012.

[11] J. Kim and C. Davis, "Comparing the consistency and distinctiveness of speech produced in quiet and in noise," *Computer Speech and Language*, vol. 28, no. 2, pp. 598 – 606, 2014, cited by: 16. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84890564861&doi=10.1016%2fj.csl.2013.02.002&partnerID=40&md5=ee7f5403484987afa0fcd3203a051fca

[12] A. J. Gully, P. Foulkes, P. French, P. Harrison, and V. Hughes, "The lombard effect in mri noise," 2019.

[13] V. L. Nguyen and J. Edmondson, "Tones and voice quality in modern northern vietnamese: Instrumental case studies," *Mon-Khmer Studies*, vol. 28, pp. 1–18, 1997.

[14] D. R. McCloy, *phonR: tools for phoneticians and phonologists*, 2016, r package version 1.0-7.

[15] J. Whitfield and A. Goberman, "Articulatory-acoustic vowel space: Application to clear speech in individuals with parkinson disease," *Journal of Communication Disorders*, vol. 51, 09 2014.

[16] C. M. Cowley, "The effects of distracting background audio on speech production," MS Thesis, Brigham Young University, 2020.

[17] L. Koenig and S. Fuchs, "Vowel formants in normal and loud speech," *Journal of Speech Language and Hearing Research*, vol. 62, pp. 1278–1295, 04 2019.

[18] G. Le, C. Shih, and Y. Tang, "Distortion in tone production due to the lombard effect," 2021, 1st International Conference on Tone and Intonation (TAI) 2021 : Tone and Intonation in a globalized, digital world, TAI2021 ; Conference date: 06-12-2021 Through 09-12-2021.