VOWEL PRODUCTION CHANGES UNDER NOISE WITH CONSIDERATION OF LOW-ORDER FORMANT MASKING

Yasufumi Uezu, Masato Akagi, Masashi Unoki

Japan Advanced Institute of Science and Technology
{y-uezu, akagi, unoki}@jaist.ac.jp

ABSTRACT

There have been numerous studies on the effects of noise on speech production. However, it is not well understood how noise, which partially obscures the amplitude spectrum of speech, affects the lower-order formants (F1 and F2) that are crucial for vowel perception and production. This study aims to investigate the effects of noise type and noise level on the vocal intensity, amplitude and frequency of F1 and F2 during a speech task of Japanese vowels /a/ and /i/ by a solitary speaker under noise conditions. The noise was designed to mask either both formants, only F1, or only F2. Our findings showed that the amplitude and frequency of the vowel formants varied systematically with the type and level of noise, which we attributed mostly to the Lombard effect and formant masking. However, some results could not be explained by these factors, suggesting that auditory spectral representation may offer further insight.

Keywords: Vowel production, Lower-order formants, Masking noise, Lombard effect

1. INTRODUCTION

Sensory feedback, such as auditory feedback, plays a crucial role in speech motor control in the process of speech production. The Lombard effect [1–4], which refers to a reflexive increase in vocal intensity and fundamental frequency (f_o) in the presence of noise, has been well-known as a phenomenon in speech production under noisy conditions.

There have been numerous studies on the Lombard effect (i.e. [5–15]) that have used different types of broadband noise (e.g., white, speech-shaped, and bubble) and various sound pressure levels (SPLs). Some studies have also investigated the effects of specific types of noise present in real-world environments [16–19], as well as the impact of noise type and noise level on speech production using broadband and filtered noise [20, 21].

Lu and Cooke [20] investigated the effects of various noise types on vocal intensity, f_o, spectral center of gravity, and first formant (F1) frequency using a sentence reading task in the presence of noise. Five types of noise (broadband speech-shaped, two low-pass filtered, and two high-pass filtered) were presented to participants through headphones at 89 dB SPL. The results showed that all of these acoustic parameters were significantly increased in all conditions compared to the quiet condition. Stowe and Golob [21] examined the effects of various types and levels of noise on vocal intensity, f_o, and duration using a picture naming task in the presence of noise. Three types of noise (broadband, notched, and bandpass) were presented to participants through headphones at 75 or 90 dB SPL. The results showed that broadband significantly increased intensity, duration, and f_o, while notched had no effect, and bandpass increased intensity and duration to a lesser extent than broadband, but had no effect on f_o.

These researches have shown that noise that partially obscures the amplitude spectrum of speech can have varying effects on speech production depending on its type and level. However, it is not well understood how noise that partially masks the amplitude spectrum of speech affects the lower formants (F1 and F2), which are an essential factor in vowel perception and production. Additionally, there have been few studies on changes in formant amplitude in speech under noise, which should be considered in conjunction with formant frequency.

The purpose of this study is to investigate the effects of noise type and level on vocal intensity, formant amplitude, and formant frequency during a speech task involving the production of Japanese vowels by a solitary speaker in the presence of noise. The noise was designed to mask one or both of the F1 and F2 on the amplitude spectrum. We expected that formant amplitude would increase with increasing vocal intensity, regardless of the vowel type. In contrast, we hypothesized that the shift in formant frequency would be influenced by the type and level of the noise, as well as the type of vowel being produced.

2. EXPERIMENT

2.1. Participants and apparatus

Ten adult male native Japanese speakers participated in the experiment (mean 25.2 ± 3.79, 22-35 years

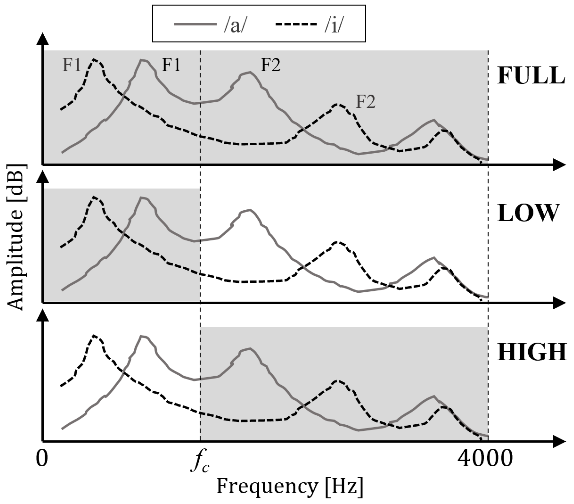

Figure 1: The amplitude spectrum of each noise type, the amplitude spectral envelopes of the vowels /a/ and /i/, and the formants (F1 and F2).

old). Prior to the experiment, all participants were screened to ensure that their speech and hearing were within normal range.

The experimental setup included a headset microphone (DPA d:fine), audio interface (Focusrite Clarett+ 4pre), mixer (YAMAHA MG10XUF), headphones (Sennheiser HD280Pro mk2), and laptop PC (Lenovo ThinkPad X1). A Z-weighted sound level meter (B&K Type2250) connected to an artificial ear (B&K Type4153) was used to calibrate the experimental setup such that a 1-kHz pure tone input to the headset microphone at 94 dB SPL was output through the headphones at 80 dB SPL. The sampling frequency was 16 kHz and the bit rate was 16 bits.

2.2. Stimuli

The back vowel /a/ and the front vowel /i/ were chosen as the speech targets among the Japanese vowels. The noise presented to the participants was generated through the following process: First, broadband (0 – 4 kHz) white noises NF75 and NF85 output through headphones with sound pressure level of 75 and 85 dB were created. Next, NL75 and NL85 were generated from NF75 and NF85 by processing them with a low-pass (0 – f_c kHz) filter, and NH75 and NH85 were generated by processing them with a high-pass (f_c – 4 kHz) filter. The cutoff frequency (f_c) for both filters were set to 925 Hz, which was the median of the mean F1 frequency and the mean F2 frequency of the vowel /a/ pronounced by 153 adult male Japanese native speakers [22]. Fig. 1 shows the amplitude spectrum of each noise type, the amplitude spectral envelope of the vowels /a/ and /i/, and the lower formants (F1 and F2). As depicted in Fig. 1, on the amplitude spectrum, the noises of LOW and HIGH types were designed to mask only F1 and F2, respectively. Table 1 shows the information (label, noise type, and noise level) on the total noise created and the L_{N} is the actual SPL of each noise presented by the headphone.

Table 1: Information on the total noise created, where L_{N} is the actual SPL of each noise presented through headphones.

Label	Noise type	Noise level	L_{N}(dB)
NF75	FULL	75	75.0
NF85	FULL	85	85.0
NL75	LOW	75	71.5
NL85	LOW	85	81.5
NH75	HIGH	75	73.0
NH85	HIGH	85	83.0

2.3. Procedure

The sequence of the trials was determined by combining the order in which the random block method was applied to all combinations of the two vowel conditions (/a/ and /i/) and the seven noise presentation conditions (six noise and no-noise) with the order of their counterbalances. Thus, all participants performed all condition combinations twice each (a total of 28 trials). The experiment was set up with a 4-second trial duration and a 3-second interval between trials. During each trial, a speech target was displayed in hiragana (“あ” or “い”) on the laptop PC monitor, while noise was presented through the headphones, except for the trials in the no-noise condition.

The experiment was conducted in a soundproof room (background noise level $L_A < 30$ dB). Participants were instructed to sit in front of the monitor, wear a headset microphone and headphones, and vocalize the vowel for as long as possible, starting immediately after the hiragana was displayed on the monitor and ending immediately after the display ended. The speech sounds recorded by the headset microphone was provided with real-time auditory feedback through headphones. The speech sound, presentation noise, and headphone output sound for each trial were recorded simultaneously on the laptop PC through the audio interface.

2.4. Data processing

We extracted the interval between one and two seconds after vocalization onset from the recorded speech and used it as speech data. The following acoustic parameters were obtained from the speech data for each trial: The F1 and F2 frequencies were

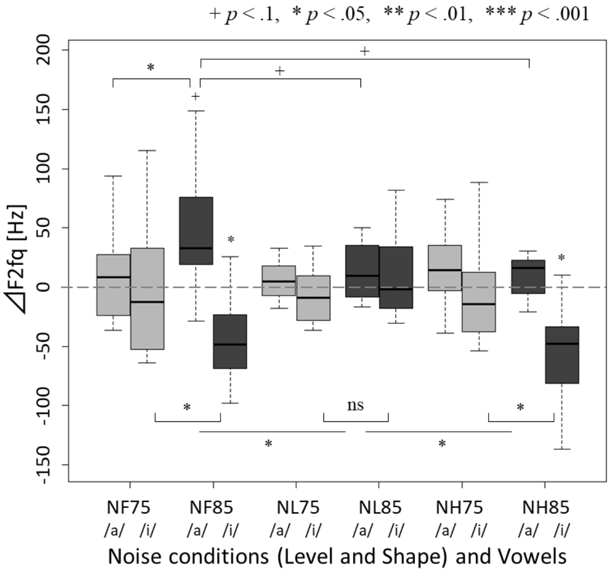

Figure 2: Changes in the F2 frequency $\Delta F2_{\mathrm{fq}}$ of the vowels /a/ and /i/ for each noise condition.

estimated using LPC analysis with an assumption of a pulse train [23] implemented as “To Formant (robust)...” in Praat [24], where the number of formants was set to 5, the formant ceiling was set to 5000 Hz, and the window length was set to 0.02 seconds. The vocal intensity in dB SPL and amplitudes of F1 and F2 in dB were calculated using MATLAB, where the window length was set to 0.02 seconds. To obtain F1 and F2 amplitudes, the cepstrum of the speech data was calculated, and peak picking around estimated F1 and F2 frequencies of the low-order cepstrum components was applied.

The changes in all acoustic parameters with and without noise (vocal intensity ΔL_{sp}, F1 amplitude $\Delta F1_{\mathrm{pk}}$, F2 amplitude $\Delta F2_{\mathrm{pk}}$, F1 frequency $\Delta F1_{\mathrm{fq}}$, and F2 frequency $\Delta F2_{\mathrm{fq}}$) were determined through the following procedure. First, the within-subject averages of the acoustic parameters were calculated for each vowel-noise condition. Next, the difference in the acoustic parameters between each noise condition and the no-noise condition was calculated.

2.5. Statistical analysis

A one-sample t-test was conducted on the change in each acoustic parameter of speech with and without noise, with p-values corrected using the Holm method (see Table 2). Based on the finding that the Lombard effect increases the vocal intensity, the ΔL_{sp}, $\Delta F1_{\mathrm{pk}}$, and $\Delta F2_{\mathrm{pk}}$ for amplitude were one-tailed tests with a change > 0.

Table 2: Results of a one-sample t-test for the amount of change in acoustic parameters of speech with and without noise.

Noise	ΔL_{sp}	$\Delta F1_{\mathrm{pk}}$	$\Delta F2_{\mathrm{pk}}$	$\Delta F1_{\mathrm{fq}}$	$\Delta F2_{\mathrm{fq}}$
NF75	$**$	$**$	$*$	$*$	ns
NF85	$*$	$*$	$*$	$***$	ns
NL75	ns	ns	ns	ns	ns
NL85	$*$	$*$	$*$	ns	ns
NH75	$*$	$*$	$+$	ns	ns
NH85	ns	$+$	ns	ns	ns

To compare the change in each acoustic parameter across noise conditions, a repeated measures analysis of variance (RM-ANOVA) with 3 (noise type) × 2 (noise level) × 2 (vowel) was performed. If the assumption of sphericity was rejected according to Mauchly’s test, the degrees of freedom were adjusted using Chi-Muller’s epsilon. The p-values for multiple comparisons were corrected using the Shaffer method.

3. RESULTS

3.1. Vocal intensity

In Table 2, the change in vocal intensity, ΔL_{sp}, was significant in all noises except NL75 and NH85. RM-ANOVA showed that the type × level interaction was significant ($F(1.35, 12.17) = 7.14, p < .05$), and the simple main effect of level on FULL and LOW types was significant or tended to be significant (FULL: $F(1,9) = 4.03, p = .07$; LOW: $F(1,9) = 17.11, p < .01$).

3.2. Formant amplitude

The changes in F1 amplitude $\Delta F1_{\mathrm{pk}}$ and F2 amplitude $\Delta F2_{\mathrm{pk}}$ exhibited a similar trend to ΔL_{sp} in Table 2. The RM-ANOVA results indicated that the type × level interaction tended to be significant for $\Delta F1_{\mathrm{pk}}$ ($F(1.2, 10.8) = 4.51, p = .05$) and the simple main effect of level on FULL and LOW type tended to be significant or had a significant (FULL: $F(1,9) = 3.92, p = .09$; LOW: $F(1,9) = 23.18, p < .001$). The type × level interaction ($F(1.51, 13.55) = 4.56, p < .05$) was significant for $\Delta F2_{\mathrm{pk}}$, and the simple main effect of level on LOW type was significant ($F(1,9) = 17.98, p < .01$). These suggest that formant amplitude increases in many noise conditions, and that it increases more with noise level when the noise type is LOW.

3.3. Formant frequencies

The change of F1 frequency $\Delta F1_{\mathrm{fq}}$ was significant for NF75 ($t[9] = 3.34, p < .01$) and NF85 ($t[9] = 6.51, p < .001$) in Table 2, while the RM-ANOVA results indicated that the simple main effect of level in FULL was significant ($F(1,9) = 11.94, p < .01$). This suggests that the F1 frequency shifts upward exclusively in the FULL type condition, irrespective of vowel type, and that the higher the noise level, the more pronounced the upward shift becomes.

Fig. 2 illustrates the change in the F2 frequency $\Delta F2_{\text{fq}}$ of the vowels /a/ and /i/ for each noise condition. In Table 2, $\Delta F2_{\text{fq}}$ was not significant in any noise condition, but the RM-ANOVA results indicated that the interaction of type × level × vowel exhibited a significant trend ($F(1.34, 12.09) = 3.24, p < .1$), thus prompting the conduct of a 3 (type) × 2 (level) RM-ANOVA for each vowel.

The type × level interaction was significant ($F(1.31, 11.79) = 5.38, p < .05$) for $\Delta F2_{\text{fq}}$ in the vowel /a/. The simple main effect of level was also significant when the noise was FULL type ($F(1,9) = 6.23, p < .05$). Additionally, the simple main effect of type was significant when the noise level was 85 ($F(1.57, 14.14) = 5.37, p < .05$), with significant trends between the FULL-LOW and FULL-HIGH conditions, respectively. This suggests that the F2 frequency in the vowel /a/ shifts upward when the noise exhibits a FULL type at a high level.

In contrast, the main effects of type ($F(1,9) = 6.89, p < .05$) and level ($F(1,9) = 8.50, p < .05$) were significant for $\Delta F2_{\text{fq}}$ in the vowel /i/. Multiple comparisons on type indicated that it was significant between the HIGH-LOW ($t[9] = 3.34, p < .05$) and FULL-LOW ($t[9] = 2.28, p < .05$) conditions, respectively. This suggests that when the noise level is higher and the spectral shape of the noise includes the F2 frequency, the F2 frequency in the vowel /i/ shifts downward.

4. DISCUSSION AND SUMMARY

In this study, we investigated changes in the vowel production of a solitary speaker under various noise conditions. The results showed that the amplitude and frequency of the formants of the vowel speech sounds systematically varied according to the type and level of noise, suggesting that humans control not only vocal intensity but also articulation in response to various noise environments. However, since sustained vowel utterances may not fully reflect actual speech processes, future work should examine speech motor control during communication with others in noisy environments similar to those in the present study.

As expected, our results showed an increase in the amplitude of low-order formants and an increase in the vocal intensity in many noise conditions. These results indicated that the Lombard effect consistently enhances the amplitude of lower-order formants and vocal intensity. Contrary to our expectations, we found that the changes in the low-order formant frequency appear dissimilar in both F1 and F2.

Under FULL noise conditions, the F1 frequency consistently shifted upward in frequency regardless of vowel type. Additionally, as the noise level increased, the F1 frequency shift also increased. As shown in Fig. 1, FULL noise effectively masks vowel speech over a wide range of frequencies, and the higher the level, the stronger the masking effect. Therefore, it is suggested that speakers would have needed to further enhancement the signal-to-noise ratio of auditory feedback for speech under FULL noise conditions, which would have resulted in increased mouth opening and lower tongue position, leading to an upward shift in the F1 frequency.

Lu and Cooke [20] showed that the F1 frequency was higher under noisy conditions compared to quiet ones, regardless of the type of noise, but this finding was not entirely congruent with our results. This discrepancy may be due to various factors such as the speech target and the type and level of noise, as well as the presence or absence of female speakers and differences in formant estimation techniques.

The direction of the F2 frequency shift depended on the vowel and noise type. The F2 frequency of the vowel /a/ only shifted upward when the FULL noise was presented at a high level. Conversely, the F2 frequency of the vowel /i/ only shifted downward when the FULL or HIGH noises were presented at high levels. These results suggest that the noise which effectively masks the F2 of the vowel causes a shift in the F2 frequency when it is input to the auditory system at high level.

It is noteworthy that the F2 amplitude increased under high level of LOW noise, despite the noise not obstructing F2, as depicted in Fig. 1. This increase in F2 amplitude was not observed at low level of LOW noise. This finding suggests the limitations of explanations based on amplitude spectra, which may be hinted at by the spectral representation (i.e., excitation patterns) in the auditory peripheral system. Further research is needed to investigate the relationship between excitation patterns and formant masking.

5. ACKNOWLEDGMENTS

This work was supported by SCOPE (201605002) from the Ministry of Internal Affairs and Communications and by a Grant-in-Aid for the Promotion of Joint International Research (Fostering Joint International Research (B)) (20KK0233).

6. REFERENCES

[1] É. Lombard, “Le signe de l’elevation de la voix,” *Annales des Maladies de L’Oreille et du Larynx*, vol. 37, no. 2, pp. 101–119, 1911.

[2] H. Lane and B. Tranel, “The Lombard Sign and the Role of Hearing in Speech,” *Journal of Speech and Hearing Research*, vol. 14, no. 4, pp. 677–709, Dec. 1971.

[3] J. Luo, S. R. Hage, and C. F. Moss, “The Lombard Effect: From Acoustics to Neural Mechanisms,” *Trends in Neurosciences*, vol. 41, no. 12, pp. 938–949, Dec. 2018.

[4] S. Uma Maheswari, A. Shahina, and A. Nayeemulla Khan, “Understanding Lombard speech: A review of compensation techniques towards improving speech based recognition systems,” *Artificial Intelligence Review*, vol. 54, no. 4, pp. 2495–2523, Apr. 2021.

[5] W. V. 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, pp. 917–928, 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. 1-2, pp. 13–22, 1996.

[7] 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, pp. 588–608, Jun. 2010.

[8] 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, no. 5, pp. 3261–3275, 2008.

[9] M. Garnier and N. Henrich, “Speaking in noise: How does the Lombard effect improve acoustic contrasts between speech and ambient noise?” *Computer Speech & Language*, vol. 28, no. 2, pp. 580–597, 2014.

[10] 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, Mar. 2014.

[11] S. Meekings, S. Evans, N. Lavan, D. Boebinger, K. Krieger-Redwood, M. Cooke, and S. K. Scott, “Distinct neural systems recruited when speech production is modulated by different masking sounds,” *The Journal of the Acoustical Society of America*, vol. 140, no. 1, pp. 8–19, Jul. 2016.

[12] J. Šimko, Š. Beňuš, and M. Vainio, “Hyperarticulation in Lombard speech: Global coordination of the jaw, lips and the tongue,” *The Journal of the Acoustical Society of America*, vol. 139, no. 1, pp. 151–162, Jan. 2016.

[13] R. Marxer, J. Barker, N. Alghamdi, and S. Maddock, “The impact of the Lombard effect on audio and visual speech recognition systems,” *Speech Communication*, vol. 100, pp. 58–68, Jun. 2018.

[14] K. Marcoux and M. Ernestus, “Pitch in native and non-native Lombard speech,” in *19th International Congress of Phonetic Sciences (ICPhS 2019)*. Australasian Speech Science and Technology Association Inc., 2019, pp. 2605–2609.

[15] C. Castro, P. Prado, V. M. Espinoza, A. Testart, D. Marfull, R. Manriquez, C. E. Stepp, D. D. Mehta, R. E. Hillman, and M. Zañartu, “Lombard effect in individuals with nonphonotraumatic vocal hyperfunction: Impact on acoustic, aerodynamic, and vocal fold vibratory parameters,” *Journal of Speech, Language, and Hearing Research*, vol. 65, no. 8, pp. 2881–2895, 2022.

[16] B. J. Stanton, L. H. Jamieson, and G. D. Allen, “Acoustic-phonetic analysis of loud and Lombard speech in simulated cockpit conditions,” in *ICASSP-88., International Conference on Acoustics, Speech, and Signal Processing*. IEEE, 1988, pp. 331–334.

[17] C. E. Mokbel and G. F. Chollet, “Automatic word recognition in cars,” *IEEE Transactions on Speech and Audio Processing*, vol. 3, no. 5, pp. 346–356, 1995.

[18] J. Lee, H. Ali, A. Ziaei, and J. H. Hansen, “Analysis of speech and language communication for cochlear implant users in noisy lombard conditions,” in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2015, pp. 5132–5136.

[19] A. J. Gully, P. Foulkes, P. French, P. Harrison, and V. Hughes, “The Lombard effect in MRI noise,” in *Proc. of the 19th International Congress of Phonetic Sciences*, 2019, pp. 800–804.

[20] 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, Sep. 2009.

[21] L. M. Stowe and E. J. Golob, “Evidence that the Lombard effect is frequency-specific in humans,” *The Journal of the Acoustical Society of America*, vol. 134, no. 1, pp. 640–647, Jul. 2013.

[22] T. Hirahara and R. Akahane-Yamada, “Acoustic characteristics of Japanese vowels,” in *Proc. 18th ICA*, 2004, pp. 3387–3290.

[23] C. H. Lee, “On robust linear prediction of speech,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 36, no. 5, pp. 642–650, 1988.

[24] P. Boersma, “Praat: Doing phonetics by computer,” *http://www. praat. org/*, 2006.