| | |
|---|---|
| Title | Influence of auditory feedback on uttering vowel speech in noisy environment |
| Author(s) | Nishigaki, Tomoya; Akagi, Masato |
| Citation | 2020 RISP International Workshop on Nonlinear Circuits, Communications and Signal Processing (NCSP 2020): 303-306 |
| Issue Date | 2020-02-20 |
| Type | Conference Paper |
| Text version | publisher |
| URL | http://hdl.handle.net/10119/16233 |
| Rights | Copyright (C) 2020 Research Institute of Signal Processing Japan. Tomoya Nishigaki and Masato Akagi, 2020 RISP International Workshop on Nonlinear Circuits, Communications and Signal Processing (NCSP 2020), 2020, pp.303-306. |
| Description | |

2020 RISP International Workshop on Nonlinear Circuits, Communications and Signal Processing (NCSP 2020)
Honolulu, Hawaii, USA, February 28 - March 2, 2020

# Influence of auditory feedback on uttering vowel speech in noisy environment

Tomoya NISHIGAKI and Masato AKAGI

Graduate School of Advanced Science and Technology, Japan Advanced Institute of Science and Technology
1-1 Asahidai, Nomi, Ishikawa 923-1292 Japan
E-mail: {s1810138, akagi}@jaist.ac.jp

**Abstract**

This paper investigates variations of frequencies and amplitudes of F1 and F2 in uttered vowels under noisy conditions, to grasp what strategies speakers use to make uttered speech more intelligible in noisy conditions. Seven types of noise were used for the noisy conditions. Five Japanese vowels (/a/, /e/, /i/, /o/, and /u/) uttered in the noisy conditions by four native Japanese speakers were recorded. Then, frequencies and amplitudes of F1 and F2 were extracted. As the results, it was found that F1 tends to mainly change its frequency and F2 changes its amplitude largely according to sound pressure level of the noises. In detail, these variations depended on the types of the noise. This study also discusses excitation patterns to take accounts of the effects of masking from the noises. These tendencies of F1 and F2 were most likely to be explained by the effects of masking from the noises.

## 1. Introduction

Our living environments are full of various noises. In such noisy conditions, humans involuntarily change their way of speaking for intelligible speech. This phenomenon is called "Lombard effect" [1]. The uttered speech due to the Lombard effect also is called "Lombard speech". In the Lombard speech, speech intensity, spectral tilt, formant frequency, fundamental frequency (F0), and duration or speaking rate are different from neutrally spoken speech. Moreover, the Lombard effect occurs not only in human but also in many other animals such as primates, birds, cats, whales, bats, and frogs [2]. With regard to formants, some previous studies found that frequency of the first formant (F1) systematically increases by Lombard effect [3]. Some studies reported that frequency of the second formant (F2) also increases, but not systematical as that of F1. Such variations in acoustic characteristics by the Lombard effect are dependent on tongue, jaw, and lip movements [4]. The tongue position of vowels in Lombard effect is on average lower than that during neutral speech [5]. Garnier et al. [5] also found correlation of tongue and lip movements not only with F1 but also with F2. Stowe and Golob [6] reported that speech intensity, duration, and F0 of uttered speech increase in bandlimited broadband noise (0.5-4 kHz band). However, with notched noise (0.5-4 kHz notch), their acoustic features did not change. These results indicated that characteristics of Lombard effect depend on properties of the noises. Matsumoto and Akagi [7] investigated what strategies speakers use to utter intelligible speech under various noisy conditions. However, this research did not grasp what strategies speakers use to make uttered speech more intelligible in the noisy conditions.

Therefore, this paper aims to grasp the tendency what strategies speakers make uttered speech more intelligible under various noisy conditions.

## 2. Experimental procedure

### 2.1 Noises

In this paper, seven types of noise were used, low-pass noise (LPN), high-pass noise (HPN), band-pass noise (BPN), notched noise (NN), and pink noise (PN). These noises were generated by Eq. (1). $N$ is generated noise. $f_L$ is lower limit of the frequency band of the generated noise. $f_U$ is upper limit of the frequency band of the generated noise. $f$ is arbitrary frequency. $T$ is time length. $t$ is time series. $\phi(f)$ is randomized phase. And, $f$ of $\phi(f)$ changes corresponding to each frequency band of noise. The NN was designed as combination LPN and HPN. In this paper, in order to verify whether the experimental environment is valid and whether the Lombard effect occurs, we investigated variations of speech when uttering in the same noise (PN) as in the previous study. PN is generated by applying a high-band attenuation filter to broadband white noise. Moreover, in PN, low frequency component are emphasized more than those in the other noises. The amplitude is calibrated to broadband noise (1-22,050 Hz) generated based on the following Eq. (1) so that the RMS value is 70 dB and 80 dB. The noises used in this paper are adjusted to have the same dB/Hz as the broadband noise.

Table:1 Formant frequency of vowels for subject 1 (adult male 1) during utterance in quiet

| Vowel | F1 [Hz] | F2 [Hz] | F3 [Hz] |
|---|---|---|---|
| /a/ | 778 | 1,224 | 3,029 |
| /e/ | 460 | 2,128 | 2,874 |
| /i/ | 332 | 2,569 | 3,366 |
| /o/ | 457 | 827 | 3,089 |
| /u/ | 371 | 1,478 | 2,602 |

$$N = \sum_{f=f_L}^{f_U} \sin\left(2\pi f \frac{t}{T} + \phi(f)\right) \qquad (1)$$

**2.2 Recording Japanese vowels in quietly condition**

In order to analyze the frequencies and amplitudes of F1 and F2, four adult speakers (2 males and 2 females) age 23 to 24 participated in the recording. They have no obstacle to hearing function. The vowel utterances of each speaker were recorded in quiet conditions in advance. Figure 1 shows the recording environment. While this recording, the speakers were asked to wear an open-air type headphone (STAX SR-L500). In this time, the noises were not presented from the headphone. Five speech data were obtained for each type of the vowels. The sampling frequency was 44,100 Hz.

**2.3 Formant analysis**

The noises for this experiment were generated to correspond to F1 and F2 frequencies of each vowel and speaker in order to compare Lombard speech with neutral speech. Therefore, formant frequencies of recording speech were calculated by Acoustic core [8]. This is a speech analysis software based on LPC. Then, this paper calculated the estimated frequencies of F1, F2 and F3 of the vowels of each speaker. Table 1 shows the analysis results of formant frequencies for each subject. The values in the Table 1 are the average of the analysis results for each vowel. In order to consider the effect of generated noises for the formants on auditory perception, this paper also calculated the excitation patterns [10] based on the previous research [9].

**2.4 Generating noise corresponding to the formant frequencies**

Based on the result of formant frequency analysis, the frequency bands of the presentation noises are determined so as to correspond to the formant frequency and vowel types of each speaker. $f_L$ and $f_U$ were set to the mid frequencies of F1-F2 and F2-F3 for each speaker and vowel. These values depend on the types of noise.

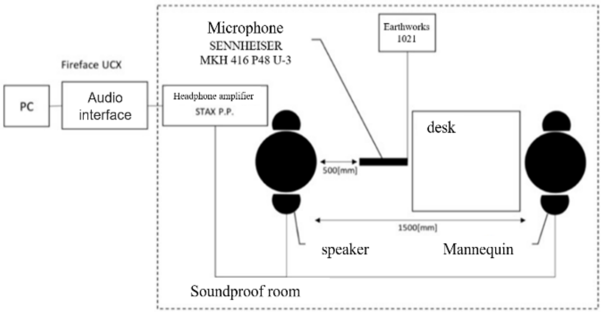

Fig 1: Recording environment

**2.5 Recording Japanese vowels under noisy conditions**

To investigate variations of the frequencies and amplitudes of formants when the speakers utter the vowels under the noisy conditions, the noises were presented from the headphone as shown in Fig. 1. In this experiment, the noises were presented for each vowel in the order of BPN (Cut-off: mid frequency of F1-F2 to mid frequency of F2-F3), LPN_F12 (Cut-off: mid frequency of F1-F2), LPN_F23 (Cut-off: mid frequency of F2-F3), HPN_F12 (Cut-off: mid frequency of F1-F2), HPN_F23 (Cut-off: mid frequency of F2-F3), BPN (Cut-off: mid frequency of F1-F2 to mid frequency of F2-F3) and PN. The noises also were presented in the order of 70 dB and 80 dB. For each vowel, 14 tasks (7 noises x 2 intensities) were prepared and a total of 70 tasks were carried out. One task is to utter the same vowel five times within 20 seconds under the noisy conditions. From this experiment, 350 speech data were obtained from each subject.

**3. Results**

**3.1 Variations of formant frequencies and amplitudes**

Formant frequencies and amplitudes of vowels uttered under the noisy conditions were analyzed and compared with that in the quiet condition. Figure 2 shows frequency and amplitude variations (%) of F1 and F2 as compared to a neutral speech uttered without noise. Frequency and amplitude variations indicate ratio of difference and difference between neutral speech and Lombard speech based on the formant frequencies of neutral speech. 70 dB and 80 dB in Figure 2 indicate the sound pressure level of each noise. The result shows, F1 tends to mainly change its frequency and F2 changes its amplitude largely according to sound pressure level of the noises. In the result of HPN_F23, the tendency described above is smaller than

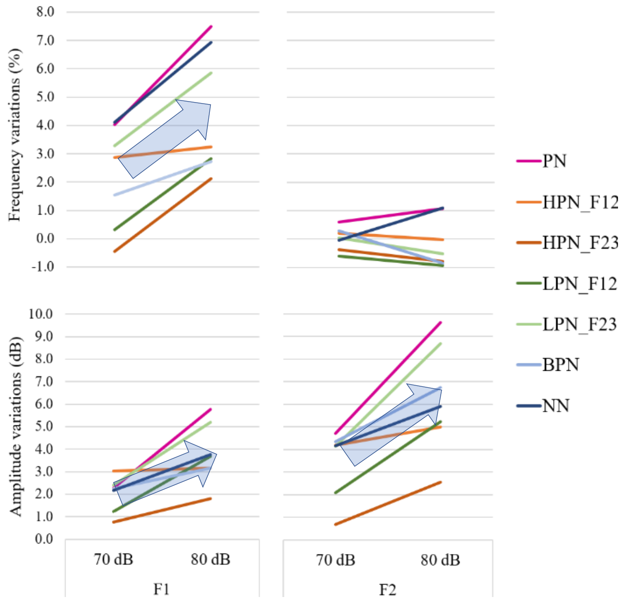


Fig 2: Frequency change rate and amplitude change of F1 and F2 as compared to a neutral speech uttered without noise

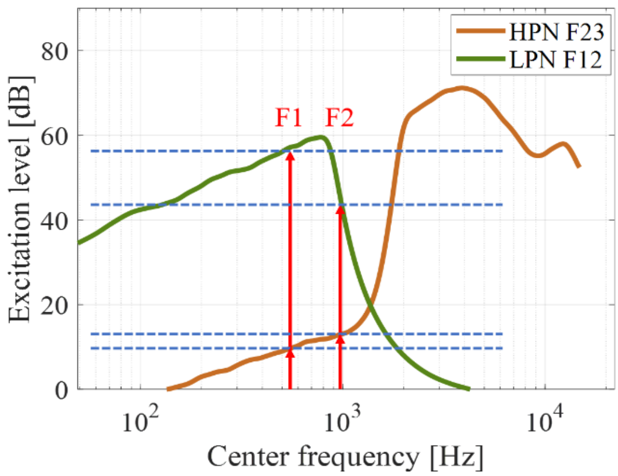


Fig 3: Excitation patterns of LPN_F12 and HPN_F23 (/a/ of subject 1)

the other noises. On the other hand, in the result of PN, the tendency is larger than the other noises. In the results of NN and PN, frequency variations of F1 by the sound pressure level are similar. However, in the result of NN, amplitude variations of F1 and F2 by sound pressure level is smaller than those in the result of PN. The result of LPN show that the found tendency is more dependent on the sound pressure level than that in the result of HPN.

### 3.2 Excitation patterns

This study focuses on auditory perception. Therefore, we observed excitation patterns [10] of the noises and the utterances. The vertical axis in Figure 3 represents excitation level. The horizontal axis represents center frequency of auditory filter. F1 and F2 in figure 3 indicate F1 and F2 of /a/ of subject 1. The dashed line indicates the effect to F1 and F2 by LPN_F12 or HPN_F23. Figure 3 shows the excitation patterns of LPN_F12 (/a/ of subject 1). According to the result, the effect to F2 is found to be present. In the result of HPN_F23, the effect of the noise to F1 is smaller than the effect to F2 in the result of LPN_F12. These noises were generated so as not to affect F2 on physical properties.

## 4. Discussion

### 4.1 Variations of utterances by noises

To grasp what strategies speakers use to utter speech more intelligible under noisy conditions, this paper investigates variations of frequencies and amplitudes of F1 and F2 in uttered vowels under noisy conditions. F1 and F2 are important to perceive the vowels [11]. Therefore, seven types of noise influencing on F1 and F2 are generated, and two types of sound pressure levels are prepared for each noise. By observing variations of frequencies and amplitudes of F1 and F2, we reveal the strategies of speakers under these noisy conditions.

According to the result of PN in Fig. 2, as in the previous study [3], the frequency of F1 increases as the sound pressure level of the background noise increases. This result indicates that the Lombard effect occurred in this experiment.

### 4.2 Variations of utterances by sound pressure level

Two types of sound pressure level are prepared for the noises in order to observe the variations of F1 and F2 amplitudes and frequencies according to the sound pressure level. As the results in Fig. 2, this study found the following tendencies that, F1 changes its frequency and F2 changes its amplitude mainly according to the sound pressure level of the noises. These tendencies were observed under all noises used in this experiment. F1 frequency depends on positions of tongue height and F2 frequency depends on the horizontal positions of tongue [4]. It is suggested that the positions of the tongue become lower as the sound pressure level of the noises increases. On the other hand, the horizontal positions of the tongue have no tendency as the sound pressure level of the noises increases. This was pointed out in previous study [3]. Increase in F2 amplitude seems to be related to the glottal-fold source signal characteristics.

### 4.3 Variations of utterances by types of noise

In order to observe the variations of utterances under various noises, seven types of noise are prepared to observe variations of F1 and F2 amplitudes and frequencies when changing the types of noise. As the results in Figs. 2 and 3, variations are dependent on types of noise. In detail, it was found that the greater the influence on F1 and F2 for the excitation patterns is, the larger the variations for utterances are.

From the results of LPN and HPN in Section 3.1, the result of LPN is more dependent on the sound pressure level than that that of HPN. From the result of HPN_F23, the variations are smaller than that of the other noises. This fact can be explained using the excitation patterns shown in Fig. 3. It is suggested that the noises in the low frequency influence on F1 and F2 more than the others when uttering vowels.

From Fig 2, the effect of PN is larger than that of other noise. PN is more emphasized in low frequency than the other noises. From the result of NN for the analyzed excitation patterns, this noise affects to F2. It is supposed that perception of NN might be similar to that of PN. As the result of analysis excitation patterns, influence on F1 and F2 by PN were larger than that of NN. As the result in Section 3.1, from the results of NN and PN, frequency variations of F1 by the sound pressure level are similar. However, amplitude variations of F1 and F2 for NN is smaller than that of PN. It is inferred that the difference depends on influence in F1 and F2.

## 5. Conclusion

This paper investigated variations of frequencies and amplitudes of F1 and F2 of vowels uttered under noisy conditions, to grasp what strategies speakers use to utter speech more intelligible in the noisy conditions. According to the results, this study found the tendency that regardless of the types of noise, F1 changes its frequency and F2 changes its amplitude mainly according to sound pressure level of the noises. These variations depend on types of noise. In detail, it was found that the greater the influence on F1 and F2 on the excitation patterns is, the larger the variations for utterance are. These results can be explained using excitation patterns of noise.

## Acknowledgement

This research was supported by JST-Mirai Program (JP-MJMI18D1).

## References

[1] Lombard, E., "Le signe de I' elevation de la voix, Ann. Mal. De L," Oreille et du Larynx, vol. 37, pp. 101–119, 1911.

[2] H. Brumm, and H. Slabbekoorn, "Acoustic communication in noise," Advances in the Study of Behavior, vol. 35, no. 9, pp. 151-209, 1991.

[3] Y. Uemura, M. Morise, and T. Nishiura, "The Lombard speech recognition based on the voice conversion towards neutral speech," ICA2010, PaperID, 167, 2010.

[4] M. Garnier, L. Bailly, M. Dohen, P. Welby, and H. Loevenbruck. "An acoustic and articulatory study of Lombard speech: Global effects on the utterance," Interspeech/ICSLP 2006, pp. 2246-2249, Pittsburgh, USA, 2006.

[5] M. Garnier, L. Mnard, and G. Richard. "Effect of being seen on the production of visible speech cues. Apilot study on Lombard speech," In 13th Annual Conference of the International Speech Communication Association (Inter Speech 2012), Portland, USA, pp. 611-614, 2012.

[6] LM. Stowe and EJ. Golob, "Evidence that the Lombard effect is frequency-specific in humans," J. Acoust. Soc. Am., vol. 134, No. 1, pp. 640-674, 2013.

[7] S. Matsumoto, M. Akagi, "Variation of Formant Amplitude and Frequencies in Vowel Spectrum uttered in Various Noisy Environments," NCSP2019, Honolulu, Hawaii, USA, Mar., pp. 4-7, 2019.

[8] Acoustic core (n.d.), Retrieved Jan 15, 2020, from http://www.acousticcore.com/

[9] ISO 532-2: "Methods for calculating loudness - Part 2: Moore-Glasberg method" Acoustics, 2017.

[10] B. C. J. Moore, R. Glasberg, "A new method of calculating auditory excitation patterns and loudness for steady sounds," Hearing Research, vol. 282, pp. 204-215, 2011.

[11] T. Chiba, M. Kajiwara, "The vowel: Its nature and structure" Iwanamisyoten, Tokyo, 2003.