Title	雑音環境下における聴覚フィードバックが母音発話に与える影響に関する研究
Author(s)	西垣，朋哉
Citation	
Issue Date	2020-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/16706
Rights	
Description	Supervisor:赤木　正人，先端科学技術研究科，修士（情報科学）

Study on influence of auditory feedback on uttering vowel speech in noisy environment

1810138 Tomoya Nishigaki

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". 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. With regard to formants, some previous studies found that frequency of the first formant (F1) systematically increases by Lombard effect. 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. The tongue position of vowels in Lombard effect is on average lower than that during neutral speech. Garnier et al. also found correlation of tongue and lip movements not only with F1 but also with F2. Stowe and Golob 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 Akagii nvestigated 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.

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). 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 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.

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. 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.

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. 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. In order to consider the effect of generated noises for the formants on auditory perception, this paper also calculated the excitation patterns based on the previous research.

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. Cut–off frequencies 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.

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. 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), NN (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.

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. The effect of PN is larger than that of other noise for atterance. PN is more emphasized in low frequency than the other noises. On the other hand, from the result of HPN_F23, the variations are smaller than that of the other noises. This noise has smaller the influence on F2 and F1 than the other noises. Therefore, from these results, it is suggested that the noises in the low frequency influence on F1 and F2 more than the others when uttering vowels. Moreover, 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. Increase in F2 amplitude seems to be related to the glottal-fold source signal characteristics.