Title	Variation of Formant Amplitude and Frequencies in Vowel Spectrum uttered under Various Noisy Environments
Author(s)	Matsumoto, Shumpei; Akagi, Masato
Citation	2019 RISP International Workshop on Nonlinear Circuits, Communications and Signal Processing (NCSP2019): 252-255
Issue Date	2019-03-06
Type	Conference Paper
Text version	publisher
URL	http://hdl.handle.net/10119/15772
Rights	Copyright (C) 2019 Research Institute of Signal Processing, Japan. Shumpei Matsumoto and Masato Akagi, 2019 RISP International Workshop on Nonlinear Circuits, Communications and Signal Processing (NCSP2019), 2019, 252-255.
Description	

2019 RISP International Workshop on Nonlinear Circuits, Communications and Signal Processing (NCSP2019)
Honolulu, Hawaii, USA, March 4-7, 2019

Variation of Formant Amplitude and Frequencies in Vowel Spectrum uttered under Various Noisy Environments

Shumpei Matsumoto 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: { s-matsumoto, akagi } @jaist.ac.jp

Abstract

In this paper, we investigate how vowels in utterance vary in various noise environments, focusing on formant frequency and amplitude on the spectrum. Noises that covered the formants of each subjects was synthesized based on the uttered voices recorded in the without any noise. We analyzed using Praat and cepstrum analysis in order to investigate the difference between the formant frequency and amplitude when several types of noise are presented, comparing with the case without noise. As a result, when uttering vowels in noise environments, humans not only increase formant amplitudes of their own voices against noise, but also change their vocal tract shapes so that their own voices are not covered by noise depending on the pattern of noise.

1. Introduction

Environment in daily life is full with various sounds. When speaking in environments with noises such as crowded places and streets with many cars, humans vary their voices according to noise in order to improve intelligibility of speech. This phenomenon is known as Lombard effect[1]. Many researches have been done on Lombard effect. Kubo *et al.*(2016) found that intelligibility of speech is increased in noisy environments with Lombard effect(Lombard speech)[2]. Stowe *et al.*(2013) found that strength, duration, and F0 of uttered speech significantly increase intelligibility under presentation of broadband noise with the same frequency bands as speech. However, any effect was not observed when using notch noise removing most of the frequency band similar to that of speech[3]. This means that Lombard effect changes its properties depending on noise.

One previous report[3] shows that Lombard effect is sensitive to frequency band of noise vital for speech.

However, details about which band is more influential on the change of utterance are not clear. This study aims to investigate how humans vary formant frequencies and amplitudes, when uttering speech in various noisy environments. Formant frequency and amplitude are one of important parameters that characterize vowels in speech communication and are closely related to vocal tract shape and glottal vibration[4]. In this paper, we investigate how Lombard effect varies in different noisy environments from the viewpoint of speech production.

2. Variation of formant according to difference of noise

2.1 Noise synthesis

In this study, three Japanese vowels /a/, /i/, and /u/ are targeted as speech to be investigated.

Formant frequencies, F1 and F2, are particularly important for vowel characteristics. These vowels draw a triangle that interpolates other vowels /e/, /o/ on the F1 - F2 plane as shown in Figure 1. By examining vowels /a/, /i/ and /u/, it is possible to understand behaviors of vowels' formants. To understand what kind of strategy each speaker is taking to enhance the F1 and F2 in noisy environments, this study investigates how spectra of the uttered vowels change when several types of noise are presented, comparing with the case without noise. We discuss relationships between the types of noises and the spectrum changes systematically.

In this study, low-pass, high-pass, band-pass, and notch noises are used to present as noises to cover the formants. These noises are synthesised based on Equation (1). N is noise to be synthesised, L is lower limit of noise frequency band, R is higher end of noise frequency band, f is an arbitrary frequency, sec is a time length, t is a sample time sequence, and $phi(f)$ is a randomized phase. f in $phi(f)$ changes corresponding to each noise pattern. For the notch noise, low-pass and high-pass noises are mixed. The amplitude was calibrated by the maximum amplitude value of the wideband noise (1 - 22050 Hz) synthesized on the basis of Equation (1) for all the synthesized noise.

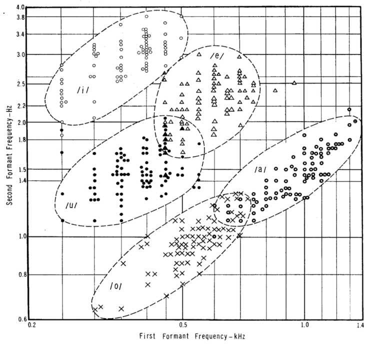

Figure 1: Second formant frequency(F2) versus first formant frequency(F1) diagram of five Japanese vowels spoken by children, youth and adults(After:[4]).

$$L \leq f \leq R$$

$$N = \sum_{f=L}^{R} \sin(2\pi f * \frac{t}{sec} + phi(f)) \tag{1}$$

2.2 Recording utterances without noise

Presenting noises are prepared according to formant's amplitudes and frequencies of each voice and speaker. To collect these for each vowel and speaker, we recorded uttered vowels in noise-free condition. The recording environment at this time is shown in Figure 2. Subjects are 2 adults (1 male and 1 female) in their twenties and have no obstacle to hearing function. The subjects were given the role of speaker and listener. The subjects wear open headphones and the speaker sat in front of a directional microphone. A panel on which one of Japanese vowels is randomly written is shown to the speaker. The task of the speaker is to read the vowel on the panel with loud voice so that the listener could hear it. The listener is given a task to write which vowel the speaker uttered on a paper put on the desk. At that time, the listener is instructed not to see the mouth of the speaker. The tasks are divided into four sections, five tasks per section, total of 20 tasks were carried out. From these tasks we obtained 20 sound data per vowel. The sampling frequency was 44100 Hz.

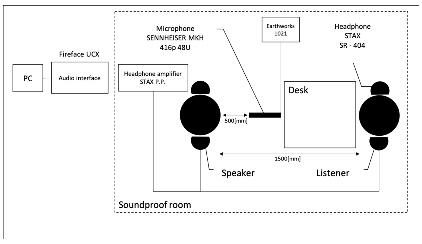

Figure 2: Recording environment

Table 1: Analysis result of the formant frequency of the subject (female)

	F1[Hz]	F2[Hz]	F3[Hz]
/a/	676	1173	2972
/i/	345	2591	3465
/u/	378	1375	2455

2.3 Analysis of formant frequency

Formant frequencies of the recorded utterances are estimated. For the estimation, we used the voice analysis tool Praat. The analysis window length is set to 0.025 s and 25% to 75% of the stable part of the utterance section is used as the analysis section. Estimated values of the first formant frequency(F1), the second formant frequency(F2), and the third formant frequency(F3) for each speaker are estimated. Table 1 shows the analysis results of the formant frequencies of the female speaker. Numerical values in the table are mean values of 20 analyzed results per vowel.

2.4 Synthesis of noise corresponding to formant frequencies

Based on the results of the formant analysis, presented noises are synthesized according to the frequency bands corresponding to the formant frequencies of each vowel and speaker. The frequency band of the noise is set with the midpoints of F1 - F2 and F2 - F3. As an example, Figures 3 to 6 show noise spectra for vowel /a/ of the subject (female).

For low-pass and high-pass noises, two patterns are prepared with thresholds at the midpoints of F1 - F2 or F2 - F3.

3. Formant variation in noisy environments

3.1 Purpose and method of experiment

We investigate how formants in uttered vowels vary when noise is presented, comparing these with the case without noise presentation. Experimental conditions are the same as those recorded for determining formant frequencies for each vowel and speaker, and the subjects were the same in Section

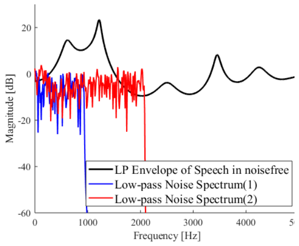

Figure 3: Low-pass noise corresponding to vowel /a/

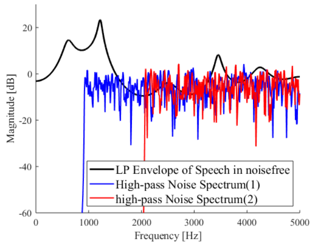

Figure 4: High-pass noise corresponding to vowel /a/

Table 2: Noise condition corresponding to the formant of the subject [female](fc = Cutoff frequency; fm = Midpoint frequency; bw = bandwidth; fmr = center frequency to remove; rw = removal width; LP = low-pass; HP = high-pass; BP = band-pass; N = notch)

	fc[Hz]	fm[Hz]	bw[Hz]	fmr[Hz]	rw[Hz]
LP(1)	925	-	-	-	-
LP(2)	2072	-	-	-	-
HP(1)	925	-	-	-	-
HP(2)	2072	-	-	-	-
BP	-	1499	1148	-	-
N	-	-	-	1499	1148

2.2. This task is different from that in section 2.2. During the period the speaker utters a vowel, we present one of noises with headphones. The condition of the presented noises is shown in Table 2. The task was divided into seven sections for each noise condition (including cases when without noise), and a total of 105 tasks were performed for 15 tasks per section.

From these tasks, 20 sound data were obtained per each condition.

3.2 Analysis of uttered vowels in noisy conditions

3.2.1 Variation of formant frequency

To investigate variations of the formant frequencies, we analyze vowel utterances under noisy conditions using the voice analysis tool Praat. Fifteen sound data per condition excluding the first utterance among tasks were used for analysis. Analysis window length is set to 0.025 s, and 25 % - 75 % of the steady-state part of the utterance section was used as an analysis section. Figure 7 shows the analysis result of the formant frequencies of subject(female) in the F1 - F2 plane. Each point in the figure shows the mean value of 15 pieces of audio data. Using one-way ANOVA analysis, variations of the formant frequencies in cases without noise and with various noises are analyzed for all types of vowels and formants (F1, F2). As a result, a significant difference in the variation of the formant frequency was confirmed under the condition excluding the second formant of the vowel /u/. Figures 8 to 11 show formant frequency changes of vowels according to noise conditions obtained by one-way ANOVA. They show significant differences are found from Figures 8 to 11. In Figure 8, F2 becomes higher when band-pass and low-pass noises covering F2 of vowel /a/. In Figure 9, if there is notch noise removing the band at the center of F2 and F3 from the center of F1 and F2 of vowel /i/ and low-pass noise covering F2, F1 of vowel /i/ tends to be higher. In Figure 10, F2 of vowel /i/ decrease when bandpass and highpass noise covering F2 of vowel /i/ are presented. In Figure 11, if there is notch noise removing the band at the center of F2 and F3 from the center of F1 and F2 of vowel /u/ and low pass noise covering F2, F1 of vowel /u/ tended to be higher. In both cases, the frequencies of the formants move away from the band in which the noise exists.

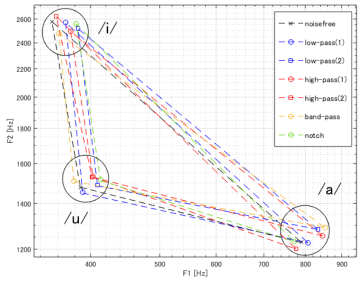

Figure 7: Second formant frequency(F2) versus first formant frequency(F1) diagram of 3 Japanese vowels spoken by the subject(female)

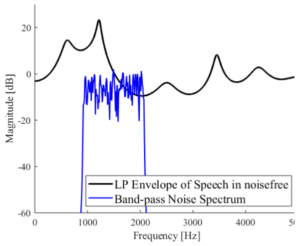

Figure 5: Band-pass noise corresponding to vowel /a/

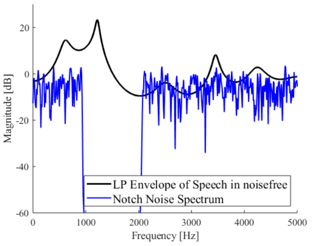

Figure 6: Notch noise corresponding to vowel /a/

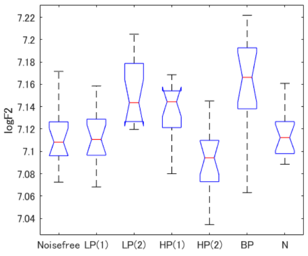

Figure 8: Variation of F2 of vowel /a/

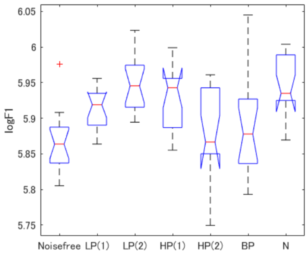

Figure 9: Variation of F1 of vowel /i/

3.2.2 Variation of formant amplitude

Variations of formant amplitudes are investigated by cepstrum analysis. Between 25% and 75% of the stable portion of uttered vowels was used as an analysis period. In addition, since this study uses stationary vowels as speech sounds, Hanning window is applied to the whole analysis period. At the time of analysis, the audio data was downsampled to 10

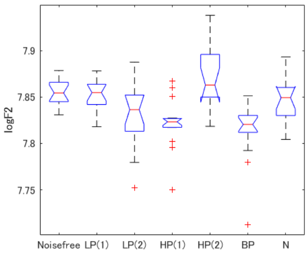

Figure 10: Variation of F2 of vowel /i/

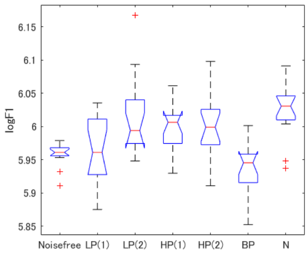

Figure 11: Variation of F1 of vowel /u/

kHz. The lifter cutoff was set to 20. One-way ANOVA was performed for the average values of the amplitude values of F1 and F2 of vowels uttered under each noise condition.

Multiple comparisons were also performed based on the obtained results. As a result, in the conditions excluding vowel /a/, a significant difference was confirmed between the formant amplitude in the case where no noise is presented and the formant amplitude in the case of presenting various noises. Figures 12 to 15 show formant amplitude changes of vowels according to noise conditions obtained by one-way ANOVA. They show significant differences are confirmed from Figures 12 to 15. From the figures, in F1 and F2, compared with the case where no noise is presented, the amplitude tends to decrease when various noise is presented.

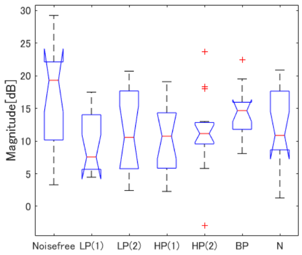

Figure 12: Variation of F1(amplitude) of vowel /i/

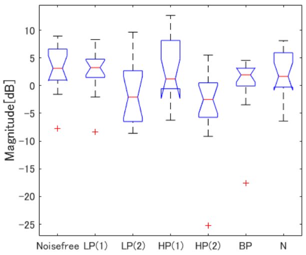

Figure 13: Variation of F2(amplitude) of vowel /i/

4. Discussion

From the analysis results, it was found that the formant frequencies and amplitudes are varied depending on the noises as a strategy to prevent human’s own voice from being covered by noise. However, depending on the type of vowels, some formant changes are not seen compared to that when no noise is presented. Also, among the obtained results, F1 was higher than that when no noise was presented, and some were approaching the band where the noise was present. One of considerable reason is that, since F1 is related to vibrations of the vocal cords, it is conceivable that F1 got higher as a result of attempting to enlarge their own voice against noise.

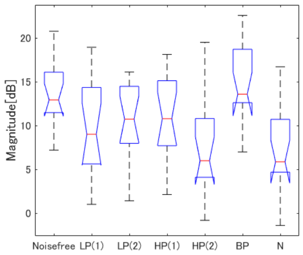

Figure 14: Variation of F1(amplitude) of vowel /u/

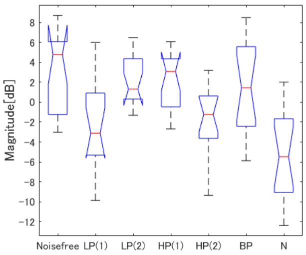

Figure 15: Variation of F2(amplitude) of vowel /u/

5. Conclusions

In this paper, we investigated how Lombard effect varies due to differences in noise conditions from the viewpoint of speech production. Noises with bands set in advance according to the individual formant frequency were synthesized. We investigated how the formants frequencies and amplitudes are affected when these noises are presented as compared with the case when without noise. As the analysis results, it was found that the formant frequencies and amplitudes are changed depending on the noises as a strategy to prevent human’s own voice from being covered by noise.

Acknowledgment

This work was supported by SECOM Science and Technology foundation and JST-Mirai Program(JPMJMI18D1).

References

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

[2] R. Kubo, D. Morikawa and M. Akagi: Effects of speaker’ s and listener’ s acoustic environments on speech in- telligibility and annoyance, Proc. Inter-Noise2016 Hamburg Germany, pp. 171–176, 2016.

[3] LM. Stowe and EJ. Golob: Evidence that the Lombard ef- fect is frequency-specific in humans, J. Acoust. Soc. Am., vol. 134, No. 1, pp. 640–674, 2013.

[4] H. Kasuya, H. Suzuki and K. Kido: Changes in Pitch and first Three Formant Frequencies of Five Japanese Vowels with Age and Sex of Speakers, The journal of the acoustical society of japan, Vol. 24, No. 6, pp. 355–364, 1968.