*International Journal of Innovative*
*Computing, Information and Control*
Volume **13**, Number **2**, April **2017**

ICIC International ©2017 ISSN 1349-4198
pp. **591–603**

# EFFECTS OF MASKING NOISE IN AUDITORY FEEDBACK ON SINGING

Satoshi Iijima, Shunsuke Ishimitsu and Masashi Nakayama

Graduate School of Information Science
Hiroshima City University
3-4-1 Ozuka-Higashi, Asa-Minami-Ku, Hiroshima 7313194, Japan
ishimitu@hiroshima-cu.ac.jp

Received August 2016; revised December 2016

Abstract. *We investigated the effects of masking noise in the auditory feedback on singing. We conducted two experiments in this study. Participants sang a steady vowel /a/ at three pitch heights C3, G3, and C4 for approximately 5 s. In the first experiment, we used a 70 dB pink noise. In the second experiment, we assessed the effects of a masking noise by combining the 70 dB pink noise with a 2,000 Hz high-pass filter. We observed changes in the fundamental frequency, sound pressure level, and formant 1 and 2 frequencies in both the experiments. We found that the sound pressure level and the formant 1 and 2 frequencies increased under the noise in both the experiments. The results of the second experiment indicated that the sound pressure level and formant 1 and 2 frequencies decreased when the 2,000 Hz high-pass filter was used, regardless of the presence or absence of the noise. Moreover, both experiments did not show any changes in the fundamental frequency of the vocalization due to the noise and filter.*

**Keywords:** Auditory feedback, Singing, Masking noise, Lombard effect, Fletcher effect

1. **Introduction.** In speech production, people receive their own voices through their auditory organs and then modulate them through feedback to their vocal organ. This feedback is called auditory feedback. Several studies have revealed that the speech process includes its production and perception.

Lombard demonstrated that the sound pressure level of the speaking voice of a person increases in a noisy environment [1]. This phenomenon is called the Lombard effect and some of its features have been proven by various studies. Korn [2] showed that the vocal sound pressure level increases by approximately 0.38 dB for every 1.0 dB increase in the noise level above 55 dB. Kano [3] investigated the effects of three types of noises on speech production, where each noise had its peak in the low, middle, and high frequencies, respectively. It was found that the rise in the vocal sound pressure level was independent of the frequency distribution of the noise and showed an average increase of 0.49 dB for every 1 dB of amplified noise up to 50-90 dB. Previous studies have reported that apart from the increase in the vocal sound pressure level, there are also other changes in the features of a voice. Uemura et al. [4] showed that the fundamental frequency, formant 1 and 2 frequencies, and power spectrum of the voice increased under noise. In addition to the increase in the fundamental frequency and the sound pressure level, Patel and Schell [5] indicated that the speech duration also increases under noise.

Further, delayed auditory feedback (DAF) has also focused on the role of auditory feedback. Lee [6] demonstrated that when speakers received their own voice after a delay, their speaking errors and speech durations increased. Fairbanks [7] clarified that the fundamental frequency and sound pressure level of the voice of the speaker increased with a DAF. Some researchers have used noise to mask their bone conduction voice during the DAF [6-9], while avoid using such a noise [10-12]. When noise is used in the DAF, the sound pressure level of the noise is approximately 70 dB and is then presented with delayed speech stimuli.

The DAF disrupts the speech production by strong interference, and it was considered to be unsuitable for investigating normal conditions of the auditory feedback. Therefore, altered auditory feedback (AAF) was proposed [13]. In AAF, participants utter a steady vowel while wearing headphones. The characteristics of the feedback voice are artificially manipulated. This manipulated voice is presented to the speaker via headphones instead of the voice transmitted naturally from the mouth to the ear during the normal course of speech. The pitch of the feedback voice is shifted by a pitch converter. The pitch converter shifts the pitch in the positive or negative direction from several cent to several hundred cent for several hundred milliseconds. The results of the pitch shift show that the compensation response, which is the response to the pitch shift in the negative or positive direction, appears with a latency of approximately 150 ms [14-16]. The response is usually not in the same direction as that of the pitch shift [15]. In addition, academic literature on AAF suggests that a masking noise can be used to mask the sounds from the surrounding environment and the air conduction sound. Various methods for masking have been discussed in past studies [13-22]. These are summarized below. Noise levels used here are from 50 to 80 dB, which mostly require a 70 or 80 dB noise to mask the original utterance. Purcell and Munhall [22] examined whether the auditory system compensates for changes in the first formant in the utterances of vowels using speech-shaped noise at 50 dB. Hafke [21] tested the increase or decrease in the unconscious control of voice intensity when the sound pressure level in the auditory feedback was increased or decreased by small measures in the presence of a pink noise of 60 dB. The remaining studies in Table 1 [13,15,17-20] investigated the reflexive control of fundamental frequency of the voice due to the pitch-shift under varying noise conditions. Summers et al. [23] reported that when spoken in variable noise environments that in quiet and in 80, 90 and 100 dB of masking noise which was white noise low-pass filtered at 3.5 kHz, speech intensity, duration, pitch and formant frequency increased as the noise increased. Although researches summarized below have been studied in varying types of noise: pink noise, white noise, multitalker babble noise and speech shaped noise, the effects of these variable noises have not been revealed yet.

On the other hand, there is much research on the effects of using noise masking [13-22] as well as the effects of not using noise masking [24-29]. In the studies that did not use noise masking, the sound pressure level of the feedback voice was generally amplified to 10 dB greater than the original voice. Therefore, each researcher in the past studies has used his or her own method, and there is a lack of a unified approach on the use of noise masking. Moreover, in AAF, the participants receive their own voice in real time. When people receive a loud feedback of their own voice, they speak in a lower voice. On the contrary, if the feedback voice becomes low, they speak more loudly. This is called the Fletcher effect [30-32]. From the above, it can be concluded that it is necessary to fundamentally assess the loudness of the masking noise and feedback voice. However, there has been little research on this, especially with respect to singing.

In the present study, two experiments were performed to investigate the effect of masking noise in the auditory feedback on singing. In the first experiment, the participants uttered a steady vowel /a/ at three pitch heights (C3, G3, and C4) with and without noise masking. In the second, we used a 2,000 Hz high-pass filter in addition to noise masking. It is believed that sounds over 2,000 Hz are important in the case of male singers. The singer's formant, which typically appears in a male operatic voice, is observed above 2,000 Hz [33]. The male singer is expected to emphasize the sounds above 2,000 Hz, thus

TABLE 1. Summary of altered auditory feedback studies using masking noise

| Authors | Voice | Noise conditions | Feedback voice conditions |
|---|---|---|---|
| Purcell and Munhall [22] | /ε/ vowel | 50 dB speech-shaped noise | Feedback at 80 dB |
| Hafke [21] | Vowel /u/ at 150 Hz for male voice, 240 Hz for female voice | 60 dB pink noise | Feedback at 75 dB |
| Hawco and Jones [19] | /a/ vowel at D4 in female voice | 70 dB pink noise | Feedback at 85 dB<br>Utterance at 75 dB |
| Sivasankar et al. [18] | /u/ vowel at a comfortable pitch in female voice | 70 dB pink noise | Feedback at 88 dB<br>Utterance at 77 dB |
| Natke and Kalveram [17] | /tatatas/ with long stressed and unstressed patterns | 70 dB low-pass-filtered white noise (fc = 900 Hz) | Calibrating a sine tone of 440 Hz and 75 dB to a feedback volume of 70 dB |
| Kawahara [13] | Sustained vowel at a comfortable pitch | 70 or 80 dB pink noise | 10 dB louder than the sidetone |
| Liu and Larson [15] | /u/ vowel at C3 and C4 for male voice, A3 and E4 for female voice | Vocal signal and 40 dB pink noise are mixed and amplified to 80 dB | Feedback at 80 dB<br>Utterance at 70 dB |
| Keough and Jones [20] | /ta/ at F4, G4, and A4 in female voice | 80 dB malti-talker babble noise | – |

affecting the auditory feedback [34]. The purpose of this study is to identify the effects of noise masking in the auditory feedback on singing [35-37]. Here, we have considered the uttering of a specific vowel at specific pitch heights as singing.

## 2. Method.

2.1. **Participants.** Healthy young males (ages 21-23) participated in the two experiments. Six males participated in experiment 1 and six males in experiment 2. Three participants took part in both the experiments. They were first required to pass a test for their sense of pitch, in which two successive tones containing unison, minor second, perfect fifth, and octave were generated using with a keyboard, and the participants were asked to answer whether the second tone was higher, lower, or the same compared to the first tone. None of participants reported a history of neurological, speech, or hearing disorders.

2.2. **Apparatus.** Participants wore headphones (Pioneer/SE-M870) in an anechoic chamber during experiment 1. They were asked to vocalize at 85 dB while looking at a sound level meter (Brüel & Kjær/Type 2250). The vocal signal from the microphone (ONOSOKKI/MI-1235) was amplified with a microphone amplifier unit (ONOSOKKI/SR-2200). A masking pink noise was presented via a PC (MacBook Air). The feedback voice and the masking noise passed through a mixer (KORG/D888). The feedback voice was amplified to 85 dB to obtain a level similar to the utterance, thus avoiding the Fletcher effect. The masking noise was amplified to 70 dB before it was presented via headphones. The sound pressure level at the ear pad of the headphones was measured with the B & K sound level meter.

The participants wore headphones in an anechoic chamber throughout the sessions in experiment 2 as well. They were asked to vocalize at 85 dB while looking at a sound level meter (RION/NL-31). The vocal signal from the microphone (Brüel & Kjær/Type 4189) was amplified with a microphone amplifier unit (ONOSOKKI/SR-2200) and passed through a high-pass analogue filter (NF Corporation MS-525) with a cut-off frequency of 2,000 Hz. The character of the high-pass filter was an 8-th order Butterworth. The frequencies lower than 2,000 Hz were attenuated by 48 dB per octave by the high-pass filter. The masking pink noise was the same as that in experiment 1. The vocal signal alone was passed through the high-pass filter. The masking pink noise was not passed through it. The feedback voice and the masking noise passed through a mixer (KORG/D888). The normal and the high-pass filtered feedback voice were amplified to 85 dB. Amplifying the high-pass filtered voice to 85 dB meant sounds over 2,000 Hz being emphasized because sounds below 2,000 Hz were attenuated. The masking noise was amplified to 70 dB and the sound pressure levels were checked at the same point as in experiment 1 with the B & K sound level meter. The signal-to-noise (S/N) ratio was 15 dB in both the experiments.

2.3. **Procedures.** In the first experiment, participants were instructed to sing a steady vowel /a/ at three pitch heights, C3 (130.81 Hz), G3 (196 Hz), and C4 (261.63 Hz), with and without noise masking. In the second experiment, the normal voice feedback and the high-pass filtered voice feedback with and without noise masking were presented to the participants. Before the experiments, several practice trials were conducted to ensure that the participants could match the notes within 100 cent. Next, the participants were asked to sustain the vowel /a/ for approximately 5 s duration at each of the three pitch heights. Sine waves, whose pitch heights were twice as much as C3, G3, and C4, were presented using the PC before each trial and after the experimenter gave the signal to begin the utterance to obtain the vocalized note. The reason we doubled the frequency of the sine waves is that the actual sounds were so low that it was difficult to confirm these pitch heights. During the utterance, the sine waves were not presented. In experiment 1, six trials constituted an experimental block, and the masking noise was either presented or not presented for each note. Each block was repeated five times for a total of 30 attempts for each participant. All trials within a block were randomized. In experiment 2, each experimental block included 12 vocalizations. For each note, the voice feedback was either normal or high-pass filtered and with or without the masking noise. These trials were randomized and performed five times each for a total of 60 trials per participant.

2.4. **Data analysis.** We used Praat [38] to analyze the voice waveforms for each participant separately in terms of the fundamental frequency, sound pressure level, and formant 1 and 2 frequencies under each experimental condition. We obtained 180 valid trials in experiment 1 and 300 valid trials in experiment 2. We averaged the data for each condition and ran a two-tailed paired t-test. In experiment 1, this test was performed in between the experimental conditions where a masking noise was used for each note and where it was not used. In experiment 2, it was performed first in between the experimental conditions where masking noise was applied without filtering (MN) and where no masking noise was applied and no filtering was done (NN). Next, it was performed in between the experimental conditions where the masking noise was applied with filtering (MF) and no masking noise was applied but filtering was done for each note (NF).

## 3. Results.

3.1. **Fundamental frequency.** Figure 1 reports the mean of the fundamental frequencies obtained with and without noise masking in experiment 1. The fundamental frequencies are expressed by converting to cent from Hz using the formula 1200log2(f1/f0) as the pitch ratio between the actually uttered voice of the fundamental frequency (f1) and the utterance fundamental frequency at which the utterance challenge (f0) (100 cent = 1 semitone). There was no significant difference in the fundamental frequencies for each note. The difference in the fundamental frequencies in the presence and absence of noise masking was 0.98 cent at C3, 0.02 cent at G3, and 2.66 cent at C4 for each note. When the masking noise was applied, all the pitch heights were lower compared to the case where no masking was applied.

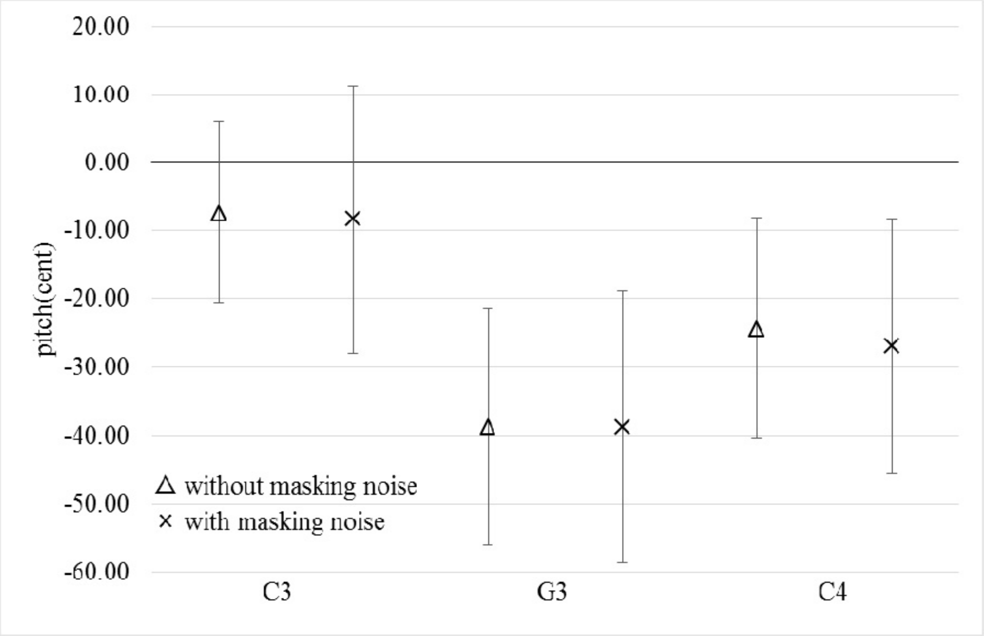

Figure 1. Mean of the fundamental frequencies with and without the masking noise for each note in experiment 1 (cent). Error bars represent the standard deviation. Thick black line at 0.00 cent indicates the target pitch heights: C3 (130.81 Hz), G3 (196 Hz), and C4 (261.63 Hz).

In experiment 2, there was again no significant difference in each note under the different experimental conditions (Figure 2). However, C3MN was higher than C3NN, and the fundamental frequency tended to decrease when the masking noise was presented regardless of whether filtering was performed.

3.2. **Sound pressure level.** Figure 3 shows the mean of the sound pressure level with and without the masking noise for each note in experiment 1. Although, participants sang while seeing the monitor, as the fundamental frequency increased, the sound pressure level increased. A significant difference appeared in C3 ($p < 0.01$) and it was observed that presenting the masking noise increased the sound pressure level for every note.

It was also observed that the higher the fundamental frequency was, the higher the sound pressure level was (Figure 4). There were significant differences between all the values except for those between C3NF and C3NN, and C4NF and C4MF ($p < 0.01$). When the masking noise was presented, the sound pressure level increased regardless of

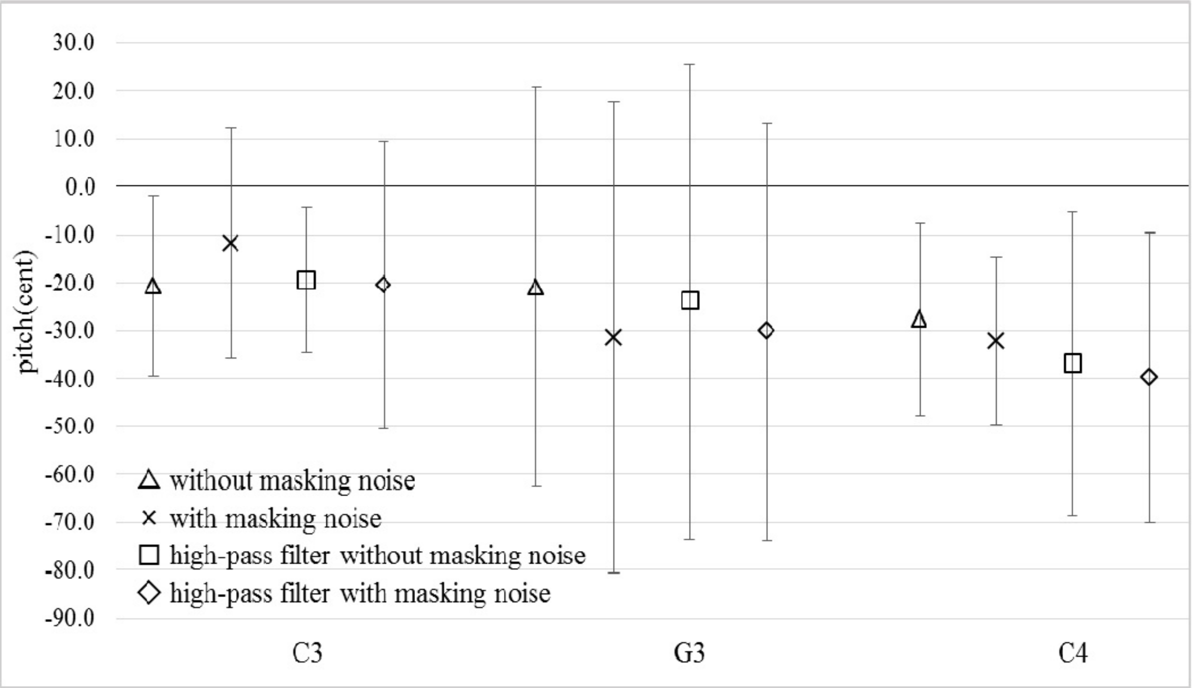

Figure 2. Mean of the fundamental frequency of the filtered and unfiltered conditions with and without the masking noise for each note of experiment 2 (cent). Error bars represent the standard deviation. Thick black line at 0.00 cent indicates the target pitch heights: C3 (130.81 Hz), G3 (196 Hz), and C4 (261.63 Hz) from left to right every four marks. The marks indicate NN, MN, NF and MF condition from left to right for each note.

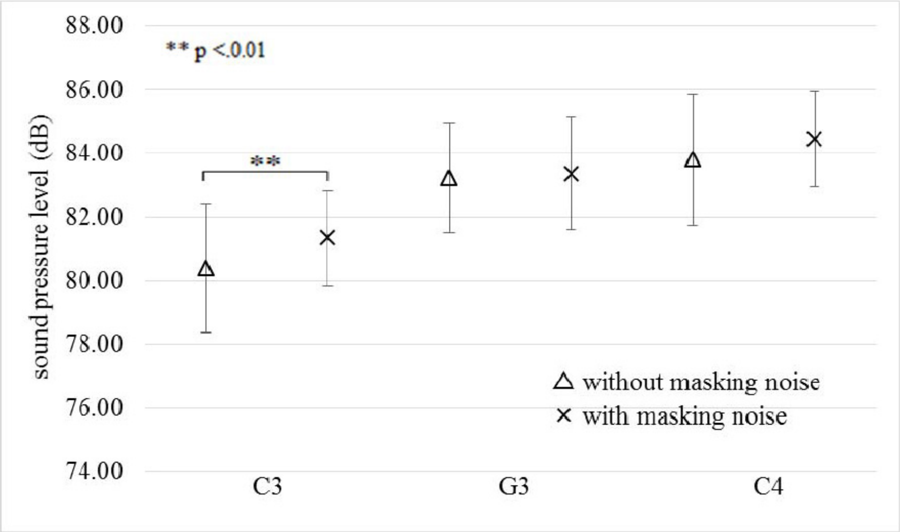

Figure 3. Mean of the sound pressure levels with and without the masking noise for each note in experiment 1 (dB). Error bars represent the standard deviation. Each set of two lines from left to right represents the target pitch height.

the filter. When the high-pass filter was applied, then the sound pressure level decreased for each note.

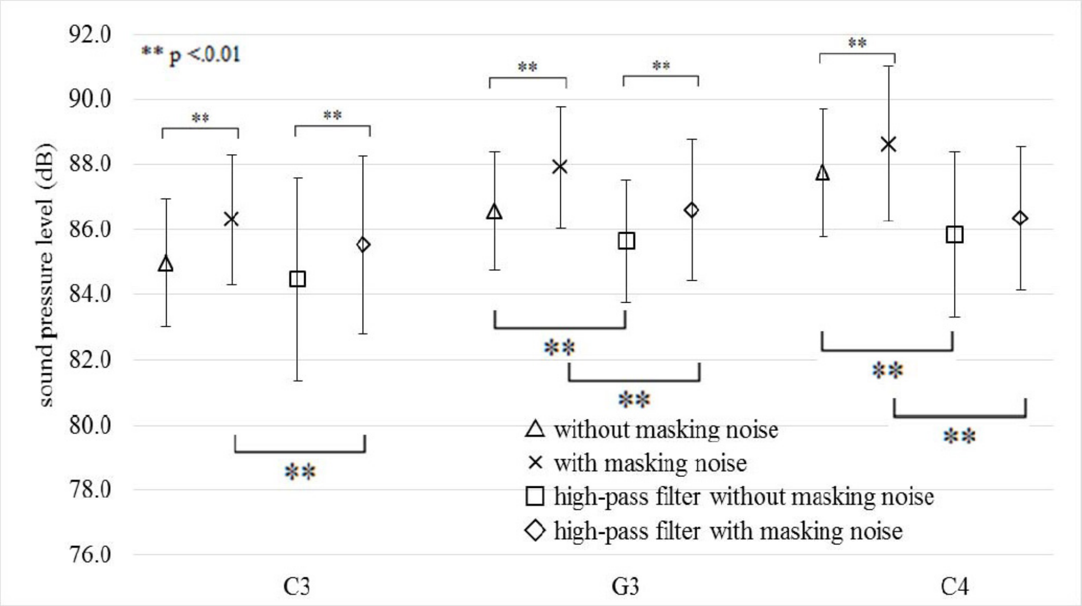

FIGURE 4. Mean of the sound pressure levels with and without filtering and with and without the masking noise for each note of experiment 2 (dB). Error bars represent the standard deviation. Each set of four lines from left to right represents the target pitch heights for NN, MN, NF, and MF.

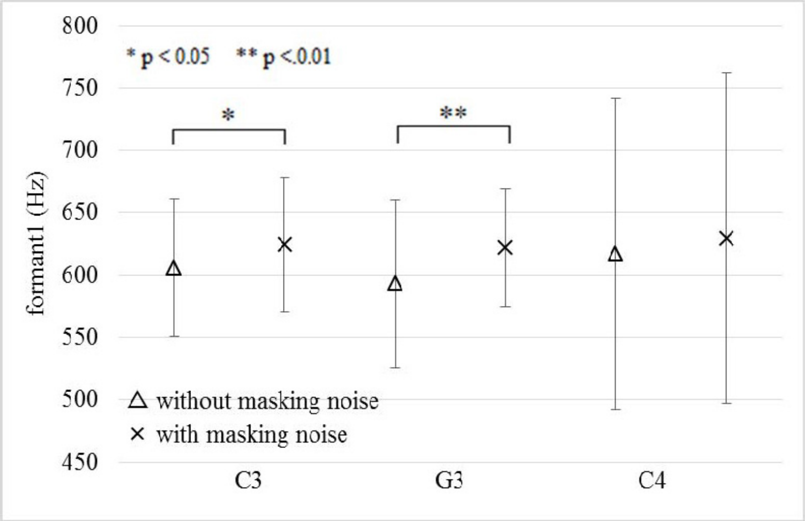

FIGURE 5. Mean of the formant 1 frequency with and without the masking noise for each note of experiment 1 (Hz). Error bars represent the standard deviation. Each set of two lines from left to right represents the target pitch.

3.3. **Formant frequency.** We analyzed formant 1 and 2 frequencies with respect to the vocalization of the vowel. In the first experiment, the formant 1 frequency increased significantly in C3 ($p < 0.05$) and G3 ($p < 0.01$) when the masking noise was applied. A participant had difficultly in uttering high notes in the C4 pitch, leading to a wide variation in the recorded data. No significant difference was observed in this case ($p = 0.143$). However, the mean of formant frequency F1 in C4 was higher when the masking noise was applied than when no masking was applied (Figure 5). Formant 2 frequency

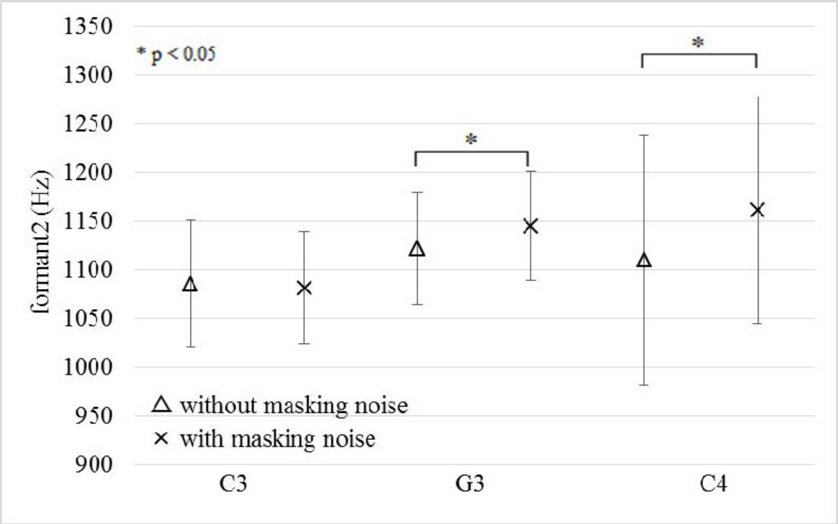

FIGURE 6. Mean of the formant 2 frequency with and without the masking noise for each note of experiment 1 (Hz). Error bars represent the standard deviation. Each set of two lines from left to right represents the target pitch heights.

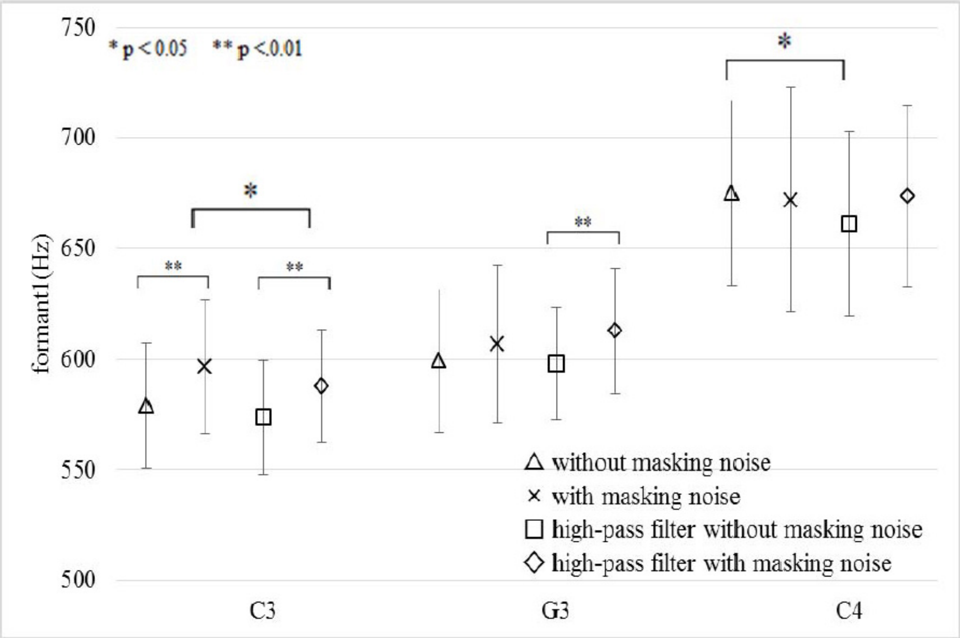

FIGURE 7. Mean of the formant 1 frequency with and without the filter and with and without the masking noise for each note in experiment 2 (Hz). Error bars represent the standard deviation. Each set of four lines from left to right represents the target pitch heights for NN, MN, NF, and MF.

increased significantly in the G3 and C4 pitches when the masking noise was applied, as shown in Figure 6. However, small differences were observed in the C3 pitch between the conditions where a masking noise was applied and where no masking noise was applied.

In the second experiment, significant differences were observed between C3MN and C3NN, C3MF and C3NF ($p < 0.01$), C3MF and C3MN ($p < 0.05$), G3MF and G3NF ($p < 0.01$), and C4NF and C4NN ($p < 0.05$) in the first formant frequency. The formant

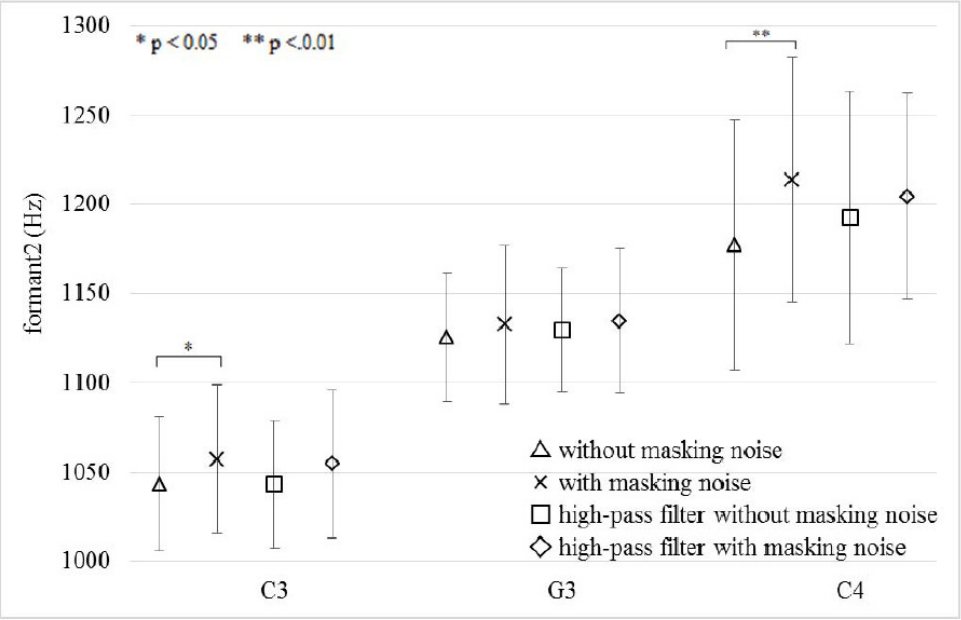

FIGURE 8. Mean of the formant 2 frequency of the filtered and unfiltered conditions with and without a masking noise for each note of experiment 2 (Hz). Error bars represent the standard deviation. Each set of four lines from left to right represents the target pitch heights for NN, MN, NF, and MF.

frequency increased when the masking noise was applied and decreased when the high-pass filter was applied. The fundamental frequency showed an increase with the increase in the first formant frequency (Figure 7). In the second formant frequency (Figure 8), significant differences were observed between C3MN and C3NN ($p < 0.05$), and C4MN and C4NN ($p < 0.01$). In addition, it was observed that the fundamental frequency increased with the increase in the second formant frequency.

## 4. Discussion.

4.1. **Fundamental frequency.** There were no statistically significant differences observed in the fundamental frequencies in both the experiments. In experiment 1, the differences between the experimental conditions where the masking noise was applied and where no masking noise was applied were 0.98 cent at C3, 0.02 cent at G3, and 2.66 cent at C4. In experiment 2, the differences under the similar situations were 1.0 cent between C3MF and C3NF, 10.6 cent between G3MN and C3NN, 6.2 cent between G3MF and G3NF, 4.6 cent between C4MN and C4NN, and 3.0 cent between C4MF and C4NF. All the results, except the difference between C3MN and C3NN, showed that the fundamental frequency decreased when the masking noise was applied. Wier et al. [39] reported that the pitch discrimination level for pure tones was 8.6 cent at 200 Hz and 4.3 cent at 400 Hz. These results suggest that our findings were considerably small. Considering the Lombard effect, it was assumed that it increased in the masked condition, but the results indicated the opposite. Letowski et al. [40] demonstrated that when a 70 dB wideband noise was presented as normal speech via earphones, the fundamental frequency increased by 12 Hz against a quiet condition for male subjects. If the fundamental frequency of conversation is assumed to be 135 Hz for males, increasing by 12 Hz would imply 147 cent ascending. It has been proposed that 70 dB of noise affects normal speech; however, no significant effects are observed during singing, which is considered to sustain a specific fundamental frequency. Mürbe et al. [37] studied the ability to control the fundamental frequency under a 105 dB band-pass noise (50-2,000 Hz) in students training to become professional solo singers. The students were tested at the beginning of their training and after three years of training. Significant differences were identified in their ability to control the fundamental frequency regardless of their period of education. This suggests that a loud noise can affect the control of the fundamental frequency during singing. However, a 70 dB pink noise is not effective level. In experiment 2, we used a 2,000 Hz high-pass filter. The fundamental frequency decreased because of the filtering. However, the difference was not significant. It is debatable whether a change in the amplification level or cut-off frequency will result in a significant decrease in the fundamental frequency. Pitch and phonetic information are not available for frequencies above 2,000 Hz. Therefore, it can be concluded that a 2,000 Hz high-pass-filtered auditory feedback does not affect the control of the fundamental frequency of human voice.

4.2. **Sound pressure level.** The results indicate an increase in the sound pressure level with increasing fundamental frequency regardless of whether the participants used the sound level meter to check their voice level while vocalizing. This increase appeared to be a natural response in which the specified sound pressure level could not be maintained as the fundamental frequency increased. It was demonstrated that in almost all the cases where the masking noise was applied, there was a significant increase in the sound pressure level. It can be concluded that a 70 dB pink noise can increase the sound pressure level while singing. When the participants received a 2,000 Hz high-pass-filtered voice amplified to 85 dB as feedback, which was louder than the original utterance, the sound pressure level decreased significantly. Human hearing is the most sensitive in the frequency band between 3,000 Hz and 4,000 Hz. When the high-pass filter was used, the sounds in the frequency band above 2,000 Hz were emphasized. It is possible that this has caused the decrease in the sound pressure level to enable the reception of the emphasized sound. This suggests that sounds above 2,000 Hz can lead to the Fletcher effect in singing. Moreover, these results indicate that the sound pressure level decreases when the subjects receive the high-pass filtered auditory feedback regardless of the presence of the masking noise.

4.3. **Formant frequency.** Uemura et al. [4] showed that the formant 1 and 2 frequencies increased under noise during speech. When noise masking was applied, we observed an increase in the first and second formant frequency. This demonstrates that the Lombard effect occurs not only in speech, but also in singing. When a 2,000 Hz high-pass filter was applied in the auditory feedback, it was observed that formant 1 frequency decreased significantly between C3MF and C3MN, and C4NF and C4NN when the participants received the emphasized high-pass filtered voice. However, the observations in C3MF and C3MN were made under the effect of a masking noise, whereas those in C4NF and C4NN were made without a masking noise. Therefore, we cannot determine whether the difference in formant 1 frequency was simply the effect of the high-pass filter or a combination of the filter and masking noise. The observations in this case merely indicate a trend. Further investigations are needed to explain the observations.

5. **Conclusion.** We assessed the effects of the masking noise in auditory feedback on singing using two experiments. The first experiment involved singing a sustained vowel /a/ at three pitch heights C3, G3, and C4 with and without noise masking using a 70 dB pink noise. The second experiment also involved singing in a similar fashion. However, in this case, a 70 dB pink noise and a 2,000 Hz high-pass filter were applied to the auditory feedback. As a result, the following was determined.

i. A 70 dB pink noise and 2,000 Hz high-pass filter do not change the fundamental frequency in singing.
ii. A 70 dB pink noise increases the sound pressure level and a 2,000 Hz high-pass filter causes the sound pressure level to decrease, considering the Lombard and Fletcher effect in singing.
iii. A 70 dB pink noise leads to an increase in the first and second formant frequencies in singing.

Thus, it is demonstrated that the Lombard and Fletcher effects can be observed in singing when presented with a masking noise and 2,000 Hz high-pass filter. Moreover, the use of noise masking in experiments on auditory feedback is questionable. Because we can observe the effects of the high-pass filter without the masking noise. The use of a masking noise can result in causing the Lombard effect, which will affect the results of the experiments. In addition, a masking noise is an auditory load. In our future work, we will conduct experiments that change the sound pressure level of the masking noise and feedback sound, and investigate the role of auditory feedback on singing.

**Acknowledgment.** This work was supported by KAWAI FOUNDATION FOR SOUND TECHNOLOGY & MUSIC. We would like to thank Editage (www.editage.jp) for English language editing.

## REFERENCES

[1] E. Lombard, Le signe de l'elevation de la voix, *Annuals Maladies Oreille Larynx, Nez, Pharynx*, vol.37, pp.101-109, 1911.
[2] T. S. Korn, Effect of psychological feedback on conversational noise reduction in rooms, *Journal of the Acoustical Society of America*, vol.26, pp.793-794, 1954.
[3] Y. Kano, A study of voice levels in noise for selection of hearing aids, *Journal of Otolaryngology of Japan*, vol.88, pp.138-147, 1985.
[4] Y. Uemura, M. Morise and T. Nishiura, Improvement of speech recognition performance based on the conversion of Lombard features, *IEICE Technical Report*, SP2010-1, pp.1-6, 2010.
[5] R. Patel and K. W. Schell, The influence of linguistic content on the Lombard effect, *Journal of Speech, Language, and Hearing Research*, vol.51, pp.209-220, 2008.
[6] B. S. Lee, Effects of delayed speech feedback, *Journal of the Acoustical Society of America*, vol.22, pp.824-826, 1950.
[7] G. Fairbanks, Selective vocal effects of delayed auditory feedback, *Journal of Speech and Hearing Disorders*, vol.20, pp.333-345, 1955.
[8] T. C. Hain, T. A. Burnett, C. R. Larson and S. Kiran, Effects of delayed auditory feedback (DAF) on the pitch-shift reflex, *Journal of the Acoustical Society of America*, vol.109, pp.2146-2152, 2001.
[9] K. T. Kalveram and L. Jäncke, Vowel duration and voice onset time for stressed and nonstressed syllables in stutterers under delayed auditory feedback condition, *Folia Phoniatrica et Logopaedica*, vol.41, no.1, pp.30-42, 1989.
[10] P. Howell, The effect of delaying auditory feedback of selected components of the speech signal, *Perception & Psychophysics*, vol.34, no.4, pp.387-396, 1983.
[11] A. Stuart and J. Kalinowski, Effect of delayed auditory feedback, speech rate, and sex on speech production, *Perceptual and Motor Skills*, vol.120, no.3, pp.747-765, 2015.
[12] A. Stuart, J. Kalinowski, M. P. Rastatter and K. Lynch, Effect of delayed auditory feedback on normal speakers at two speech rates, *Journal of the Acoustical Society of America*, vol.111, no.5, pp.2237-2241, 2002.
[13] H. Kawahara, Interactions between speech production and perception under transformed auditory feedback, *IEICE Technical Report*, HC93-70, pp.33-40, 1994.
[14] R. Yanaga and H. Kawahara, F0 control dynamic parameters in speech and singing test results using transformed auditory feedback with a revised procedure, *IPSJSIG Technical Report*, vol.2003, no.82, pp.71-76, 2003.
[15] H. Liu and C. R. Larson, Effects of perturbation magnitude and voice F0 level on the pitch-shift reflex, *Journal of the Acoustical Society of America*, vol.122, no.6, pp.3671-3677, 2007.

[16] M. Okada, K. Mori, S. Okazaki, C. Cai, A. Nishikido and K. Ochi, Differences between males and females in voice F0 control to pitch-shifted auditory feedback, *IEICE Technical Report*, SP2011-167, pp.65-69, 2012.

[17] U. Natke and K. T. Kalveram, Effects of frequency-shifted auditory feedback on fundamental frequency of long stressed and unstressed syllables, *Journal of Speech, Language, and Hearing Research*, vol.44, no.3, pp.577-584, 2001.

[18] M. Sivasankar, J. J. Bauer, T. Babu and C. R. Larson, Voice responses to changes in pitch of voice or tone auditory feedback, *Journal of the Acoustical Society of America*, vol.117, no.2, pp.850-857, 2005.

[19] C. S. Hawco and J. A. Jones, Control of vocalization at utterance onset and mid-utterance: Different mechanisms for different goals, *Brain Research*, vol.1276, pp.131-139, 2009.

[20] D. Keough and J. A. Jones, The sensitivity of auditory-motor representations to subtle changes in auditory feedback while singing, *Journal of the Acoustical Society of America*, vol.126, no.2, pp.837-846, 2009.

[21] H. Hafke, Nonconscious control of voice intensity during vocalization, *Archives of Acoustics*, vol.34, no.4, pp.407-414, 2009.

[22] D. W. Purcell and K. G. Munhall, Compensation following real-time manipulation of formants in isolated vowels, *Journal of the Acoustical Society of America*, vol.119, no.4, pp.2288-2297, 2005.

[23] W. van Summers, D. B. Pisoni, R. H. Bernacki, R. I. Pedlow and M. A. Stokes, Effects of noise on speech production: Acoustic and perceptual analyses, *Journal of the Acoustical Society of America*, vol.84, pp.917-928, 1988.

[24] P. Liu, Z. Chen, J. A. Jones, D. Huang and H. Liu, Auditory feedback control of vocal pitch during sustained vocalization: A cross-sectional study of adult aging, *PloS One*, vol.6, no.7, 2011.

[25] H. Liu, R. Behroozmand, M. Bove and C. R. Larson, Laryngeal electromyographic responses to perturbations in voice pitch auditory feedback, *Journal of the Acoustical Society of America*, vol.129, no.6, pp.3946-3954, 2011.

[26] R. Behroozmand, N. Ibrahim, O. Korzyukov, D. A. Robin and C. R. Larson, Functional role of delta and theta band oscillations for auditory feedback processing during vocal pitch motor control, *Frontiers in Neuroscience*, vol.9, 2015.

[27] R. Behroozmand, O. Korzyukov, L. Sattler and C. R. Larson, Opposing and following vocal responses to pitch-shifted auditory feedback: Evidence for different mechanisms of voice pitch control, *Journal of the Acoustical Society of America*, vol.132, no.4, pp.2468-2477, 2012.

[28] H. Liu, M. Meshman, R. Behroozmand and C. R. Larson, Differential effects of perturbation direction and magnitude on the neural processing of voice pitch feedback, *Clinical Neurophysiology*, vol.122, no.5, pp.951-957, 2011.

[29] R. Behroozmand and C. R. Larson, Error-dependent modulation of speech-induced auditory suppression for pitch-shifted voice feedback, *BMC Neuroscience*, vol.12, no.1, 2011.

[30] H. Fletcher, G. M. Raff and F. Parmley, Study of the effects of different sidetones in the telephone set, *Western Electrical Company, Report*, (19412), 1918.

[31] H. Lane and B. Tranel, The Lombard sign and the role of hearing in speech, *Journal of Speech, Language, and Hearing Research*, vol.14, no.4, pp.677-709, 1971.

[32] G. M. Siegel and H. L. Pick Jr, Auditory feedback in the regulation of voice, *Journal of the Acoustical Society of America*, vol.56, no.5, pp.1618-1624, 1974.

[33] J. Sundberg and K. Sakakibara, *Utagoe no Kagaku (Science of the Singing Voice)*, Tokyo Denki University Press, Tokyo, 2007.

[34] H. Sawada and K. Kakehi, The effect of feedback speech information on fluent speech production, *IEICE Technical Report*, SP2002-192, pp.25-30, 2003.

[35] T. A. Pruitt and P. Q. Pfordresher, The role of auditory feedback in speech and song, *Journal of Experimental Psychology: Human Perception and Performance*, vol.41, no.1, pp.152-166, 2015.

[36] S. Iijima, S. Seikoba, S. Ishimitsu and M. Nakayama, Effects of high-pass filtered and pitch-shifted auditory feedback on a singer's pitch, *ICIC Express Letters, Part B: Applications*, vol.6, no.12, pp.3285-3289, 2015.

[37] D. Mürbe, F. Pabst, G. Hofmann and J. Sundberg, Effects of a professional solo singer education on auditory and kinesthetic feedback – A longitudinal study of singers' pitch control, *Journal of Voice*, vol.18, pp.236-241, 2004.

[38] P. Boersma and D. Weenink, *Praat: Doing Phonetics by Computer*, http://www.fon.hum.uva.nl/praat/.

[39] C. C. Wier, W. Jesteadt and D. M. Green, Intensity discrimination as a function of frequency and sensation level, *Journal of the Acoustical Society of America*, vol.61, pp.178-184, 1977.

[40] T. Letowski, T. Frank and J. Caravella, Acoustical properties of speech produced in noise presented through supra-aural earphones, *Ear and Hearing*, vol.14, no.5, pp.332-338, 1993.