ICIC Express Letters
Part B: Applications
Volume **10**, Number **1**, January **2019**

pp. **63–69**

EFFECTS OF GAZE ANGLE ON VOCALIZATION

Satoshi Iijima, Shunsuke Ishimitsu and Masashi Nakayama

Graduate School of Information Science
Hiroshima City University
3-4-1, Ozuka-Higashi, Asa-Minami-Ku, Hiroshima 731-3194, Japan
s-iijima@hiroshima-cu.ac.jp

Received July 2018; accepted October 2018

Abstract. *The present study investigates how changes in gaze angle affect vocalization. When people learn speech and singing, the direction of the student's eyes is considered to be important. 'Look far away', 'raise your line of sight', and 'do not look down too much' are typical instructions used by voice coaches. To test the effectiveness of these instructions, we recorded participants vocalizing a steady vowel /a/ at the pitch G3 (196 Hz) while staring at three different gaze angles, −35°, 0° and +35° in the vertical direction. The results indicate that the fundamental frequency and sound pressure level were not affected by the participant's gaze angle. Higher formant frequencies were somewhat affected by changes in the gaze angle, suggesting that eyelid or eyeball motion affects the third and fourth formant frequencies.*
Keywords: Speech production, Vocalization, Gaze angle

1. **Introduction.** Speech is produced using various sensory modes, including auditory, kinesthetic and visual feedback. Auditory feedback has obvious and well-known effects on speech production. When people speak, they sense their own voices through auditory organs and unconsciously compensate for various factors with feedback to the vocal organs. Lombard demonstrated that the sound pressure level (SPL) of a person's speaking voice increases in a noisy environment [1]. This phenomenon is called the Lombard effect. After this finding, several studies investigated the properties of this effect, and there have been many studies that have focused on the interaction between speech production and perception. Several studies have shown that speech production is possible even in very noisy environments [2-4]. Singing in noisy environments is also possible [5]. These results suggest that the role of auditory feedback diminishes after acquiring speech and singing skills; subsequently, the role of other feedback modalities (e.g., kinesthetic feedback for motor control) increases [6]. Furthermore, visual feedback has also been shown to affect speech perception, in phenomena such as the McGurk effect [7,8]. Additionally, pronunciation may be more easily learned by watching the tongue's movement for visual feedback [9]. Above all, many organs and functions are related to speech skills; however, the effect of gaze angle on speech production and vocalization has not been explored as yet.

When students learn speech or singing skills, the gaze angle is often considered important. Voice teachers often issue instructions, including 'look far away', 'raise your line of sight' and 'do not look down too much'. The use of these coaching phrases anecdotally suggests that the gaze angle affects speech production. There have been some studies that investigated how sounds affected eye gaze [10-12]. Moreover, there have been studies investigating to extract emotions from the gaze [13,14]. However, these studies are not about how the gaze affects human's activities but how they are affected, and few studies

DOI: 10.24507/icicelb.10.01.63

have addressed the effects of gaze angle on vocalization. Hence, we designed an experiment to test the effect of gaze angles on the waveform produced by participants without speech and hearing disorders.

In the experiment, participants were asked to vocalize the vowel /a/ for approximately five seconds at the G3 pitch (196 Hz) while gazing up, straight ahead or down. The gaze angles were set to $-35°$, $0°$ and $+35°$. The $0°$ visual target was adjusted to the eye level of each participant. We then analyzed the voice waveforms for each participant in terms of the fundamental frequency (F_0), sound pressure level (SPL) and the first, second, third and fourth formant frequencies (F_1, F_2, F_3 and F_4). If a change in the gaze angle affects how the participants form the note, we should observe some changes in the recorded voice waveforms.

In this paper, we present findings from the experiment on the effects of gaze angle on vocalization. At first, we reviewed some of the early studies as introduction. Then, we showed our experimental methods in the next chapter. After that, we indicated some of the data taken from the experiment and discussed about the results. Finally, we summarized the present study.

2. **Methods.** We focused on the change of the vocal quality when participants changed their gaze angles in this experiment. The participants were asked to vocalize sustained vowel at a decided pitch height while changing their gaze angles in vertical direction without moving their heads. It was considered that the restriction, without moving head, could make it possible to assess the effects of the gaze angle on vocalization.

2.1. **Participants.** Six healthy young males (aged 21-23) participated in the experiment. None of the participants reported a history of neurological, speech or hearing disorders, and none had received any professional vocal training.

2.2. **Apparatus.** Figure 1 shows the experimental setup. The experiments were conducted in an anechoic chamber. The vocal signal was fed from a microphone (ONOSOKKI /MI-1235) to a pre-amplifier (ONOSOKKI/MI-3111) and then to a microphone amplifier unit (ONOSOKKI/AU-2200). Recordings were collected at a sampling frequency of 48 kHz with 16-bit quantization using a recording unit (ONOSOKKI/AU-4100A). A protractor and a laser pointer were used to mark the visual target to direct the participants' gaze angle.

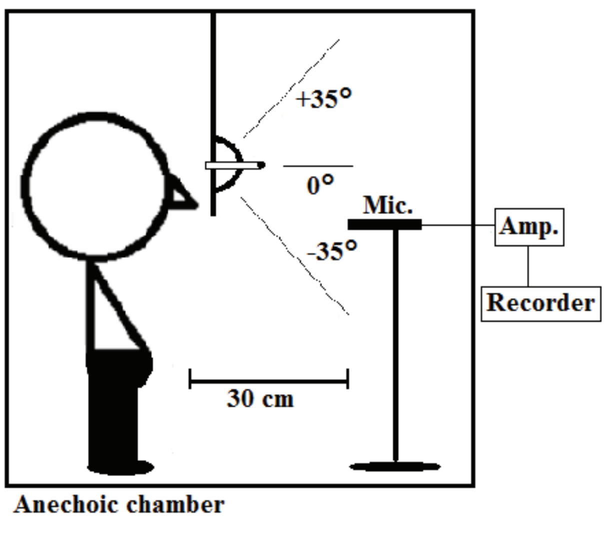

FIGURE 1. Experimental setup

2.3. **Procedure.** The participants were instructed to vocalize a steady vowel /a/ at the G3 pitch (196 Hz) for approximately five seconds. The participants were instructed to sound the note at whatever volume felt comfortable. Before the experiment, several practice trials were conducted to ensure that the participants could match the note within 100 cent. A sine wave, at a pitch double that of G3, was presented to the participants using a PC (MacBook Air) before each trial. We doubled the frequency of the reference sine wave because actual G3 was too low for most of the participants to match easily. The sine wave was not presented when the participants vocalized in the trials. The gaze angles were set to $-35°$, $0°$ and $+35°$ in the vertical direction. The $0°$ target was adjusted to the eye level of each participant. The participants sung the note while standing in the anechoic chamber and were instructed to change their gaze angle without moving their heads. The gaze targets were indicated using a laser pointer affixed to a protractor. The participants performed the vocalizations while gazing in all three directions as one set. The order of the trials was randomized three times and each participant repeated the task thrice, thereby producing a total of 9 trials per participant.

2.4. **Data analysis.** We recorded 45 valid trials (5 participants $\times$ 3 directions $\times$ 3 sets) from the experiment. One of the participants was excluded from the analysis as his vocal pitch was unstable. The software Praat [8] was used to analyze the voice waveforms for each participant separately in terms of F_0, SPL and F_1, F_2, F_3 and F_4 for each experimental condition. The formants were analyzed using the Burg method with a maximum of five formants. In the analysis, we used vocal signals that were recorded from 0.5 s to 3.5 s after the beginning of the utterance, and we averaged the data for each condition.

3. Results.

3.1. **Fundamental frequency.** Figure 2 reports the mean F_0 for the three different gaze angles. The thick solid black line along 0 cent indicates the target pitch of G3. The error bars represent 95% confidence intervals. The F_0 are expressed by converting to cent from Hz (100 cent = 1 semitone). The results show that the mean F_0 were -34.14 cent at $-35°$, -31.57 cent at $0°$ and -27.58 cent at $+35°$.

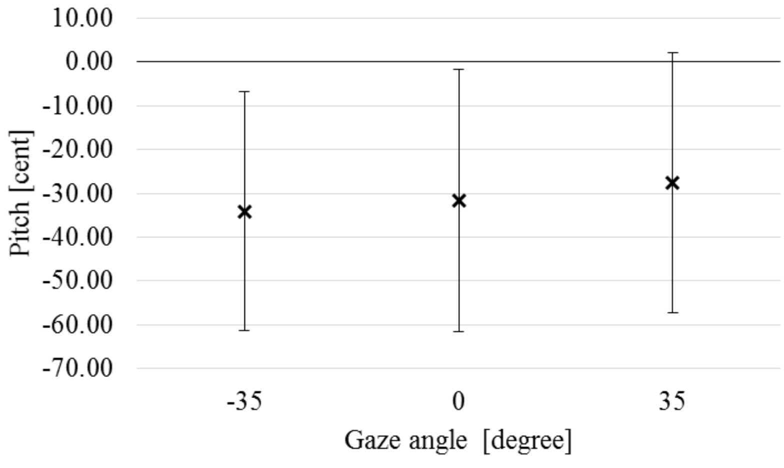

FIGURE 2. Mean F_0 for the three gaze angle conditions. The error bars represent 95% confidence intervals.

3.2. **Sound pressure level.** Figure 3 plots the mean SPL for the three gaze angle conditions. The error bars represent 95% confidence intervals. The mean SPLs were 72.36 dB at $-35°$, 72.85 dB at $0°$ and 72.62 dB at $+35°$.

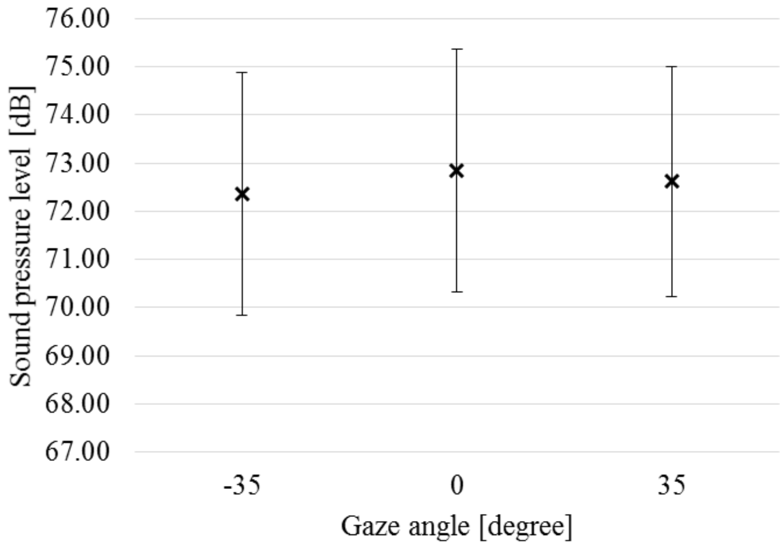

FIGURE 3. Mean SPL for the three gaze angle conditions. The error bars represent 95% confidence intervals.

3.3. **Formant frequency.** Figures 4, 5, 6 and 7 show the mean F_1, F_2, F_3 and F_4, respectively, for the three gaze angle conditions. The error bars represent 95% confidence intervals. The data showed that the mean F_1 were 603 Hz at $-35°$, 601 Hz at 0° and 601 Hz at +35°; the mean F_2 were 1,150 Hz at $-35°$, 1,136 Hz at 0° and 1,132 Hz at +35°; the mean F_3 were 2,757 Hz at $-35°$, 2,765 Hz at 0° and 2,834 Hz at +35°; and the mean F_4 were 3,574 Hz at $-35°$, 3,567 Hz at 0° and 3,621 Hz at +35°.

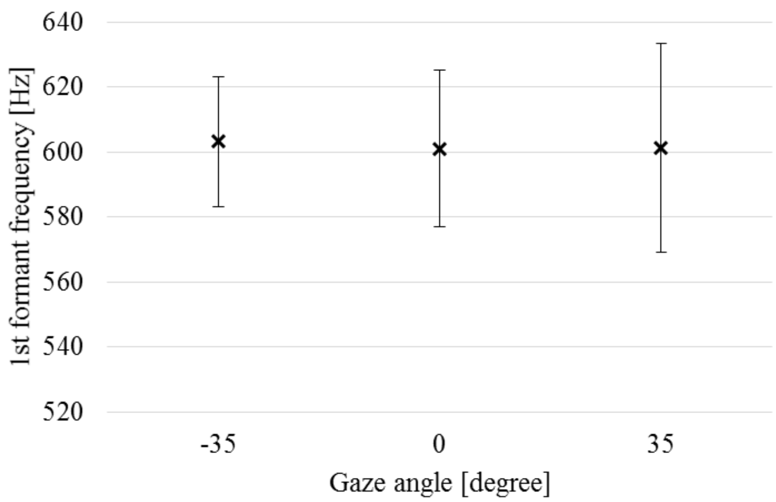

FIGURE 4. Mean F_1 for the three gaze angle conditions. The error bars represent 95% confidence intervals.

4. **Discussion.** Participants produced stable pitches when gazing at all three angles in the experiment. Katz et al. reported that the pitch-discrimination level for pure tones is 8.6 cent at 200 Hz and 4.3 cent at 400 Hz [9]. The mean F_0 in the present tests were -34.14 cent at $-35°$ and -27.58 cent at +35°; hence, the difference of the mean F_0 between $-35°$ and +35° was 6.56 cent. Following the result of Katz et al., this difference in pitch is imperceptible. This suggests that a change in gaze angle does not perceptibly affect the vocal pitch produced. Likewise, the SPLs indicated almost the same dB at all three gaze angles. Therefore, a change in gaze angle does not seem to affect the vocal SPL.

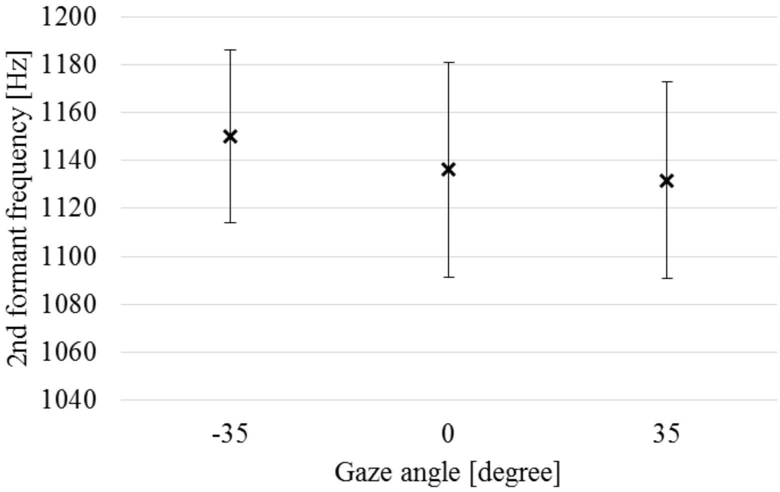

Figure 5. Mean F_2 for the three gaze angle conditions. The error bars represent 95% confidence intervals.

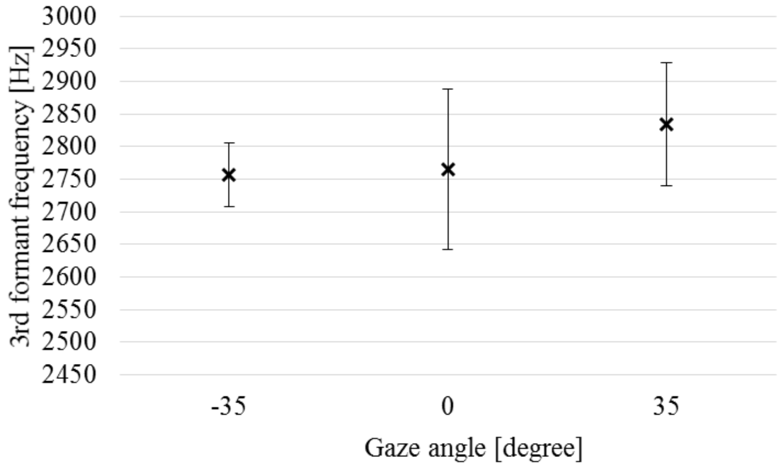

Figure 6. Mean F_3 for the three gaze angle conditions. The error bars represent 95% confidence intervals.

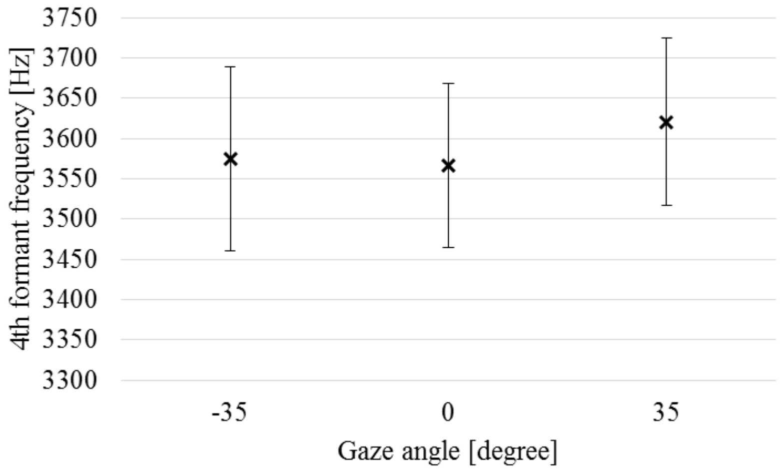

Figure 7. Mean F_4 for the three gaze angle conditions. The error bars represent 95% confidence intervals.

We found that the stable F_1 and F_2 at all three gaze angles and formant frequencies at $+35°$ were a little higher than those at $-35°$ and $0°$ for F_3 and F_4. The F_3 and F_4 are generally understood to depend on the extent of the oral cavity and vocal tract length [10]. The suprahyoid muscle affects the vertical motion of the hyoid and larynx [11]. The suprahyoid group comprises many muscles including the stylohyoid, digastric, mylohyoid and geniohyoid muscles, and they are controlled by various nerves. For example, the stylohyoid is controlled by the facial nerve, the mylohyoid is controlled by the trigeminal nerve and the geniohyoid is controlled by the hypoglossal nerve. As people look upwards without moving their heads, their upper eyelids rise up. The eyelid is controlled by the facial nerve; hence, the gazing-up motion may affect the participants' control of the stylohyoid muscle, which may affect the higher formants. The relationships between the muscles and nerves around the eyes, mouth and throat are extremely complex [12]. The present evidence suggests that the eye motion may affect the F_3 and F_4, and the complexity of the interactions between facial muscles leaves this possibility open for further work.

5. **Conclusions.** This study investigated the effect of gaze angle on the waveform produced in a simple vocalization. The participants vocalized a steady vowel /a/ at the G3 pitch (196 Hz) while gazing at three different angles, namely, $-35°$, $0°$ and $+35°$ in the vertical direction. Recordings of these vocalizations showed that neither F_0 nor SPL were affected by the change in the gazing angle. The results suggest the possibility of the eyelid or eyeball affecting the F_3 and F_4. Further research is needed to test this hypothesis.

REFERENCES

[1] E. Lombard, The sign of the elevation of the voice, *Annuals Maladies Oreille Larynx*, Nez, Pharynx, vol.37, pp.101-109, 1911.
[2] 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.
[3] P. Bottalico, I. I. Passione, S. Graetzer and E. J. Hunter, Evaluation of the starting point of the Lombard effect, *Acta Acustica United With Acustica*, vol.103, no.1, pp.169-172, 2017.
[4] F. X. Brajot, D. Nguyen, J. DiGiovanni and V. L. Gracco, The impact of perilaryngeal vibration on the self-perception of loudness and the Lombard effect, *Experimental Brain Research*, vol.236, no.6, pp.1713-1723, 2018.
[5] 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.
[6] J. A. Jones and K. G. Munhall, Learning to produce speech with an altered vocal tract: The role of auditory feedback, *Journal of the Acoustical Society of America*, vol.113, no.1, pp.532-543, 2003.
[7] H. McGurk and J. MacDonald, Hearing lips and seeing voices, *Nature*, vol.264, no.5588, pp.746-748, 1976.
[8] J. Keil, N. Müller, N. Ihssen and N. Weisz, On the variability of the McGurk effect: Audiovisual integration depends on prestimulus brain states, *Cerebral Cortex*, vol.22, no.1, pp.221-231, 2011.
[9] W. Katz, T. F. Campbell, J. Wang, E. Farrar, J. C. Eubanks, A. Balasubramanian and R. Rennaker, Opti-speech: A real-time, 3D visual feedback system for speech training, *The 15th Annual Conference of the International Speech Communication Association*, 2014.
[10] H. Meng, D. Huang, H. Wang, H. Yang, M. AI-Shuraifi and Y. Wang, Depression recognition based on dynamic facial and vocal expression features using partial least square regression, *Proc. of the 3rd ACM International Workshop on Audio/Visual Emotion Challenge*, pp.21-30, 2013.
[11] J. R. Williamson, E. Godoy, M. Cha, A. Schwarzentruber, P. Khorrami, Y. Gwon and T. F. Quatieri, Detecting depression using vocal, facial and semantic communication cues, *Proc. of the 6th International Workshop on Audio/Visual Emotion Challenge*, pp.11-18, 2016.
[12] A. Lazarov, Z. Ben-Zion, D. Shamai, D. S. Pine and Y. Bar-Haim, Free viewing of sad and happy faces in depression: A potential target for attention bias modification, *Journal of Affective Disorders*, 2018.
[13] G. Song, D. Pellerin and L. Granjon, Different types of sounds influence gaze differently in videos, *Journal of Eye Movement Research*, vol.6, no.4, 2013.

[14] E. Frid, R. Bresin, E. L. S. Pysander and J. Moll, An exploratory study on the effect of auditory feedback on gaze behavior in a virtual throwing task with and without haptic feedback, *Sound and Music Computing (SMC)*, pp.242-249, 2017.
[15] P. Boersma and D. Weenink, *Praat*, http://www.fon.hum.uva.nl/praat/, 2018.
[16] 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.
[17] J. Sundberg, *The Science of the Singing Voice*, Northern Illinois University Press, Dekalb, IL, 1987.
[18] J. M. Palmer and D. A. LaRusso, *Anatomy for Speech and Hearing*, Harper & Row, 1972.
[19] H. Saigusa, Anatomy of the suprahyoid muscles, *Oto-Rhino-Laryngology*, vol.53, pp.246-253, 2010.