ACOUSTIC CHARACTERISTICS OF THE SINGING VOICE AND VOCAL TRACT CHANGES DURING SINGING INSTRUCTION

Jun Takahashi[1], Itsuki Shishime[2], Natsuki Toda[2], Hironori Takemoto[2]

[1] Osaka University of Arts, Osaka, Japan
[2] Chiba Institute of Technology, Chiba, Japan

***Keywords*: Singing Voice; Vocal Tract; Real-Time MRI; Singer's Formant**

Abstract:

Objectives / Introduction: In singing instruction, internal body movements during singing cannot be observed directly. Therefore, teachers judge the state of the vocal organs based on the singing voice and use words to describe internal body movements (such as "open your throat"). Recently, real-time magnetic resonance imaging (rtMRI) has made it possible to capture moving images. This approach can be used to observe the movements of the vocal organs during singing [1]. In this study, we evaluated changes in acoustic characteristics and body movements during singing after one year of singing instruction.

Methods: The vocal participant was a student with no previous singing training. Singing instructions were provided by a professional vocalist who taught at the music academy. A total of 24 lessons were conducted over the course of one year. The subject underwent rtMRI and voice recording before beginning training. Voice recordings were then made monthly. Finally, rtMRI and voice recording were performed at the end of the study period. The singing task involved the vowel /a/ at any pitch. Images of the vocal tract were obtained using an MRI system (Siemens MAGNETOM Prisma Fit 3T). The subject sang in the supine position. Movies were captured for 50 s at a slice thickness of 10 mm and a frame rate of 10 fps. The pixel size was 1 × 1 mm. The contours of the vocal tract were extracted as follows: line 1 was the shortest line between the upper and lower lips, line 2 was the shortest line between the contact point of the soft palate and the posterior wall of the pharynx to the tongue, and line 3 connected top of the arytenoid region to the bottom of the vallecula. The vocal tract bound by line segments 1 and 2 was determined as the width of the oral cavity, and that

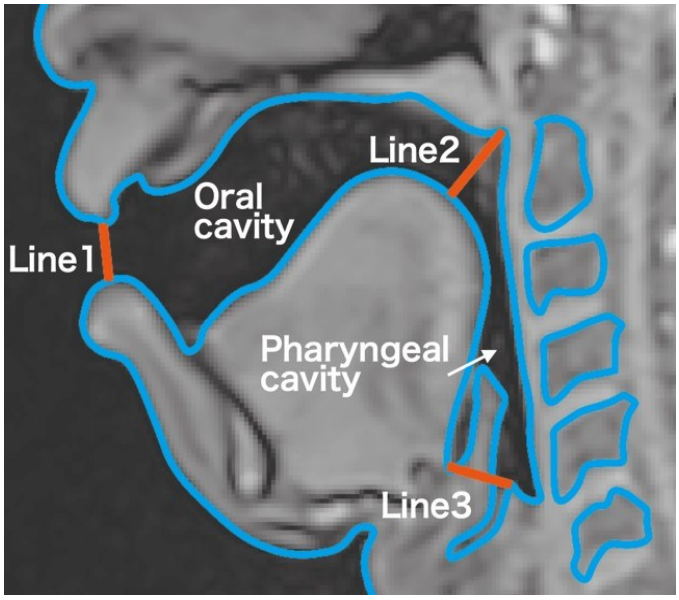

Fig. 1 Methods for extracting vocal tract contours bound by line segments 2 and 3 was determined as the width of the pharyngeal cavity in pixels (Fig. 1). The Singer's formant (SF), a common acoustic characteristic, is observed at approximately 3 kHz in the spectrum of the singing voice; the more pronounced it is, the higher it is evaluated by the listener. In this experiment, we used STRAIGHT [2] to analyze singing voices. In the STRAIGHT analysis, a singing voice is decomposed into a fundamental frequency and a smoothed spectrum that represents the characteristics of the vocal tract. The sum of the power in the 0–12 kHz band and the sum of the power in the SF band (2–4 kHz) were determined. The value of SF power divided by the total power was defined as "SF occupancy." Thus, if the total power was in the 2–4 kHz band, the SF occupancy was 1 (maximum). The vocal participant was evaluated by 18 professional singers and instrumentalists (three sopranos, three tenors, six baritones, five pianists, and one violinist). The singing voice was rated on a 5-point Likert scale (Good, 5; Bad, 1). Mean ratings of the singing voice before and after training were compared using a t-test.

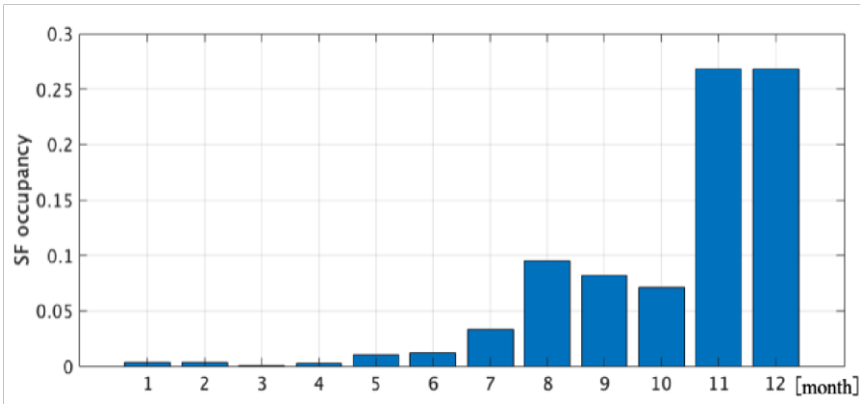

Fig. 2 SF occupancy rate by month

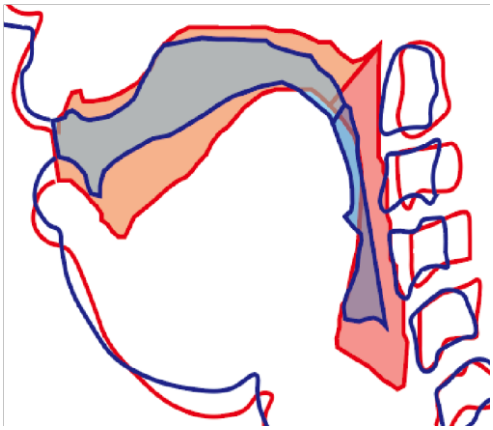

Fig. 3 Vocal tract shape before and after training

Results: The mean ratings of the singing voice were 2.22 before training and 3.44 after training, and their difference was significant ($t(17) = -3.42$, $p = 0.0033$), suggesting that singing voices improved after training. As shown in Fig. 2, SF occupancy rates of the singing voice were 0.0033 before training and 0.2680 after training, indicating an increase after training. These results explain the increase in singing voice ratings after training. It can be observed that the SF occupancy rate increased gradually after the training. Fig. 3 shows the traced and layered vocal tract shapes of the vowel /a/ before and after training. The oral and pharyngeal cavities were larger after training (red) than before training (blue). Before training, the oral cavity was 1298 pixels and the pharyngeal cavity was 383 pixels. After training, the sizes were 2183 and 985 pixels, respectively. These results indicate that one year of singing training expanded the oral and pharyngeal cavities. The expansion of the oral cavity could be achieved mainly by widely opening the mouth, while the expansion of the pharyngeal cavity mainly by lowering the larynx. The shape of the vocal tract, particularly the large cross-sectional area ratio of the larynx to pharynx (which is necessary for SF synthesis [3]), is related to the acoustic characteristics of the singing voice. Therefore, the expansion of the pharyngeal cavity after training increased the SF occupancy of the singing voice.

Conclusions: In this study, we examined how the acoustic characteristics of a student's singing voice and vocal tract changed after one year of singing training. Evaluation by 18 professional singers and instrumentalists rated the singing voice higher after training than before. This suggests that one year of singing training caused a change in the singing voice of the subject, resulting in a higher evaluation. A singing voice analysis also showed that the SF occupancy of the singing voice was higher after the training than before, providing insight into the improvement in singing voice after training. Furthermore, the oral and pharyngeal cavities were larger after training than before training. The shape of the vocal tract is related to the acoustic characteristics of the singing voice, and the expanded pharyngeal cavity after training increases the SF occupancy of the singing voice.

Acknowledgements:
This work was supported by JSPS KAKENHI (Grant Numbers 22K13773 and 23K11172). We also thank all of the participants in the experiments.

References:

[1] Takemoto, H., Goto, T., Hagihara, Y., Hamanaka, S., Kitamura, T., Nota, Y., & Maekawa, K. (2019). *Speech organ contour extraction using real-time MRI and machine learning method*. Proceedings of Interspeech 2019, 904-908. Graz, Austria.
[2] Kawahara, H., Masuda-Katsuse, I., & de Cheveigné, A. (1999). *Restructuring speech representations using a pitchadaptive time–frequency smoothing and an instantaneous-frequency-based F0 extraction: Possible role of a repetitive structure in sounds.* Speech Communication, 27, 187-207.
[3] Sundberg, J. (1987). *The Science of the Singing Voice*. Northern Illinois University Press.