INTERSPEECH 2021
30 August – 3 September, 2021, Brno, Czechia

A Simplified Model for the Vocal Tract of [s] with Inclined Incisors

Tsukasa Yoshinaga[1], *Kohei Tada*[1], *Kazunori Nozaki*[2], *Akiyoshi Iida*[1]

[1]Toyohashi University of Technology, Japan
[2]Osaka University Dental Hospital, Japan

yoshinaga@me.tut.ac.jp, tada@aero.me.tut.ac.jp, knozaki@dent.osaka-u.ac.jp, iida@me.tut.ac.jp

Abstract

To examine the effects of inclined incisors on the phonation of [s], a simplified vocal tract model is proposed, and the acoustic characteristics with different maxillary incisor angles are predicted by the model. As a control model, a realistic vocal tract replica of [s] was constructed from medical images, and the angle of the maxillary incisor was changed from the original position up to 30°. The simplified model was constructed with a rectangular flow channel using the average dimensions of the vocal tracts for five Japanese subjects. Both geometries were set in an anechoic chamber, and sounds generated from the geometries were recorded with a microphone. The results showed that amplitudes of the sound generated by the realistic geometry were decreased by increasing the incisor angle, and this tendency agreed well with the simplified model. Moreover, the slope value of the decrease in overall pressure levels estimated by the model was consistent with that of the realistic geometry, indicating the capability of estimating the effects of inclined incisors with dental prostheses on the production of [s] by using the simplified model.

Index Terms: speech production, vocal tract model, incisor, sibilant fricative

1. Introduction

The sibilant fricative [s] is one of the consonants and is known to be pronounced by a turbulent flow in the front part of the vocal tract. The turbulent flow is generated by a narrowing of the vocal tract passage at the alveolar ridge, and an aeroacoustic sound source is formed by the impingement of the jet flow on the incisors and lips [1-3].

Since the fricative [s] is pronounced by the turbulent flow near the incisor walls, the speech difficulty due to the inclined incisors in orthognathic surgery has been reported in the field of dentistry [4-7]. For example, Runte et al., [4] constructed a maxillary denture and studied the effects of the inclination angle of the central incisors on the production of [s] by varying the angle in the range of −30° to +30°. In addition, the effects of incisor positions were investigated by measuring the jaw movement during the phonation of [s] [8]. The effects of dental prostheses on the vocal tract geometry of [s] and its acoustic properties have also been examined by simultaneous measurements of a microphone and contact sensors in the mouth [9]. However, since the flow configurations in the vocal tract could not be observed during the production of [s], it is still unclear how the deformed geometry affects the generation of the aeroacoustic sound source.

The production mechanisms for vocal tracts without speech difficulty have been investigated by modeling the geometry with simplified flow channels. Shadle [2] proposed a simplified vocal tract model with a constricted flow channel and a teeth-like obstacle in a cylinder, and the model reproduced the acoustic characteristics of English fricatives. In addition, numerical flow simulations were applied to the simplified model and the effects of geometrical differences on the turbulent flow sources were investigated [10-12].

Recently, our group examined the flow configuration in the vocal tract of [s] by constructing a vocal tract replica using computed tomography (CT) images [13] and developed simplified vocal tract models of [s] and [ʃ] based on flow velocities at five cross-sections from throat to lips [3, 14]. By considering the flow configuration at five points in the vocal tract, acoustic characteristics of the subject of medical images were reproduced up to 15 kHz. However, to the author's knowledge, there are very few models to simulate speech difficulty due to the incorrect formation of the vocal tract including effects of orthognathic surgery and dental prostheses. Models estimating the effects of varied vocal tract geometries on [s] would enable improvements in dental treatment operations. In addition, in a process of constructing a model and reproducing the sound, we can understand the aeroacoustic phenomena occurring in the vocal tract with the inclined incisors and generalize the phenomena in different vocal tract geometries of each patient.

Therefore, in this study, we propose a simplified vocal tract model of [s] considering inclined incisors which enables to predict the effects of maxillary incisor angles on the acoustic characteristics of [s]. To parametrically change the incisor angle, a realistic vocal tract replica was first constructed using the CT images, and the angle of incisors was modified. Then, a simplified model was constructed to reproduce the acoustic characteristics of the realistic geometry. The sounds generated by both vocal tracts were experimentally measured and compared.

2. Materials and Methods

2.1. Realistic vocal tract geometry

The vocal tract geometry was reconstructed from CT images of a subject sustained [s] without vowel context [13]. The subject is a 32-year-old Japanese male with a normal dentition of angle class I and has no speech disorders. The resolution of the CT images is $0.1 \times 0.1 \times 0.1$ mm^3 and surfaces of the vocal tract were extracted using the software itk-SNAP [15].

The extracted vocal tract geometry is shown in Fig. 1. The red color in the figure indicates the subject's upper incisor face which was inclined in this study. The axes are defined as

http://dx.doi.org/10.21437/Interspeech.2021-231

follows: anterior-posterior direction is *x*; superior-inferior is *y*; transverse direction is *z*. The throat downstream from the pharynx was extracted, and the upstream part was replaced with an air tube. We confirmed that the exclusion of the geometry upstream from the pharynx is negligible for the flow and acoustic characteristics of [s] [16]. The previous experiments showed that the replica constructed by a 3D-printer with the reconstructed geometry can reproduce the subject's acoustic characteristics up to 16 kHz [13].

2.2. Modification of the incisor angle

For the realistic geometry, the subject's maxillary incisor angle was artificially modified in the three-dimensional polygon data using the software Meshmixer (Autodesk Inc., USA). The incisor part is including four teeth from the left and right maxillary central incisors to the left and right lateral incisors as depicted in red in Fig. 1. The original position of the incisor is defined as angle 0°, and the incisor part was rotated every 10° from 0° to 30° by positioning the rotating axis at the alveolar ridge in *z*-direction. The mid-sagittal plane of the inclined incisors is depicted in Fig. 2.

The horizontal overlap (overjet) between the upper and lower incisors was varied from 2.3 to 2.5 mm when the angle was changed from 0° to 30°, whereas the vertical overlap (overbite) was varied from 0.3 to –1.7 mm. The negative value indicates the vertical gap between the upper and lower incisors. The measured values were within the range of the clinical measurement [8].

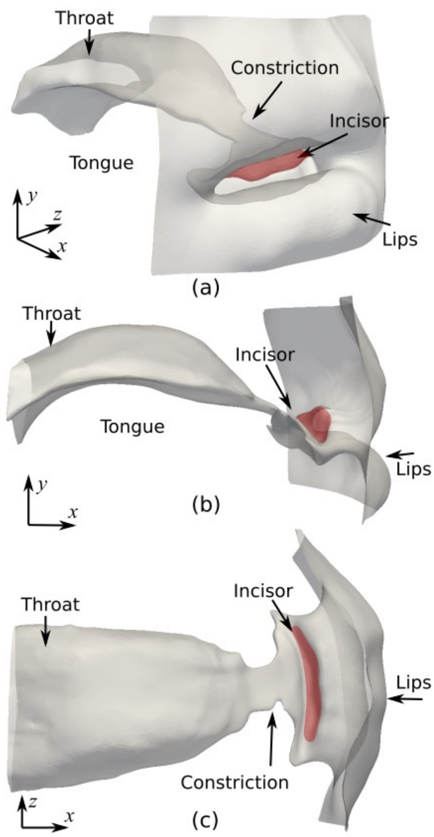

Figure 1: *Realistic vocal tract geometry reconstructed from CT scan: (a) front view; (b) side view; (c) top view. The upper incisor was indicated in red.*

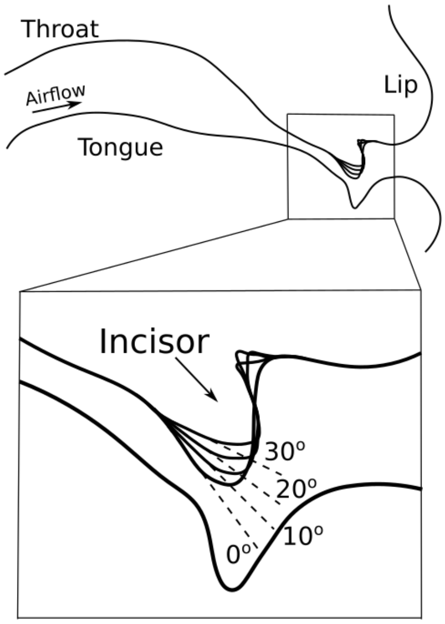

Figure 2: *Mid-sagittal plane of the realistic vocal tract and inclined incisor. The original position is at angle 0° and the inclination angle was varied every 10° from 0° to 30°.*

2.3. Simplified model

A simplified vocal tract model of [s] considering the effects of inclined incisors was constructed based on the previous simplified models designed with the medical images including CT and MRI [3, 14]. The model consists of four parts: throat, tongue constriction, upper and lower incisors, and lip cavity. The geometry and dimensions of the model are shown in Fig. 3. The upstream flow channel of the throat has a length of 150 mm and a cross-section of 6.8 × 23.6 mm. The tongue constriction with a height of 1.3 mm and a width of 6.1 mm is positioned at the front part of the model. At the downstream from the tongue, upper and lower incisors with a thickness of 1 mm are placed, and the lip cavity has a length of 6.4 mm. All these dimensions were determined based on the average values of measurements for five Japanese subjects pronouncing sustained [s] [3].

The model was designed to reproduce the flow velocity measured at five parts: throat, constriction, a space behind the upper incisor, the gap between the incisor, and lip cavity. Therefore, the cross-sectional areas of each part were matched to the measured values of the medical images so that the flow velocity is reproduced at each part. With this model, the individual acoustic characteristics of sustained [s] and [ʃ] were reproduced for five Japanese subjects in the frequency range up to 20 kHz.

To mimic the teeth angle 0° in the realistic geometry, the upper incisor was rotated with an angle of 15°. We define this incisor position as 0° for the simplified model. Then, the upper incisor was rotated every 10° from 0° to 30° (*i.e.* the angle varied from 15° to 45° from the vertical axis *y*). The horizontal overlap between the upper and lower incisors was varied from 1.8 to 3.2 mm whereas the vertical overlap was varied from 0.0 to –0.5 mm when the angle was changed from 0° to 30°.

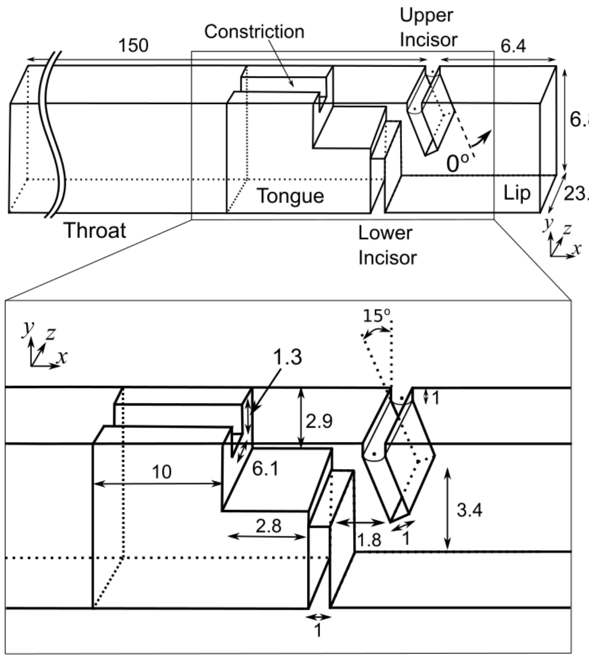

Figure 3: *Simplified vocal tract model of [s] with inclined upper incisor. The unit of dimensions is mm.*

2.4. Experimental setups

The experimental setups are depicted in Fig. 4. The replica of the realistic geometry was built using a 3D printer (Form3, Formlabs, USA), whereas the simplified model was constructed using acrylic boards. The steady airflow was inserted into the replicas by an air compressor (SOL-2039, Misumi, Japan) through a flow valve (IR2000-02, SMC, Japan) and a mass-flow meter (PFM750-01, SMC, Japan) to reproduce a physiological flow rate of 400 cm^3/s [3].

Far-field sounds were measured by a 1/4 inch microphone (Type 4939, Bruel & Kjaer, Denmark) at 300 mm along axis *x* from the outlet of the lips in an anechoic chamber (V = 8.1 m^3). The sound pressure signals were recorded for 10 s with a sampling frequency of 40 kHz by a data acquisition system (PXIe-6351, National Instruments, USA). The spectrum of the sound was calculated using discretized Fourier transform (DFT) with 2048 signal points and Hann window. The time average of DFT was calculated with 389 sets of 50% overlapped windows. The sound pressure level (SPL) was calculated based on the reference level 20×10^{-6} Pa.

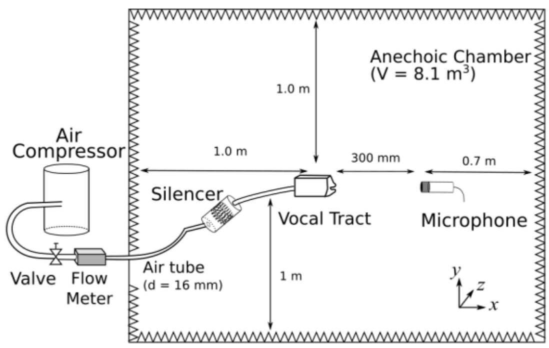

Figure 4: *Experimental setups of the sound recording.*

3. Results and Discussion

The spectra of measured sounds for the realistic replica and the simplified model are plotted in Fig. 5 (a) and (b), respectively. For both geometries, the maximum peak of spectra appeared at approximately 5.5 kHz, and the overall spectral shape of the simplified model agreed well with that of realistic geometry. The peak amplitude at 5.5 kHz with the teeth angle of 0° was approximately 50 dB for both geometries. Although the simplified model was designed with the average dimensions of five Japanese subjects, the spectrum shape and peak frequency of the model matched with those of the realistic geometry. This is probably because the averaged dimensions were accidentally similar to those of the subjects, and the dimension of the model needs to be adjusted when the subject's acoustic characteristics were significantly different.

With the increase of the teeth angle, amplitudes in the frequency range above 5 kHz were decreased for both geometries. The amplitude at 10 kHz of the realistic replica decreased by 5 dB with the increase of every 10°, whereas the decrease of the amplitude at 10 kHz was varied from 2 to 5 dB in the range of 0° to 30° for the simplified model.

The realistic geometry with the angle of 0° produced the general acoustic characteristics of [s] which are broadband noise above the characteristic peak frequency of 4 to 6 kHz [1-3]. In contrast, by increasing the angle up to 30°, the amplitudes above 5 kHz were decreased and the spectral shape became closer to that of [ʃ] in Japanese [3]. However, the characteristic peak frequency was not changed from the range of [s] and was different from that of [ʃ] which is around 2 to 3 kHz. This result is consistent with the previous study of the

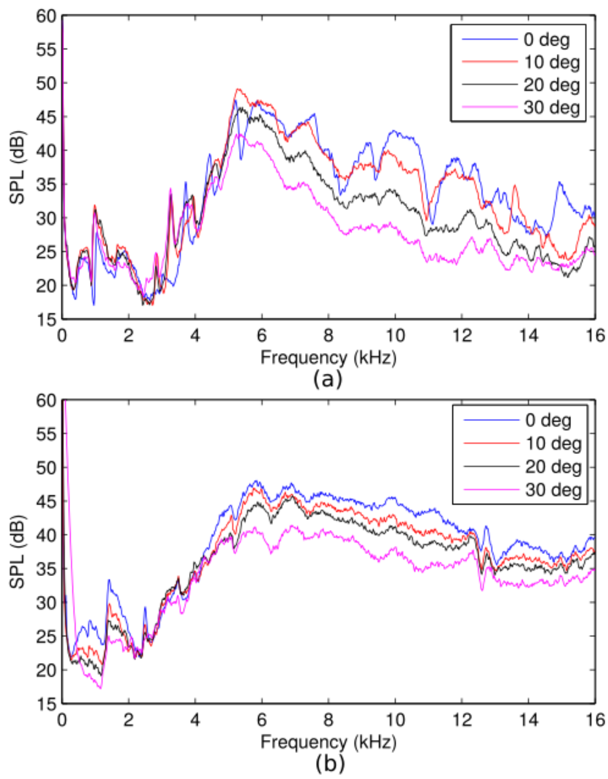

Figure 5: *Sound spectra measured at 300 mm from lips for (a) realistic replica and (b) simplified model.*

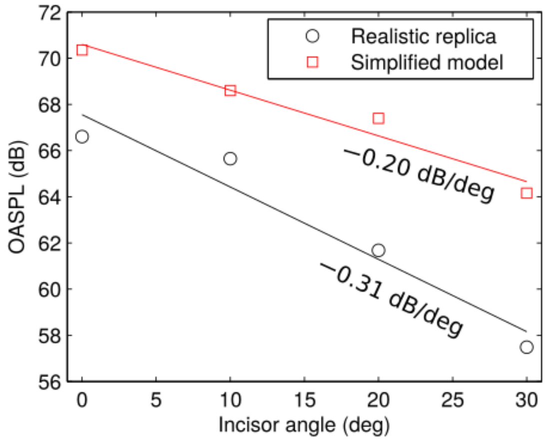

Figure 6: *Overall sound pressure level (OASPL) with each inclined incisor angle. Regression lines were calculated for the realistic replica and the simplified model.*

maxillary denture [4] and indicates that the peak frequency is mainly determined by the length between the constriction and the lip outlet as reported in [2, 3]. The sound with inclined incisors is probably recognized as [s] with smaller amplitudes compared to the original geometry. However, it needs to be further evaluated by a perception test with Japanese listeners, since the sibilant [s] has language-specific acoustic characteristics, like a first spectral moment in Bark scale [17-18].

To compare the amplitudes above the characteristic peaks, the overall sound pressure level (OASPL) in the frequency range 6 to 16 kHz was calculated for each spectrum and plotted in Fig. 6. In addition, a regression line was calculated and the slope for each geometry was obtained. The OASPL of the realistic geometry decreased from 67.5 to 57.5 dB, whereas the OASPL of the simplified model decreased from 70.4 to 64.2 dB by changing the incisor angle from 0° to 30°. The slope values were −0.31 and −0.20 dB/deg for the realistic and simplified geometries, respectively.

While the OASPLs of the simplified vocal tract model overestimated those of the realistic vocal tract, the slope value of the simplified model agreed with that of the realistic one. This result suggests that the proposed model enables to predict the effects of inclined incisors on the acoustic characteristics of [s]. According to the parametric study of the simplified model of [s], the overall amplitudes are mainly depending on the size of the constriction and flow rate which determines the jet velocity and aeroacoustic source magnitudes [2-3, 14]. Therefore, by further adjusting the parameters around the constriction, the overestimation of OASPLs in the simplified model might be improved. Moreover, the slight difference of the slope values indicates that the tendency in changes of the source with inclined incisors of the simplified model was different from that of the realistic geometry. This needs to be improved by applying the numerical flow simulation [16] on the vocal tract geometry and examining the aeroacoustic source configuration with the inclined incisors.

For future research, it is expected to apply the model to the actual patient with the maxillary denture in the dental hospital and predicting the acoustic characteristics of [s] with the dental prostheses. However, since the proposed model was applied only for one subject, the model needs to be further validated by increasing the number of subjects for the realistic geometry. Then, we can propose a criterion for the incisor angles which won't affect the pronunciation of [s].

4. Conclusions

In this study, the experimental measurements with the realistic vocal tract replica of [s] were conducted to clarify the effects of maxillary incisor angle on the acoustic characteristics. Furthermore, the simplified vocal tract model was proposed to predict the acoustic characteristics of [s] with inclined incisors. The results showed that the simplified model reproduced the overall spectral shape of the sound generated by the realistic geometry with varied incisor angles from 0° of the original geometry to 30°. Although the OASPL of the realistic geometry was overestimated by the simplified model, the slope value of the decrease in the OASPLs of the model was consistent with that of the realistic geometry. These results indicate that the tendency of the effects of inclined incisors of the vocal tract [s] can be predicted by the proposed model. Further validation of the model with different subjects and development for the clinical application for the dentists are anticipated.

5. Acknowledgements

This work was supported by MEXT as "Priority Issue on Fugaku supercomputer" (hp200123, hp200134), JSPS KAKENHI (JP19H03976, JP20K14648), and JSPS Grant-in-Aid for Scientific Research on Innovative Areas (JP20H04999).

6. References

[1] K. N. Stevens, "Airflow and turbulence noise for fricative and stop consonants: Static considerations," *J. Acoust. Soc. Am.*, 50(4), 1180-1192 (1971).

[2] C. H. Shadle, "The acoustics of fricative consonants," *Ph.D. Thesis*, Massachusetts Institute of Technology (1985).

[3] T. Yoshinaga, K. Nozaki, and S. Wada, "Aeroacoustic analysis on individual characteristics in sibilant fricative production," *J. Acoust. Soc. Am.*, 146(2), 1239-1251 (2019).

[4] C. Runte, M. Lawerino, D. Dirksen, F. Bollmann, A. Lamprecht-Dinnesen, and E. Seifert, "The influence of maxillary central incisor position in complete dentures on /s/ sound production," *J. Prosthet. Dent.*, 85(5), 485-495 (2001).

[5] A. S. Lee, T. L. Whitehill, V. Ciocca, and N. Samman, "Acoustic and perceptual analysis of the sibilant sound /s/ before and after orthognathic surgery," *J. Oral Maxillofac. Surg.*, 60(4), 364-372 (2002).

[6] R. Liu, A. Hu, L. Niu, F. Wang, Y. Lu, and Q. Zhou, "Association of incisal overlaps with /s/ sound and mandibular speech movement characteristics," *Am. J. Orthod. Dentofacial Orthop.*, 155(6), 851-859 (2019).

[7] E. Fonteyne, L. Van Doorne, L. Becue, C. Matthys, E. Bronckhorst, and H. De Bruyn, "Speech evaluation during maxillary mini-dental implant overdenture treatment: A prospective study," *J. Oral Rehabil.*, 46(12), 1151-1160 (2019).

[8] C. A.Burnett and T. J. Clifford, "Closest speaking space during the production of sibilant sounds and its value in establishing the vertical dimension of occlusion," *J. Dent. Res.*, 72(6), 964-967 (1993).

[9] S. L. Hamlet, B. L. Cullison, and M. L. Stone, "Physiological control of sibilant duration: Insights afforded by speech compensation to dental prostheses," *J. Acoust. Soc. Am.*, 65(5), 1276-1285 (1979).

[10] A. Van Hirtum, X. Grandchamp, X. Pelorson, K. Nozaki, and S. Shimojo, "Les and "in vitro" experimental validation of flow around a teeth-shaped obstacle," *Int. J. Appl. Mech.,* 2(02), 265-279 (2010).

[11] J. Cisonni, K. Nozaki, A. Van Hirtum, X. Grandchamp, and S. Wada, "Numerical simulation of the influence of the orifice aperture on the flow around a teeth-shaped obstacle," *Fluid Dyn. Res.*, 45(2), 025505 (2013).

[12] M. Arnela and O. Guasch, "Finite element simulation of /asa/ in a three-dimensional vocal tract using a simplified aeroacoustic source model," in *Proceedings of the 23rd International Congress on Acoustics*, pp. 1802–1809 (2019).

[13] K. Nozaki, T. Yoshinaga, and S. Wada, "Sibilant /s/ simulator based on computed tomography images and dental casts," *J. Dent. Res*., 93(2), 207-211 (2014).

[14] T. Yoshinaga, K. Nozaki, and S. Wada, "Effects of tongue position in the simplified vocal tract model of Japanese sibilant fricatives /s/ and /S/," *J. Acoust. Soc. Am.* 141, EL314–EL318 (2017).

[15] P. A. Yushkevich, J. Piven, H. C. Hazlett, R. G. Smith, S. Ho, J. C. Gee, and G. Gerig. "User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability." *Neuroimage* 31(3), 1116-1128 (2006).

[16] T. Yoshinaga, K. Nozaki, and S. Wada, "Experimental and numerical investigation of the sound generation mechanisms of sibilant fricatives using a simplified vocal tract model," *Phys. Fluids*, 30(3), 035104 (2018).

[17] F. Li, J. Edwards, and M. Beckman. "Spectral measures for sibilant fricatives of English, Japanese, and Mandarin Chinese," In *Proceedings of the XVIth international congress of phonetic sciences,* Vol. 4, pp. 917-920 (2007).

[18] F. Li, B. Munson, J. Edwards, K. Yoneyama, and K. Hall. "Language specificity in the perception of voiceless sibilant fricatives in Japanese and English: Implications for cross-language differences in speech-sound development," *J. Acoust. Soc. Am.*, *129*(2), 999-1011 (2011).