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

Vocal-tract models to visualize the airstream of human breath and droplets while producing speech

Takayuki Arai
Sophia University, Tokyo, Japan
arai@sophia.ac.jp

Abstract

Due to the COVID-19 pandemic, visualizing the airstream of human breath during speech production has become extremely important from the viewpoint of preventing infection. In addition, visualizing droplets and the larger drops expelled when we speak consonantal sounds may help for the same reason. One visualization technique is to pass a laser sheet through the droplet cloud produced by a human speaker. However, the laser poses certain health risks for human beings. Therefore, we developed an alternative method to passing a laser against a human body in which we utilize physical models of the human vocal tract. First, we tested a head-shaped model with a lung model from our previous study to visualize the exhaled breath during vowel production (with and without a mask). Then, we implemented an extended version of the anatomical-type vocal-tract model introduced in our previous study. With this newly developed model, lips are made of the same flexible material that was used to form the tongue part in the previous model. We also attached these lips to another previous model for producing sounds including /b/. Finally, the lip models were tested to visualize the droplet cloud including expelled drops present while producing a bilabial plosive sound.

Index Terms: COVID-19, visualizing droplet cloud, vocal-tract models, head-shaped model, anatomical-type model, lip model

1. Introduction

In 2020, the World Health Organization (WHO) declared the worldwide COVID-19 pandemic [1], and we now face the problem of the spread of infection. One of the crucial transmissions of the virus is through droplets. Such droplets are expelled by humans not only when we cough or sneeze but also during singing and speaking [2]. Another way the infection spreads is through aerosol transmission, and the aerosol produced while speaking can contain viruses and thereby become the source of infection [3].

During speech production, the droplets and the aerosol are not able to been seen, so a visualization technique is needed. One such technique is passing a laser against the droplet cloud produced by a speaker during speech production [2]. However, the laser poses certain health risks for human beings. In particular, a laser beam can contain a high energy that damages the human body, including the eyes, which sometimes results in blindness. In this study, as an alternative to passing a laser against a human body, we utilized physical models of the human vocal tract.

We first applied a head-shaped (HS) model with a lung model from our previous studies [4–6]. The HS + lung models were originally developed for students and the general public to explain human speech production from breathing, phonation, and articulation. When the air in the lungs goes out through an artificial larynx, a glottal sound is produced that results in resonances inside the vocal tract of the HS model. Finally, a spectral envelope is characterized by the resonances for each vowel depending on the configuration of the vocal tract. These HS and lung models were used in the current study to visualize the aerosol from the mouth (and nostril) with a special air with and without a facial mask.

Second, we implemented an extended version of our anatomical-type model of the human vocal tract. This model is useful not only for pedagogical purposes but also for clinical applications and pronunciation training. The anatomical-type models we developed in 2017, 2018, and 2019 all look like anatomical models for medical students. In addition to the appearance, they also produce speech-like sounds when feeding a glottal sound. The 2017 model produces the vowel /a/, as the vocal-tract configuration is fixed for that vowel [7]. The 2018 model has a tongue part fabricated with a flexible material so that it can produce different types of vowels by changing the tongue position [8]. The 2019 model has a flexible tongue and also a movable mandible, so the jaw opens and closes depending on the target sounds [9]. In 2020, we developed yet another anatomical-type model that is based on the 2019 model with the movable mandible. The newly developed 2020 model has lips constructed with a flexible material so that the shape of the lips can be changed for certain sounds. Due to the flexibility of the lip model, we are now able to test a bilabial plosive sound to visualize a droplet cloud including expelled drops present during the production of a sound. In addition, the same lip model can be attached to the BMW-RL model for producing consonants /b/, /m/, /w/, /r/, and /l/ [10]. Therefore, we also tested the lip model with the BMW-RL model for the same visualization purpose.

2. Visualizing human breath during vowel production

2.1. Head-shaped and lung models

Figure 1 shows the HS model and the lung model [4–6]. The lung model consists of an acrylic body, a rubber membrane on the bottom, and two balloons. The acrylic body imitates a sealed thoracic cavity and its volume is controlled by the bottom membrane mimicking a diaphragm. A whistle-type artificial larynx (the red arrow in Fig. 1) is attached to the single end of a Y-shaped tube of the lung model. A glottal sound is produced during the exhalation. When the HS model is placed on top of the artificial larynx, vowel /a/ is produced because, in the case of Fig. 1, the configuration of the vocal tract simulates the one during vowel /a/. This HS model also has a nasal cavity and a movable velum so that different degrees of nasalization can be simulated.

http://dx.doi.org/10.21437/Interspeech.2021-449

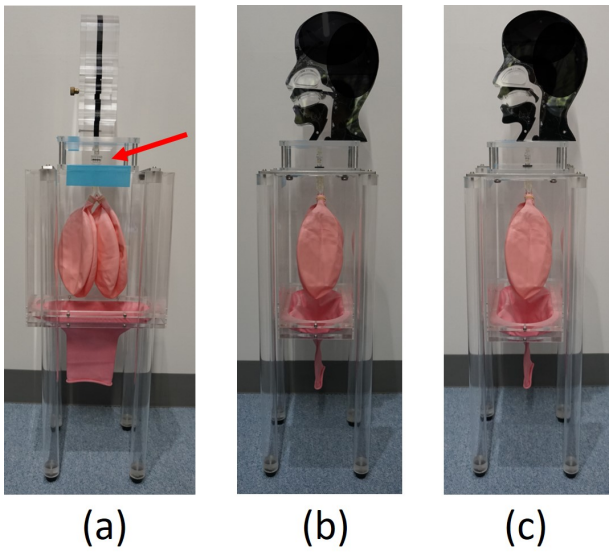

Figure 1: *HS and lung models [4]. (a) Front view (red arrow indicates location of the whistle-type artificial larynx). (b) Side view with the velopharyngeal port closed. (c) Side view with the velopharyngeal port open.*

2.2. Method

The aerosol is visualized by using air with oil mist. Dioctyl sebacate (DOS) was used for the oil [11]. The oil mist is generated by two generators: the PIV Part14 (PIVTech), where the mean diameter of the particles is 1–2 μm, and the CTS-1000 (Seika Digital Image), where the mean diameter of the particles is 3–4 μm. The brightness due to Mie scattering is proportional to the square of the particle size.

During video and audio recording, a laser sheet is passed over the midsagittal plane of the models. When the diaphragm is pulled down, the air moves into the lungs and the two balloons are inflated. During an inhalation period, the air with the oil mist is sucked from the mouth into the vocal tract, trachea, and lungs. During an exhalation period, the air with the oil mist goes out through the mouth (and nostril). The video recordings were taken by a high-speed camera (Vision Research, Phantom T1340 72GB) at 1000 frames/s. The audio recordings were done using a sound level meter (Rion, NA-28) and an audio interface (RME, Fireface UC) with a sampling frequency of 48 kHz. The microphone was placed 1 m away from the mouth of the model. A trigger signal was generated to start video recording, and the signal was also recorded on the sub channel of audio recording, whereas the audio signal from the microphone was recorded on the main channel.

Table 1 lists the six measurements we conducted. For each, the maximum A-weighted sound pressure level was measured ("SPL" column in the table) in decibels. The "Nasal" column indicates with ("1") or without ("0") the nasalization (when nasalized, the velopharyngeal port was open; when not nasalized, the port was closed). The "LP" column indicates the lung pressure during exhalation: "H" for high pressure (approximately 10 cm H_2O) and "L" for low pressure (approximately 5 cm H_2O). The "Mask" column indicates whether a mask was placed on the top of the mouth of the HS model ("1") or not ("0"). The mask was made of nonwoven fabric (Sharp, MA-1050). The "Particle" column indicates the size of the particles, where "S" means 1–2 μm and "L" means 3–4 μm.

2.3. Results

Figure 2 shows snapshots of the measurement results (the numbers correspond to each row in Table 1). From the measurements, we can observe several points.

1) By comparing Nos. 1 and 2, it becomes clear that how far the aerosol reaches does not depend on the lung pressure. When the lung pressure was low, the sound pressure level decreased more than 5 dB. However, even with low lung pressure, the aerosol reached as far as 400 mm, which is the size of the pictures in Fig. 2, or even more.

Table 2: *List of measurements for HS and lung models.*

No.	SPL (dB)	Nasal (0/1)	LP (H/L)	Mask (0/1)	Particle (S/L)
1	81.2	0	H	0	S
2	75.9	0	L	0	S
3	78.8	1	H	0	S
4	78.5	1	H	1	S
5	78.9	0	H	1	S
6	81.1	0	H	0	L

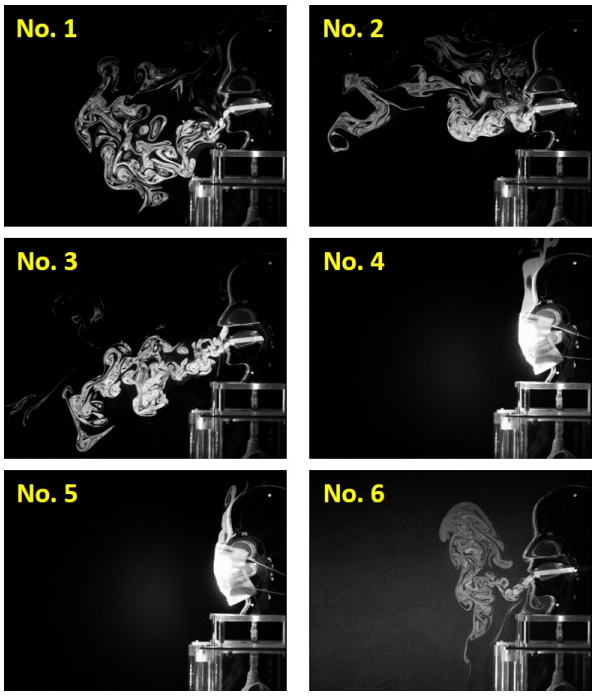

Figure 2: *Snapshots for each measurement listed in Table 1 with the HS and lung models.*

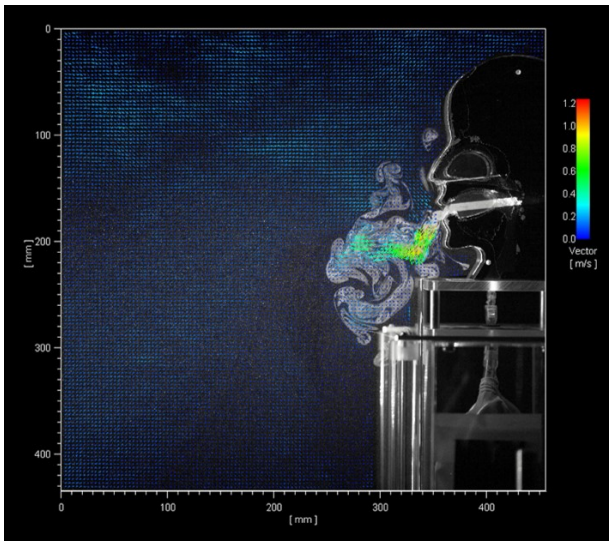

Figure 3: *Snapshot of PIV results based on measurement No. 6.*

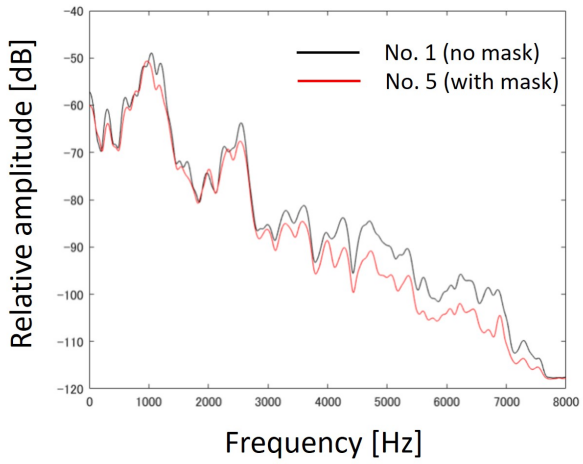

Figure 4: *Spectra of measurement Nos. 1 (with no mask) and 5 (with mask).*

2) By comparing Nos. 1 and 3, we can see that the two air streams from the mouth and the nostril merged and formed a more directional pattern.
3) The snapshots of Nos. 4 and 5 demonstrate the effectiveness of wearing the nonwoven fabric mask. The only aerosol observed in these pictures related to upward movements, as the mask did not fit a little on the top of the nose.

2.4. PIV

Measurement No. 6 was used to calculate particle velocities by applying a particle image velocimetry (PIV) technique [12]. We used the program for PIV computation provided by Seika Digital Image. Figure 3 shows a snapshot of the computational results based on measurement No. 6. The color scale indicates the velocity in meters per second. As we can see, the particles were sometimes more than 1 m/s.

2.5. Acoustic measurements

Figure 4 shows two spectra calculated from the measurement results of No. 1 with no mask and No. 5 with mask. The difference was small in the low frequency range. However, they were slightly different in the frequency range above 4 kHz, where the difference was approximately 8 dB maximum. This result is consistent with the previous reports (e.g., [13]).

3. Visualizing droplets during consonant production

3.1. Vocal-tract models with the lip model

Figure 5 shows the newly developed anatomical-type vocal-tract model, namely, the 2020 model. This model is based on the 2019 model [9] featuring the mandible that can be opened and closed and the tongue part made of a flexible material (a polyethylene-styrene copolymer) with two degrees of hardness (2 and 4 in ASKER-C hardness). The same material was used to make the lips of the 2020 model, as shown in the figure. Due to the flexibility of the lip model, we are now able to test a bilabial plosive sound to visualize the droplet cloud including expelled drops present during the production of a sound.

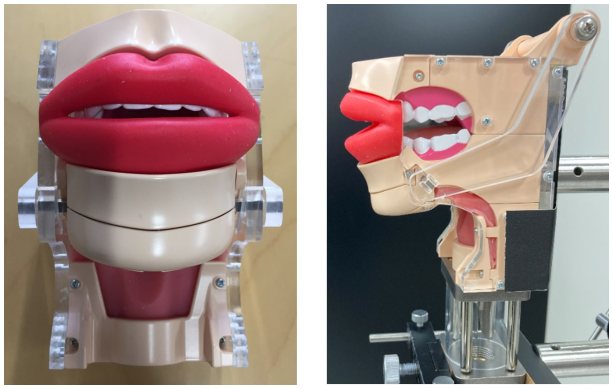

Figure 5: *2020 model with flexible tongue and lips. Left: front view. Right: side view (mounted on the reed-type sound source).*

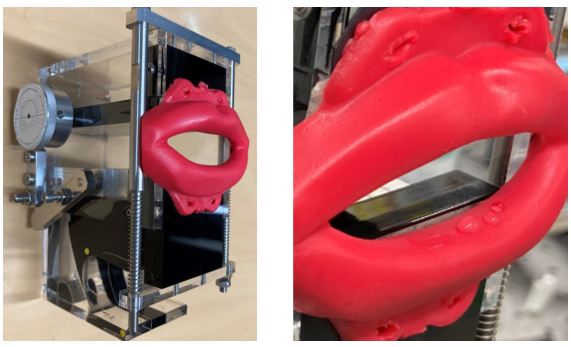

Figure 6: *BMW-RL model with the lip model from the 2020 model. The simulated saliva solution was placed on the top surface of the lower lip (right).*

The same lip model can also be attached to the BMW-RL model for producing consonants /b/, /m/, /w/, /r/, and /l/ [10]. Figure 6 shows the BMW-RL model with the lip model used in the 2020 model. It was also tested for the same visualization purpose.

3.2. Method

The droplets are visualized by using a "simulated saliva" solution [14]. This solution is artificially simulated human saliva made from sodium chloride (12 g), pure glycerin (76 g), and distilled water (1L). Approximately 30 µL of the simulated saliva solution is placed on the top surface of the lower lip. After placing the saliva solution, the mandible is raised to form closed lips. During video and audio recording, a laser sheet is passed over the midsagittal plane of the models. The vocal-tract models are attached to the house of the reed-type sound source, which is connected to an air pump. When the air pump is pressed, the air stream moves through the reed-type sound source and produces a glottal sound. At the same time, the lips are closed and the pressure builds up inside the oral cavity. When the lip closure is released with the mandible open, a bilabial plosive sound followed by a vowel is produced. The video recordings were made by a high-speed camera (Vision Research, Phantom T1340 72GB) at 1000 frames/s. The audio recordings were done using a sound level meter (Rion, NA-28) and an audio interface (RME, Fireface UC) with a sampling frequency of 48 kHz. The microphone was placed 1 m away from the mouth of the model.

3.3. Results

Figures 7 and 8 show a time sequence of four snapshots of the droplet cloud measured with the 2020 model and the BMW-RL model + lip model, respectively. In both cases, the maximum A-weighted sound pressure level was approximately 92 dB. As we can see in Fig. 7, the droplets were clearly expelled when a bilabial plosive sound, such as /b/, was produced with the 2020 model. The BMW-RL + lip models also showed the droplets at the production of a bilabial plosive sound, although it was not that clear in Fig. 8. Unlike the diffusion of the aerosol shown in Fig. 2, the droplets fell down after being expelled from the lips. This is due to the size of the droplets, which were much heavier than the particles in the aerosol. We also analyzed a sound produced by the 2020 model. Figure 9 shows (a) a waveform and (b) its sound spectrogram up to 8 kHz.

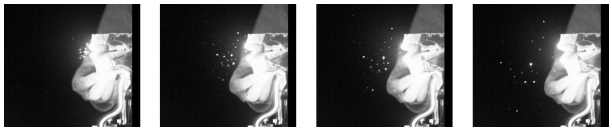

Figure 7: *Time sequence of snapshots (from left to right) with the 2020 model.*

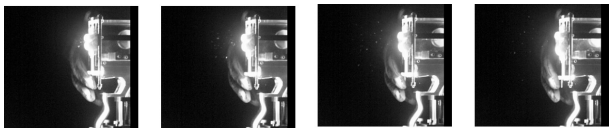

Figure 8: *Time sequence of snapshots (from left to right) with the BMW-RL model + lip model.*

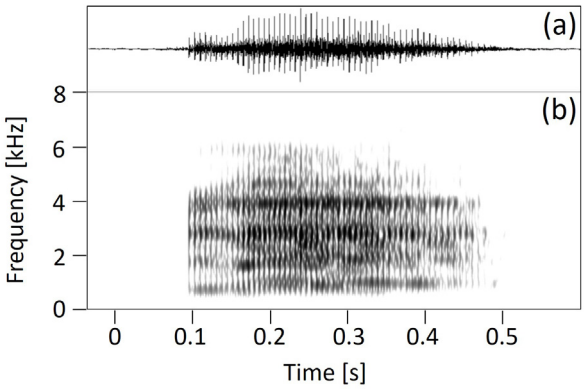

Figure 9: *(a) Waveform and (b) spectrogram when the 2020 model produced a plosive sound.*

4. Conclusion

In this study, we applied physical models of the human vocal tract for laser-based visualization of the airstream of human breath and droplet clouds. A head-shaped model with a lung model from our previous study was first applied to visualize the exhaled breath during vowel production with and without a mask. As a result, we were able to repeatedly observe the airstream of the human breath. In addition, we observed the aerosol spread even with low lung pressure. The PIV technique showed the qualitative speeds of particles during vowel production. The effectiveness of wearing a mask was also confirmed with a slight reduction of power in high frequencies above 4 kHz.

We then implemented an extended version of our previous anatomical-type vocal-tract model (2020 model). With this model, both the lips and the tongue part were made of a flexible material. The same lip model was also attached to the BMW-RL model. With both models, we were able to repeatedly visualize the droplets during production of a bilabial plosive sound.

Due to the risks to the human body, the implementation of measurements like this with human subjects is limited, especially in terms of repeating the same procedure many times. As an alternative, vocal-tract models are helpful for avoiding the risks and simulating human speech production at the same time. There are various improvements to be done in the future. For example, the 2020 model with the flexible lips has a slight problem with air leakage. In addition, while the tongue of this model is flexible, it is not able to make a complete closure at either the alveolar or velar position. If we can enable oral closure, we will be able to perform additional measurements besides bilabial plosive sounds. Besides, we would also like to actually measure droplets in terms of the size (health people produce droplets between the size of 0.1 and 10 µm [15]) and the count.

5. Acknowledgements

This work was partially supported by JSPS KAKENHI Grant Numbers 18K02988/21K02889 and Sophia University Special Grant for Academic Research (Research in Priority Areas).

6. References

[1] D. Cucinotta and M. Vanelli, "WHO declares COVID-19 a pandemic," *Acta Biomed.*, 91(1), pp. 157–160, 2020.

[2] P. Anfinrud, V. Stadnytskyi, C. E. Bax, and A. Bax, "Visualizing speech-generated oral fluid droplets with laser light scattering," *N. Engl. J. Med.*, 382, pp. 2061–2063, 2020.

[3] E. L. Anderson, P. Turnham, J. R. Griffin, and C. C. Clarke, "Consideration of the aerosol transmission for COVID-19 and public health," *Risk Analysis*, 40(5), pp. 902–907, 2020.

[4] T. Arai, "Education system in acoustics of speech production using physical models of the human vocal tract," *Acoust. Sci. Tech*., vol. 28, no. 3, pp. 190–201, 2007.

[5] T. Arai, "Education in acoustics and speech science using vocal-tract models," *J. Acoust. Soc. Am.*, vol. 131, no. 3, Pt. 2, pp. 2444–2454, 2012.

[6] T. Arai, "Vocal-tract models and their applications in education for intuitive understanding of speech production," *Acoust. Sci. Tech*., vol. 37, no. 4, pp. 148–156, 2016.

[7] T. Arai, "Vocal-tract model with static articulators: Lips, teeth, tongue, and more," *Proc. of INTERSPEECH*, pp. 4028–4029, 2017.

[8] T. Arai, "Flexible tongue housed in a static model of the vocal tract with jaws, lips and teeth," *Proc. of INTERSPEECH*, pp. 171–172, 2018.

[9] T. Arai, "Two different mechanisms of movable mandible for vocal-tract model with flexible tongue," *Proc. of INTERSPEECH*, pp. 1366–1370, 2020.

[10] T. Arai, "Integrated mechanical model for [r]-[l] and [b]-[m]-[w] producing consonant cluster [br]," *Proc. of INTERSPEECH*, pp. 979–983, 2017.

[11] T. Inamura, H. Yanaoka, and T. Kawada, "Visualization of airflow around a single droplet deformed in an airstream," *Atomization and Sprays*, 19(7), pp. 667–677, 2009.

[12] G. N. Sze To, M. P. Wan, C. Y. H. Chao, L. Fang, and A. Melikov, "Experimental study of dispersion and deposition of expiratory aerosols in aircraft cabins and impact on infectious disease transmission," *Aerosol Sci. Technol.*, 43(5), pp. 466–485, 2009.

[13] M. Magee, C. Lewis, G. Noffs, H. Reece, J. C. S. Chan, C. J. Zaga, C. Paynter, O. Birchall, S. R. Azocar, A. Ediriweera, K. Kenyon, M. W. Caverlé, B. G. Schultz, and A. P. Vogel, "Effect of face masks on acoustic analysis and speech perception: Implications for peri-pandemic protocols," *J. Acoust. Soc. Am.*, 148, pp. 3562–3568, 2020.

[14] M. P. Wan, C. Y. H. Chao, Y. D. Ng, G. N. Sze To, and W. C. Yu, "Dispersion of expiratory droplets in a general hospital ward with ceiling mixing type mechanical ventilation system," *Aerosol Sci. Technol.*, 41(3), pp. 244–258, 2007.

[15] H. Zhang, D. Li, L. Xie, and Y. Xiao, "Documentary research of human respiratory droplet characteristics," *Procedia Engineering*, 121, pp. 1365-1374, 2015.