Hands-on Speech Science Exhibition for Children at a Science Museum

Takayuki Arai[1], *Kanae Amino*[2], *Mee Sonu*[1,3], *Keiichi Yasu*[1], *Takako Igeta*[1], *Kanako Tomaru*[1], *and Marino Kasuya*[1]

[1]Faculty of Science and Technology, Sophia University, Tokyo, Japan
arai@sophia.ac.jp
[2]National Research Institute of Police Science, Chiba, Japan
[3]JSPS Postdoctoral Fellowships for Foreign Researchers, Japan

Abstract

In previous studies, we developed several physical models of the human vocal tract, reporting that they are intuitive and helpful for students studying acoustics and speech science. Furthermore, we designed a sliding vocal-tract handicraft model at a science workshop, enabling children to make their own vocal-tract model with a sound source. Additionally, at various science museums, we supervised several exhibitions where children were presented with simple speech production demonstrations using physical models of the human vocal tract. In addition to these hands-on activities, we arranged an exhibition at another science museum where children could learn more about speech by analyzing their own voices, observing sound spectrograms, and synthesizing a speech sound by concatenating pre-printed, short duration spectrograms using Digital Pattern Playback (DPP). In this paper, we reported and discussed another hands-on speech science exhibition for children. In this exhibition, children 1) produced vowels using vocal-tract models, 2) observed a waveform and its spectrogram, and 3) used their own voices with DPP. We confirmed that this combination has a synergistic effect on education in acoustics and speech science.

Index Terms: science museum, hands-on exhibition, speech science, physical models of the human vocal tract, sound spectrogram, digital pattern playback

1. Introduction

Several physical models of the human vocal tract have been developed for educational purposes in our previous studies, and they have proven to be intuitive and helpful for students in acoustics and speech science [1-4]. All of our models demonstrate 1) the relationship between vocal-tract configuration and vowel quality, and 2) the source-filter theory of speech production [5]. Even with the sliding-three-tube (S3T) model [3,6], one of the simplest types, we can demonstrate these two concepts.

These physical models are not only useful for college and graduate students, but also for small children. Furthermore, the S3T model is used for a handicraft project at a science workshop, where children can make their own vocal-tract model for use with a sound source [3]. Producing vowels with physical models of the human vocal tract is so intuitive, children can understand the basic phenomena without understanding the more complicated underlying speech science theories.

At the Shizuoka Science Museum in Shizuoka-City, Japan, the first author supervised an exhibition of vowel production using the cylinder and plate-type models [1,2]. The science museum "Exploratorium" in San Francisco also has a similar exhibition [7]. At another science museum, "Sony ExploraScience" [8], we supervised an exhibition where children experienced a simple demonstration using the physical models of the human vocal tract, learned more about speech by analyzing their own voices, observed sound spectrograms, and synthesized a speech sound by concatenating pre-printed, short duration spectrograms with Digital Pattern Playback (DPP) [9]. This exhibition was very successful and many visitors of all ages visited the exhibition.

Every summer, an event called "Science Square" is held at the National Museum of Nature and Science, Tokyo [10]. During this month-long event, several booths are open every day, each of which is organized by a different committee or organization for a different length of time. The Committee of Education in Acoustics of the Acoustical Society of Japan has participated in this event since 2007. Every summer, we set up our booth for two days. One of our demonstrations is entitled "Let's get familiar with sound and vibration." In this booth, we have several activities covering such topics as the visualization of sounds, resonance, and a handicraft of a toy with a plate that rotates by means of vibration [10]. Since 2011, we rearranged the exhibition, adding speech "corners" staffed by members of Arai Lab. at Sophia University. In this paper, we discuss our Science Square speech science exhibitions, giving particular attention to the one we organized in 2011 (Fig. 1).

Figure 1: *The exhibition booths at the Science Square at the National Museum of Nature and Science, Tokyo (Japan).*

Figure 2: *Cylinder-type models of the human vocal tract for producing vowels. The vowels are Japanese /i/, /e/, /a/, /o/, and /u/ from left to right (the glottis end is the bottom, the lip end is the top). The material is acrylic resin. (See more details in [1,2].)*

2. Hands-on speech science exhibition

The exhibition we organized at the 2011 Science Square consisted of these three speech corners: 1) producing vowels with physical models of the human vocal tract, 2) analyzing one's own speech with two types of speech analysis software, and 3) reproducing a speech sound with DPP. The target age of this exhibition is 8 through 18 years old; however, most of the participants are around between 6 and 12 years old. A detailed description of each of the three areas follows.

2.1. Vowel production with physical models of the human vocal tract

The vowel production corner is the first area visitors see when they enter our exhibition. In this section we used the cylinder-type model [1,2] which was based on Chiba and Kajiyama's measurements [11] as shown in Fig. 2. An electrolarynx was used as a sound source. When feeding a buzz-like sound from the electrolarynx into any one of the models in Fig. 2, a vowel sound was produced, depending on the shape of the vocal tract model chosen. Children were able to test this demonstration themselves.

2.2. Speech analysis with software

As participants proceed through our exhibition, they come next to the speech analysis corner, where the children can look at their own voices using software, such as "WaveSurfer" [12] and "Praat" [13]. WaveSurfer was used for visualizing speech with sound spectrograms, the time-frequency representations of speech sounds. Using this software, children tested the following, either in real time or by recording speech, looking at its spectrogram, and replaying it through a loudspeaker afterwards:

- High-frequency tones, such as a high-pitched whistle, appear in the upper part on the display, whereas low-frequency tones appear in the lower part of the display.
- Frication noises, such as /s/, appear in the upper part, whereas vowels, with their horizontal dark bands (or yellow/red bands when in color) appear in the lower part.
- Different steady vowels show different horizontal patterns in darkness (or color patterns); as a result, a vowel sequence, such as /aiai.../, shows a modulated pattern along the horizontal axis.
- A repeated consonant-vowel sequence, such as, /atatata.../, shows repeated vertical gaps on the display.
- A word or a child's name shows a combination of the sound patterns explained above.

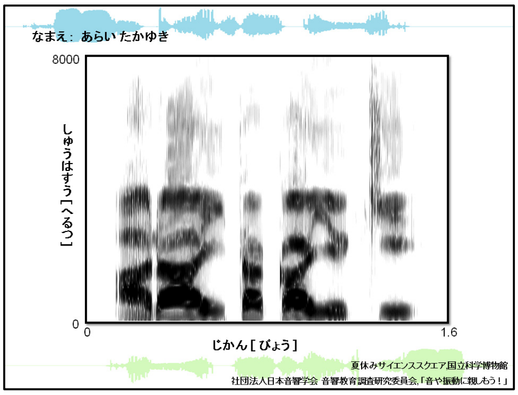

Figure 3: *Printed image of a sound spectrogram we provided for each visitor.*

First, an exhibitor uttered his/her own speech sounds and explained them by showing the resulting spectrogram to the children. Then, children uttered a phrase or their name, and the exhibitor helped them read the spectrogram.

The Praat software was mainly used for recording the child's name and printing its spectrogram on a 3.5 x 5 inch card. An example of the printed image is shown in Fig. 3. After the spectrogram was printed on a card, the card was used in the third activity. Afterwards, the card was given to the visitor.

2.3. Reproduction of speech from the printed spectrogram

The third speech corner of our exhibition used DPP [14-16], which is the digital version of Pattern Playback (PP) [17-19]. Pattern Playback was originally developed by Cooper and his colleagues from Haskins Laboratories in the late 1940s for converting a spectrogram into sound. PP has contributed greatly to the development of research in speech science. We developed the digital version by using digital signal processing, so that the system can be implemented on a PC. The system consisted of a video (or web) camera, PC, and a loudspeaker with an amplifier.

In this corner, children were able to listen to a sound reproduced from either the child's name spectrogram from the previous corner (described in Section 2.2) or from a pre-printed spectrogram containing a simple phrase or sentence. In the former case, the recorded name printed on the card, such as "Arai Takayuki," was heard with a monotonic sound with a constant fundamental frequency (f0). With the pre-printed cards, children could quiz themselves on what was said.

In 2012, we applied the variable f0 from [15] instead of the constant f0 contour we used in 2011. In this version, the colored f0 contour is laid over a spectrogram, so that DPP can extract the f0 contour by recognizing the color. We could draw an arbitrary f0 contour directly on a spectrogram or on a transparency, but we prepared a red wire instead, bent it along the f0 contour, and played with different intonations on a single spectrogram.

3. Discussion

During the two-day exhibition in 2011, many children visited our booth. Interestingly, not only children but also many parents showed interest in the exhibition. In this section, we will discuss what we observed through this event.

Figure 4: *Vocal tract model produces a vowel.*

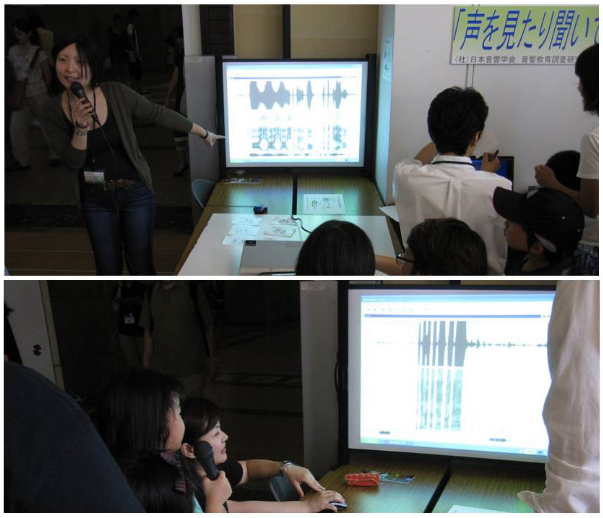

Figure 5: *Waveform and its spectrogram.*

3.1. Vocal tract model

In the first corner, with the vocal-tract models, we observed the following (Fig. 4):

- Children were interested in the vocal-tract models, because the physical models are tangible, visible, and audible.
- At the beginning, almost none of the children had any idea what the terms vocal tract, vibration, or resonance mean. However, by playing with the vocal-tract models, they began to understand these concepts as they explored the relationship between the sounds produced by the models, the sound source and the shape of the models, as well as the shape of their mouth as they produced vowel sounds.
- There were many children who wondered why similar round bottle shapes could produce different vowel-like sounds.
- Children liked to experiment with the vocal-tract models and see how they differed from one another.
- When we asked children to guess what vowels they heard, brothers and sisters competed against each other.
- We asked children to investigate what causes different vowels to be produced, and they tried very hard to get the answer.
- Many children were interested in the electrolarynx.

Something that really impressed us was that children easily recognized that the difference in vowel quality came from the shape of the vocal tract. And, we were able to get to the main point, “this is what’s happening inside your mouth,” very quickly. We think this is an easy, but effective way for children to learn the systematic relationship between vowel quality and vocal-tract shape.

3.2. Waveform and spectrogram

In the second corner, with the speech analysis software, we observed the following (Fig. 5):

- When we demonstrated the repetition of the vowel sequence /a/ and /i/, for example, children were surprised because speech sounds became visible and they observed unique patterns in the spectrogram.
- Children got so interested in the visualization of speech, especially their own speech, that many of them did not want to give up the microphone.
- Some noticed that the amplitudes varied depending on intensity, and they enjoyed whispering and shouting, etc.
- Some of the children understood the metaphor of a "music score" when explaining the axes of a spectrogram (the horizontal axis is time and the vertical axis is tonal height).
- Printing a spectrogram of their own voice was very popular because they could take it home.
- Children tried to find a difference between their own spectrograms and the others’, when they were with friends and/or family members.
- We also helped children to identify which part of the spectrogram corresponded to which phonemes in their names.

Some children were too shy to say their names loud enough, and others were too excited to record, so our recording endeavors met with varying degrees of success. However, when successful, it was a welcome surprise for them to actually see what they said on paper. The children were generally very cooperative while exploring their formant patterns with us.

Not only the children, but also the parents were excited about seeing the speech. The parents helped the children find their names in the spectrogram. When parents were cooperative, it was easier for the children to concentrate on the spectrogram. It often seemed that the most exciting part of this demonstration for children was to get a picture of what they had just said, and they were less interested in learning to locate the phonemes. Few children ever recorded an additional word to see whether the spectrogram looked different from that of their name. Perhaps they had already learned that things which sound different, look different. If so, we believe that if children are able to see the difference between words, they will later be able to see the difference between one phoneme and another.

3.3. DPP

In the third speech corner, using DPP, we observed the following (Fig. 6):

- Many children were impressed when they heard speech sounds played from the visual information on a card.
- Children were interested when they realized the "DPPed" speech with a constant f0 sounded like a robot.
- It seemed to surprise participants when a child’s voice was converted into a male voice by applying DPP with a low f0.
- Some children wondered how speech sounds would change if the orientation of the spectrogram were changed, so they flipped the cards around to investigated this.

Figure 6: *Printed spectrogram (left) was used for the DPP (right).*

- The quiz with the pre-printed spectrograms was very popular. Once they knew the answer, they claimed that they had really heard that particular phrase or sentence.
- In the 2012 version, children were able to compare the DPPed sounds with flat f0 contour and different intonations many times by changing the shape of the wire. We were also able to show them the difference between affirmative sentences and questions by using a single spectrogram with different f0 contours.

4. Conclusions

In this paper, we described a hands-on speech science exhibition for children that we organized at a science museum. In this exhibition, children were able to produce vowels by feeding a sound source into vocal-tract models, observe a waveform and its spectrogram, and reproduce speech sounds from the spectrogram by using DPP. One of our most important goals is to encourage children to ask the question "why?" Through the activities of our exhibition, we believe we succeeded in that goal.

There are several ways we can improve future exhibitions. First, we would like to add new models of the vocal tract, especially smaller models, for children's voices. We have already developed shorter versions of the S3T model and confirmed that they can produce child-like vowels with a sound source of higher f0 frequency [20]. We would also like to add a model with jaw movement. Children seemed to understand that there is a relationship between jaw movement and vowel sound change. However, they had no opportunity to learn about this explicitly. It is important that children be able to see a model with jaw movement, as this would provide a trigger to help them think more completely about how human sound is produced. Bent-tube models, as opposed to straight-tube models, might also help children to imagine where the vocal-tract is situated within the head [2]. In 2012, vowel production corners contained vocal-tract models, speech analysis and software in the same place, and this had a synergetic effect. Finally, we would like to consider introducing a worksheet in the future, as children may gain a better understanding through following the instructions and/or quizzes on the worksheet.

5. Acknowledgments

We would like to thank members of the National Museum of Nature and Science and members of the Committee of Education in Acoustics, the Acoustical Society of Japan (especially, Prof. Fumiaki Sato, the chair of the Committee of Education in Acoustics). This work was partially supported by Grants-in-Aid for Scientific Research (21500841, 24501063) from the Japan Society for the Promotion of Science.

6. References

[1] Arai, T., "The replication of Chiba and Kajiyama's mechanical models of the human vocal cavity," *J. Phonetic Soc. Jpn.*, 5(2), 31-38, 2001.
[2] Arai, T., "Education system in acoustics of speech production using physical models of the human vocal tract," *Acoust. Sci. Tech.*, 28(3), 190-201, 2007.
[3] Arai, T., "Education in acoustics and speech science using vocal-tract models," *J. Acoust. Soc. Am.*, 131(3), Pt. 2, 2444-2454, 2012.
[4] Arai, T., "Gel-type tongue for a physical model of the human vocal tract as an educational tool in acoustics of speech production," *Acoust. Sci. Tech.*, 29(2), 188-190, 2008.
[5] Fant, G., *Theory of Speech Production*, Mouton, The Hague, Netherlands, 1960.
[6] Arai, T., "Sliding three-tube model as a simple educational tool for vowel production," *Acoust. Sci. Tech.*, 27(6), 384-388, 2006.
[7] Exploratorium: http://www.exploratorium.edu/
[8] Sony ExploraScience: http://www.sonyexplorascience.jp/
[9] Arai, T. "Speech exhibition at a science museum," *Proc. Spring Meet. Acoust. Soc. Jpn.*, 1387-1390, 2009 (in Japanese).
[10] Satoh, F., Yano, H., Tachibana, H., Ueno, K., Yokoyama, S., Sakamoto, S., Nakamura, K. and Arai, T., "Demonstrations at the "SCIENCE-SQUARE" held by the National Museum of Nature and Science," *Proc. Spring Meet. Acoust. Soc. Jpn.*, 1385-1386, 2009 (in Japanese).
[11] Chiba, T. and Kajiyama, M., *The Vowel: Its Nature and Structure*, Tokyo-Kaiseikan Pub. Co., Ltd., Tokyo, 1942.
[12] WaveSurfer: http://www.speech.kth.se/wavesurfer/
[13] Boersma, P. and Weenink, D. Praat: doing phonetics by computer [Computer program]. Retrieved from http://www.praat.org/
[14] Arai, T., Yasu, K. and Goto, T., "Digital pattern playback," *Proc. Autumn Meet. Acoust. Soc. Jpn.*, 429-430, 2005 (in Japanese).
[15] Arai, T., Yasu, K. and Goto, T., "Digital pattern playback: Converting spectrograms to sound for educational purposes," *Acoust. Sci. Tech.*, 27(6), 393-395, 2006.
[16] Arai, T., "Digital pattern playback for education in digital signal processing and speech science," *Proc. of IEEE ICASSP*, 2769-2772, Kyoto, 2012.
[17] Cooper, F. S., Liberman, A. M. and Borst, J. M., "The interconversion of audible and visible patterns as a basis for research in the perception of speech," *PNAS*, 37, 318-325, 1951.
[18] Cooper, F. S. Delattre, P. C. Liberman, A. M. Borst, J. M. and Gerstman, L. J., "Some experiments on the perception of synthetic speech sounds," *J. Acoust. Soc. Am.*, 24(6), 597-606, 1952.
[19] Borst, J. M., "The use of spectrograms for speech analysis and synthesis," *J. Audio Eng. Soc.*, 4, 14-23, 1956.
[20] Arai. T., "Vowels produced by sliding three-tube model with different lengths," *Proc of INTERSPEECH*, Portland, 2012.