A “voice” instrument based on vocal tract models by using soft material for a 3D printer and an electrolarynx

Fou Yoshimura
Graduate School of Design,
Kyushu University, Fukuoka, Japan
ysmr.fou4@gmail.com

Kazuhiro Jo
Faculty of Design,
Kyushu University, Fukuoka, Japan
YCAM, Yamaguchi, Japan
jo@jp.org

ABSTRACT
In this paper, we propose a “voice” instrument based on vocal tract models with a soft material for a 3D printer and an electrolarynx. In our practice, we explore the incongruity of the voice instrument through the accompanying music production and performance. With the instrument, we aim to return to the fact that the “Machine speaks out.” With the production of a song “Vocalise (Incomplete),” and performances, we reveal how the instrument could work with the audiences and explore the uncultivated field of voices.

Author Keywords
Voice, Vocal Tract, Electrolarynx, Music Production, Performance

CCS Concepts
• **Applied computing → Sound and music computing**; Performing arts;

1. INTRODUCTION
What is an instrument of “voice”? A classic form of a voice instrument is Von Kempelen's talking machine [4]. Modeling a human oral organ for creating speech is a starting point for showing that synthetic speech is a technique to reproduce each state of the vocal organs [2]. When the phonograph was released, the audience felt a sense of incongruity with the fact that the machine gave a speech. [8] When Vocoder by Homer Dudley came out in 1928 [5], there was a problem with the pitch; it was difficult to say that it was a human voice; it was a very disgusting sound. Dudley also invented a manual talking machine, Vodar, in 1939 [6]. It utters consonants and vowels with a pitch-modulator-like function [12]. Again, people considered it equally impractical to a vocoder until their applications in music and entertainment after military use. After that, such incomplete voices turned into technologies that reconfirm the identity of voice to the listener in the field of music.

For over a century after the VOCALOID [9] came out, there was another criticism that the mechanical singing voice is harsh and uncomfortable, though it is exciting with a niche demand [15]. Otamatone, by Meiwa Denki[1], has gained popularity due to another approach, with its cute appearance and tone-deafness caused by the difficulty of pitch control. Masahiro Miwa and Nobuyasu Sakonda formed the Formant brothers[2] in 2000. By developing the *Brother’s Button-to-Phoneme Transfer Standard for International languages* [11] and many other experimental works, they explored philosophical, aesthetical, musical, and, technical raison d'être of machine singing in the 21st century.

By following the genealogy of the incongruity of “voice” instruments, we propose an instrument based on vocal tract models with a soft material for a 3D printer and an electrolarynx. The instrument aims to be free from restrictions of specific age and gender based on our previous research [16]. Through the accompanying music production and performance, we explore the incongruity of “Machine speaks out” to people habituated to synthesizing speech [14].

2. A “VOICE” INSTRUMENT
Based on the influence of previous practices, we produce a “voice” instrument based on vocal tract models with a soft material for a 3D printer and an electrolarynx.

2.1 Vocal tract models
A vocal tract model is the path of the human voice in a sound tube. Because the characteristics of sound change with resonance on the vocal tract, it produces various sounds by changing its shape [7]. Chiba and Kajiyama measure the vocal tract shape for the five Japanese vowels and realize a vocal tract model with oil clay [3]. Arai restored the model with acrylic resin [1] in two types, viz., pipe and plate, with similar sound quality [10]. For our instrument, we use the 3D data of the model[3] with their permission. For the ease of performing and carrying, we chose the pipe type for the building. As an instrument of “voice,” we hypothesize that the primitive shapes and vowels of the model could lead us to a root of the discomfort of “Machine speaks out.”

Figure 1. Vocal tract model of the five Japanese vowels with the soft material.

[1] Otamatone, Meiwa Denki, https://www.maywadenki.com/otamatone/
[2] Formant Brothers, http://formantbros.jp/
[3] Vocal tract model, Arai Lab, Sophia University http://splab.net/Vocal_Tract_Model/

[4] Karaoke with an electrolarynx with a pitch frequency control function https://www.youtube.com/watch?v=uuyJUrnPITk

NIME’19, June 3-6, 2019, Federal University of Rio Grande do Sul, Porto Alegre, Brazil.

For building the instrument, we employ a soft material, a flexible resin for the Form2 SLA 3D printer to produce the five Japanese vowels (Fig.1). Instead of the rigid material used for a standard 3D printer, the material allows us to reconfigure the vowel shape, like the Otamatone, which cannot be attained by the vocal tract of a human being. With the ambiguity of the voice, we aim to realize "the incongruity of voice."

2.2 Electrolarynx

We use an electrolarynx with a pitch control function [13] "Yourtone" to actuate the soft vocal tract models. An electrolarynx is a substitute vocalization device for people who have had a laryngectomy surgery mainly due to laryngeal cancer. Though there are difficulties in the use of the device, there are people who appreciate Karaoke with the device[4]. In our practice, we regard the difference with the vocal cord and the electrical vibration as an essential character of the instrument.

2.3 How to perform

To perform with the instrument, we use our dominant hand to operate the slide switch and use the reverse hand to operate the vocal tract models. During a performance, we change the mode (i.e., conversation and singing), and the presence of intonation. We also deform the vocal tract model by applying a force from various directions to utilize its softness fully. We adjust the grip from grasp to squeeze to produce diverse ranges of voices.

3. MUSIC PRODUCTION AND PERFORMANCE

3.1 Music production

For the instrument, one of the authors composes "Vocalise (Incomplete)." This is an electronic acoustic work composed for synthesized speech to make full use of the voices of our instrument. Based on an "Incomplete" foundation with various voice colors of VOCALOID, the performance overlaps improvisations with several "voice" instruments, including Pocket Miku, Otamatone, the "voice" instrument, a voice of the performer with Electrolarynx, and the performer herself to bear an adjusted "complete" music. "Vocalise" originally means a singing law without lyrics in classical music, in which the skill of a singer is tried. In the piece, we followed the original concept by limiting the voice of all the instruments entirely to vowels. With the limitation, we aimed to make the listeners sensitive to the perception of the voice itself, to be aware of the difference between expression and experience, as it is an instrument producing sounds different from the voice of a living person.

3.2 Performance

We conducted three performances of "Vocalise (Incomplete)" at FREQ2018 - Sounds and Music in the Early 21st Century on March 30, 2018, Tsukuruto! vol.5 on July 15, 2018, and COPY CONT ROLL 03 on November 29, 2018.

4. DISCUSSION

In the second and third performance, we have found high interest for the vocal tract models among the audience. In the last one, a spectator (@emarikko) reacted on Twitter as follows:

"The performer firstly used Otamatone, then played with Electrolarynx. I cannot explain this easily in my vocabulary ... The attempt itself was quite interesting. The vowel-shaped object (what I was talking about ...) was also very intriguing. In order to make a difference between 'A' and 'Ka', I need elements other than shape."

The last sentence raises a question regarding the handling of consonants, which applies to all the voice instruments, as we previously mentioned.

Through the practice of instrument building, music production, and performance, the utilization of a vocal tract model facilitates the primitive simulation of human organs with changes in voice colors, as we could observe in "Von Kempelen's talking machine." The results suggest that the listener could notice an occurrence where the "Machine speaks out." At the last performance, the combination of two voices, an electrolarynx with the performer's vocal trait and the performer's voice itself, points out an alternative incongruity of "voice" in which two different pitches of voice came from a vocal tract.

ACKNOWLEDGMENTS

This work was supported by JSPS KAKENHI Grant Number JP17H04772.

Video from the latest performance on November 29, 2018.
https://photos.app.goo.gl/daxMvZHtM92tuVLS6

REFERENCES

[1] T. Arai. The Replication of Chiba and Kajiyama's Mechanical Models of the Human Vocal Cavity, *J. Phonetic Soc. Jpn*, 5(2) (2001), 31-38.
[2] J. P. Cater. *Electronically Speaking: Computer Speech Generation*: Howard M. Sams & Co., Indianapolis, 1983.
[3] T. Chiba, and M. Kajiyama. *The Vowel ' Its Nature and Structure*, Tokyo-Kaiseikan Pub. Co., Ltd., Tokyo, l941.
[4] H. Dudley, and T. H. Tarnoczy, The speaking machine of Wolfgang von Kempelen. *J. Acoustical Soc. Am.*, 22, 2 (1950), 151-166.
[5] H. Dudley. The vocoder—Electrical re-creation of speech. *J. Soc. Motion Pict.Eng.*, 34, 3 (1940), 272-278.
[6] H. Dudley, R. R. Riesz, and S. S. A. Watkins. A synthetic speaker. *J. Franklin Inst.*, 227, 6 (1939), 739-764.
[7] H. K. Dunn. The calculation of vowel resonances, and an electrical vocal tract. *J. Acoust. Soc. Am.*, 22, 6 (1950), 740-753.
[8] R. Gelatt. *The Fabulous Phonograph: 1877—1977*, London: Cassell, 1977.
[9] H. Kenmochi, and H. Ohshita, Vocaloid-commercial singing synthesizer based on sample concatenation. *In Eighth Annual Conference of the International Speech Communication Association*, 2007.
[10] E. Maeda, N. Usuki, T.Arai, N.Saika, and Y. Murahara. Comparing the characteristics of the plate and cylinder type vocal tract models. *Acoust. Sci. Technol.*, 25, 1 (2004), 64-65.
[11] M. Miwa, and N. Sakonda. The long and straight road to "brother's button-to-phoneme transfer standard for international language," *Jpn. Soc. Sonic Arts*, 5, 2 (2013), 13-16 (in Japanese).
[12] D. Tompkins. *How To Wreck A Nice Beach: The Vocoder from World War II to Hip-Hop*, Melville House Publishing, 2011.
[13] N. Uemi, T. Ifukube, M. Takahashi, and J. I. Matsushima. Design of a new electrolarynx having a pitch control function. *In* Proceedings of 1994 3rd IEEE International Workshop on Robot and Human Communication*, 1994 (RO-MAN'94) (Nagoya, Japan, 18-20 July 1994)*, 198-203.
[14] A. Van Den Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, ... K. Kavukcuoglu. Wavenet: A generative model for raw audio. *CoRR* abs/1609.03499. 2016
[15] B. Werde, Could I Get That Song in Elvis, Please? *New York Times Online*, 2003. https://www.nytimes.com/2003/11/23/arts/music-could-i-get-that-song-in-elvis-please.html
[16] F. Yoshimura, Perceptual impressions of the age and the gender for voices, *BA Thesis*, Kyoto City University of Arts, Kyoto, Japan, 2017.