V^3: an Interactive Real-Time Visualization of Vocal Vibrations

Rebecca Kleinberger *
MIT Media Lab

Abstract

Our voice is an important part of our individuality but the relationship we have with our own voice is not obvious. We don't hear it the same way others do, and our brain treats it differently from any other sound we hear [Houde et al. 2002]. Yet its sonority is highly linked to our body and mind, and deeply connected with how we are perceived by society and how we see ourselves. The V^3 system (Vocal Vibrations Visualization) offers a interactive visualization of vocal vibration patterns. We developed the hexauscultation mask, a head set sensor that measures bioacoustic signals from the voice at 6 points of the face and throat. Those signals are sent and processed to offer a real-time visualization of the relative vibration intensities at the 6 measured points. This system can be used in various situations such as vocal training, tool design for the deaf community, design of HCI for speech disorder treatment and prosody acquisition but also simply for personal vocal exploration.

1 Introduction and Motivations

Auscultation is the act of listening to internal sounds from the body. Our sensor listens at the skin surface to the voice from 6 different locations. According to the source-filter model of vocal production, the filter part of the voice, called the vocal tract shape, defines parts of the sound quality. The differences between two sounds, /æ/ and /ɘ/ for example, are controlled by muscle contraction, opening and closing of cavities and placement of the tongue. All these elements shape the vocal tract and create a unique resonant characteristic that results in a unique formant pattern, which is to say, in a unique sound. The V^3 project gives access to some upstream physical control parameters of the voice by making their origin more accessible. This point of view offers an alternative consideration of a potential fullness of the voice that is not guided by grammar or specific language but by the physical control offered by our body. In doing so this project aims to increase the awareness of the physicality and embodiment of people's voice.

2 Interaction Model and System Design

The user wears the two sides of the mask on his face supported by the ears. For every sound pronounced, the system shows in real-time the amplitude of the vibrations at the six locations as size-varying concentric waves. We call vibration patterns the 6*1 vector formed by the signals. The system contains 2 viewing modes: absolute mode, where the size of the ring corresponds directly to the intensity; and relative mode, where the vibration pattern is normalized by the mean of all the points to become independent from loudness. When using a webcam, the system also offers a mirror mode in which the vibration rings are displayed on top of a real-time video feed. In bioacoustics, the importance of limiting air-borne sensitivity of sensors to only measure the tissue-borne signal pledges in favor of piezzo-electric sensor instead of other accelerometer-type technologies or microphones[Zanartu et al. 2009].Our sensors are piezzo-based phantom-powered vibration sensors. We used impedance matching boards. For the Hexauscultation mask we used 1/4 inch piezzo sensors whose channels are preprocessed by a phantom power preamplifier board. The output is read by a 6 input channel audio interface (we use a TASCAM US-1800). The board is chassis-grounded on the metal thread and the whole device is isolated and shielded. For questions of safety, comfort, convenience and aesthetics, we wrap each sensor into shrink-wrap. The sensors are mounted on flexible metal wires that are shaped to support the sensors at the right place, in contact with the skin and supported behind the ears. The real-time audio processing part of the system is done using Max/MSP; we pre-filter the 6 input channels and extract their envelopes. Then we assign a different coefficient on the three levels of measurement: throat, cheek and forehead. Finally, we send the 6 envelopes in real-time by Open Sound Control (OSC) to the visualization part of the system implemented in Java. The locations of the 6 points are by default displayed in a pseudo circle but can be modified manually.

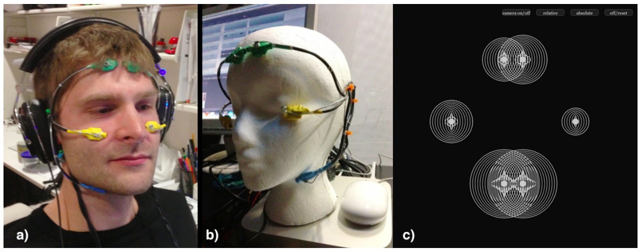

Figure 1: *a) and b) Hexauscultation mask c) Example of vibration patern observed with the V^3 system*

3 Evaluation and Applications

After testing the accuracy of the system, we invited 45 participants to try it and give comments. The system proved efficient at ignoring facial motions when the user moves and articulates sounds. Individual differences in skin thickness and dryness, hairiness and ear position seemed to cause slight differences in the signal while remaining in a reasonable range. Users' comments about the structure reveal that people find the mask very light and not constraining. Even though it is sometimes painstaking to put on, the participants reported they forgot about it after a couple of minutes. Participants reported that they enjoyed the system and that it helped them produce unusual sounds. Some people reported being frustrated that they were not able to make the top of their head vibrate more. Professional singers reported that the device helped them connect sound and body and thought it could be helpful for vocal training. This system can also be used in various situations such as tool design for the deaf community, design of HCI for speech disorder treatment and prosody acquisition but also simply for personal vocal exploration.

References

HOUDE, J. F., ET AL. 2002. Modulation of the auditory cortex during speech: an meg study. *Journal of cognitive neuroscience 14*, 8, 1125–1138.

ZANARTU, M., ET AL. 2009. Air-borne and tissue-borne sensitivities of bioacoustic sensors used on the skin surface. *Biomedical Engineering, IEEE Transactions on 56*, 2, 443–451.

---

*e-mail:rebklein@mit.edu

SIGGRAPH 2015 Posters, August 09 – 13, 2015, Los Angeles, CA.
ACM 978-1-4503-3632-1/15/08.
http://dx.doi.org/10.1145/2787626.2792624