Live voice transformations for web conferencing using the browser

Gerard Llorach*, Mattes Ohlenbusch*, and Thomas Brand
Dept. of Medical Physics and Acoustics, Universität Oldenburg
{gerard.llorach.to, mattes.ohlenbusch, thomas.brand}@uni-oldenburg.de

INTRODUCTION

This demo consists of two elements: the web technology for real-time vocoding and its application for live voice transformation in web conferencing. This is an interesting approach, now more than ever in times of COVID-19, for modifying your speech in virtual conferences. We present a software-based solution that should be available to anyone doing web conferencing.

WEB-BASED VOCODER

In this demo we show a simple vocoder that works on live input. The technologies behind are developed with Audio Worklets, which is a replacement for the deprecated ScriptProcessorNode in the Web Audio API. The vocoder works with LPC analysis and synthesis [1]. The signal is synthesized with an overlap and add routine. In our demonstration, we do speech transformations by modifying the vocal tract length [2]. Other voice transformations such as pitch and voiced/unvoiced synthesis are also possible. We developed a 2D interface to test some of these voice transformations (see Figure 1). This 2D interface allows the user to control the voice transformation in real-time intuitively. The code is available at: https://github.com/gerardllorach/webbasedvocoder. The demo works only in Chrome and it can be seen at: https://gerardllorach.github.io/Fortgeschritenne/.

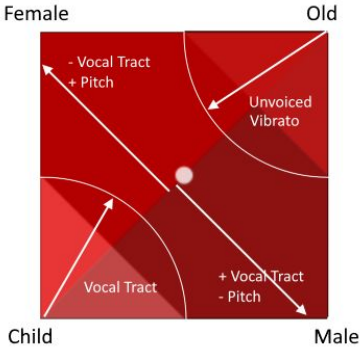

Figure 1. 2D interface to control the voice transformations. In white, the speech modifications are shown, depending on the position of the white cursor.

Licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0). **Attribution:** owner/author(s).

Web Audio Conference WAC-2021, July 5–7, 2021, Barcelona, Spain.

© 2021 Copyright held by the owner/author(s).

*These authors contributed equally to this work.

VOCODER FOR WEB CONFERENCING

Instead of reproducing the vocoder's synthesized signal through the loudspeakers or headphones, the output of the web-based vocoder is connected to the actual audio input of a web conferencing software. Then, the recipients of the virtual conference hear the synthesized live signal. The voice transformations can be modified in real-time during the virtual conference through the vocoder's web interface. We tried this system with Windows, but the method should be possible with other operating systems, e.g., Linux. In Windows, the StereoMix creates a virtual source and this virtual source can then be connected to another software's input. Generally speaking, every audio loopback software can be utilized for this. The audio input from the user (via microphone) is passed through the web application and then forwarded to the software used for web conferencing. Because this re-routing is entirely virtual, no additional hardware is required. In our demo, we used Chrome for the voice transformations and Firefox for the virtual conference. This demo can be demonstrated virtually, as we will be able to show this technology live.

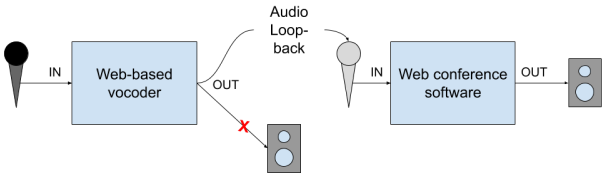

Figure 2. Schema of the audio routing for voice transformations. On the left, we have the physical microphone capturing the speech signal. The signal goes into the web-based vocoder (Chrome). Its output is then connected to the virtual microphone of the web conference software.

ACKNOWLEDGMENTS

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 352015383 – SFB 1330 A1.

REFERENCES

[1] J. D. Markel and A. H. Gray, 1976. Linear Prediction of Speech. New York, Springer.

[2] L. Rabiner, 2012. Linear Predictive Coding (LPC)-Lattice Methods, Applications. University of California-Santa Barbara. Microsoft Powerpoint presentation.