SIGNAL CHARACTERIZATION OF OCCLUDED IN-EAR VERSUS FREE-AIR VOICE PICKUP ON HUMAN SUBJECTS

Antoine Bernier, and Jérémie Voix
École de Technologie Supérieure, Université du Québec, 1100 Notre-Dame Ouest, Montréal (QC), Canada, H3C 1K3
antoine.bernier@polymtl.ca jeremie.voix@etsmtl.ca

1. INTRODUCTION

While speaking, a human subject will perceive his own voice mostly through the free-air radiation path. Nevertheless, the vibration of his vocal tracks and mouth cavity are also inducing a bone (skull) conducted vibration that will excite the cochlea. This phenomenon is inherently part of one's perception of his voice.

When the ear canal is occluded with an in-ear device (such as an earplug) providing a good acoustic seal (from the external environment) while leaving the ear canal walls free to vibrate (device shallow inserted), the regenerated sound pressure level inside the occluded ear canal can be very clearly perceived by the human subject (it is called the occlusion effect), while the free-air radiation path will no longer be significant to the resulting perception.

While the occlusion effect is often perceived negatively, it also enables the pickup of one's own voice from within his occluded ear canal. A small microphone placed in a well occluded ear will provide a strong signal, free from external disturbances. However, the resulting voice-pickup will be perceptually different, usually colored and not as "clear" sounding as a free-air propagating voice taken under the same conditions, unless extra signal processing is applied.

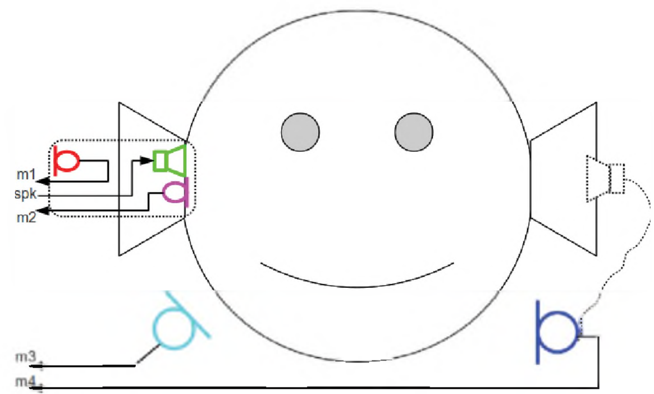
The purpose of the current research study is to characterize the differences in the signal characteristics of the occluded ear voice pickup relatively to the free-air voice pickup, in the form of a transfer function between the two signals: A first step to identify the signal processing parameters that would be required to alter the in-ear voice pickup to have both signal perceptually similar to the human ear.

2. METHOD

2.1 Test procedure and Data Acquisition

Nine subjects (index *s01* to *s09*) of different age, sex and physiological characteristics were guided through to same procedure to acquire the test signals: The subjects were to read seconds a text in their first language (French, denoted *fr* or English, denoted *en*) for 40 seconds while wearing a special in-ear device. Three bilingual subjects were retested in their second language. The in-ear device used is a new communication earpiece developed by Sonomax [1] and designed to instantly custom-fit all ear canals. For this experiment, the same earpiece was reused and, in order to achieve a good seal for each subject, a layer of disposable and malleable wax was added around the curved cone-shaped end of the earpiece.

The original in-ear device, illustrated in Fig. 1, enclosed two microphones: a first microphone (*m1*) inside the earpiece for in-ear voice-pickup through a sound-bore, the second (*m2*) on the outer faceplate of the in-ear device, pointing outward for free-air pick-up. A high fidelity free-field reference microphone (*m3*) is also present. The subjects wore the device on their right ear, and were equipped on the other ear with a commercial hand-free cell phone apparatus composed of an earphone (in dotted line in Fig. 1) with a wire-hanging microphone (*m4*). Although present in the original earpiece, the speaker (*spk*) was not used for this experiment.



All the 4 microphone signals were recorded through National Instruments PXI 4410 data acquisition cards at sample frequency 44.1 kHz with 24 bits resolution. The NI cards were interfaced with Matlab [2] that ran all the post-processing of the signals.

2.2 Post-Processing

To obtain the transfer function between the signals of interest, a dual channel FFT analysis [3, 4, 5] was conducted. To compute the FFT, the 40 seconds long signals were divided into the smallest power of 2 frame length offering experimental stability (from where larger frame length no longer modified the results), and the resulting FFT were used to obtain the Cross-spectra and Auto-spectra. These spectra were then averaged and used to compute the best transfer function estimate. The algorithm was tested and validated using signals altered by known transfer functions. Validity of the transfer function estimate over the speech frequency bandwidth was monitored by calculating the coherence between the signals. Presence of coherence would mean a given speech signal frequency component was picked up by both microphones, meaning it is carried by both the free-air and the bone conducting path, hence permitting a valid transfer function estimation at this point. A lower or no coherence (values lower than 0.1) would either mean there is no signal to analyze (outside of the speech frequency bandwidth) or that it is absent on one of the microphone (if the bone conducting path only partially transmit voice inducted vibrations, or the microphone is not sensitive enough).

Due to the space contains of this conference proceedings, only the transfer functions between *m1* and *m4* are presented in this paper. Microphones *m4* and *m1* are commercial and pre-commercial MEMS audio microphones with similar sizes both designed to be surface-mounted but produced by different –undisclosed- manufacturers. Their frequency response is similar and almost flat at +0/-6 dB over the 20 Hz – 20 kHz frequency range, but their nominal sensitivity is slightly different, with *m1* at -38 dB (0 dB reference is 1V/Pa)) and *m4* at -40 dB. Neither the 2 dB difference in electrical sensitivity, nor pre-gain used in signal conditioning are accounted for in the data presented here-below.

3. RESULTS

The magnitude of the transfer function between external microphone *m4* and internal microphone *m1* is presented in Figure 2 for the 9 subjects, over the 12 tests performed. The group average and the standard deviation are also computed for every 1/12th octave band: Fig. 2 shows in overlay the mean as well as the mean ± standard deviation. It first can be seen that large inter-subject variations can occur in the magnitude of the transfer function, but that such shift is most often quite constant over a large frequency range. For example, the transfer function magnitude measured for English-speaking subject 5 (mustard color curve with legend *s05-en* in Fig. 2) is consistently larger than the other transfer function (hence shifted upwards), but shows the same frequency pattern than the other transfer function for the other subjects and is hence looking similar to the group average transfer function. It could also be seen, for the bilingual subject tested (*s04*, *s05* and *s09*), that the measured transfer function is not much sensitive to the language used.

The coherence functions are presented in Fig. 3 for the same group of subject, together with the group average. It first can be seen that the coherence value equals or exceeds 0.3 for the 100 Hz – 2 kHz frequency range and that a 60 Hz – 3.5 kHz range can be achieved, allowing for the coherence to be as low as 0.1. Such frequency range is to be compared to the classical 300 Hz – 3.5 kHz used in traditional phone line systems. An abnormal lower coherence value around 1 kHz can also be observed, it may well be related to a know physical/acoustical resonance of the earpiece at that frequency.

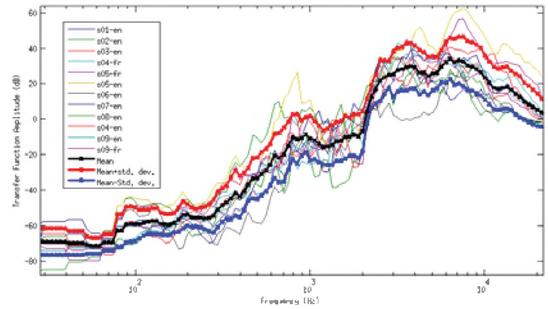


Figure. 2. Measured Transfer Functions

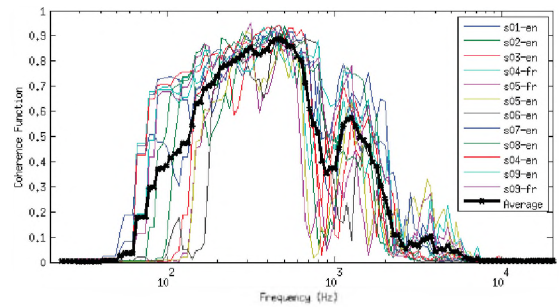


Figure. 3. Measured Coherence Functions

If the average transfer function magnitude is now considered over the meaningful frequency range (i.e where the coherence is higher than 0.1), it can be seen that the typical correction to be applied to an in-ear microphone signal can be quite simple, as it can be in first-approximation considered to be a monotonic rising slope. Such rising pattern can be for example, obtained from an analog high-pass filter used in its cutting-band and is quite straightforward to achieve in the analog or digital domain, depending of the application envisioned.

REFERENCES

[1] Sonomax Technologies (www.sonomax.com)
[2] Mathworks (www.mathworks.com)
[3] Thrane, N. (1979). The Discrete Fourier Transform and FFT Analysers. B&K Technical Review. No 1.
[4] Herlufsen, H. (1984). Dual Channel FFT Analysis (Part I). B&K Technical Review. No 1.
[5] Herlufsen, H. (1984). Dual Channel FFT Analysis (Part II). B&K Technical Review. No 2.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the financial and logistic support from CRITIAS, the Sonomax-ETS Industrial Research Chair in In-Ear Technologies.

AUTHOR NOTES

Some of the work presented here, was conducted while Antoine Bernier was completing his electrical engineering internship at Sonomax Technologies inc, as a bachelor degree student from École Polytechnique de Montréal.