ESTIMATING TRANSFER FUNCTION FROM AIR TO BONE CONDUCTION USING SINGING VOICE

Sook Young Won *Jonathan Berger*

Center for Computer Research in Music and Acoustics,
Department of Music,
Stanford University

sywon@ccrma.stanford.edu brg@ccrma.stanford.edu

ABSTRACT

Singers often notice a disparity between how they perceive their own voice while they are singing and how they perceive their voice when hearing a recorded playback. This research aims to produce a digital filter which will take a sound sung by a singer and render it to a processed sound which is a close match to that heard by the singer during performance. One reason for the disparity arises from the fact that the skull conveys sound to ears when people speak or sing, and then the sound that is delivered blends with the air-conducted ones at the cochlea. In this paper, we investigate the spectral difference between the air-conducted sound and the bone-conducted sound, and estimate the transfer function from air conduction to bone conduction by signal processing methods. Then we transform the original recording by air transmission using the filter coefficients of the transfer function and compare the filtered sound to the original recording by bone transmission.

1. INTRODUCTION

Over many decades, several experiments on the subject of bone transmission have been performed in various ways: estimating bone conduction transfer functions using oto-acoustic emissions [1], measuring vibration responses by means of miniature accelerometers onto a dry skull [2], and investigating the free-damped natural frequencies (i.e. resonance frequencies) of the living human skull with skin penetrating titanium implants [3].

The researchers have revealed certain characteristics of bone transmission and confirmed the well-known phenomenon of bone conduction: the lateralization effect[1]. However, a clear and complete model of skull vibrations has not been suggested; the bone conduction mechanism is very complex and the variations among the inter-subjects of experiments are large, probably due to individual variations in the skull geometry and in the mechanical parameters. Furthermore, it is impossible to access the movement of the basilar membrane and the cochlear of a living human in the real-time because of ethical reasons. [2]

We thus conduct our own experiment for recording sounds propagated through air and bone, analyze and compare two different types of sounds, and then implement the bone-conducted sound by filtering. The schematics of this research are the following:

2. EXPERIMENT

2.1. The Measurement Setup[2]

The first step of the experiment is setting up a reliable recording tool for picking skull vibrations. We build our own contact microphone consisting of a 'Piezo Transducer 273-073A' from RadioShack, a quarter-inch phone jack, and audio cables.[3] Picking up sounds conducted by air is simpler; we use a common microphone (Sennheiser ME65/K6) with the super-cardioid pick-up pattern. The bone-conducted sound captured by the piezo and the air-conducted sound captured by the microphone each go into the right and the left channel of portable direct CD recorder 'Marantz Professional CDR300.'

2.2. Recording

A human cannot sing impulses, white noise, or the golay code, a useful sound source in obtaining the impulse response of the medium. Therefore we attempt to investigate not the transfer function of the human skull itself, but the function from the air-conducted sound to the bone-conducted sound.

To acquire two types of sound simultaneously, we attach the piezo contact microphone on the flat region close to the ears, and place the Sennheiser ME65/K6 in front of the mouth of the subject.[4] The subject then speaks a single vowel and sings various patterns such as a single note, a chirp and a scale. The Marantz CDR300 records the sounds in the form of a wave file directly. The recordings have been performed twice, namely on Recording 1 and Recording 2.

Recording 1 on October 27th: speech and singing in head voice, speech in chest voice - 21 pairs of samples.
Recording 2 on November 13th: singing chirp pattern and scale pattern in head voice - 10 pairs of samples.

---

[1] The lateralization effect arises when the stimulation of one cochlea is either temporally ahead of or at a higher level than the other cochlea, even if the stimulation point is essentially closer to one of the cochleae.[2]

[2] We assume that frequency responses of all recording equipments are ideally flat.

[3] http://home.earthlink.net/~erinys/contactmic.html

[4] The subject is a classical soprano singer with vocal control of head voice or chest voice.

All the recording equipments and conditions are the same in both experiments except for the piezo microphone used for Recording 2, which has a soundproofing cover to prevent the microphone from sensing the sound transmitted by air.

3. ANALYSIS AND IMPLEMENTATION

For the first step, the recorded time domain signals are transformed into frequency domain signals, as computed by taking Discrete Fourier Transform [5]. Then we extract the spectral envelopes of signals by Linear Predictive Coding and Cepstral Smoothing.

3.1. Linear Predictive Coding

LPC is a powerful technique for representing a spectral envelope using a linear predictive model [6], [7], [8], [9].

The linear prediction of the future value *y(n)* from a linear function of previous values can be represented as the following;

$$y(n) = -a_1 y(n-1) - a_2 y(n-2) - \cdots - a_M y(n-M) + e(n) \quad (1)$$

where a_i and $e(n)$ denote the prediction coefficients and the prediction error respectively. We assign the value of 80 to the order of M in practice.

The prediction coefficients are computable from the autocorrelation function.

$$r_{y_m}(l) \triangleq \sum_{n=-\infty}^{\infty} y_m(n) y_m(n+l) = \mathrm{DFT}^{-1}|Y_m|^2 \quad (2)$$

Matlab's `lpc` function finds the predicted filter coefficients from the input parameters including signals and the value of the order, and `freqz` function returns the complex frequency response vector with the given filter coefficients.

3.2. Cepstral Smoothing

The *cepstrum* is a spectral technique that represents the frequency distribution of the spectrum curve fluctuations [6], [10], [11].

$$Y_m = \mathrm{DFT}\,[w \cdot \underbrace{\mathrm{DFT}^{-1}\log(X_m)}_{\text{complex cepstrum}}\,] \quad (3)$$

where w is a lowpass window.

$$w(n) = \begin{cases} 1, & |n| < n_c \\ 0.5, & |n| = n_c \\ 0, & |n| > n_c \end{cases} \quad (4)$$

3.3. Estimating the Transfer Function

The transfer function $H(w_k)$ from the air-conducted sound to the bone-conducted sound can be obtained (by a simple division) as the following:

$$H(\omega_k) \triangleq \frac{B(\omega_k)}{A(\omega_k)} \quad (5)$$

where $A(w_k)$ and $B(w_k)$ are the spectra of the air-conducted sound and of the bone-conducted sound respectively.

We first yield the geometric means of $A(w_k)$ and $B(w_k)$ in all the recording samples per the experiment date, and find the geometric mean of $H(w_k)$ by division.

3.4. Filtering the Recordings

The `invfreqz` function in Matlab gives the real feedback and feedforward coefficients from the transfer function. These coefficients feed into the filter for simulating the artificial bone-conducted sound with the recorded air-conducted sound.

4. RESULTS

We apply the whole signal processing procedure to each recording sample. The following figures are the spectra - the gray line - and the spectral envelopes – the solid line is the one by cepstral smoothing technique and the dash-dot line is the one by lpc technique - of one example, in which a pair of recorded singing voice has a chirp pattern.

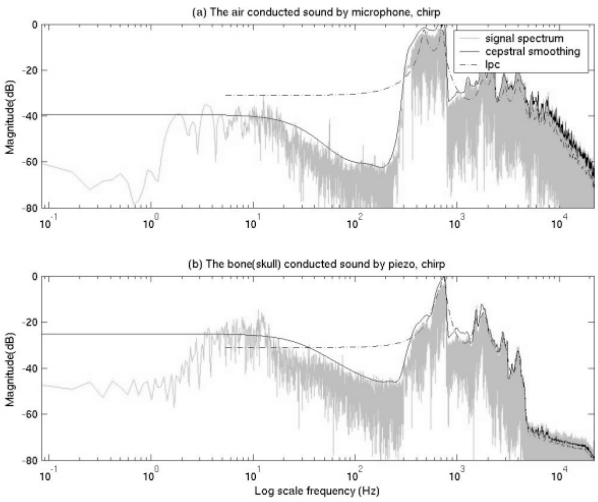

Figure 1. (a) The air-conducted sound by the microphone. (b) The bone-conducted sound by the piezo contact microphone.

For comparing the spectral differences according to the transmission medium, we plot the two spectral envelopes of the air-conducted sound and the bone-conducted sound.

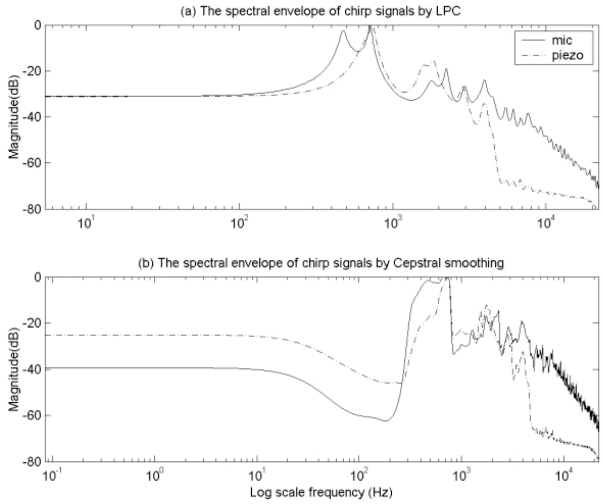

Figure 2. (a) The spectral envelope of a chirp signal by the LPC technique. (The solid line is the spectrum of the air-conducted sound and the dash-dot line is the one of the bone-conducted sound.) (b) The spectral envelope of a chirp signal by the Cepstral Smoothing technique.

As you can see, the spectrum of the bone-conducted sound (the dash-dot line) has larger amplitude below 3000Hz and smaller amplitude above 3000Hz to the air-conducted sound (the solid line). This tendency is distinct in the figure 3 and the figure 4, showing the average spectral envelopes and the estimated transfer functions using the average values.

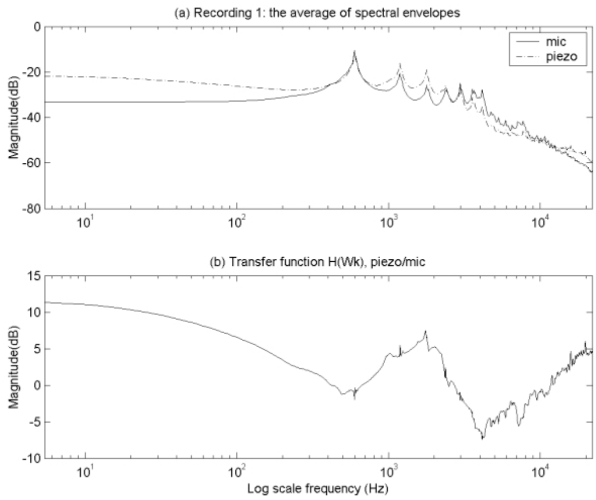

Figure 3. (a) The average spectrum of Recording 1 (b) The transfer function from the air-conducted sound to the bone conducted sound of Recording 1.

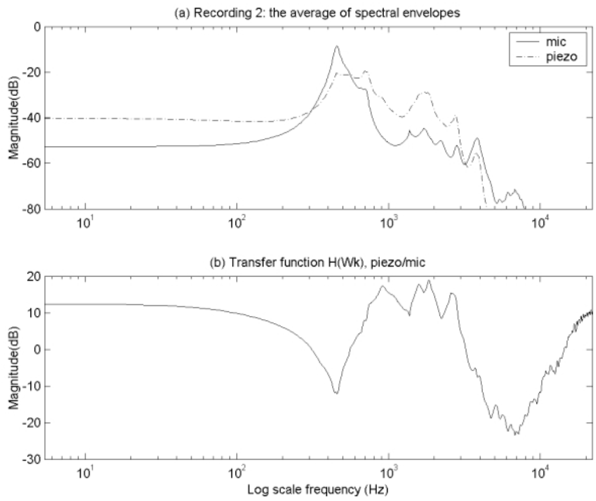

Figure 4. (a) The average spectrum of Recording 2 (b) The transfer function from the air-conducted sound to the bone conducted sound of Recording 2.

As evident on the figure 5, the overall shape of each transfer function is similar; having set 3000Hz as the threshold, an adjustment is made for the low and the high frequency regions, such that below 3000Hz, the magnitude is boosted, and above it the magnitude is cut. However the gradient of the curve is different; the slope of the dash-dot line is much steeper, where the peak goes up to 19dB and the trough goes down to -23dB. We presume this result is caused by the fact that the piezo contact microphone used for Recording 2 has a soundproofing cover, thereby cutting the air transmission effectively.

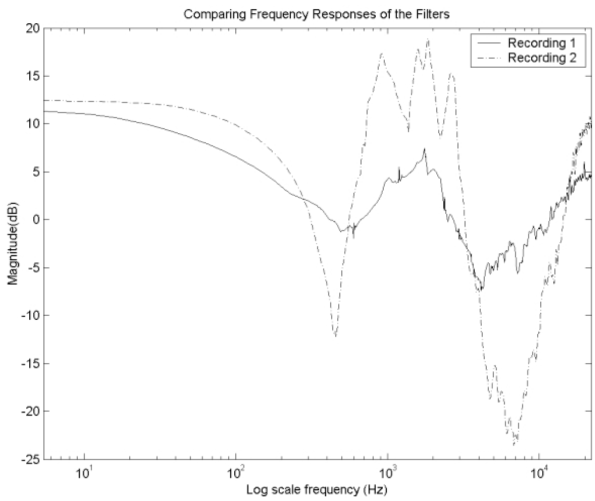

Figure 5. The transfer functions from the two different experiments: Recording 1 (indicated by the solid line) and Recording 2 (indicated by the dash-dot line).

The last stage is filtering the original recording by air transmission and comparing the two filtered sounds to the original recording by bone transmission. The figures 6 (a) and (b) show the original recordings by the microphone and the piezo. The figures 6 (c) and (d) show the two filtered sounds simulating the bone-conducted sound.

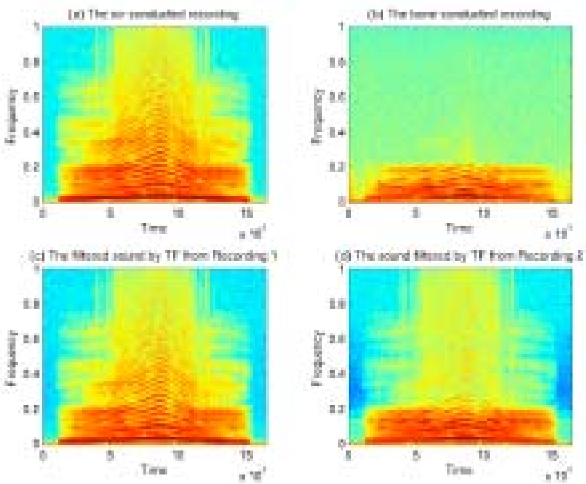

Figure 6. (a) the air-conducted sound, singing a chirp pattern (b) the bone-conducted sound, singing a chirp pattern (c) the filtered sound by the transfer function of Recording 1 (d) the filtered sound by the one of Recording 2. -> error!

From the figures 6 (c) and (d), we find that (d), with a steeper transfer function, has a more similar spectrum to the one in (b). This is also testable by the perception through ears.

5. CONCLUSIONS

In this paper, we find that the human skull boosts low frequencies and cuts high frequencies when propagating singing voice, with 3000Hz set as the threshold (graphing the magnitude vs. the frequency, we obtain a graph that is similar to a sine graph). These characteristics are confirmed by two consecutive experiments with various types of recording samples, the findings of which are in correspondence to the common sense that solid materials cut high frequencies because of the damping of the materials. They also explain the tendency for people to prefer the singing timber of the internal hearings more than the one from the recordings since the boost in low frequencies can conceal defects of singing techniques such as the inconsistence of pitch and irregular vibration, and make the singing timber rich.

However, this result must be regarded as a case study since the subject consists of only one female singer. Thus we plan to perform the same experiments with more subjects, in order to make these findings more concrete and hopefully obtain enhanced results.

This paper is the precursor of the main research, which aims to simulate the internal hearing. As the next step, we plan to examine the perception test for finding the internal hearing ratio of the air-conducted sound to the bone-conducted sound, and then implement the application that generates the internal hearing sound from the microphone input by air transmission in STK [12].

6. ACKNOWLEDGMENTS

We would like to thank professor Julius Smith for providing all the knowledge about digital signal processing and the advice on research procedure. We also would like to thank professor Chris Chafe for his help in finding the method to pick up skull vibrations.

7. REFERENCES

[1] Purcell, D. W., Kunov, H., and Cleghorn, W., " Estimating bone conduction transfer functions using oto-acoustic emissions", *Journal of the Acoustical Society of America,* 2003.

[2] Stenfelt, S., Hakansson, B., and Tjellstrom, A.. "Vibration characteristics of bone conduction sound in vitro", *Journal of the Acoustical Society of America,* 2000.

[3] Hakansson, B., Brandt, A., Carlsson, P., and Tjellstrom, A., "Resonance Frequency of the human skull in vivo", *Journal of the Acoustical Society of America,* 1994

[4] Wever, E. G., *Theory of Hearing*, John Wiley & Sons, Ic., New York, 1961.

[5] Smith, J. O., *Mathematics of the Discrete Fourier Transform (DFT)*, W3K Publishing, 2003.

[6] Smith, J. O., *Music 420 Course Reader: Spectral Audio Signal Processing*, CCRMA - Stanford University, 2004.

[7] Makhoul, J., "Linear Prediction: A Tutorial Review", *in Proceedings of the 1975 IEEE*, 1975

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

[9] Rabiner, L. and Schafer, R. W., *Digital Processing of Speech Signals*, Prentice-Hall, New Jersey, 1978

[10] Oppenheim, A. V. and Shafer, R. W., *Digital Signal Processing,* Prentice-Hall, New Jersey, 1975

[11] Thibault, F., *High-level Control of Singing Voice Timbre Tansformations*, online draft, http://www.music.mcgill.ca/~thibault/Thesis, 2004.

[12] Cook, P. R. and Scavone, G. P., "The Synthesis ToolKit," *in Proceedings of the 1999 International Computer Music Conference*, Beijing. 1999.