A study on an MTF-based blind restoration method for bone-conducted speech

Kota Kinugasa (710021)

School of Information Science,
Japan Advanced Institute of Science and Technology

February 5, 2009

Keywords: Modulation Transfer Function, speech intelligibility, bone-conducted speech, filterbank, blind restoration method.

It is very difficult for us to accomplish speech communications in extremely noisy environments such as factory and disaster sites. One of the solutions is to record the speech with a bone-conducted microphone because bone-conducted (BC) speech can be recorded by this microphone without interference of external noises. However, sound quality and intelligibility of BC speech are very poor and low [1]. Therefore improving these losses of BC speech is needed for speech communications using BC speech and it is a challenging topic.

Generaly, the attenuation of BC speech is stronger than that of air-conducted (AC) speech at higher frequencies. A straightforward method of restoring BC speech is to compensate these attenuated frequency components by using high-pass filtering. Since these attenuations varied in complex manner depending upon BC pickup-points, speakers, and pronounced syllables, it is very difficult to design one unique type of high-pass filtering with these variations. There are various methods of deriving inverse filtering such as cross-spectrum [2] and long-term Fourier transform methods [3], however, these yield restored speech signals with an artifact such as echoes. Therefore, there is less improvement in voice quality. Furthermore, information on AC speech is needed to construct the inverse filtering in these models [2] [3].

On the other hand, considering relationship between AC and BC speech signals as the transfer function, we had been studied on a common strategy based on the source-filter model for improving intelligibility and sound quality of BC speech. As the results, we found that the filter characteristics are more important than the source characteristics for improving sound quality and intelligibility of BC speech, based on source-filter model. Vu *et al.* thus proposed the LP-based blind BC speech restoration method that originates from the idea of the source-filter model in frequency-domain [4]. This method can restore the BC speech blindly, but this method required learning AC-LP coefficients. Kimura *et al.* proposed BC speech restoration method based on modulation transfer function (MTF) concept in time-domain [5]. This method compensates the reduced modulation index of each temporal power envelope in the filterbank model. Because of MTF relating to speech intelligibility, this method can improve the intelligibility of BC speech directly. However, information on AC speech is needed for restoring the BC speech in their method.

Since it is very important to improve the loss of speech intelligibility as well as voice quality for speech communications, we focused on the MTF concept and the aim of this study is to propose an MTF-based method for blindly restoring BC speech.

In this paper, we model the MTF by fitting three low-pass models to the MTF derived from the database. Then we assumed that the MTF relation $e_h^2(t)$ can be represented as follows:

$$e_h^2(t) = a^2 \exp(-2bt) \tag{1}$$

This model has two parameters, a and b. The parameter a can control the gain. The parameter b can control attenuation characteristic of the MTF filter. In previous work, the method for estimating b has been already proposed by Unoki and Hiramatsu[6]. In this study, we analyzed the characteristics between all of AC and BC power envelopes in the AC/BC speech database. From these results, all as in the channels can be approximated bt regression curve as $a_n = -cn^{-1} + d$, where c and d are parameters depend on measurement point, and n is Number of channel of filterbank. Since all as in the channels can be determined from these results and all bs, an MTF-based method for blindly restoring BC speech was realized.

We carried out simulations to evaluate the proposed method using the AC/BC database. The correlation and SNR of the power envelopes of AC and the restored speech signals or the power envelopes of AC and BC speech signals were used to evaluate the improvement of restoration of the power envelopes. Log spectral distortion (LSD), linear predictive LSD (LP-LSD), cepstrum distance (CD), and MER cepstrum distance (MCD) were used to evaluate the improvements of speech quality. Slope of regression line of MTF, RMS of MTF, and intelligibility weighted LSD[7] were used to evaluate the improvements of speech intelligibility. The results of evaluations suggested that the proposed method is effective for blind BC speech restoration.

References

[1] Kitamori, S., and Takizawa, M., "An analysis of bone conducted speech signal by articulation tests," *IEICE Trans.*, vol. J72-A, no. 11, pp. 1764-1771, Nov. 1989.

[2] Ishimitsu, S., Kitakaze, H., Tsuchibushi, Y., Yanagawa, H., and Fukushima, M., "A noise-speech recognition system making use of body-conducted signals," *Acoust. Sci. &, Tech.*, vol. 25, no. 2, pp. 166-169, 2004.

[3] Tamiya, T., and Shimamura, T., "Reconstruction filter design for bone-conducted speech," *Proc. ICSLP2004*, vol. II, pp. 1085-1088, Oct. 2004.

[4] Vu, T. T., Seide, G., Unoki, M., and Akagi, M., "Method of LP-based blind restoration for improving intelligibility of bone-conducted speech," *Proc. Interspeech2007*, pp. 966-969, Antwerp, Belgium, Aug. 2007.

[5] Kimura, K., Unoki, M., and Akagi, M., "A study on a bone-conducted speech restoration method with the modulation filterbank," pp. 411-414, *NCSP05*, Hawaii, USA, Mar. 2005.

[6] Hiramatsu, S., and Unoki, M. "A Study on the Blind Estimation of Reverberation Time in Room Acoustics," *J. Signal Processing*, **12**(4), 323-326. July 2008.

[7] ANSI S3.5-1997, "American National Standard Methods for Calculation of the Speech Intelligibility Index," 1997.