2009 International Workshop on Nonlinear Circuits and Signal Processing
NCSP'09, Waikiki, Hawaii, March 1-3, 2009

An MTF-based blind restoration method for improving intelligibility of bone-conducted speech

Kota Kinugasa, Masashi Unoki, and Masato Akagi

School of Information Science, Japan Advanced Institute of Science and Technology
1–1 Asahidai, Tatsunokuchi, Nomi, Ishikawa, 923–1292 Japan
Phone/Fax: +81–761–51–1699 ex, 1391/+81–761–51–1149
Email: {k-kota, unoki, akagi}@jaist.ac.jp

Abstract

Bone-conducted (BC) speech is useful for speech communications in extremely noisy environments. However, sound quality and intelligibility of BC speech are very poor and low. In this paper, we propose a blind restoration method for improving BC speech based on the concept of modulation transfer function (MTF). We first investigate the relationship between the power envelope of air-conducted (AC) and BC speech signals by analyzing AC/BC database. We then model these relations by fitting some models to the MTF derived from the database. The best-fitted model has two parameters. One is the gain factor, and another is the attenuation factor. We then propose a method for determining the parameters of model without the AC speech. We finally evaluate the proposed method using SNR, correlation, RMS of MTF, PESQ, and LSD. As the results, we showed the proposed method is effective for blind BC speech restoration.

1. Introduction

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 use the speech with a bone conduction microphone because 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, compensating these losses of BC speech is needed for speech communications using BC speech and it is a challenging topic.

Generally, the power attenuation of BC speech is stronger than that of AC speech at higher frequencies. A straightforward method for restoring BC speech is to compensate these attenuated frequency components by using high-pass filtering. Since these attenuations vary in a complex manner depending on 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 the restored speech signals with artifacts, i.e., echoes. Therefore, there is less improvement in voice quality. Furthermore, the AC speech is needed to design the inverse filtering in these models [2, 3].

On the other hand, by considering the relationship between AC and BC speech signals as the transfer function, we have been studied a common strategy based on the source-filter model for improving the 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 the source-filter model [4, 5, 6]. Vu *et al.* thus proposed the LP-based blind BC speech restoration method that originates from the idea of the source-filter model in the frequency domain [4]. Although this can restore the BC speech blindly, this method required learning AC-LP (AC-LSF) coefficients. Kimura *et al.* proposed a BC speech restoration method based on MTF concept in the 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, AC speech is needed for restoring the BC speech in their method. Moreover, their method was exactly effective to restore the BC speech, but it is debatable to determine what type of model of MTF is the most useful to restore the BC speech.

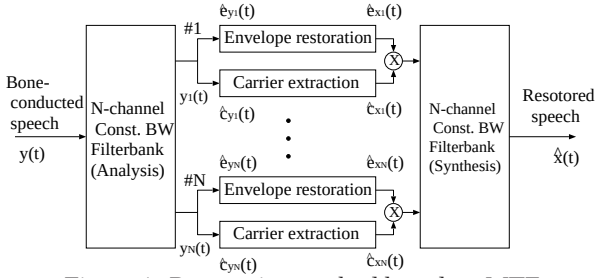

Figure 1: Restoration method based on MTF.

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

In this paper, first, we investigate the relationship between power envelopes of AC and BC speech. Secondly, the MTF derived from the database is modeled by fitting. Thirdly, we propose the method determining the parameters of the model without the AC speech. Finally, we propose an MTF-based method for blindly restoring BC speech.

2. Concept of MTF-based restoration method

The MTF concept was proposed by Houtgast *et al.* [7] to predict the speech intelligibility. Drullman revealed [8] that a temporal envelope information appears to be more important for speech intelligibility than carrier information. We think

that differences between AC and BC temporal envelopes affect speech intelligibility and sound quality significantly.

The BC speech restoration method in the filterbank is shown in Fig. 1. AC and BC speech signals are decomposed into the temporal envelopes, $e_x(t)$ and $e_y(t)$, and the carriers, $c_x(t)$ and $c_y(t)$, by N-channel bandpass filterbanks with a constant bandwidth (40-Hz). Here, we assume that the signals,$x(t)$ and $y(t)$, can be represented as

$$x(t) := \sum_{n=1}^{N} x_n(t) = \sum_{n=1}^{N} e_{x_n}(t) \cdot c_{x_n}(t), \quad (1)$$

$$y(t) := \sum_{n=1}^{N} y_n(t) = \sum_{n=1}^{N} e_{y_n}(t) \cdot c_{y_n}(t), \quad (2)$$

where, $x_n(t)$ and $y_n(t)$, are the bandpass signals. Power envelope and carrier of the BC speech are calculated as

$$e^2_{y_n}(t) = \mathrm{LPF}[|y_n(t) + j\mathrm{Hilbert}(y_n(t))|^2], \quad (3)$$

$$c_{y_n}(t) = y_n(t)/e_{y_n}(t), \quad (4)$$

where Hilbert($\cdot$) is the Hilbert transform and LPF[$\cdot$] denotes the low-pass filtering with a 20-Hz cut-off frequency. $e_x(t)$ and $c_x(t)$ can also be calculated from $x(t)$ usnig the same method. Then, inverse filtering, $E_h^{-1}(z)$ is used to restore the BC speech as follows

$$E_h^{-1}(z) = E_x(z)/E_y(z), \quad (5)$$

where $E_h(z)$, $E_x(z)$, and $E_y(z)$ are respectively the z-transform of $e^2_h(t)$, $e^2_x(t)$, and $e^2_y(t)$. The impulse response between AC and BC power envelopes, $e_h(t)$, was defined in previous paper [5] as $e_h(t) = a\exp(-bt)$, where a is the gain factor and b is the factor to control attenuation. However, it was not sure whether an MTF model as $a\exp(-bt)$ is the best model or not to restore the BC speech.

3. Characteristics of bone conduction

A question we have is how to design the inverse filter to restore the BC speech. We analyze the characteristics of bone conduction to design the inverse filtering, $E_h^{-1}(z)$.

3.1. AC/BC speech database

We used the AC/BC speech database [6] to analyze the characteristics between AC and BC power envelopes. BC speech was collected at five measurement points (1: mandibular angle, 2: temple, 3: philtrum, 4: forehead, and 5: calvaria). Different microphones were used at points from 1 to 4 and at the point 5. 25 Japanese words of each degree of familiarity were chosen from NTT-database [9]. Speakers were 5 males and 5 females.

3.2. Analysis of Characteristics of bone conduction

We analyzed the characteristics between AC and BC power envelopes with the AC/BC speech database using the following measures: (1) correlation of AC and BC power envelopes, (2) SNR (S: $e^2_x(t)$, N: $e^2_x(t) - e^2_y(t)$, (3) MTF $\left(M(\omega) = \left|\int_0^\infty e^2_h(t)\exp(-j\omega t)dt \Big/ \int_0^\infty e^2_h(t)dt\right|\right)$, (4) transfer function ($|\mathcal{F}[y(t)]/\mathcal{F}[x(t)]|$), and (5) power ratio between AC and BC power envelopes. Here $\mathcal{F}[\cdot]$ is the long-term Fourier transform. Kimura *et al.* had been analyzed the characteristics for the measurement at point 5 (calvaria). We analyze the characteristics for the measurements at all points, and we carefully investigate the characteristics of bone conduction.

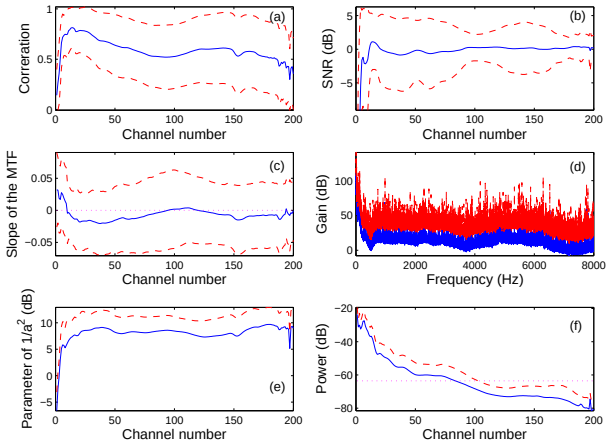

Figure 2: Analysis results of all dataset (solid line: mean, dashed line: mean ± standard deviation). (a) correlation, (b) SNR, (c) slope of MTF, (d) transfer function, (e) mean of power ratio of power envelope (parameter $1/a$ of model) and regression curve, and (f) mean of $e^2_y(t)$ of each channel.

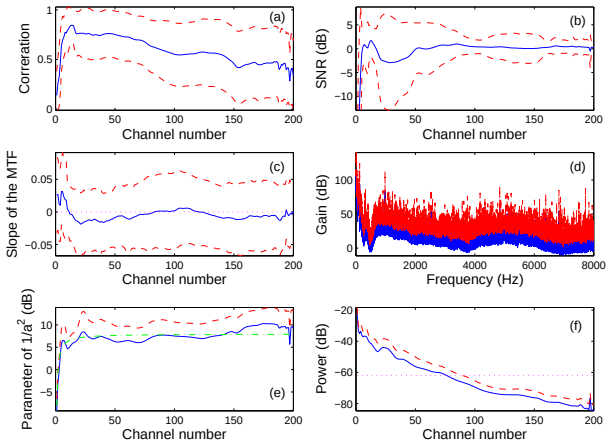

Figure 3: Analysis results at measurement point 2. Format is the same that in Fig. 2.

3.3. Results and discussion

Figure 2 shows the results of all dataset and Figure 3 shows the results at measurement point 2. Solid lines show the mean and dashed lines show the mean ± standard deviation. Figures 2 (a) and (b) show the distortion of BC power envelope. Figure 2 (c) shows the slope of regression line of MTF (1 to 10 Hz). The reason to limit the range of MTF from 1 to 10-Hz is that MTF of at the higher modulation frequencies is influenced by the internal noise such as blood flow, and noise of transmission-line, and noise flooring. If the value of slope is negative, the MTF has low-pass characteristics and if the value of slope is positive, the MTF has high-pass characteristics. This result shows the MTF has low-pass characteristics

in most channels. Figure 2 (d) shows that the characteristic of bone conduction is low-pass. Figure 3 (e) shows the mean of power ratio of power envelope of BC signal to AC signal in each channel. We fitted some curves to the mean of power ratio and investigated that the power ratio can be approximated by regression curve as $-c\omega^{-1} + d$, where c and d are parameters depend on measurement point. These characteristics were also shown in the results at the other measurement points in our analysis. We can approximate the power ratio of BC power envelope to AC power envelope of each channel by changing the parameter of regression curve. Also, the analysis results show that the regression curve does not depend on the difference of speaker and difference of syllable so much.

3.4. Modeling the MTF between AC and BC power envelopes.

In the previous method [5], MTF was represented as an exponential model. However, it was not evident whether MTF can be represented by an exponential curve. We confirmed that the MTF has low-pass characteristics from the analysis result. Thus, we model the MTF by fitting three low-pass models (one of exponential curve $e_h(t) = at\exp(-bt)$, the model of MTF used in previous method $e_h(t) = a\exp(-bt)$, and $e_h(t)$ = low-pass filter) to the MTF derived from the database. Figure 4 shows a case of the MTF derived from the database and MTF that removed internal noise derived from database and three models. The shape of the MTF derived from the database seems to be rippled. Because MTF without internal noise did not fluctuate, influence of internal noise might ripple the shape of MTF. The model $a\exp(-bt)$ is especially suitable in this case. Figure 5 shows the mean and the standard deviation of the analysis results using $a\exp(-bt)$. Upper panel shows the slope of regression line of model that fitted MTF derived from the database. If the value of slope is zero, the MTF does not influence the BC speech. Lower panel shows the root mean squared (RMS) differences between model and that of MTF derived from database. Because the RMS difference for $a\exp(-bt)$ is the smallest in all models, we assumed that the MTF relation, $e_h(t)$, can be represented by the model as $e_h(t) = a\exp(-bt)$. Then, we can model the MTF by fitting models to the MTF derived from the database and finally determined inverse filtering as follow

$$E_h^{-1}(z) = \frac{1}{a^2}\left\{1 - \exp\left(-\frac{2b}{f_s}\right)\right\}, \qquad (6)$$

where f_s is the sampling frequency of 16 kHz.

4. MTF-based blind restoration method

The previous method [5] needs AC speech to determine the parameter of MTF model and to set restoring condition. From the analysis results, we improve the previous method from the following two standpoints.

4.1. Parameter determination

One of the stand point is how we determine the two parameters of model, a and b, for restoring the BC speech. For the method for determining parameter a, Figure 3 (e) shows the parameter a can be approximated by the regression curve and this curve depends only on the measurement point. We can

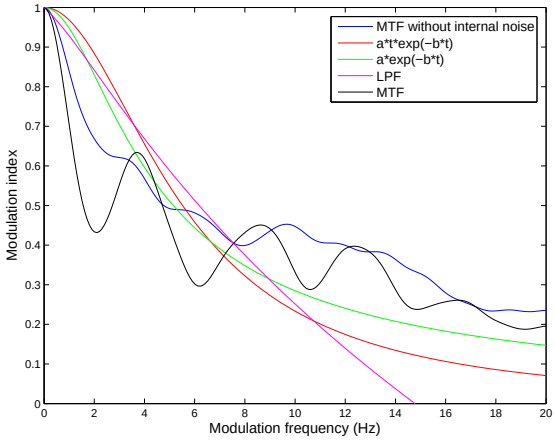

Figure 4: Comparison between MTFs derived from the database and model. MTF without internal noise: MTF that removed internal noise derived from database. $at\exp(-bt)$: one of exponential curve. $a\exp(-bt)$: the model of MTF used in previous method. MTF: MTF derived from database. LPF: low-pass filter.

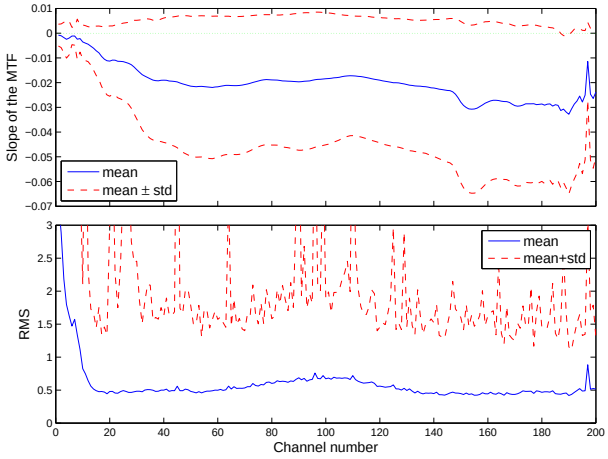

Figure 5: Analysis results of MTF by fitting model. Upper: Slope of regression line of model that fitted MTF derived from database. Lower: RMS of difference between model and that fitted MTF derived from database.

determine the parameter a without AC speech to study each measurement of the regression curve.

To estimate the parameter b, we used the method proposed by Hiramatsu and Unoki [10] which was originally introduced for estimating reverberation time based on MTF concept. The utilization of this method is that our research and their work are based on exactly the same model. The parameter b can be estimated as

$$\hat{b} = \arg\min_b \left(\left|\hat{E}_y(0)\cdot|M(f_{dm},b)|/\hat{E}_y(f_{dm})\right|\right), \qquad (7)$$

where f_{dm} is dominant frequency of BC power envelope and $\hat{E}_y(\cdot)$ is power envelope of the restored speech.

4.2. Modification in recovering condition

Another stand point is that the previous method restores the BC speech when the correlation between AC and BC power

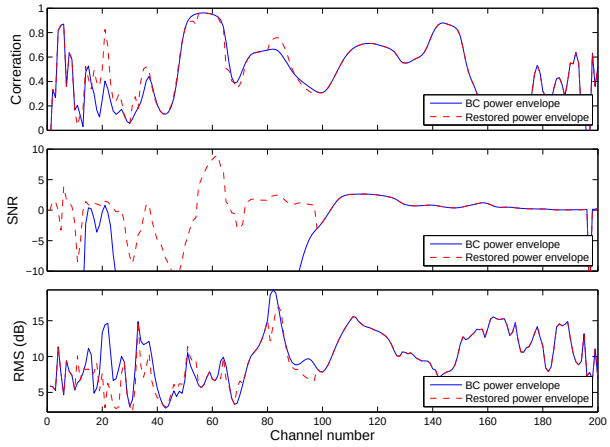

Figure 6: Improving of correlation, SNR and RMS of MTF. Solid line: BC speech. Dashed line: restored speech by proposed method.

envelopes is not over 0.8 and the relative power of AC power envelope is over −20 dB. To restore the BC speech without AC speech, we change these conditions to an appropriately condition that the method restores the BC speech when the relative power of BC power envelope is over −40 dB. The reason we set such a condition is that the internal noise in the BC speech appears when the relative power of BC power envelope is not over −40 dB. Because internal noise increased the power of DC of MTF, the method cannot estimate the true value of parameter b.

Based on the above results, an MTF-based method for blindly restoring BC speech was proposed.

5. Evaluations

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. Perceptual evaluation of sound quality (PESQ) [11] that was recommended by ITU-T P. 862, and log spectral distortion (LSD) were used to evaluate the improvements of speech quality. RMS of MTF was used to evaluate the improvement of intelligibility of the restored speech. RMS of MTF means RMS difference between MTF that is attenuation and MTF that is not attenuation. If the all modulation frequency of MTF is 0 dB, the MTF does not influence the BC speech. Therefore, the closer to 0 the value of RMS of MTF, the better the improvement of MTF. Figure 6 shows an example of the SNR, correlation, and RMS of MTF. The solid lines show the result of the BC speech and the dashed lines show the result of the restored speech by the proposed method using parameter a derived from regression curve. These results showed that the power envelopes can be adequately restored by the proposed method and RMS of MTF showed the improvement of intelligibility. LSD value has been decreased by 1.5 dB and PESQ value has been increased by 0.7 points. These results showed that the sound quality can be improved by the proposed method. The results of evaluations showed that the proposed method is effective for blind BC speech restoration.

6. Conclusion

We analyzed the characteristics between the AC and BC power envelopes with the AC/BC speech database. As the results, we found that the characteristic of MTF is the low-pass filtering and each measurement point of power ratio can be approximated by regression curve. We then modeled the MTF as $a\exp(-bt)$ and proposed the methods to determine the parameters of model without the AC speech. As the results, we proposed the MTF-based blind restoration method. Finally, we evaluated the proposed method and showed that the proposed method was effective for blind BC speech restoration.

As the next step in our research, we will carry out the comprehensive evaluation and investigate the performance of the proposed method.

Acknowledgment

This work was supported by a Grant Program by the YAZAKI Memorial Foundation for Science and Technology It was also partially supported by the Strategic Information and COmmunications R&D Promotion ProgrammE (SCOPE) (071705001) of the Ministry of Internal Affairs and Communications (MIC), Japan.

References

[1] Kitamori, S., and Takizawa, M., “An analysis of bone conducted speech signal by articulation tests,” *IEICE Trans.*, **J72-A**(11), 1764-1771, Nov. 1989.

[2] Ishimitsu, S., Kitakaze, H., Tsuchibushi, Y., Yanagawa, H., and Fukushima, M., “A noise-robust speech recognition system making use of body-conducted signals,” *Acoust. Sci. &, Tech.*, **25**(2), 166-169, Mar. 2004.

[3] Tamiya, T., and Shimamura, T., “Reconstruction filter design for bone-conducted speech,” *Proc. ICSLP2004*, **II**, 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*, 966-969, Belgium, Aug. 2007.

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

[6] Vu, T. T., Kimura, K., Unoki, M., and Akagi, M., “A study on restoration of bone-conducted speech with MTF-based and LP-based models,” *J. Signal Processing*, **10**(6), 407-417, Nov. 2006.

[7] Houtgast, T., and Steenken, H. J. M., “The Modulation Transfer Function in Room Acoustics as a Predictor of Speech Intelligibility,” ACUSTICA. **54**, 557, Aug. 1973.

[8] Drullman, M. “Temporal envelope and fine structure cues for speech intelligibility,” *J. Acoust. Soc. Am.*, **97**, 585-592, Jan. 1995.

[9] Database for speech intelligibility testing using Japanese word lists. NTT–AT, Mar. 2003.

[10] 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.

[11] Yi, H., and Philipos, C. L., “Evaluation of objective measures for speech enhancement,” *Interspeech2006*, 1447-1450, Pittsburgh, Pennsylvania, Sept. 2006.