Combination of Bone-Conducted Speech with Air-Conducted Speech Changing Cut-Off Frequency

Tetsuya Shimamura and Fumiya Kato
Graduate School of Science and Engineering
Saitama University
255 Shimo-Okubo, Sakura-Ku, Saitama, 338-8570
JAPAN
shima@sie.ics.saitama-u.ac.jp http://www.sie.ics.saitama-u.ac.jp

Abstract: - In this paper, we combine bone-conducted speech with air-conducted speech to improve the quality of speech in noisy environments. The bone-conducted speech is low-pass filtered, and the air-conducted speech is high-pass filtered by changing the cut-off frequency in a commonly used transition band. Through experiments, the performance of the combination filter is investigated.

Key–Words: - Bone-conducted speech, air-conducted speech, cut-off frequency, combination filter

1 Introduction

For a variety of fields of speech applications, the quality of speech degrades due to additive noise. Once speech and noise are added, however, it is not easy to separate them from a mixture of both. While many approaches exist to reduce additive noise in noisy speech [1][2], we do not have a promising one to provide the resulting speech with sufficient quality.

Recently, a great deal of attention has been paid to picking up the vibrations of bone in noisy environments [3][4]. The transmission of voice on bones is referred to as bone conduction [5]. The voice waveforms transmitted from the voice source (vocal cord) through the vocal tract wall and skull do not confront directly with noise. This is the reason why the bone-conducted speech is effectively utilized in noisy environments. However, the intelligibility of the bone-conducted speech is low. Even in a noiseless environment, the intelligibility of the bone-conducted speech is lower than that of the corresponding air-conducted speech, which is the speech transmitted through air. This is caused by the fact that the high frequency components of bone-conducted speech are attenuated due to impedance mismatching between bones and skins.

In [4], a reconstruction filter to emphasize the high frequency components of bone-conducted speech is designed to obtain quality improvement. In [4], to design the reconstruction filter, the air-conducted speech is utilized with the bone-conducted speech. However, both are unnecessary to obtain from an air-conducted microphone and a bone-conducted microphone simultaneously. This means that a common sentence of pronunciation is not required for both the air-conducted speech and bone-conducted speech.

In this paper, we set out to design a digital filter to combine the bone-conducted speech with the air-conducted speech, where a common sentence is required simultaneously. Two digital filters; one is a low-pass filter and the other is a high-pass filter, are used and each output is added to produce a speech signal with improved quality. The bone-conducted speech is used as input to the low-pass filter, and the air-conducted speech is done to the high-pass filter. Both digital filters have the common cut-off frequency in each transition band. Experiments are conducted and the cut-off frequency dependency on the performance of the combination filter is investigated.

2 Bone-Conducted Speech and Air-Conducted Speech

In this section, bone-conducted speech and air-conducted speech are compared. Figure 1 shows a waveform comparison of both speech signals. The pronunciation is commonly /a/ /i/ /u/ /e/ /o/ of vowels. Both waveforms have been measured from a speaker using a bone-conducted microphone and an air-conducted (normal) microphone simultaneously. It is observed that the relative amplitude of the bone-conducted speech is different from that of the air-conducted speech.

Each vowel part was detected and the power spectrum of that was calculated. Figures 2-6 show a comparison of the power spectrum of the bone-conducted and the corresponding air-conducted speech vowels,
respectively. We see that the high-frequency components of the bone-conducted vowels are attenuated commonly.

![Figure 1: Waveforms of bone-conducted and air-conducted speech signals](image1)

![Figure 2: Power spectrum of /a/](image2)

![Figure 3: Power spectrum of /i/](image3)

![Figure 4: Power spectrum of /u/](image4)

![Figure 5: Power spectrum of /e/](image5)

3 Combination Filter Design

Figure 7 shows a block diagram to implement the combination filter to be proposed in this paper.

We set out to combine the bone-conducted speech with the air-conducted speech simultaneously. As shown in Fig.7, two digital filters; one is a low-pass filter and the other is a high-pass filter, are used and each output is added to produce a speech signal with improved quality. Both filters are constructed by an
infinite impulse response (IIR) filter in practice. More specifically, a Butterworth filter with 4-th orders is designed based on setting the cut-off frequency in the transition band. One commonly used cut-off frequency is set to design the low-pass and high-pass filters. The bone-conducted speech and air-conducted speech are inputted to the designed low-pass filter and high-pass filter, respectively.

Figure 8 shows an example of the spectrogram of the resulting speech signal. It is observed that the high frequency components of the output from the combination filter are compensated adequately.

4 Experiments

In this section, the performance of the combination filter is investigated through experiments.

4.1 Speech Data

Speech data used in the experiments are a set of bone-conducted speech and air-conducted speech of Japanese two males and two females, which are sampled by a sampling frequency of 8 kHz. Both the bone-conducted speech and air-conducted speech data were pronounced from a common speaker and recorded simultaneously. We used Temco HG-17 and SHURE SM58 as the bone-conducted and air-conducted microphones, respectively.

Noises used are a car noise, an exhibition noise and a factory noise, which are also sampled by a sampling frequency of 8 kHz. At first, only a white noise was generated from a loud speaker in a sound isolated room and two kinds of noise data were measured through a bon-conducted microphone and an air-conducted (normal) microphone, respectively. The ratio of the bone-conducted noise to the air-conducted noise was then calculated. According to the calculated ratio, noises were added to the bone-conducted speech and air-conducted speech data, and a set of noisy bone-conducted speech and noisy air-conducted speech data was prepared. Table 1 summarizes the speech data used for the experiments.

<table>
<thead>
<tr>
<th>Speech and Noise</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Speakers</td>
<td>two males and females</td>
</tr>
<tr>
<td>Speech</td>
<td>Japanese sentence (10 sec)</td>
</tr>
<tr>
<td>Sampling frequency</td>
<td>8kHz</td>
</tr>
<tr>
<td>Noise</td>
<td>car, exhibition and factory</td>
</tr>
<tr>
<td>SNR</td>
<td>0dB,10dB</td>
</tr>
</tbody>
</table>
4.2 Preliminary Experiment

As a preliminary experiment, the output of the proposed combination filter was investigated by changing the cut-off frequency used in filter design. The weighted spectral slope measure (WSSM) \[6\] was used for speech quality assessment of the resulting speech.

Filter banks are used to evaluate the WSSM. For \(i\)-th frame, the WSSM is obtained as
\[
WSSM(i) = M_{spl}(M - \hat{M}) + \sum_{m=1}^{25} w_a(m)(P(m) - \hat{P}(m))^2
\]
(1)

where \(M, \hat{M}\) and \(M_{spl}\) are parameters to evaluate the WSSM and \(w_a(m)\) corresponds to the weighted coefficients for each filter bank. Slope of the power spectrum of the original speech, \(P(m)\), are compared with that of the processed speech, \(\hat{P}(m)\), at \(m\)-th band. For the parameters and coefficients of the WSSM, the values shown in \[7\] were set for the experiments. The final WSSM was evaluated by averaging the WSSM for several frames as
\[
WSSM = \frac{1}{L} \sum_{l=1}^{L} WSSM(l),
\]
(2)

and the improvement in the resulting WSSM was used for assessment as
\[
WSSM_{imp} = WSSM_{input} - WSSM_{output}
\]
(3)

where \(WSSM_{input}\) and \(WSSM_{output}\) correspond to the WSSMs of the input speech and output speech, respectively.

Figure 9 shows the improvement in WSSM by changing the cut-off frequency. Figure 9 suggests that the best performance of the combination filter is obtained with different setting of the cut-off frequency. This is because the frequency characteristics of each noise are different. Figures 10-12 show power spectra of each noise data. In Figs.10-12, the power spectrum of the noise contaminated in the bone-conducted microphone is compared with that in the air-conducted microphone for 10 consecutive frames.

4.3 Comparison Experiment

To validate the effectiveness of the combination filter, the reconstruction filter derived by Tamiya et al. \[4\] was compared. Additionally, the iterative spectral subtraction method \[8\] was implemented as the post-processing of the reconstruction filter so that the noise remained by the reconstruction filter was further suppressed.

Figure 10: Spectra of car noise included in air-conducted and bone-conducted microphones

Figure 11: Spectra of exhibition noise included in air-conducted and bone-conducted microphones

Figure 12: Spectra of factory noise included in air-conducted and bone-conducted microphones
Figures 10-12 show the spectrogram of the output of the proposed combination filter in each noise for the case of signal-to-noise ratio of 0 dB, where the iterative spectral subtraction (iterative SS) [8], Tamiya method (tamiya) [4], Tamiya method followed by the iterative spectral subtraction (tamiya and iterative-SS) and the proposed combination filter (air-bone) are compared. The cut-off frequency used for the combination filter is commonly 1600 Hz. It is observed that the output of the proposed combination filter best approaches to the original clean speech by suppressing the noise as well as by keeping the harmonics of the speech.

5 Conclusions

In this paper, we have combined bone-conducted speech with air-conducted speech with the cut-off frequency changed to improve the quality of speech in noisy environments. The resulting combination filter provides better speech quality than the conventional reconstruction filter for bone-conducted speech.

References:

Figure 13: Spectrogram in car noise
Figure 14: Spectrogram in exhibition noise

Figure 15: Spectrogram in factory noise