HAMBURG 2016

A Study on Quality Improvement of Bone Conducted Speech by AM Method

Tomohiro MINAMI; Shigeaki AOKI[1]

Kanazawa Institute of Technology, Japan

ABSTRACT

As one of the methods for sending a speech in telecommunications under a high-noise environment, a bone conductive microphone is effective. Since the bone conductive microphone records speech vibrations transmitted in the skull, its anti-noise characteristic is superior to that of air conducted sound transmitted in the air. However, bone conducted speech has the drawbacks of reduced sound quality and articulation score. It is considered that these drawbacks are mainly caused by the loss of high-frequency components in transmission of the skull. Therefore, it is expected that the quality of bone conducted speech is improvable by appropriately generating these high-frequency components. A method of restoring the lost high-frequency components of bone conducted speech by utilizing an alias spectrum was proposed in our previous report. We conducted a listening test and confirmed the effectiveness of the proposed method. As the restored high-frequency components were generated by the alias spectrum, the structure of spectrogram was not natural. In this paper, a new method of restoring the lost high-frequency components of bone conducted speech by utilizing a duplicate that is shifted from low-frequency components to high-frequency components is proposed. Moreover, we conduct a listening test to confirm the effectiveness of the proposed method.

Keywords: Bone conducted speech, Speech quality, Background noise I-INCE Classification of Subjects Number(s): 71.1

1. INTRODUCTION

As one of the methods for sending a speech in telecommunications under a high-noise environment, a bone conductive microphone is effective. Since the bone conductive microphone records speech vibrations transmitted in the skull, its anti-noise characteristic is superior to that of air conducted sound transmitted in the air. However, bone conducted speech has the drawbacks of reduced sound quality and articulation score. It is considered that these drawbacks are mainly caused by the loss of high-frequency components in transmission of the skull. Therefore, it is expected that the quality of bone conducted speech is improvable by appropriately generating these high-frequency components.

A method of restoring the lost high-frequency components of bone conducted speech by utilizing an alias spectrum was proposed in our previous report [1-3]. The method is denoted as ASM (Alias Spectrum Method) in this paper. We conducted a listening test and confirmed the effectiveness of the proposed method. As the restored high-frequency components were generated by the alias spectrum, the structure of spectrogram was not natural. It seemed to be a reason of insufficient recovery of sound quality.

In this paper, a new method of restoring the lost high-frequency components of bone conducted speech by utilizing a duplicate that is shifted from low-frequency components to high-frequency components is proposed. Moreover, we conduct a listening test to confirm the effectiveness of the proposed method.

2. PROPOSED METHOD

2.1 Effective frequency region

A basic structure of speech is characterized by formants filtered through the prominent vocal tract, and the fundamental frequency and harmonic components that are generated by the vocal fold

[1] aoki_s@neptune.kanazawa-it.ac.jp

oscillation. The bone conducted speech loses the high-frequency components as shown in Figure 1. Nevertheless, it maintains the information of low-frequency components, for example, harmonic components. On the assumption that there is an analogy between the low-frequency and high-frequency components, a method that utilizes the effective low-frequency components as the lost high-frequency components is proposed.

As the proposed method utilizes the effective frequency components that include the enough individual speech characteristics, it is expected that it can simply and easily recover the lost high-frequency components and maintain naturalness and individuality of speech.

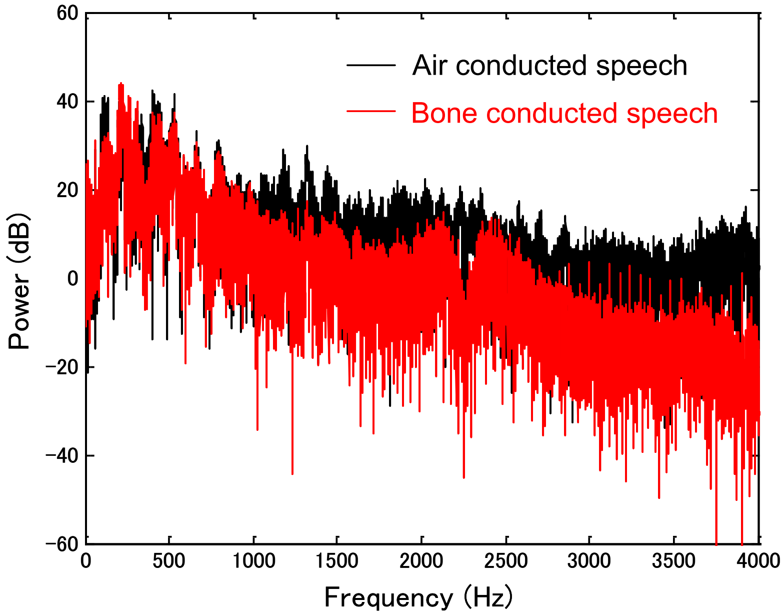


Figure 1 – Typical example of frequency characteristic of bone conducted speech and air conducted speech (male speech)

2.2 Application of a modulated processing

A system block diagram is shown in Figure 2. The effective low-frequency components are filtered from an original bone conducted speech using LPF. Simultaneously the low-frequency components that should be shifted to high frequency components are extracted using BPF. The method is denoted as MSM (Modulated Spectrum Method). The high-frequency components are generated from the extracted low-frequency components by upper single sideband amplitude modulation process shown in Figure 3.

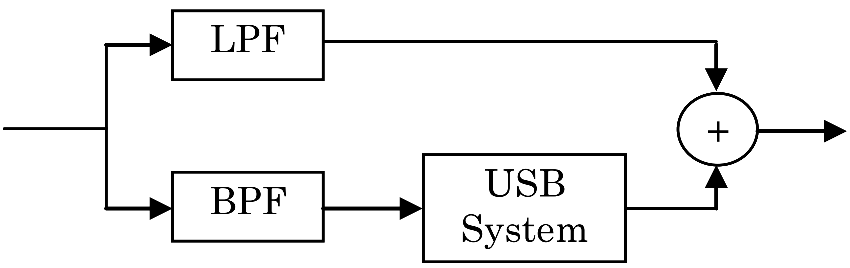


Figure 2 – System block diagram. UAB System is upper single sideband amplitude modulator

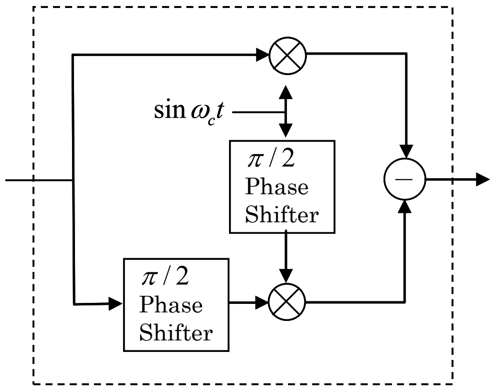
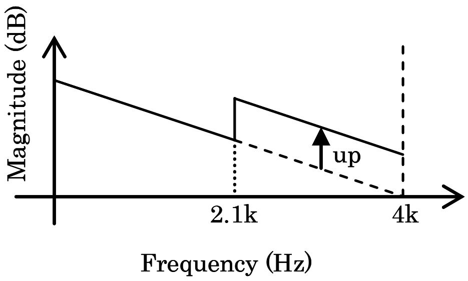
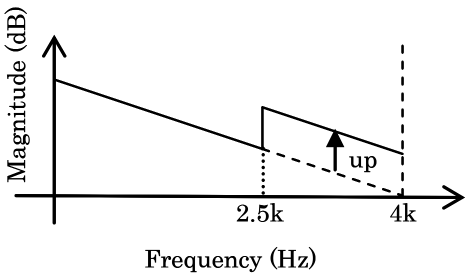


Figure 3 – Block diagram of UAB System

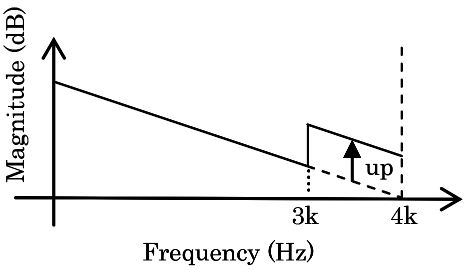
In applying the modulation process, a boundary frequency is defined as the boundary between the effective low-frequency region and the generated high-frequency region. The boundary frequency 2.1, 2.5 and 3.0 kHz are focused on. These schematic frequency characteristics are shown in Figure 4. In comparison between the characteristics of boundary frequencies of 2.1, 2.5 and 3 kHz, it is found that the frequency characteristics of the high-frequency region differ and the degree of emphasis of the high-frequency region can be adjusted.



(a) Boundary frequency 2.1 kHz



(b) Boundary frequency 2.5 kHz



(c) Boundary frequency 3.0 kHz

Figure 4 – Schematic frequency characteristics

For reference, the previous proposed method which applies the spectrum extrapolation method using the alias spectrum caused by re-sampling to the bone conducted speech is shown in Figure 5. The difference between the present and previous methods is clear in comparing with Figure 6.

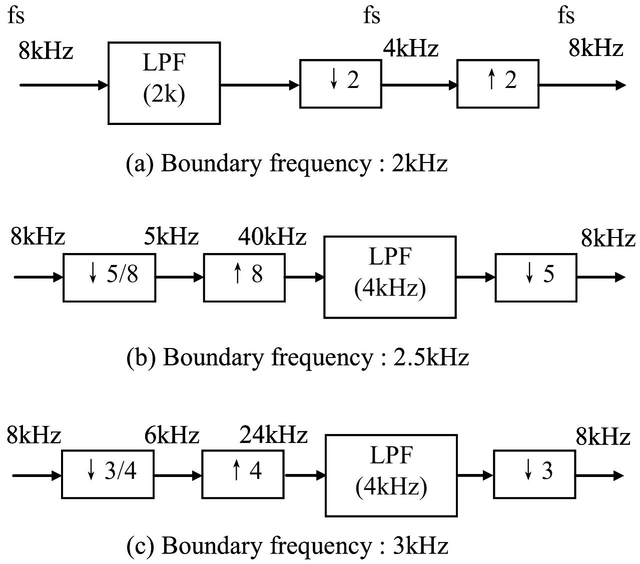


Figure 5 – Block diagrams of the previous proposed system configuration.

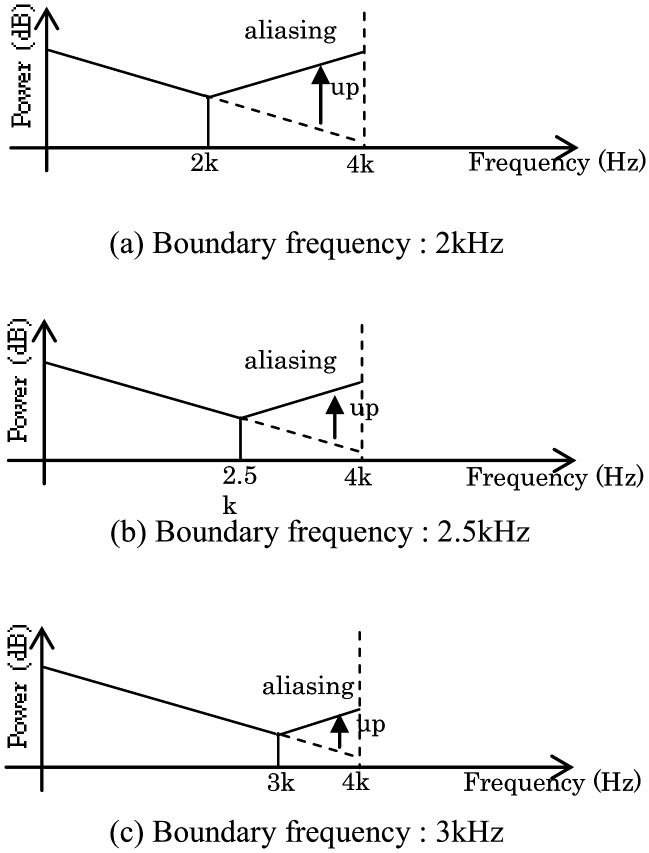


Figure 6 – Schematic frequency characteristics of the previous method

3. Paired Comparison Listening Test

3.1 Test Method

Listening tests using the method of paired comparisons are conducted on a total of four different signals. They are an original bone conducted speech and three different processed speeches. The first and second processes are ASM [1-3] and MSM. The last process is an ordinary filtered process using long-term spectra of original air conducted and bone conducted speech [4]. The last method is denoted LSM (Long-term Spectra Method). The boundary frequencies are unified into 2.5 kHz.

The speeches used for the tests are recordings of three texts read by five adult males that are recorded using the bone conductive microphone (G-450). The three texts are from phonetically balanced speech data distributed by NTT-AT.

Test signals are presented to the left ear with headphones (HAD-200). Five subjects answer by the question, "which sound has better quality?" In addition to the abovementioned processing, all test signals are passed through a band pass filter. The bandwidth is 0.2 - 3.8 kHz.

3.2 Test Results and Consideration

The results of paired comparisons are shown in Table 1, where A is the original bone conducted speech and B, C and D are the processed speeches by the ASM, MSM and LSM. The boundary frequencies of ASM, MSM are 2.5 kHz, respectively. The analyzed result with the Thurstone's paired comparisons is shown in Figure 7.

It is clear that the percentage of win rate of C was the highest and that of D was the lowest. The percentage of win rate of B was inferior to that of C. The reversal of frequency components at the boundary frequency had a serious influence on the sound quality. According to introspection reports, the low win rate of D was caused by back ground noise. It was also found that the effectiveness tends to be dependent of speakers.

Table 1 –Results of comparative judgments (win rate). A: original bone conducted speech, B: restored speech (previous proposed method ASM, boundary frequency: 2.5 kHz), C: restored speech (proposed method MSM, B. F.: 2.5 kHz), D: restored speech (ordinary Long-term Spectra method LSM)

	A	B	C	D
A		0.37	0.29	0.57
B	0.63		0.38	0.59
C	0.71	0.62		0.65
D	0.43	0.41	0.35	

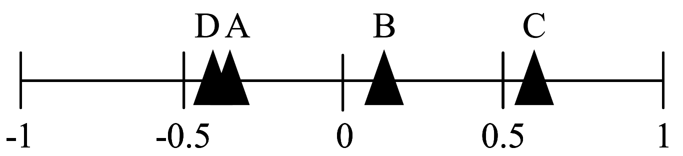


Figure 7 – Analyzed result with the Thurstone's paired comparisons. A, B, C and D are the same as in Table 1

4. SPECTRUM DISTORTION

In addition to the subjective tests, the spectral distortion (SD) is calculated to evaluate objectively the difference between the original bone conducted speech and three different processed speeches. The calculated result is shown in Figure 8. It was confirmed that each method improve similarly the SDs and the standard deviations of B and C are smaller than that of D.

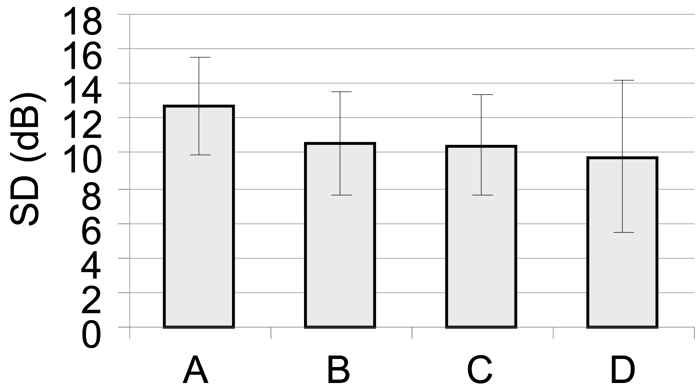


Figure 8 – Spectral distortion (SD). A, B, C and D are the same as in Table 1

5. CONCLUSIONS

In this paper we proposed a new method for improving sound quality of bone conducted speech by utilizing amplitude modulation. The subjective test was conducted in comparison with an original bone conducted speech and three different processed speeches, that is, the present proposed method (Modulated Spectrum Method), the previous proposed method (Alias Spectrum Method) and the ordinary filtered method (Long-term Spectra Method). The results of paired comparison listening test showed the best validity of sound quality improvement by the present proposed method. In addition to the subjective tests, the spectral distortion (SD) was calculated to evaluate objectively the difference between test signals. It was confirmed that each method improved similarly the SDs and the standard deviations of our present and previous methods were smaller than that of the ordinary filtered method.

REFERENCES

1. Yasukawa, H., Spectrum extrapolation of band limited signals using adaptive method. Technical Report of IEICE, CAS95-11, 1995
2. Minami T., Aoki S., "A Study on Quality Improvement of Bone Conducted Speech," INTER-NOISE 2012
3. Minami T, Nishida K, Ujike T, Aoki S, Study on quality improvement of bone-conductive voice Acoust. Sci. & Tech. 2012 ;33(6):376-378
4. Tamiya T. Shimamura T., Reconstruct Filter Design for Bone-Conducted Speech Proc. ICSLP2004 vol. II, 1085-1088