PITCH CHARACTERISTICS OF BONE CONDUCTED SPEECH

M. Shahidur Rahman and Tetsuya Shimamura

Graduate School of Science and Engineering, Saitama University
255 Shimo Okubo, Sakuraku, Saitama 338-8570, Japan
phone: + (81) 048-858-3496, fax: + (81) 048-858-3716, email: {rahmanms, shima}@sie.ics.saitama-u.ac.jp
web: www.sie.ics.saitama-u.ac.jp

ABSTRACT

This paper investigates the pitch characteristics of bone conducted speech. Pitch determination of speech signal can not attain the expected level of accuracy in adverse conditions. Bone conducted speech is robust to ambient noise and it has regular harmonic structure in the lower spectral region. These two properties make it very suitable for pitch tracking. Few works have been reported in the literature on bone conducted speech to facilitate detection and removal of unwanted signal from the simultaneously recorded air conducted speech. In this paper, we show that bone conducted speech can also be used for robust pitch determination even in highly noisy environment that can be very useful in many practical speech communication applications like speech enhancement, speech/ speaker recognition, and so on.

1. INTRODUCTION

Presence of additive noise makes speech processing a challenging problem. Speech enhancement is thus required in many practical applications of speech communication. Though the enhancement techniques are somewhat successful in reducing noise at moderately high SNR (speech to noise ratio) conditions, they are not effective at low SNR conditions. Again, handling non-stationary noises has been another difficult problem in speech enhancement. In non-stationary noisy environment, determining the state of the speaker (speaking or not) is one of the most difficult issue. Since the speech is itself non-stationary in nature, it is extremely difficult to detect and remove background noises from just one channel information of speech signal. To this end, utilization of bone conducted speech has been studied by many researchers [1, 2], where the bone conduction pathways have been used to record the talker's voices by placing a bone-conductive microphone on the talker's head. An important advantage of a bone-conductive microphone over an air boom microphone is that it is less susceptible to background noise. Researchers utilized this advantage to detect and remove non-speech background noise [3] from the simultaneously recoded air conducted speech. Recently, a Microsoft research group proposed a hardware device that combines regular air-conductive microphone with bone-conductive microphone [4]. They reported to be able to eliminate more than 90% of the background speech. Utilization of bone conducted speech in combination with the air conducted speech has also been shown to improve accuracy in speech recognition [5]. Although bone conduction is always induced by sound vibrations arriving at the head of the listener, its effectiveness is less than that of air conduction. This is because sounds at higher frequencies are impeded more during the transmission of bone conduction. This leads many researchers to concentrate on improving the quality of bone conducted speech by incorporating the higher frequency components derived from the air conducted speech [6, 7, 8, 9].

In summary, bone conducted speech has great potential in many applications. In this paper, we identified its ability in determining the pitch (fundamental frequency) contours particularly in noisy environments. A reliable tracking of pitch contour is critical for many speech processing tasks including prosody analysis, speaker identification, speech synthesis, enhancement and recognition. Thousands of pitch determination algorithms have been developed, none of which, however, is perfect in adverse environmental conditions. Speech signals in public places, like railway station, crowded street, industrial working area and battle field, become very noisy. In contrast, bone-conductive microphone works by capturing the bone vibrations which is expected to be immune from air conduction. Experiments have been set up to measure the noise induction to bone-conductive microphone from surrounding environments. Though bone conducted speech is noticed to be little affected, the quantity is insignificant especially for pitch determination. When the pitch estimation using the heavily affected normal speech signal gets deteriorated, bone conducted signal still yields accurate pitch estimates that could be very useful for robust speech processing applications including the quality enhancement of both the bone conducted speech and the simultaneously recoded air conducted speech.

2. SPECTRAL CHARACTERISTICS OF BONE CONDUCTED SPEECH

Conduction of sound through the vocal tract wall and bones of the skull is referred to as bone conduction. A headset directly coupled with the skull of the talker is usually used to capture the bone conduction. Researchers sought to identify the effective head locations for bone vibrations [1, 2]. Forehead, temple, mastoid and vertex are reported to be more suitable for picking up bone vibrations. Study has revealed that effectiveness of the bone conducted speech is 40 dB less than that of air conduction [10]. Sounds at lower frequencies are impeded less during bone conduction transmission than sounds at higher frequencies. This causes the higher

frequency components attenuated in case of bone conducted speech. Since the range of human-pitch usually resides within 50–400 Hz, bone conducted speech signal is thus appropriate for pitch tracking without making any spectral improvement (e.g. [6]). Spectrograms of bone conducted speech spoken by a Japanese male and a female speaker are shown in Fig. 1. Frequency components up to 3 kHz are shown, where the pitch harmonics are obvious in the lower portion. The sampling frequency used here is 12 kHz.
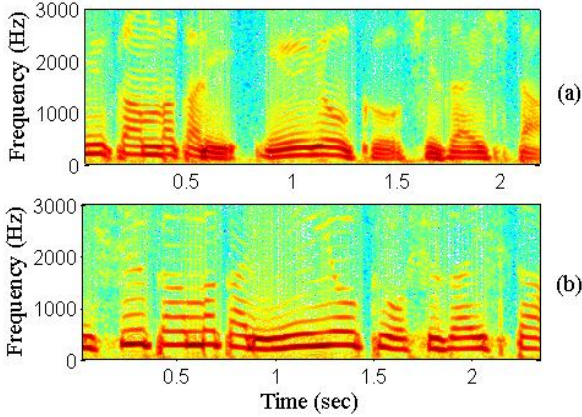

Figure 1: Frequency contents (< 3000 Hz) of bone conducted speech. a) Derived from MALE speech, b) derived from FEMALE speech.

In this study, we used Temco Japan HG-17 bone-conductive microphone comprising three receivers, two of which are placed on left and right temples and the other on the top of the head (vertex).

3. PITCH TRACKING OF AIR AND BONE CONDUCTED SPEECH IN NOISELESS CONDITIONS

As seen in the spectrograms in Fig. 1, though the higher pitch harmonics are mostly absent, harmonics of both the male and female speech signals are clearly evident in the lower part of the spectrograms. This facilitates pitch determination using bone conducted speech. Pitch contours estimated from simultaneously recorded air and bone conductive speech signals for four Japanese sentences are shown in Figs. 2 and 3. A standard Panasonic RP-VK25 microphone is used for recording air conducted speech. The weighted autocorrelation method described in [11] is applied for pitch determination. Pitch frequency is estimated from 30 ms-long frame with a frame shift of 10 ms. Both type of speech signals are band limited to 2 kHz before pitch determination.

The left and right panel in Figs. 2 and 3 corresponds to the pitch contours estimated from the air and bone conducted speech, respectively, in noiseless environment. The center panel corresponds to the spectrogram estimated from air conducted speech. In case of bone conducted speech, shape of the estimated

pitch contours match almost perfectly with the pitch harmonics apparent in the respective spectrogram. This can be attributed to the regular harmonic structure in the the lower region of the spectrum. On the other hand, pitch tracking is sometimes erroneous in case of clean air conducted speech, it deteriorates further in noisy conditions.

The four Japanese sentences used in this study are,
$s1$: arayuru genzitsuwo subete zibun no houhe nezimage tanoda
$s2$: isshyukan bakari nyuyoku wo shuzaishita
$s3$: bukka no hendouwo kouryo shite kyuhusuizyunwo kimeru hitsuyou ga aru
$s4$: irohanihoheto tirinuruwo

The speech is recorded in a noise-isolated laboratory room. We analyzed forty such sentences (four sentences spoken by five male and five female speakers). Results presented in Figs. 2 and 3, is typical to our analysis outcome.

4. NOISE SUSCEPTIBILITY OF AIR AND BONE CONDUCTED SPEECH

Unlike air conduction, the bone conduction pathway does not directly confront with outside environment. Even so, noise induction to bone conducted speech can not be ignored in highly noisy environment [12, 13]. The technology of bone- conductive microphone is still in the early stage of development, it therefore lacks perfection. An experiment has been conducted (in the laboratory where the authors are affiliated with) to investigate the extent at which bone conducted speech is induced from noisy environments. Artificial noisy conditions have been created for white and babble noise (noise due to human crowd) by playing noise signal during the recording process. The relative amplitude of noise induced with both the air and bone conducted speech are shown in Figs. 4(a) and 4(b), respectively. The top and bottom panels represent the case of white and babble noise, respectively. Recordings of male and female speech are accomplished separately but in the same noisy condition. The SNRs calculated from the noise corrupted air and bone conducted speech are shown in Table 1.

Table 1: SNRs of Air and Bone conducted speech

Speaker	Speech type	White (dB)	Babble (dB)
Male	Air	0.18	1.31
	Bone	5.26	20.30
Female	Air	-1.02	2.12
	Bone	16.23	31.27

Two observations can be made from Fig. 4 and from the data in Table 1. Firstly, bone conducted speech is affected much less than air conducted speech. Secondly, higher SNR values in the last row signify that female voice is more viable to produce intelligible bone conducted speech than male voice.
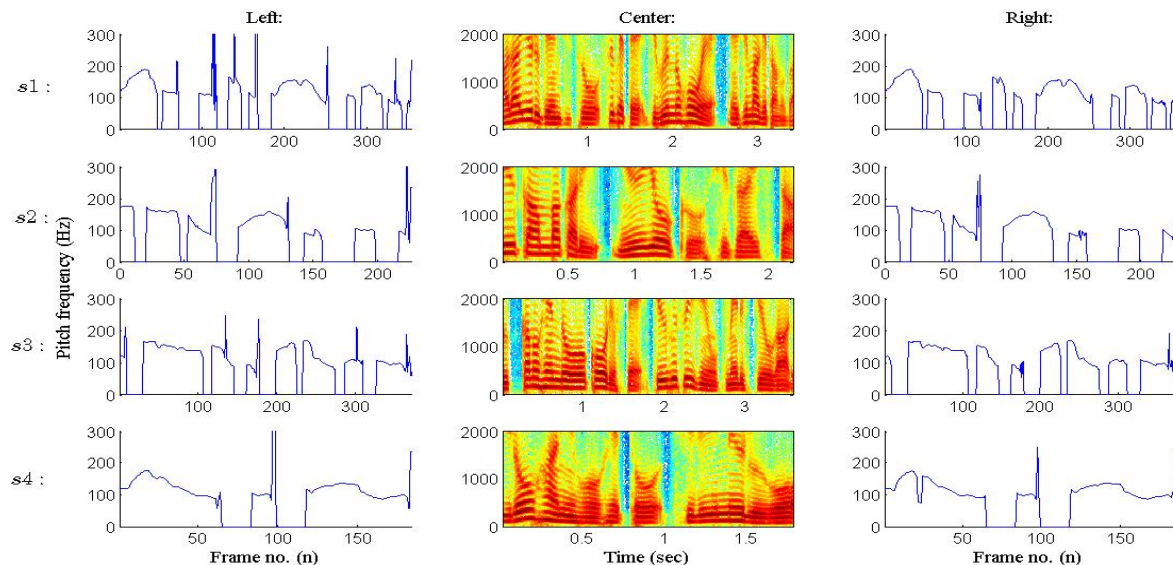
Figure 2: Pitch tracking of bone conducted speech spoken by a MALE speaker in noiseless condition. Left: Using air conducted speech, Center: it's spectrogram, Right: using bone conducted speech.
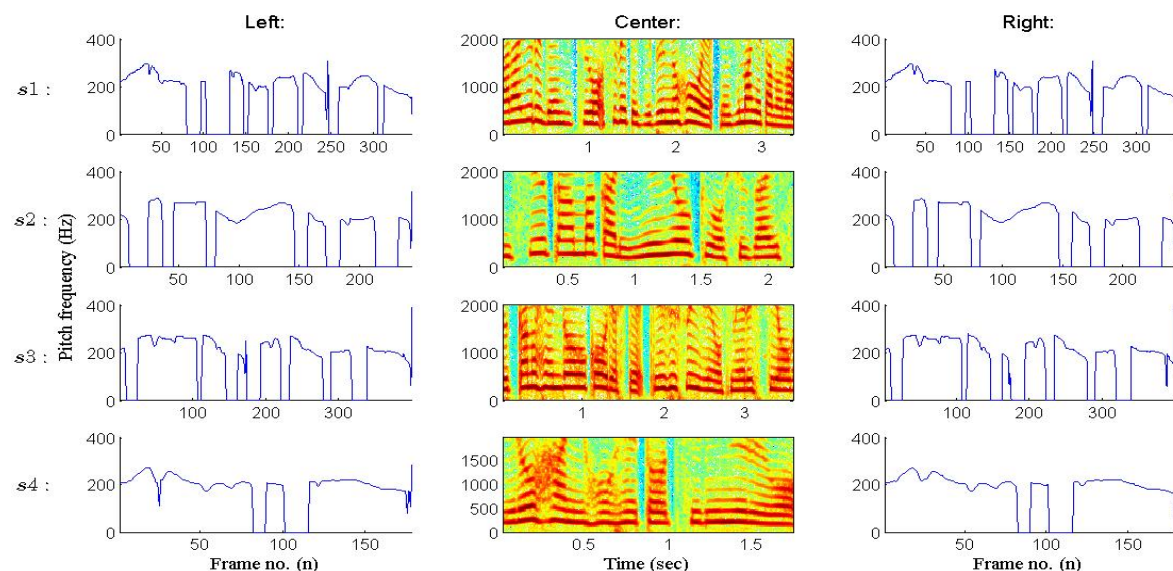

Figure 3: Pitch tracking of bone conducted speech spoken by a FEMALE speaker in noiseless condition. Left: Using air conducted speech, Center: it's spectrogram, Right: using bone conducted speech.

5. PITCH TRACKING OF AIR AND BONE CONDUCTED SPEECH IN NOISY CONDITIONS

Satisfactory performance of most pitch detection algorithms are limited to clean speech. Due to the difficulty of dealing with noise intrusions, the design of a robust pitch detector has proven to be very challenging. On the contrary (as it is obvious in the previous section), bone conducted speech is much insensitive to environmental adversity. This facilitates robust pitch tracking using bone conducted speech in highly noisy conditions. Pitch contours estimated using air and bone conducted speech for sentence $s3$ corrupted by white noise are shown in Figs. 5 and 6, for male and female speakers, respectively.

The same as in Figs. 5 and 6 are repeated in Figs. 7 and 8 in case of babble noise. The signal for sentence $s3$ is used here again to have a better comparison.

As depicted in Figs. 5 through 8, numerous octave errors are introduced in the pitch contours determined from noise corrupted normal speech signal. On the other hand, pitch contours determined from the cor-
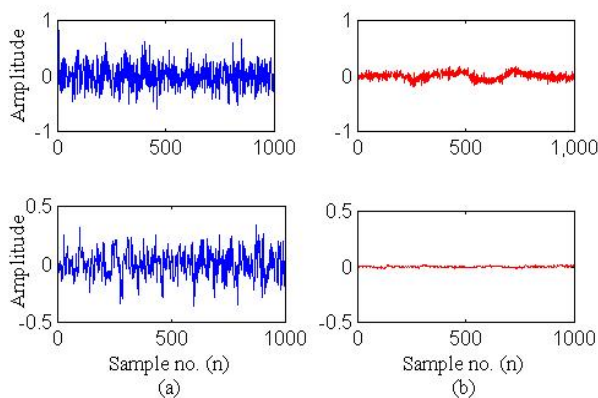
797

Figure 4: Relative amplitude of noise induced with speech (Top: White noise, Bottom: Babble noise). a) In case of air conducted speech, b) in case of bone conducted speech.
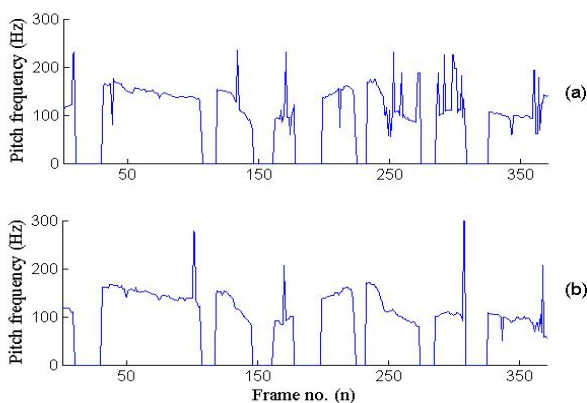

Figure 5: Pitch contours estimated from speech spoken by a MALE speaker when corrupted by WHITE NOISE. a) Using air conducted speech, b) using bone conducted speech.

rupted bone conducted speech almost resembles with those estimated in noiseless conditions as in Figs. 2 and 3 that indicates its robustness against ambient noise.

6. CONCLUSION

Bone conducted speech has to go a lot far to be suitable for independent utilization. It is, however, applied successfully together with normal speech for enhancement that resulted in improved speech recognition and speaker identification. In this paper, we have studied another important property of bone conducted speech, namely pitch frequency, both in noiseless and noisy environments. Experiments have been conducted to see the degree at which bone conducted speech is corrupted. It became evident that when the noise corrupted normal speech fails to yield accurate pitch contours, bone conducted speech can be employed for much greater accuracy.
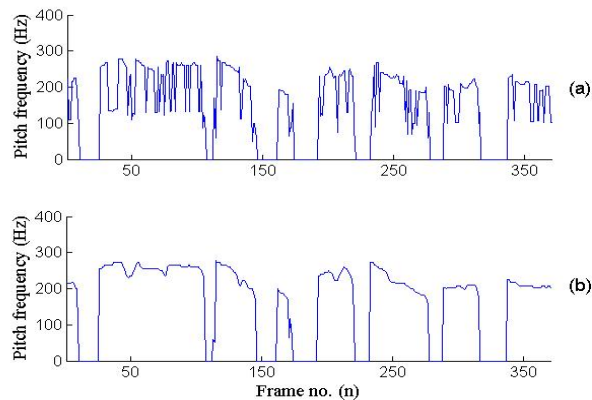

Figure 6: Pitch contours estimated from speech spoken by a FEMALE speaker when corrupted by WHITE NOISE. a) Using air conducted speech, b) using bone conducted speech.
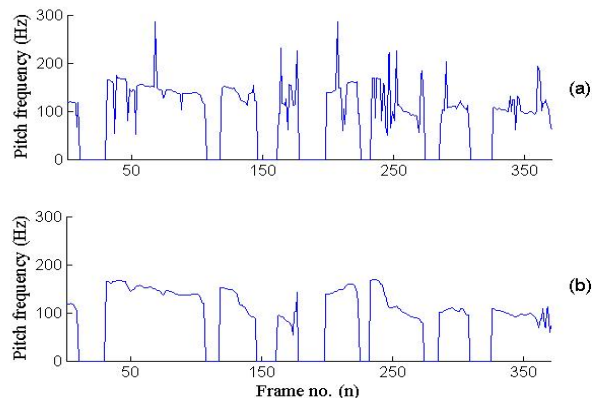

Figure 7: Pitch contours estimated from speech spoken by a MALE speaker when corrupted by BABBLE NOISE. a) Using air conducted speech, b) using bone conducted speech.
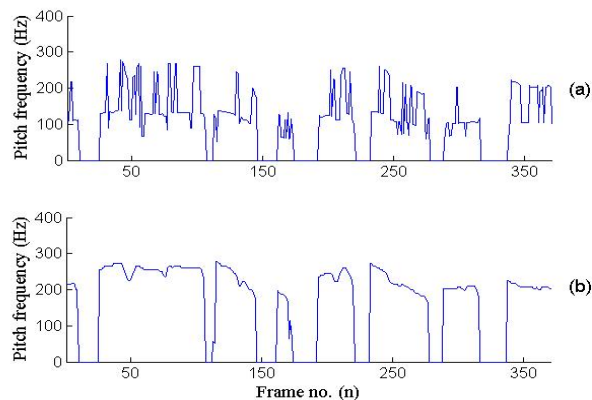

Figure 8: Pitch contours estimated from speech spoken by a FEMALE speaker when corrupted by BABBLE NOISE. a) Using air conducted speech, b) using bone conducted speech.

REFERENCES

[1] H.M. Moser and H.J. Oyer, "Relative intensities of sounds at various anatomical locations of the head and neck during phonation of the vowels," *Journal of Acoustical Society of America*, vol.30, no.4, pp.275–277, 1958.

[2] P. Tran, T. Letowski, and M. McBride, "Bone Conduction Microphone: Head Sensitivity Mapping for Speech Intelligibility and Sound Quality," in *International Conference on Audio, Language and Image Processing*, Shanghai, China, July 7-9, 2008, pp. 107–111.

[3] M. Zhut, H. Jit, F. Luott, and W. Chent, "A Robust Speech Enhancement Scheme on The Basis of Bone-conductive Microphones," in *The Third International Workshop on Signal Design and Its Applications in Communications*, Chengdu, China, September 2327. 2007, pp. 353–355.

[4] Y. Zheng., Z. Liu, Z. Zhang, M. Sinclair, J. Droppo, L. Deng, A. Acero, and X. Huang, "Air- and bone-conductive integrated microphones for robust speech detection and enhancement," in *IEEE Automatic Speech Recognition and Understanding Workshop*, St. Thomas, US Virgin Islands, November 30-December 3, 2003, pp. 249–254.

[5] Z. Liu, Z. Zhang, A. Acero, J. Droppo, and X. Huang, "Direct Filtering for Air- and Bone-Conductive Microphones," in *Proc. IEEE Workshop on Multimedia Signal Processing*, Siena, Italy, September 2004, pp. 363–366.

[6] T. Shimamura and T. Tamiya, "A Reconstruction Filter for Bone-Conducted Speech," in *IEEE International Midwest Symposium on Circuits and Systems*, Cincinnati, Ohio, August 7-10, 2005, pp. 1847–1850.

[7] T. Shimamura, J. Mamiya, and T. Tamiya, "Improving Bone-Conducted Speech Quality via Neural Network," in *IEEE International Symposium on Signal Processing and Information Technology*, Vancouver, Canada, August 27-30, 2006, pp. 628–632.

[8] V. T. Thang, K. Kimura, M. Unoki, M. Akagi, "A study of restoration of bone conducted speech with MTF-based and LP-based Model," *Journal of Signal Processing*, vol. 10, no. 6, pp. 407–417, November 2006.

[9] K. Kondo, T. Fujita, and K. Nakagawa, "On equalization of bone conducted speech for improved speech quality," in *IEEE International Symposium on Signal Processing and Information Technology*, Vancouver, Canada, August 27-30, 2006, pp. 426–431.

[10] M. Gripper, M. McBride, B. Osafo-Yeboah, and X. Jiang, "Using the Callsign Acquisition Test (CAT) to compare the speech intelligibility of air versus bone conduction," *International Journal of Industrial Ergonomics*, vol. 37, pp. 631-641, May 2007.

[11] T. Shimamura and H. Kobayashi, "Weighted Autocorrelation for Pitch Extraction of Noisy Speech," *IEEE transactions on speech and audio processing*, vol. 9, no. 7, pp. 727–730, October 2001.

[12] K. Kamata, K. Yamashita, T. Shimamura, and T. Furukawa, "Restoration of Bone-Conducted Speech with the Wiener Filter and Long-Term Spectra," in *RISP International Workshop on Nonlinear Circuits and Signal Processing*, Shanghai, China, March 3-6, 2007, pp. 583–586.

[13] Z. Liu, A. Subramanya, Z. Zhang, J. Droppo, and A. Acero, "Leakage Model And Teeth Clack Removal For Air- and Bone-Conductive Integrated Microphones," in *Proc. ICASSP*, Philadelphia, USA, March 2005, pp. 1093–1096.