



A STUDY TO ANALYZE ENHANCEMENT TECHNIQUES ON SOUND QUALITY FOR BONE CONDUCTION AND AIR CONDUCTION SPEECH PROCESSING

PUTTA VENKATA SUBBAIAH* AND HIMA DEEPTHI V.†

Abstract. In recent years, enhancing speech signal for communications become more complex due to the environmental noises for military and navel applications. Many researchers analyzed and developed complex algorithms and expressed ways to enhance speech signal for both Bone Conducted Speech (BCS) and Air Conducted Speech (ACS) processing techniques. BCS signal is vigorous for outer environmental noises and with low quality signal. ACS signal is sensitive for environmental noises but rigid in amplitude and high frequencies. Here we have reviewed and investigated various articles to explore their ideas in speech processing for enhancing intelligibility of sound quality.

Key words: Bone Conducted speech, Air Conducted speech, Sound quality, Signal Enhance, Speech intelligibility, Cochleae implant, Deep neural network, Linear prediction, Denoising.

AMS subject classifications. 92B20, 94A12

1. Introduction. Speech is primary tool and form to communicate by human beings. Few issues make the speech to restrict in outer environmental areas. Researchers utilized some hearing aids and speakers to overcome from the restrictions. There are two types of ways to collect the speech data from humans. Air Conducted Speech (ACS) signals have high efficiency, collected through speakers and sensitive with outer environmental noises. Bone Conducted Speech (BCS) signals have high amplitude, robust to the outer environmental issues by restricted with high frequency voice signals.

To enhance the intelligibility of speech in noise environment, automatic speech recognition systems are introduced with the help of complex algorithms to reduce or cancel the low level noises, but ineffective for high level of noises due to algorithm and equipment limitations. To overcome this microphones are utilized to record the speech signals through the bone conducted speech [BCS] process.

There are 85000 of deaf people can not sense any sounds with normal hearing aids. Normal human hearing range should be less than 24 KHz. Speech processing researchers fall on various bone conducted speech techniques and developed Ultrasonic Bone Conduction techniques, can be used to hear higher frequencies above 24 KHz. Several methods of BCS signals retrieving can be done by cochleae Implant (CI) i.e., a microphone will be placed behind skin by surgery. Several hearing aids rapidly explored, came to exist for BC signal retrieving. On implanting the hearing aids under skin, the performance of microphone in CI should not be attenuated by acoustic damping effect, Additionally due to the vibrations arrived of various body parts like mouth, skull and other the unexpected biological signals may generated and effects the performance of hearing aids microphone. To enhance the speech signal various conventional methods, fully implantable hearing devices (FIHDs) are developed for sensorineural hearing impaired people. Now a days a few approximately 5% of people are going for CI methods to embed in the skin, to outperform this Bone Conducted Ultrasonic hearing Aids are used for profoundly deaf sense people. Wearable electronic devices are used for military and navel applications to overcome from outer noises for ACS signals recording. ACS can go for high frequency values but sensitive with environmental noises but have better performance for high frequencies.

*Research Scholar, Vel Tech Rangarajan Dr.Sagunthala R&D Institute of Science and Technology, Assistant Professor in Department of Electronics and communication Engineering, Vel Tech Multi Tech Dr.Rangarajan Dr.Sakunthala Engineering College, No. 42, Avadi-Vel Tech Road, Avadi, Chennai-600062, India. (venky806@gmail.com).

†Professor in Department of Electronics and Communication Engineering, Vel Tech Rangarajan Dr.Sagunthala R&D Institute of Science and Technology, No. 42& 60, Avadi-Vel Tech Road, Avadi, Chennai-600062, India. (himadeepthi@veltech.edu.in).

This paper explores various methodologies developed by researchers to enhance in speech processing techniques by comparing or by combining the BC Speech signal and AC Speech signal. The scope of paper is as follows. Section 1 deals with basic information and ideology of speech processing issues, Section 2 with a study on researcher's propagation to enhance speech processing techniques, section 3 with a discussion on escalations obtained from Section 2, Section 4 analyzing all sections.

2. Brief review on approaches. To recover the profoundly deaf sense of auditory sense the 5% or less of them utilizes the hearing aid Cochleae Implant [CI] for its unfinished concert and a surgery needed to embed behind skin. According, Nakagawa et al.[1] to overcome this Bone conducted ultrasonic hearing aid developed for profoundly deaf sense audition is utilized without surgery with 20% of words were able to extricate successfully with further developments in epoch assembly confidently, and approximately 47% of profoundly deaf themes were not able to recognize the sounds through this aid. According to Dong-Wook Kim et al. [11], Cochleae aid's performance mainly effected by various body vibrations causing biological noise done while eating, playing and other. This can be reduced by restructuring the microphone and signal processing of FIHDS.

According to Yishan Luo et al. [5], the comparison of restoring speech signals of normal hearing people to the profoundly deaf people, Bone conducted ultrasonic hearing aid (BCUHA) will perform very efficiently by comparing using Finite Difference Time Domain with the restoration signals with Conventional bone conducted hearing aid (BCHA) and Cochleae implant. The conventional BCHA have limited output energy and dampening effect in frequency response, and BCHA can't adopt contralateral cochlea unilaterally so Bilateral BCHA should be used. Where in as BCUHA unilateral is sufficient and will have great significance to stimulate both cochlea.

According to Aminaka et al. [12], amyotrophic lateral sclerosis patients suffering from ear-stacking-syndrome have difficulty in auditory stimuli perception, by removing the obstructed auditory AC channel employing bone conduction auditory BCI for testing a few Morse-code auditory bone conducted patterns for speller collected from BCI 2000 bundle. 80% of test results through online spelling interfacing explores the technique is efficient.

Shunsuke Ishimitsu expressed in his book [25] as in noisy of marine engine systems for body conducted speech with small repetitions of utterance through the effectiveness of adaptive processing, above 50% of recognition rates at 98dB SPL were successfully achieved with in one utterance of speech recognition rate of higher value of 95% were attained. In a few experiments on healthy people retrieved speech, high frequency of formant information of speech signals were retrieved and possible applied for conversion and to estimate natural speech retrieval for speech disorders people.

According to Yamasaki et al. [9], speech processing technology is used for digital speech coding, spoken language, text to speech, automatic speech recognition (ASR). Information can also be extracted from speech. Yamasaki et al. proposed two new techniques for robust bone conducted speech in speaker authentication systems in highly noise environments. These proposed transformations methods shows improved equal error rate in conventional speech transformation.

According to Satoki Ogiso et al. [14], the characteristics of BC sound generated in human head to be studied through computer interfaced simulation in Finite Element Method (FEM), results the sound wave propagates in form of longitudinal and curving wave via skull of approximatively 1500 M/S and 260 M/S. The propagation curve wave in the skull is very slower in speed with high amplitude than of longitudinal wave. According to Zhang et al. [21], perfect frequency estimation in noisy environment is done by coalescing weighted auto correlation function (WACF) with Cepstrum (CEP) method for enhancing the speech in ARS and BCS in white noise environment. At variable SNR conditions, to improve the pitch signal in white noise environment the compared, clear voice ACS to the BCS signal is utilized. The proposed CEP scheme for female speech value shows better performance with lower error rate at high SNRs than ACF and WACF schemes. The experimental conditions shown in Table 2.1 chosen for 4 male and 4 female speech at same time with BC microphone were recorded are shown below.

According to Shimamura et al. [2], by passing speaker output of bone conducted speech signal a long term spectra reconstructed filter design, the arena of bone conducted and reconstructed speech from normal signal could be highly utilized. By passing normal speech signal through linear process with a transfer function of bone conduction can be obtained, to identify the properties of numerous sounds, spectral shapes by removing the dissimilarities. In Noisy environments, speech enrichment is done by using Bone conduction, to reconstruct

TABLE 2.1
The experimental conditions

Speaker	Male 4, Female 4
Flame Shift	64
Flame Length	256
Sampling Frequency	12 kHz
FFT Point	1024
Bone - Microphone	Temco HG - 17
Air - Microphone	BOSH ECM - 17
SNRs (db)	∞ , 10,5,0,-5

TABLE 2.2
Comparison Table of MSE for FIR and FLANN recovered signals

SNR (dB)		-3	-1	0	1	10	100
Female	FLANN	1.8103	1.8023	1.7801	1.7702	1.7604	1.7528
	FIR	1.9481	1.9287	1.9128	1.9063	1.8922	1.8895
Male	FLANN	1.7806	1.6774	1.5883	1.4962	1.3142	1.2936
	FIR	1.9754	1.7980	1.7336	1.6871	1.5143	1.4898

the bone conducted speech signal from speaker a linear phase impulse response filter will be utilized effectively. The designed linear phase FIR filter improves the quality of bone conducted speech with speech enhancement method, at around filter length 64 can be tolerably utilized without effected by filter length [3].

According to Ran Xaio et al. [19], artificial neural network based linear and non-linear adaptive noise cancellers (ANC) is utilized to improve the speech de-no semantic for BCS and ACS signals. At noisy conditions, high frequency components are recovered by the proposed ANC and outperforms its links for FIR filter, FLANN and other. Watanabe et al. [20, ?, ?].

Singh et al. [24], by regenerating sounds from Mel frequency cepstrum [MFC] coefficients, the mitigated frequency components have no effect, to overcome this effect and to improve of low frequency components of BCS signal, introduced a wavelet transform for the removal of noise from same speech by utilizing anti phase high frequency components in speech.

In noise environments like live military scenarios, ships and other, the air conducted speech will be highly distorted if the outer noise is very harsh. To overcome this a wearable bone conducted speech system was implemented and signals were collected from various skull positions which have very high frequency components and compared with air conducted speech signals. According to Boyan Haung et al. [22], in order to enhance the Bone and Air Conducted speech signal, employing FIR filter alone may not shows better performance, using trigonometric expansions of FLANN filter combined together be the better signal enhance. The proposed scheme have smaller MSE with recovered speech signal and clean speech signals this shown in Table 2.2. This can be achieved by taking largely effected ACS signal as primary with BCS as reference signal [17]. The experimental results for recovered signal from FIR and FLANN filter are compared each with clean signal, noisy signal and bond conducted is shown in Figure 2.1.

According to Thang Tat Vu et al. [4], due to its stability against highly noisy environment, bone conducted speech is highly advantageous however to mend the quality of restoration speech signal of bone conducted method, LP-Linear Perdution is used to improve the voice quality. The proposed linear perdution (LP) based blind restoration is completely distinguished and compared with various models by measuring their intelligibility of bone conducted speech with LSD, MCD and LCD, JWITs- Japanese word intelligibility tests and Vietnamese word intelligibility, tests proven better performance and lower degradation of signal in conduction of bone transduction [6].

According to Tsuge et al. [7], the speech data collected from the transducers for recognition through Bone conducted and Air conducted speech, by comparison proves that performance of AC is higher than BC speech. By combining the both AC Speech and BC Speech, similarities are calculated and noise will be eliminated certainly, results reduced equal error rate of BC speech by 71.7% and AC speech 16%.

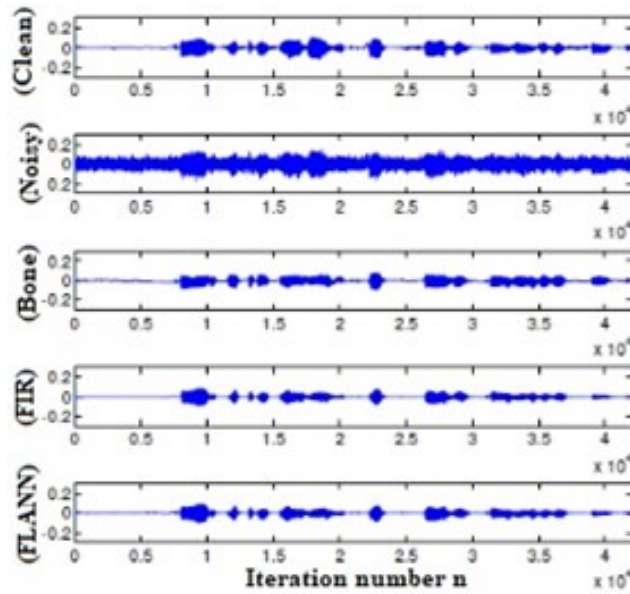


FIG. 2.1. The female speech subject chosen for experimental for comparing the signal at noisy, bone conducted, FIR filter (recovered speech), FLANN filter recovered speech at SNR= -1dB with clean speech

According to Rahman et al. [8, 10], the speech intelligibility can be enhanced with no distortion is employed by improper noise estimation, BC speech is healthy to ambient noise by improving speech tainted low frequency band noise. Using time domain filtering the BC speech can be performed efficiently for noise destruction at lower signal to noise ratios. Regularly speech travel in two formants, speech directly to air from mouth, i.e. air conducted speech and another is bone conducted speech travel through vocal tract and skull bone, subjective by vocal tract shapes and pathways can be apprehended by bone conducted microphone on talkers head. The second formant have very sensitive amplitude with higher value than the second formant [13].

According to Minghi Li et al. [16], for clear speech formants in short time Fourier transform domain, geometric harmonics and Laplacian pyramid were utilized and analyzed by log spectral measure and spectrograms.

According to Rachel E. Bouserhal et al. [18], in order to improve the hearing and radio communication experience for workers in noisy environments requires high in signal to noise ratio with high bandwidth. By investigating the three uniquely signals received from three microphones placed at outer the ear, inside ear [occluded ear], and in front of mouth using Gaussian mixture approach low and high frequency signals varies at signal to noise ratio with extended bandwidth. According to Steven L. Grant et al. [15], by coupling bone vibrators and bone conduction microphones will enhances the transmitted signal to noise ratio at ambient air-born noise issues. Echo cancellation performed effectively and short by 10Ms in path is linear enough through bone conduction echo path. Double talk control is designed in echo path canceller and tests outperforms in quality.

According to Irwansyah et al. [23], the cross-talk should compensate when sound transfers from transducers, wave moves from either left ear or right ear unswervingly to both ears of cochlea. It also increases the ability of sound sense in bone conduction headphone. This compensation filter can be assumed by otoacoustic emission (OAE), can be done by placing a microphone at outer ear canal, anti-sound produced by it using transfer functions results by filter-x least mean square algorithm reduces the prevised loudness.

3. Analysis. On reconstruction of speech signal BCS and ACS have different restrictions, ACS signals have high amplitude and contains outer noises. By the same time BCS are rigid to outer noises with low quality of retrieval signal. So most of the researchers concentrated on BCS signal processing techniques to enhance the signal by utilizing wavelets, deep neural networks and filters. Some researchers utilized ACS signal and BCS signal by combining and measured the retrieval signal strength. They found some crosstalk signal, need

to compensate, utilized OAE to enhance reconstructed signal.

Artificial neural networks used for improve speech de-no semantic in ACS speech and BCS speech by noise cancellers. Some study says voice transferring from vocal card should contains some vibrations occurred by skull, mouth and other may have impact on BCS. The speech wave travelling from mouth to outer air have two formants, longitudinal with lower amplitude and curve wave with slow travelling nature. For ACS, the signal to noise ratio should be high by increasing the band width utilizing gaussian mixture approach. ACS can be measured.

4. Conclusion. Firstly we investigated on various methods of speech enhancement techniques and their escalations caused due to current scenarios. On both BCS signal and ACS signal quality of the speech can be retrieved by removing noises, utilizing filters or cancellers from outer environments noises and inner vibrations. Combining the both signals have performed well than by measuring the individual Bone and Air conducted. BCS signal have high stability strength and rigid against ambient noise.

REFERENCES

- [1] S.NAKAGAWA., M. YAMAGUCHI ET AL. , 2002, *Development of Bone – Conducted Ultrasonic Hearing Aid for Profoundly Deaf*, SICE , pp. 2859-2862.
- [2] TETSUYA SHIMAMURA., TAKESHI TOMIKURA, *Quality Improvement of Bone-Conducted Speech*
- [3] TETSUYA SHIMAMURA., TAKESHI TOMIKURA., 2005. *A Reconstruction Filter for Bone-Conducted Speech*. IEEE, pp. 1847 - 1850.
- [4] THANG TAT VU, MASASHI UNOKI., MASATO AKAGI., 2006. *A Study on an LP-based Model for Restoring Bone-Conducted Speech*. IEEE, pp. 294 - 299.
- [5] YISHAN LUO., YUANYUAN WANG., 2007. *Simulation studies for comparison of Bone-conducted Ultrasound and Bone-conducted Audible Sound*. IEEE/ICME International Conference on complex Medical Engineering, pp. 1546-1549.
- [6] THANG TAT VU, MASASHI UNOKI., MASATO AKAGI., 2008. *An LP-based blind Model for Restoring Bone-conducted Speech*. IEEE, pp 212-217.
- [7] SATORU TSUGE., DAICHI KOIZUMI., MINORU FUKUMI., SHINGO KUROIWA., 2009. *Speaker verification method using bone-conduction and air-conduction speech*. IEEE/International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS 2009), pp 449-452.
- [8] M. SHAHIDUR RAHMAN., TETSUYA SHIMAMURA., 2010. *Pitch Characteristics of Bone Conducted Speech*. EURASIP, ISSN 2076-1465, pp 795-799.
- [9] NAOTO YAMASAKI., TETSUYA SHIMAMURA., 2010. *Accuracy Improvement of Speaker Authentication in Noisy Environments Using Bone-Conducted Speech*. IEEE, pp 197-200.
- [10] M. SHAHIDUR RAHMAN., ATANU SAHA., TETSUYA SHIMAMURA., 2011. *Low - Frequency Band Noise Suppression Using Bone Conducted Speech*. IEEE, pp 520-525.
- [11] DONG-WOOK KIM, JANG-WOO LEE, SEONG-TAK WOO, KI-WOONG SEONG., JIN-HO CHO., 2011. *Effects of Bone-conducted Signal on the Implantable Microphone*. IEEE, pp 112-114.
- [12] DAIKI AMINAKA., KOICHI MORI., TOSHIE MATSUI., ET AL., 2013. *Bone-conduction based Brain Computer Interface Paradigm - EEG Signal Processing, Feature Extraction and Classification*. IEEE/International Conference on Signal-Image Technology & Internet-Based Systems, pp 818-824.
- [13] M. SHAHIDUR RAHMAN., TETSUYA SHIMAMURA., 2013. *A Study on Amplitude Variation of Bone Conducted Speech Compared to Air Conducted Speech*. IEEE. ISBN: 978-986-9406-0-4.
- [14] SATOKI OGISO., KOICHI MIZUTANI., ET AL., 2014. *Analysis of Sound Propagation in Human Head for Bone-Conduction Headphones Using Finite Elements Method*. IEEE 3rd Global Conference on Consumer Electronics, pp 573-576.
- [15] MOHAMMAD HOSSEIN BEHGAM, STEVEN L. GRANT., 2014. *Echo Cancellation for Bone Conduction Transducers*. IEEE, pp 1629-1632.
- [16] MINGZI LI., ISRAEL COHEN, SAMAN MOUSAZADEH, 2014. *Multisensory Speech Enhancement In Noisy Environments Using Bone-Conducted and Air-Conducted Microphones*. IEEE, pp 1-5.
- [17] BOYAN HUANG., YEGUI XIAO., ET AL., 2014. *Speech enhancement based on FLANN using both bone – and air – conducted measurements*. IEEE/APSIPA.
- [18] RACHEL E. BOUSERHAL., TIAGO H. FALK., JEREMIE VOIX., 2015. *On the Potential for Artificial Bandwidth Extension of Bone and Tissue Conducted Speech: A Mutual Information Study*. IEEE, pp 5108-5112.
- [19] RAN XIAO., YEGUI XIAO., ET AL., 2016. *Speech enhancement using bone – and air-conducted signals and adaptive GFLANN filter*. IEEE.
- [20] DAIKI WATANABE., YOSUKE SUGIURA ., ET AL., 2017. *Speech Enhancement for Bone-Conducted Speech Based on Low-Order Cepstrum Restoration*. IEEE/International Symposium on Intelligent Signal Processing and Communication Systems, November 6-9, 2017, pp 212-216.
- [21] SHIMING ZHANG, YOSUKE SUGIURA., ET AL., 2017. *Fundamental Frequency Estimation Combining Air-Conducted Speech with Bone-Conducted Speech in Noisy Environment*. IEEE/International Conference on Electrical, Computer and Communication Engineering (ECCE), February 16-18, 2017, pp 244-247.

- [22] BOYAN HUANG., YIHAN GONG., ET AL., 2017. *A Wearable bone-conducted speech enhancement system for strong background noises*. IEEE/18th International Conference on Electronic packaging Technology, pp 1682-1684.
- [23] IRWANSYAH, TSUYOSHI USAGAWA., 2017. *Application of Active Control Technique on a Bone Conducted Headphone for Estimating a Cross-talk Compensation Filter*. IEEE/TENCON, November 5-8, 2017, pp 3099-3104.
- [24] PREMJEET SINGH., MANOJ KUMAR MUKUL., 2018. *Enhancement of Bone Conducted Speech Signal by Wavelet Transform*. IEEE, pp 317-321.
- [25] SHUNSUKE ISHIMITSU 2010, *Body – Conducted Speech Recognition and its Application to Speech Support System* In tech open, Advances in Speech Recognition, Book edition by: Noam R. Shabtai, ISBN 978-953-307-097-1, pp.164, September 2010, Sciyo, Croatia.

Edited by: Swaminathan JN

Received: Sep 28, 2019

Accepted: Nov 11, 2019