A Study on Amplitude Variation of Bone Conducted Speech Compared to Air Conducted Speech

M. Shahidur Rahman* and Tetsuya Shimamura†
*Department of Computer Science and Engineering, Shahjalal University of Science and Technology, Sylhet 3114, Bangladesh.
E-mail: rahmanms@sust.edu Tel: +880-821-713491
† Department of Information and Computer Sciences, Saitama University, Saitama 338-8570, Japan.
E-mail: shima@sie.ics.saitama-u.ac.jp Tel: +81-48-858-3486

***Abstract*—This paper investigates the amplitude variation of bone conducted (BC) speech compared to air conducted (AC) speech. During vocalization, vibrations travel through the vocal tract wall and skull bone, which can be captured by placing a bone-conductive microphone on the talker's head. Amplitude of this recorded BC speech is influenced by the mechanical properties of bone conduction pathways. This influence has relation with the vocal tract shape that determines the resonances of the vocal tract filter. Referring the vocal tract output as AC speech for simplicity, amplitude variation of BC speech can be described with respect to the location of the formants of AC speech. In this paper, amplitude variation of BC speech of Japanese vowels and long utterances have been investigated by exploiting the locations of first two formants of AC speech. Our observation suggests that when the first formant is very low with higher second formant, the relative amplitude of BC speech is amplified. As opposed to this, relatively higher first formant and lower second formant of AC speech cause reduction of the relative BC amplitude.**

I. Introduction

When we speak, voice signal is transmitted via two different ways. Air conduction is the normal path of the sound that exits from mouth and transmitted through air. Bone conducted component, on the other hand, travels as vibrations through vocal tract wall and skull bone on its way to the cochlea. A bone conductive microphone (BCM) equipped with a vibration sensor can capture these vibrations and convert to electrical signal. Unlike air conduction, bone conduction does not confront with ambient disturbances. This apparent advantage leads many researchers to study BC speech in case of severe noisy conditions where performance of the conventional speech processing systems substantially deteroriates. Communication headsets using BC speech have already been developed and made commercially available for military, rescue, and security operations where it is inappropriate to communicate using AC speech.

Although there has been few decades of research on hearing by bone conduction [1], [2], [3], [4], study of its application for speech processing in noisy environment has just been emerging. One main concern of BC speech is its lower intelligibility because it lacks higher frequency components [5], [6]. Most of the recent works therefore concentrate on enhancement of speech intelligibility [7], [8], [9], [10]. However, spectral envelop based transfer function [7], [8] can not be generalized among speakers because individualized details are compromised. A number of studies also provide information regarding the appropriate location for picking up the bone vibrations [11], [12]. The current work aims to study the time-domain amplitude variation of BC speech compared to AC speech based on the spectral property (particularly the first and second formant) of the later. Though the comparison of BC and AC speech aiming at a transfer function is not new [3], [13], this particular analysis of dependence on formant frequencies is sufficiently novel. Adjustment of the formant frequencies of the underlying BC speech based on the variation of its relative amplitude can lead to the enhancement of the intelligibility of BC speech, which will be very useful for speech processing in noisy environment. Secondly, since the transfer function based enhancement technique suffers from generalization, the proposed amplitude-variation based formant modification can be employed in addition to achieve individualized adjustments.

The excitation signal, after being modified by the glottis and the vocal tract filter, travels as vibrations through the skull bone. The mechanical impedance of the bone conduction pathways, which is mainly the skull bone and skin, influences the amplitude and frequency contents of the recorded BC speech. This influence can be related with the vocal tract shape and thus with the formant frequencies of the vocal tract filter. Considering the output of the vocal tract filter as AC speech, a study is conducted based on formant locations of five Japanese vowels spoken by male and female speakers. We found that amplitude of BC speech is very much sensitive with the locations of the first formant (F_1) of AC speech. Particularly, voice with very low F_1 and higher second formant (F_2) is more suitable for bone conduction in terms of time-domain amplitude, which leads to the amplification of relative BC amplitude compared to AC speech. On the other hand, reduction of relative BC amplitude takes place for relatively higher F_1 and lower F_2 values. This observation is also found valid when examining the long natural utterances. The phenomenon is studied primarily using a commercially available BCM, Temco HG-17. Theoretically flat microphone response is then simulated to obtain the

condition of microphone independence.

II. Modeling the Air and Bone Conducted Speech

Ignoring the effect of lip radiation for simplicity, AC speech can be modeled as

$$x = u * v \tag{1}$$

where u and v represent glottal waveform and vocal tract impulse response, respectively. The operator $*$ stands for convolution. For voiced sound, the source is quasiperiodic puffs of the airflow through the glottis vibrating at a certain fundamental frequency. The vocal tract impulse response is determined by the shape of the vocal tract, which, in turn, determines its resonances. Stevens [14] developed rules for mapping changes in vocal tract shape to formant transitions based on physical principles. BC speech, on the other hand, can be modeled as

$$y = u * v * b * k * m \tag{2}$$

where, b, k and m represent skull bone, skin and microphone impulse response, respectively. Eq. (2) can be rewritten using Eq. (1) as

$$y = x * b * k * m. \tag{3}$$

Eq. (3) indicates that compared to AC speech x, BC speech y is influenced additionally by the bone and skin properties. When sound vibrations propagate through the skull bone, these need to overcome the bone's opposition to transfer energy caused by its impedance. Two impedance measures, frequently referred to as the skull impedance and the skin impedance, are important in bone conduction. Several investigators have attempted to measure the impedance of the skull with and without the skin present [15]. Though the coupling between the skull and the skin is not yet well understood, they have certain effect on the amplitude of x. Both types of impedances have been described to be affected based on the frequency characteristics of the input (i.e., x) [15], [16]. This suggests that amplitude variation of BC speech is related with the spectral properties of AC speech. In this study, we attempt to explain the amplitude variation of y in terms of the formants estimated from x. Further, unlike air-boom microphone, response of the BCM is not flat which may have additional effect on the recorded speech. Amplitude variation of BC speech is reported taking or without taking the microphone's effect into account.

III. Amplitude Behavior of Bone Conducted Speech Using Temco HG-17 Microphone

As discussed in Sec. II, the combined effect of b, k, and m changes the amplitude of AC speech x. Vowel sounds and natural continuous utterances are used here to explain the amplitude variation of BC speech with respect to the AC counterpart. All the speech materials are recorded at 48 kHz rate which is then down sampled to 8 kHz for processing. A standard Panasonic RP-VK25 microphone is used for recording AC speech and a Temco HG-17 microphone is used for capturing BC speech where the vibration sensor is originally positioned on the top of the head (vertex). The recording of all speakers are completed in the same settings. Since the sensor types and hardware (e.g., amplifiers) used for capturing AC and BC speech are different, the sequence of utterances is normalized so that the maximum occurring amplitude in the utterance is set to +/-1.

A. Experiments with vowel sounds

Preliminary experiments have been conducted on five Japanese vowel sounds. AC and BC speech signals and corresponding auto-regressive spectra of /a/, /i/, /u/, /e/ and /o/ spoken by a male speaker are shown in Figs. 1 and 2, respectively. As can be seen in Fig. 1, the relative amplitude

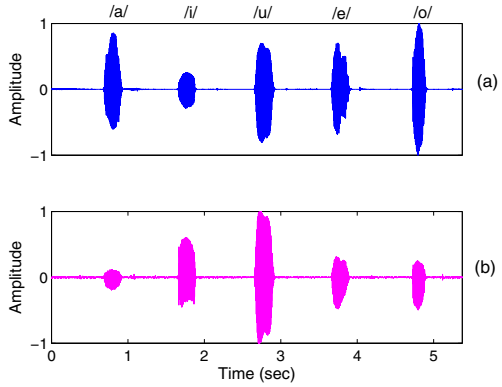

Fig. 1. Relative amplitude of five vowel sounds spoken by a male speaker. a) In case of AC speech, b) In case of BC speech.

of BC speech /a/ is reduced and the same of /i/ and /u/ are amplified compared to the AC counterpart. In close vowels (also referred to as high vowels), such as /i/ and /u/, the tounge is positioned high in the mouth, whereas in open vowels (also referred to as low vowels), such as /a/, the tounge is positioned low in the mouth. The lower the F_1 value, the more close the vowel [17], [18]. As seen in Fig. 2, the first formant of /i/ and /u/ occurs at roughly 322 Hz and 331 Hz, respectively. According to our observation, when F_1 is very low as in the case of the close vowels /i/ and /u/, the relative BC amplitude is notably amplified compared to AC speech. On the contrary, the higher the F_1 value, the more open the vowel. As seen in Fig. 2, the first formant of /a/ occurs at 725 Hz. When F_1 is higher as in the case of open vowel /a/ the relative BC amplitude is reduced. Moreover, in front vowels, such as /i/, the tounge is positioned forward in the mouth, whereas in back vowels, such as /u/, the tounge is positioned towards the back of the mouth. The higher the F_2 value, the fronter the vowel and the lower the F_2 value, the more retracted the vowel. When the F_2 value is substantially higher, relative BC amplitude of a close vowel is more amplified. For example, relative BC amplitude of /i/ with higher F_2 is amplified more compared to /u/ with lower F_2 value. Again, both /e/ and /o/ are mid vowels (where the tounge is positioned between high and low in the mouth), but /e/ is located fronter than /o/. Relative BC amplitude of front vowel /e/ is therefore less

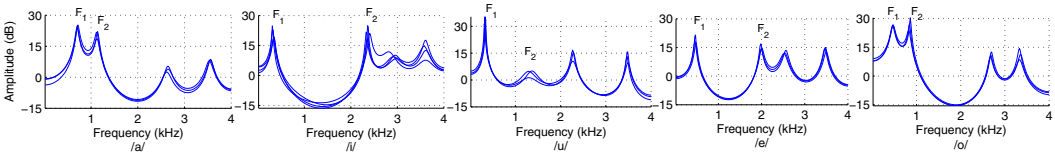

Fig. 2. Spectra of five vowel sounds spoken by a male speaker shown in Fig. 1

reduced than the back vowel /o/. This is evident in the the short-time magnitude ratio of BC and AC speech as shown in Fig. 3, where the scores indicate the degree of amplification (> 1) or reduction (< 1). This plot is produced using the BC and AC speech signal given in Fig. 1. Here, the short-time magnitude is computed as $\sum_{n=0}^{N-1} |s(n)|$, where N is the frame length and s refers to the speech signal. AC and BC speech signals of five vowels for a female speaker are shown in Fig. 4. Clearly, the amplification and reduction of BC speech compared to AC speech in Fig. 4 are similar as in the case of Fig. 1. In both cases, amplitude variation of BC speech is mainly determined by the F_1 value (i.e., vowel height), while the frontness and backness (i.e., F_2 value) play an additional role in amplification and reduction, respectively.

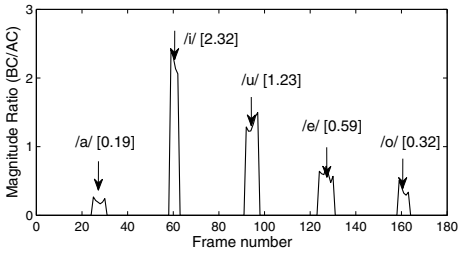

Fig. 3. Short-time magnitude ratio of BC and AC speech given in Fig. 1.

To illustrate the phenomenon in greater detail, the F_1 and F_2 values with short-time magnitude ratio of BC and AC speech spoken by eight speakers (four male MA1$\sim$MA4 and four female FE1$\sim$FE4) for close vowels /i/ and /u/ and open vowel /a/ are shown in Table 1. Only the close and open vowels are chosen because they produce very obvious results in amplitude variation. Here, both the F_1 and F_2 values represent the mean value obtained from a number of frames spanning over the middle portion of every vowel sound estimated using the *Praat* system [19]. Similarly, the rational value BC/AC, shown in Table 1, represents the mean magnitude as indicated in Fig. 3. As it is obvious from Table 1 that maximum amplification of BC speech amplitude is observed for the close and front vowel /i/ with lowest F_1 and higher F_2/F_1 value and maximum reduction of BC amplitude is observed for the open and back vowel /a/ with highest F_1 and lower F_2/F_1 value. However, since formant locations of voiced speech are related to physiology and vary among speakers, as will the bone and skin impedances, the extent of the observed results could be different among speakers. In spite of the similarity of amplitude behavior of a particular vowel sound, individualized details are also evident in varying BC/AC values in Table 1.

TABLE I

F_1 and F_2 with short-time magnitude ratio of BC and AC speech spoken by four male (MA1 $\sim$ MA4) and four female (FE1 $\sim$ FE4) speakers

Vowel	Speaker	F_1 (Hz)	F_2 (Hz)	F_2/F_1	BC/AC
/a/	MA1	725	1152	1.59	0.19
	MA2	622	1021	1.64	0.57
	MA3	634	1088	1.72	0.51
	MA4	713	1375	1.93	0.39
	FE1	627	1275	2.03	*0.59*
	FE2	933	1310	1.40	*0.49*
	FE3	999	1407	1.41	*0.27*
	FE4	747	1196	1.60	*0.77*
/i/	MA1	322	2346	7.28	2.32
	MA2	255	2367	9.28	2.56
	MA3	265	2169	8.18	1.85
	MA4	293	2410	8.22	1.64
	FE1	314	2773	8.83	*1.96*
	FE2	301	2723	9.04	*2.50*
	FE3	371	2848	7.67	*2.94*
	FE4	322	2944	9.14	*2.70*
/u/	MA1	331	2153	6.50	1.23
	MA2	354	1893	5.34	1.31
	MA3	290	1604	5.53	0.55
	MA4	309	1393	4.51	1.02
	FE1	432	1486	3.43	*1.23*
	FE2	368	1596	4.33	*2.38*
	FE3	501	1724	3.44	*2.22*
	FE4	452	1297	2.87	*1.41*

B. Experiments with Natural Continuous Utterance

Unlike separate vowel sounds, continuous speech always exhibits complex characteristics due to faster transitions of vocal tract shapes. We attempt to observe this phenomenon on a Japanese utterance “arayuru genzitsuwo” spoken by a female speaker. The AC speech segment, its spectrogram and formant tracks along with the corresponding BC speech signal are shown in Fig. 5. Again, the spectrogram and formant tracks are obtained using the *Praat* system [19]. As discussed earlier in Sec. III(A), every time the vocal tract shape is changed, the transitions of formants take place. Particularly, the variation of F_1 value displays clear sensitivity. To illustrate the amplitude variation of BC speech compared to AC speech, the F_1 track is marked with a number of locations using English letters A and R in Fig. 5(b). The F_1 and F_2 values on the selected locations along with short-time magnitude ratio BC/AC are shown in Table 2. The locations $R1$ through $R4$ in Fig. 5 are manifested with higher F_1 value causing reduction in the

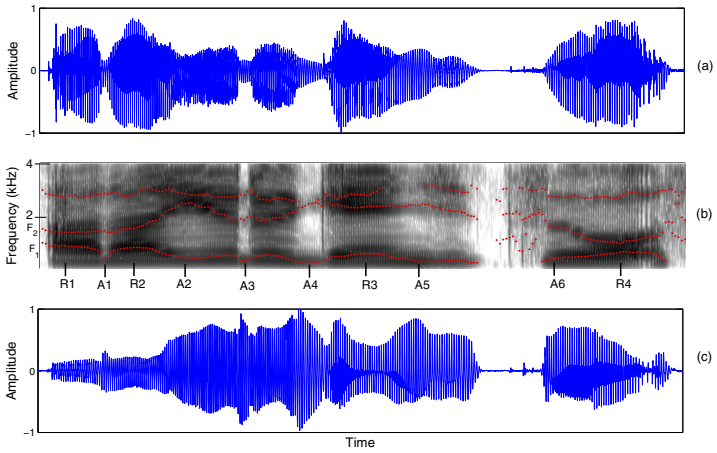

Fig. 5. a) Long segment of AC speech, b) Spectrogram and formant tracks obtained from speech signal in (a), c) Corresponding BC speech segment.

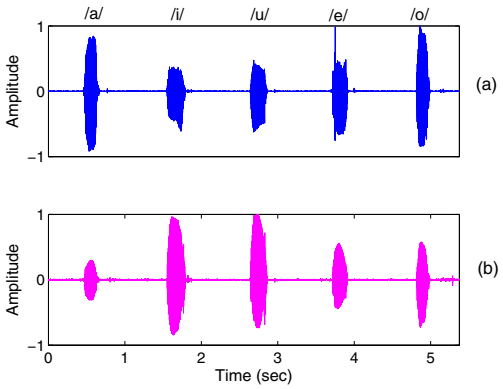

Fig. 4. Relative amplitude of five vowel sounds spoken by a female speaker. a) In case of AC speech, b) In case of BC speech.

TABLE II
F_1 and F_2 with short-time magnitude ratio of BC and AC speech at selected locations in a long natural utterance spoken by a female speaker.

Locations	F_1 (Hz)	F_2 (Hz)	F_2/F_1	BC/AC
R1	860	1378	1.60	0.50
A1	423	1417	3.35	1.05
R2	803	1529	1.90	0.41
A2	405	2476	6.11	2.13
A3	274	1845	6.73	4.09
A4	265	2454	9.26	4.21
R3	563	2346	4.17	0.75
A5	275	2426	8.82	1.73
A6	397	1552	3.91	2.19
R4	630	956	1.52	0.80

relative amplitude of the corresponding BC speech segments. This is also evident in Table 2 showing BC/AC value less than 1. On the other hand, the locations $A1$ through $A6$ are manifested with very low F_1 value causing amplification of BC amplitude compared to the AC counterpart. This is shown in Table 2 with BC/AC values greater than 1. Though the variation due to F_2 transition is not as obvious as that of F_1, BC speech signals with higher F_2/F_1 values are mostly amplified more. For example, BC amplitude at locations $A3$ and $A4$ with higher F_2/F_1 values are more amplified compared to those at $A1$ and $A6$ with lower F_2/F_1 values. In summary, variation of relative BC amplitude for continuous utterance is also straightforward when change of vocal tract shape results in clear F_1 transitions. We analyzed a number of Japanese utterances and noticed the same phenomenon.

IV. Amplitude Behavior of Bone Conducted Speech for Flat Microphone Response

According to Eq. (3), isolation of microphone effect is necessary to study the exact amplitude behavior of BC speech. Unlike normal air conductive microphone (ACM) which exhibits a nearly flat frequency response, BCM is designed to emphasize the high frequency components. Though the frequency response of BCM can be obtained by artificial modeling [20], a simple experiment is conducted here to approximate the characteristics of Temco HG-17 microphone. A subject wearing a bone conductive headset sits in front of a recording system without uttering any sound. Artificial white noise is played in a noise isolated room. Since the subject does not speak, the signal induced through BCM is only due to the background white noise. The spectra of the original noise and the recorded noise with ACM and BCM are shown in Fig. 6. When the spectrum of the recorded BC noise in Fig. 6 is compared to that of AC noise, which is known *a priori* as flat, the BCM is observed to emphasize the higher frequency components. This property of emphasizing higher frequency components is parameterized with a finite impulse response filter, inverse of which is then applied on the BC speech. This results in a theoretically flat response of the Temco HG-17 microphone and thus the filtered BC speech is now free from microphone effect. It is, however, noted that the effect

of sound field is also included with the white noise as input to the BCM and thus the proposed inverse-filtering approach actually provides an approximated result in isolating the BCM effect. The filtered version of the vowel sounds is shown in Fig. 7 together with the primarily recorded BC signal as in Fig. 4. According to Fig. 7, isolation of the microphone effect due to high-frequency emphasis does not affect the relative amplitude behavior of BC speech except slight change in the overall amplitude. This indicates that amplitude behavior of BC speech described in Sec. III is essentially due to the impedance variations of bone conduction pathways (i.e., skull bone and skin).

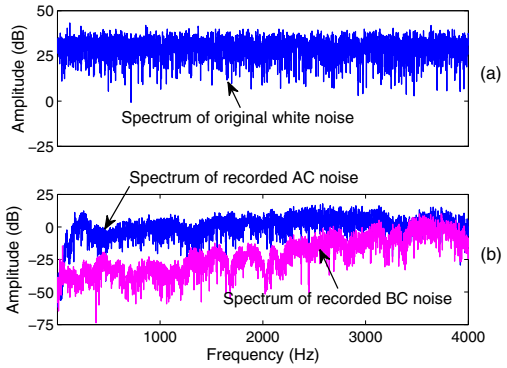

Fig. 6. Spectrum of white noise. a) Obtained from original white noise, b) Obtained from recorded noise by air and bone conductive microphone.

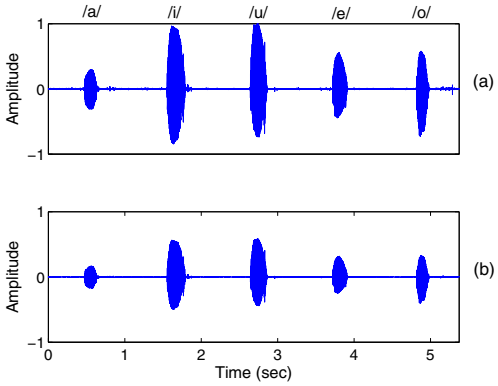

Fig. 7. a) Primarily recorded BC speech, b) Filtered version of (a) in case of flat microphone response.

V. Conclusion

This paper presented the results of our study on the amplitude behavior of BC speech compared to AC speech. Our observation suggests that relative BC amplitude is very sensitive with the location of the first formant of AC speech. When the first formant is very low with a higher second formant, relative BC amplitude of the underlying segment is amplified and it is reduced for relatively higher first formant with lower second formant. Though the second formant plays an additional role in amplification or reduction, its effect is rather less sensitive.

Acknowledgements

This work has been supported by the Japan Society for the Promotion of Science.

References

[1] Békésy, G. V., "The structure of the middle ear and the hearing of one's own voice by bone conduction," *The Journal of the Acoustical Society of America*, Vol. 21, pp. 217–232, 1949.

[2] Porschmann, C., "Influences of bone conduction and air conduction on the sound of one's own voice," *Acta Acustica united with Acustica*, Vol. 86(6), pp. 1038–1045, 2000.

[3] Reinfeldt, S., Östli, P., Håkansson, B., and Stenfelt, S., "Hearing ones own voice during phoneme vocalization-Transmission by air and bone conduction," *The Journal of the Acoustical Society of America*, Vol. 128, pp. 751–762, 2010.

[4] Stenfelt, S. and Goode, R. L., "Bone-conducted sound: physiological and clinical aspects," *Otology & Neurotology*, Vol. 26, No. 6, pp. 1245–1261, 2005.

[5] Acker-Mills, B., Houtsma, A. and Ahroon, W., Speech Intelligibility with Acoustic and Contact Microphones, in Proceedings of New Direction for Improving Audio Effectiveness, RTO-MP-HFM-123, paper 7, Neuilly-sur-Seine, France, pp. 1-14, 2005.

[6] Gripper, M., McBride, M., Osafo-Yeboah, B. and Jiang X., "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, 2007.

[7] Liu, Z., Zhang, Z., Acero, A., Droppo, J. and Huang, X., "Direct filtering for air- and bone-conductive microphones," *Proc. IEEE Workshop on Multimedia Signal Processing*, Siena, Italy, pp. 363–366, 2004.

[8] Shimamura, T. and Tamiya, T., "A reconstruction filter for bone-conducted speech," Proc. *Proc. IEEE International Midwest Symposium on Circuits and Systems*, pp. 1847–1850, 2005.

[9] Uchino, E., Yano, K. and Azetsu, T., "A self-organizing map with twin units capable of describing a nonlinear input-output relation applied to speech code vector mapping," *Information Sciences*, Vol. 177, pp. 4634–4644, 2007.

[10] Rahman, M. S., Atanu, S. and Shimamura T., "Low-frequency band noise suppression using bone conducted speech." *Proc. IEEE Pacific Rim Conference on Communications, Computers and Signal Processing (PacRim)*, pp. 520–525, 2011.

[11] McBride, M., Tran, P., Letowski, T. and Patrick, R., "The effect of bone conduction microphone locations on speech intelligibility and sound quality," *Applied Ergonomics*, Vol. 42, pp. 495–502, 2011.

[12] Stanley, R. M. and Walker, B. N., "Intelligibility of bone-conducted speech at different locations compared to air-conducted speech," *Proc. Annual Meeting of the Human Factors and Ergonomics Society*, 2009.

[13] Stanley, R. M., and Walker, B. N., "Towards a transfer function used to adjust audio for bone-conduction transducers," *The Journal of the Acoustical Society of America*, Vol. 123, no. 5, pp. 3565–3565, 2008.

[14] Stevens, K. N., *Acoustics Phonetics*, MIT Press, Cambridge, MA, 1998.

[15] Håkansson, B., Carlsson, P. and Tjellstrm, A., "The mechanical point impedance of the human head, with and without skin penetration," *Journal of the Acoustical Society of America*, Vol. 80, pp. 1065–1075, 1986.

[16] Stenfelt, S. and Goode, R., "Transmission properties of bone conducted sound: measurements in cadaver heads," *Journal of the Acoustical Society of America*, Vol. 118, pp. 2373–2391, 2005.

[17] Ladefoged, P., *A Course in Phonetics (Fifth Edition)*, Boston, MA: Thomson Wadsworth, p. 189, 2006.

[18] O'Shaughnessy, D., *Speech Communications: Human And Machine*, IEEE press, p. 60, 2000.

[19] Boersma, P. and Weenink, D., *Praat: Doing Phonetics by Computer (Ver 5.1.32)*.

[20] MacDonald, J. A., Henry, P., and Letowski, T. R., "Spatial audio through a bone conduction interface," *International Journal of Audiology*, Vol. 45, pp.595–599, 2006.