DOI: 10.61782/fa.2023.0476

EFFECT OF BONE CONDUCTION MICROPHONE LOCATION AND MOUTH OPENING ON TRANSFER FUNCTION BETWEEN ORAL CAVITY SOUND PRESSURE AND SKIN ACCELERATION

J. RICHARD[1,2]* **V. ZIMPFER**[1] **S. ROTH**[2]
[1] French-German Research Institut of Saint-Louis, 68300 Saint-Louis, France
[2] Laboratoire Interdisciplinaire Carnot de Bourgogne, site UTBM,UMR 6303 CNRS
Univ. Bourgogne Franche-Comté, France

ABSTRACT

Bone Conduction Microphones (BCM) allow recording the speech signal which has traveled through solid material composing the human body (bone, cartilage, soft tissues). The main advantage of these microphones is their ergonomics since they can easily be placed in head equipment. However, their drawback is their lack of intelligibility, which depends on several parameters, including the location of the BCM. In order to understand the changes induced by switching the BCM placement, the transfer function difference between the oral cavity sound pressure and the signal recorded by the BCM for two different positions has been estimated. An oral sound source has been designed in order to obtain reproducible measurements. Moreover, the influence of the mouth opening on the result has also been estimated since the mouth opens more or less to radiate the sound while speaking. Finally, an intelligibility test has been conducted at the same locations. The results show a better wave transmission from oral cavity to forehead than to chin in the frequency band [800;2500]Hz giving a first explanation to the greater intelligibility observed at the forehead.

Keywords: *Bone conduction microphone - Bone-conducted speech - Intelligibility - Transfer function*

1. INTRODUCTION

Speech signal travels through the air but also bones, cartilage, and soft tissues. When the signal propagates only through air, it is called AC (Air-Conducted) speech. When the signal travels through biological tissues, it is known as BC (Bone-Conducted) speech. BC speech can be experienced by speaking while covering one's ears. Different pathways of BC speech to the auditory system have been determined thanks to the work of Von Bekesy [1], Tonndorf [2], and Stenfeldt [3]. The bone was, for a long time, the medium considered most important in wave propagation through the body. Hosoi [4], and Nishimura [5] highlighted the importance of cartilage conduction, and Sohmer [6] studied extensively the soft tissue conduction which balances the exclusive contribution of bone conduction. It is common to record AC speech but rare to record BC speech. Assumptions made on the propagation paths of BC speech allowed the development of microphones that capture BC speech as the in-ear microphone or the Bone Conduction Microphone (BCM).

The BCM, which is a contact microphone, is considered the first one developed and used to record BC speech. Initially, this microphone was placed at the throat position and was called the throat microphone [7]. This microphone has been and continues to be developed because it presents two strong advantages over a conventional aerial microphone. First, since it records BC speech, it is less disturbed by environmental sound. BC speech has a higher signal-to-noise ratio (SNR) than AC speech recorded by one aerial micro-

---

**Corresponding author*: julien.richard@isl.eu.

phone [8]. Second, BCM is an ergonomics solution to record speech. It can be placed in head equipment and can be worn simultaneously with a mouth-covering protective device such as a gas mask. However, compared to an aerial microphone, the BCM's drawbacks are poor intelligibility and sound quality [9, 10]. Therefore research tries to understand and improve these limitations.

More or less complex signal treatments have been applied to improve these limitations, with some based on transformation filter [11–15] and others based on machine learning process [16–19]. These methods try to increase the intelligibility and sound quality of the BC speech by using AC speech as the target sound. The simplest treatment consists of filtering the BC speech by the estimated transfer function between AC speech and BC speech [12–14]. These methods have the advantage of being low in machine cost compared to the one using advanced machine learning but do not lead to satisfactory results. All these methods are based on signal processing, and no physical investigation of BC speech is made to improve it. Higher knowledge of BC speech mechanism and physical parameters influencing BC speech transmission is needed to increase these two metrics more efficiently.

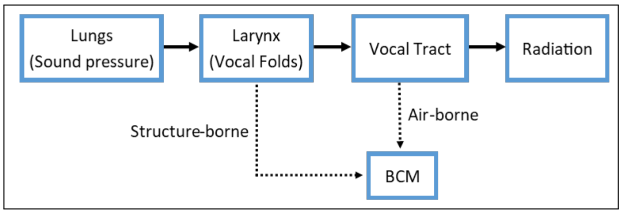

Figure 1. Speech production model and assumed signals recorded by BCM for voiced sound.

Previous studies pointed out that intelligibility and sound quality of the BCM are influenced by the microphone's position [20–22] and depend on the talker [23]. The position where the recorded speech has the better quality and intelligibility does not correspond to the location with the higher speech intensity [24, 25]. Assumption can be made that BCM record air-borne sound originating from the vocal tract and structure-borne sound originating from the larynx/vocal folds. This hypothesis is presented in Figure 1. The reason for the difference observed regarding intelligibility depending on the position is still unclear. This article aims to explain the intelligibility difference at two positions by determining the change in the transfer function "oral cavity/BCM". The influence of the mouth on these transfer functions is also investigated since the variation of the amplitude of bone-conducted speech with respect to the amplitude of Air-conducted speech is phoneme dependent [26, 27].

2. MATERIAL AND METHODS

2.1 Material

2.1.1 Recording materials & loudspeaker

The experiments took place in an audio-metric booth. They implied a loudspeaker and various recording materials given in Table 1.

Table 1. Recording material used and its details

Material	**Details**
Loudspeaker	Knowles HE-31751-000
Reference (aerial) microphone	B&K 1/2" type 2669
BCM	Knowles BU-23173-000 (accelerometer with an adapted electronic circuit as used by Zimpfer et al. [10])
Data recorder	TEAC Lx-10

2.1.2 Subjects

To determine the effect of BCM placement and the mouth opening on the wave transmission between the oral cavity and BCM (estimation of the transfer functions), ten volunteers aged from 23 to 57 participated in the experiment. Among the subjects, three were female.
To conduct the intelligibility test, listeners and talkers were needed. Four talkers aged from 25 to 49, whose half were female, took part in the experiment. All talkers were native French speakers, and none of them presented speaking disabilities. Twenty listeners who are also native French speakers accepted to do the intelligibility test. They were aged from 18 to 59, and none demonstrated

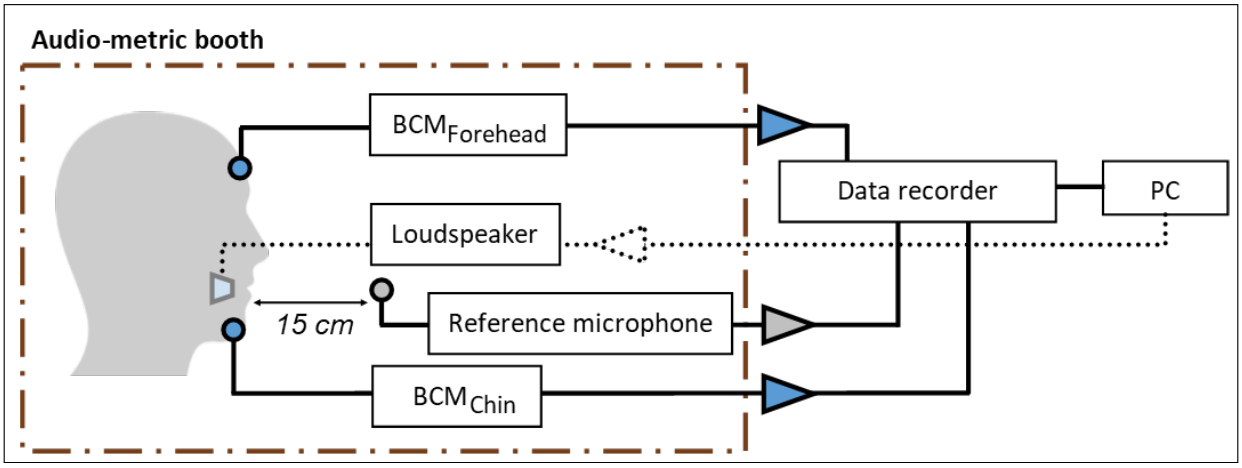

Figure 2. Scheme of the experimental setup.

hearing disorders.

Before starting the experiment, all subjects were aware of the complete protocol and agreed to participate. The experiments did not cause any physical damage to the subjects.

2.2 Methods

2.2.1 *Estimating change in transfer function between oral cavity sound pressure and skin acceleration without recording sound inside the mouth*

Recording sound inside the oral cavity is complicated since placing a microphone inside the mouth is difficult, and it records a local sound pressure not representative of the mean sound pressure in the mouth. To avoid these problems, we decided to record the speech sound at 15cm from the mouth. As the mouth radiation is the same when recording simultaneously at different head positions, it still enables us to see the influence of changing the microphone position on the transfer function between the oral cavity and BCM.

2.2.2 *Effect of BCM placement and mouth opening*

To evaluate the effect of the BCM placement, BCM were positioned on the subject at two locations: forehead and chin. The positions were chosen among the ones explored by Tran et al. [20, 24], which shows that placing BCM at the forehead should lead to higher intelligibility than the chin but weaker signal energy. The subjects equipped with the microphone sat in an audio-metric booth, and an aerial reference microphone was placed 15 cm away from their mouth. A data recorder controlled by a computer was used to record the signal of all microphones simultaneously with a sampling frequency of 48 kHz. The subjects put the loudspeaker inside their mouth at 1.8 cm behind the teeth and shaped their mouth like pronouncing the vowel [a]. The subject breathed with the nose bringing the palatal velum in contact with the tongue. The loudspeaker emitted a series of 0.4-second tonal sounds. The emitted frequency corresponded to the central frequencies of the sixth-octave band, starting from 250 Hz to 4490 Hz. A scheme of the experimental setup is presented in Figure 2.

The recorded signals were post-treated. For each signal frame containing a unique tone, the Fourier transform at the tone's frequency was determined. The spectra of the transfer functions between the reference microphone and the BCM were then derived by interpolating between the emitted frequencies.

To determine if the mouth opening influences the transfer functions, the experiment was conducted twice with the subject opening ”slightly” and ”normally” the mouth but keeping it in a [a] position. The two openings corresponded to a distance between the upper and lower incisors approximately equal to 1 cm and 2 cm.

2.2.3 *Intelligibility Test*

To conduct the intelligibility test, the same setup as in Figure 2 was used, except that no loudspeaker was placed in the mouth of the talker. The two mostly used intelligibility tests, the MRT [28] and the DRT [29], are rhyme tests, and both methods are described in [30]. Both methods use single-syllable word stimuli, which need to be recognized among six words and two words for the MRT and the DRT test, respectively. Both evaluate consonant discrimination in the initial position in single-syllable words but also in the final position in MRT. The advantage of the DRT over the MRT is that each choice involves single consonant distinctive feature discrimination, and the recognition score can be attributed to this difference. These features were first studied by Jakobson, Fant, and Halle [31]. Since bone conduction can be seen as a low-pass filter [32], the two features present in the French language that should be the most impacted were thought to be graveness and compactness. Therefore, the DRT was chosen since it allowed focusing on these two features and having a short intelligibility test. The paired consonants with these opposed features tested in the DRT are given here :

- Grave/Acute: [m]/[n]; [f]/[s]; [p]/[t]; [v]/[z]; [b]/[d]
- Compact/ Diffuse: [ʃ]/[s]; [k]/[p]; [k]/[t]; [ʒ]/[z]; [g]/[b]; [g]/[d]; [ʁ]/[l]

For each feature, a list of 12-word pairs was constituted. The apparition of the consonant pairs was homogeneously distributed, and nine phonetic vowels (six oral vowels and three nasal vowels) were used. All 24-word pairs were read by the four talkers and recorded by all the microphones simultaneously.

The recorded sounds were equalized with respect to their maximum amplitude before being presented to the listeners. The latter heard all the words randomly and chose among the two words composing the pair which one they recognized. They listened to all the sounds recorded by the different microphones, which made it possible to calculate the recognition rate for each microphone.

3. RESULTS

3.1 Transfer functions

Figure 3 and 4 present the transfer function between the reference microphone and the BCM placed at the forehead and chin, respectively. In both cases, the mouth was slightly opened. The mean transfer function for the ten subjects and the interval of two standard deviations centered on the mean were drawn in both figures. Figure 3 has a mean transfer function that is flat (varies in an interval of 5 dB) below 1 kHz and above the amplitude decrease by around 12 dB/octave. Figure 4 has a mean transfer function that decreases by 8 dB/octave below 1.6 kHz and becomes flatter at higher frequencies. The average standard deviation is around 6 dB for both transfer functions.

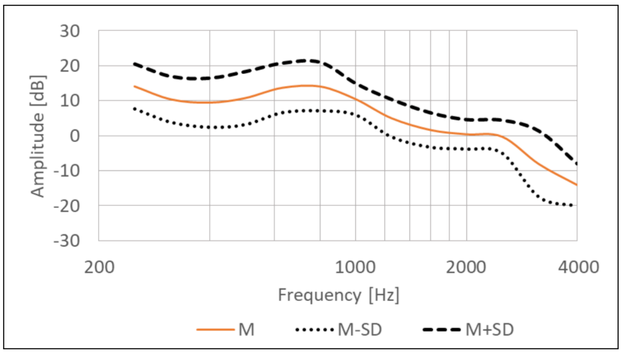

Figure 3. Transfer function between reference microphone and BCM at the forehead with the slightly opened mouth (M: Mean; SD: Standard Deviation).

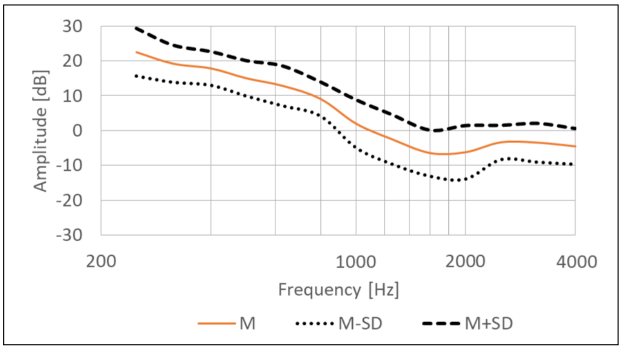

Figure 4. Transfer function between reference microphone and BCM at the chin with the slightly opened mouth (M: Mean; SD: Standard Deviation).

The same transfer functions were derived with the opened mouth. Figure 5 and 6 present the transfer function with the BCM placed at the forehead and at the chin, respectively. A decreasing trend can be observed in these two figures, and the average standard deviation is around 6 dB for both. Figure 5 has a mean transfer function that is flat (varies in an interval of 5 dB) between 315 Hz and 1 kHz. The amplitude decreases by around 11 dB/octave at higher frequencies. Figure 4 has a mean transfer function that decreases by 7 dB/octave below 1.6 kHz and becomes flatter at higher frequencies.

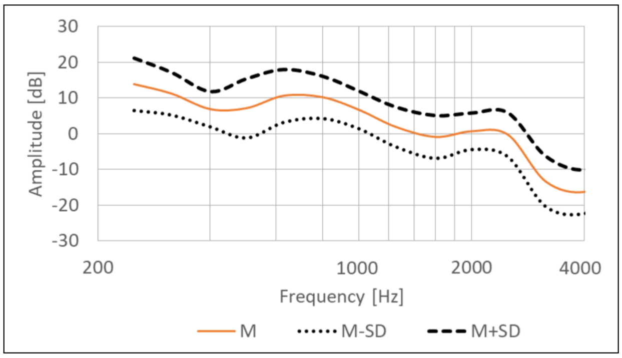

Figure 5. Transfer function between reference microphone and BCM at the forehead with the normally opened mouth (M: Mean; SD: Standard Deviation).

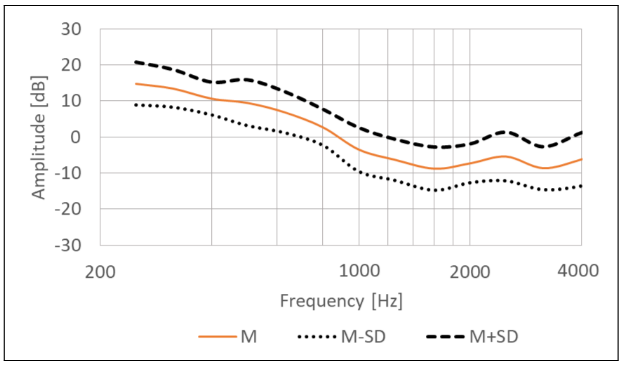

Figure 6. Transfer function between reference microphone and BCM at the chin with the normally opened mouth (M: Mean; SD: Standard Deviation).

3.2 Intelligibility Test

The subjective intelligibility test was conducted, and the recognition scores of 96 words (4 talkers * 24 words) of the 20 subjects depending on the microphone are represented by boxplots in Figure 7. The really high recognition score of the reference microphone confirms that the listener understands the test and that no mistake was made during the words recording process. The signals being recorded simultaneously, the consistency of the result given by the listeners are therefore not questioned. The number of correct answers when recording through reference microphone is for all the listeners higher than BCM microphones. It correlates with the results of the literature [9].

The reference microphone has the smallest spread of results, and the BCM at chin has the biggest spread among the three microphones. The mean recognition rate is 99% for the reference microphone, 86% for the microphone at the forehead, and 77% for the microphone at the chin. For 75% of the listener, the recognition rate when recording with the BCM at the forehead and at the chin is higher than 83% and lower than 82%, respectively. This result confirms that with the microphone used in this study, placing the BCM at the forehead location allows recording a more intelligible speech than at the chin, as shown for another BCM in [20, 22]

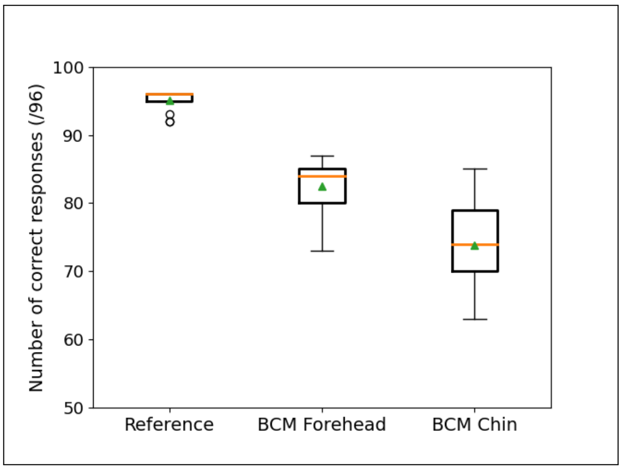

Figure 7. Recognition score of the single-syllable words depending on the microphone used to record speech (The green triangle represents the mean of the series).

4. DISCUSSION

The four transfer functions between the reference microphone and the BCM (Figure 3, Figure 4, Figure 5, Figure 6) can be seen as a low pass filter with different cut-off frequency and order depending on the position. This observation is in line with the result in the literature [32]. Indeed it is usually considered that above 2 kHz or 3 kHz, the BCM does not record enough speech signals. The influence of the microphone's position was observed by calculating the differences in dB of the two transfer functions of each subject with the slightly opened mouth. This result also refers to the transfer function between the skin acceleration at the chin and the forehead, which can be observed in Figure 8.

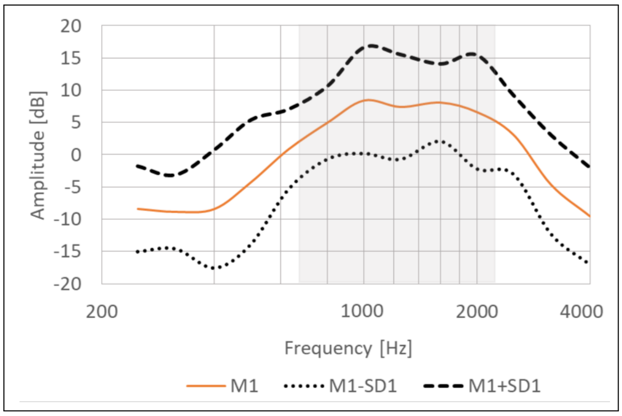

Figure 8. Transfer function between the skin vibration at the chin and the one at the forehead. Thin-darkened frequency band corresponds to the bands where the mean is above 3 dB (M: Mean; SD: Standard Deviation).

Figure 8 shows that the mean transfer function between skin vibration at the chin and the forehead is above 3 dB between 800 Hz and 2500 Hz. It demonstrates that changing the position of the BCM from the chin to the forehead leads to better recording of the sound frequency between 800Hz and 2500 Hz. These frequencies are central for the speech signals since much speech information is contained in this frequency band. This result demonstrates with a physical explanation that the forehead is a BCM's location to prefer over the chin. This better transmission could be due to the thinner layer of soft tissues between the BCM and the bone at this location [33]. Another reason could be that the wave propagates easier from the oral cavity to the skull through the palatal bone than to the mandible. The subjective intelligibility test confirms that the speech is more intelligible when recorded with the BCM placed at the forehead than the chin and could be explained by this better transmission. However, it is only one of the reasons, and others, as the structural vibration of the vocal fold that could deteriorate the recorded speech signal, should also be studied.

Estimation of the transfer function between skin vibration at the chin and the forehead is also conducted for the normally opened mouth case to observe the influence of the mouth opening. Figure 9 gathers in one graph the transfer functions corresponding to the two mouth opening with index 1 (thin lines) corresponding to the slightly opened mouth and index 2 (thick lines) to the normally opened mouth. The modification of the mouth opening doesn't change the previous result, as seen in Figure 9. Moreover, the change in the mouth opening does not change the trend of the transfer function between the reference microphone and BCM. Some differences that seem to appear between Figure 3 and Figure 5 and Figure 4 and Figure 6 are assumed to be due to the change in the radiation effect. This assumption seems fair considering Figure 9. Therefore the transfer function between oral cavity sound pressure and skin acceleration does not seem to be influenced by the mouth opening, which indicates that mouth opening does not affect the wave transmission between the oral cavity and BCM position.

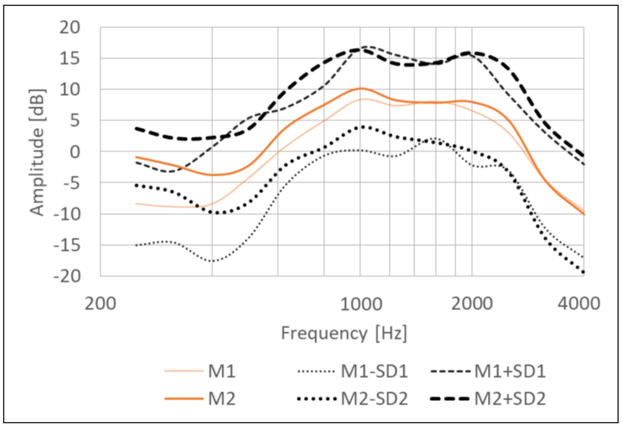

Figure 9. Influence of mouth opening on the transfer function between the skin vibration at the chin and the one at the forehead (M: Mean; SD: Standard Deviation; Index 2 corresponds to a bigger mouth opening than Index 1).

5. CONCLUSION

In this paper, we investigated the difference in the transfer function between oral cavity sound pressure and BCM response implied by changing two parameters: the bone conduction microphone position (chin and forehead) and the mouth opening. The results demonstrate that the wave transmission from the oral cavity to the forehead in the frequency band [800; 2500] Hz is higher than from the oral cavity to the chin. It could explain the result of the subjective intelligibility test conducted in this paper which shows a higher intelligibility score when recording at the forehead than at the chin. This relative transmission of the oral cavity to the forehead compared to the chin was unchanged when modifying the mouth opening. The latter was finally considered to have no influence on the propagation efficiency from the oral cavity to BCM locations. Other locations could be studied and compared to the present one to determine the best BCM location based on the objective criteria: better transmission from the oral cavity to BCM in the speech frequency band.

6. ACKNOWLEDGMENTS

The authors would like to express their sincere gratitude to all the volunteers who took their time to make this research possible.

7. REFERENCES

[1] G. v. Békésy, "Note on the definition of the term: Hearing by bone conduction," vol. 26, pp. 106–107, jan 1954.

[2] J. Tonndorf, "Bone conduction. studies in experimental animals.," *Acta oto-laryngologica*, pp. Suppl 213:1+, 1966.

[3] S. Stenfelt, "Acoustic and physiologic aspects of bone conduction hearing," pp. 10–21, KARGER, 2011.

[4] H. Hosoi, "Approach in the use of cartilage conduction speaker," *Japanese patent*, vol. 4, 2004.

[5] T. Nishimura, H. Hosoi, O. Saito, R. Miyamae, R. Shimokura, T. Matsui, T. Yamanaka, and H. Levitt, "Is cartilage conduction classified into air or bone conduction?," vol. 124, pp. 1214–1219, nov 2013.

[6] H. Sohmer, "Soft tissue conduction: Review, mechanisms, and implications," *Trends in Hearing*, vol. 21, p. 233121651773408, oct 2017.

[7] B. Stuart, "Throat microphone." Patent US2165124, 1939.

[8] S. Zhang, Y. Sugiura, N. Yasui, and T. Shimamura, "Quantifying noise robustness of bone-conducted speech," in *2020 IEEE 63rd International Midwest Symposium on Circuits and Systems (MWSCAS)*, IEEE, aug 2020.

[9] B. E. Acker-Mills, A. J. M. Houtsma, and W. A. Ahroon, "Speech intelligibility with acoustic and contact microphones," 2005.

[10] V. Zimpfer and K. Buck, "Ambiguity in the recognition of phonetic vowels when using a bone conduction microphone," *The Journal of the Acoustical Society of America*, vol. 123, pp. 3067–3067, may 2008.

[11] K. Kondo, T. Fujita, and K. Nakagawa, "On equalization of bone conducted speech for improved speech quality," in *2006 IEEE International Symposium on Signal Processing and Information Technology*, IEEE, aug 2006.

[12] T. Shimamura and T. Tamiya, "A reconstruction filter for bone-conducted speech," in *48th Midwest Symposium on Circuits and Systems, 2005.*, IEEE, 2005.

[13] T. Shimamura and T. Tomikura, "Quality improvement of bone-conducted speech," in *Proceedings of the 2005 European Conference on Circuit Theory and Design, 2005.*, IEEE.

[14] S. Y. Won and J. Berger, "Estimating transfer function from air to bone conduction using singing voice," in *ICMC*, 2005.

[15] M. S. Rahman and T. Shimamura, "Intelligibility enhancement of bone conducted speech by an analysis-synthesis method," in *2011 IEEE 54th International Midwest Symposium on Circuits and Systems (MWSCAS)*, IEEE, aug 2011.

[16] M. T. Turan and E. Erzin, "Enhancement of throat microphone recordings by learning phone-dependent mappings of speech spectra," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, IEEE, may 2013.

[17] D. Shan, X. Zhang, C. Zhang, and L. Li, "A novel encoder-decoder model via NS-LSTM used for bone-conducted speech enhancement," *IEEE Access*, vol. 6, pp. 62638–62644, 2018.

[18] H.-P. Liu, Y. Tsao, and C.-S. Fuh, “Bone-conducted speech enhancement using deep denoising autoencoder,” *Speech Communication*, vol. 104, pp. 106–112, nov 2018.

[19] Q.-H. Nguyen and M. Unoki, “Bone-conducted speech enhancement using vector-quantized variational autoencoder and gammachirp filterbank cepstral coefficients,” in *2022 30th European Signal Processing Conference (EUSIPCO)*, IEEE, aug 2022.

[20] P. Tran, T. Letowski, and M. McBride, “Bone conduction microphone: Head sensitivity mapping for speech intelligibility and sound quality,” in *2008 International Conference on Audio, Language and Image Processing*, IEEE, jul 2008.

[21] S. Nanri, T. Shinobu, S. Otsuka, and S. Nakagawa, “Intelligibility of bone-conducted speech detected on the scalp assessed by mono-syllable articulation and speech transmission index,” in *2021 43rd Annual International Conference of the IEEE Engineering in Medicine Biology Society*, IEEE, nov 2021.

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

[23] K. A. Pollard, P. K. Tran, and T. Letowski, “Morphological differences affect speech transmission over bone conduction,” *The Journal of the Acoustical Society of America*, vol. 141, pp. 936–944, feb 2017.

[24] P. K. Tran, T. R. Letowski, and M. E. McBride, “The effect of bone conduction microphone placement on intensity and spectrum of transmitted speech items,” *The Journal of the Acoustical Society of America*, vol. 133, pp. 3900–3908, jun 2013.

[25] 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,” *The Journal of the Acoustical Society of America*, vol. 30, pp. 275–277, apr 1958.

[26] M. S. Rahman and T. Shimamura, “A study on amplitude variation of bone conducted speech compared to air conducted speech,” in *2013 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference*, IEEE, oct 2013.

[27] M. S. Rahman and T. Shimamura, “Amplitude variation of bone-conducted speech compared with air-conducted speech,” *Acoustical Science and Technology*, vol. 40, pp. 293–301, sep 2019.

[28] A. S. House, C. E. Williams, M. H. L. Hecker, and K. D. Kryter, “Articulation-testing methods: Consonantal differentiation with a closed-response set,” *The Journal of the Acoustical Society of America*, vol. 37, pp. 158–166, jan 1965.

[29] W. Voiers, “The present state of digital vocoding technique: A diagnostic evaluation,” *IEEE Transactions on Audio and Electroacoustics*, vol. 16, pp. 275–279, jun 1968.

[30] International Telecommunication Union, “Recommendation ITU-T P.862: Subjective test methodology for assessing speech inteligibility,” 2016.

[31] R. Jakobson, G. Fant, and M. Halle, *Preliminaries to Speech Analysis*. MIT Press, 1951.

[32] M. Kumashita, T. Shimamura, and J. Suzuki, “Property of voice recorded by bone-conduction microphones,” in *1996 spring meeting on Acoust. Soc. Jpn*, pp. 269–270, 1996.

[33] H. Huempfner-Hierl, A. Bohne, A. Schaller, G. Wollny, and T. Hierl, “Does facial soft tissue protect against zygomatic fractures? results of a finite element analysis,” *Head Face Medicine*, vol. 11, jun 2015.