Interspeech 2018
2-6 September 2018, Hyderabad

Bone-Conduction Sensor Assisted Noise Estimation for Improved Speech Enhancement

Ching-Hua Lee, Bhaskar D. Rao, and Harinath Garudadri

Department of Electrical and Computer Engineering
University of California, San Diego
{chl438, brao, hgarudadri}@ucsd.edu

Abstract

State-of-the-art noise power spectral density (PSD) estimation techniques for speech enhancement utilize the so-called speech presence probability (SPP). However, in highly non-stationary environments, SPP-based techniques could still suffer from inaccurate estimation, leading to significant amount of residual noise or speech distortion. In this paper, we propose to improve speech enhancement by deploying the bone-conduction (BC) sensor, which is known to be relatively insensitive to the environmental noise compared to the regular air-conduction (AC) microphone. A strategy is suggested to utilized the BC sensor characteristics for assisting the AC microphone in better SPP-based noise estimation. To our knowledge, no previous work has incorporated the BC sensor in this noise estimation aspect. Consequently, the proposed strategy can possibly be combined with other BC sensor assisted speech enhancement techniques. We show the feasibility and potential of the proposed method for improving the enhanced speech quality by both objective and subjective tests.

Index Terms: bone-conduction (BC), noise power spectral density (PSD) estimation, speech presence probability (SPP), speech enhancement

1. Introduction

Most of the modern speech enhancement systems are developed in the time-frequency (T-F) domain via the short-time Fourier transform (STFT) [1]. In general, the goal is to estimate the clean speech STFT coefficients from the noisy observation [2]. Performance of the clean speech spectral estimator is heavily dependent on the given noise power spectral density (PSD) estimate [3]. For single-microphone systems, many noise PSD estimation techniques have been proposed [4]. The earliest methods rely on certain mechanism to detect speech/non-speech segments often referred to as voice activity detection (VAD) [5]. Martin [6] proposed the minimum statistics approach which estimates the noise PSD without a VAD. Later, Cohen [7] introduced the concept of speech presence probability (SPP) to realize a soft-decision update of the noise PSD estimate called the minima controlled recursive averaging (MCRA). Rangachari and Loizou [8] proposed an MCRA-based approach for highly non-stationary environmental noise, referred to as MCRA-2. More recently, Gerkmann and Hendriks [9] proposed an unbiased minimum mean square error (MMSE) noise power estimator that also utilizes the SPP, referred to as MMSE-SPP.

Though many approaches have been proposed, it is still difficult to track the background noise in a highly non-stationary, low signal-to-noise ratio (SNR) environment, especially from only one channel of audio signals. Even the state-of-the-art SPP-based techniques still suffer from inaccurate estimation when the noise characteristics change abruptly. To further improve the speech enhancement system, additional information such as that provided by a different type of sensor is needed for better tracking of noise power.

The bone-conduction (BC) sensor is one of the special sensors that can be used to improve noise suppression in adverse environments. Different from the regular air-conduction (AC) microphone, the BC sensor is comparatively insensitive to the environmental noise since it collects the vibration of sounds through bones of the skull, instead of through the air. As an example for demonstrating the difference, we show in Figure 1 the spectrograms of a sentence recorded simultaneously by an AC microphone and a BC sensor in a noisy environment. We can see that the BC signal is relatively noise-free as compared to the AC signal. However, the main drawback of using the BC sensor is that the high frequency components (>4kHz) are significantly attenuated due to transmission loss. This makes in a noiseless environment, the BC signal has worse quality than the AC signal due to distortion caused by the absence of the high frequency portion. The resulting BC sensor signal would sound muffled and unnatural and thus is not suitable for direct use.

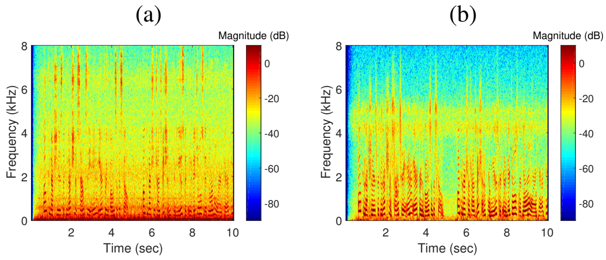

Figure 1: *Spectrograms of (a) the AC microphone signal and (b) the BC sensor signal.*

Several techniques have been proposed to utilize the BC sensor for speech enhancement [10]. There are basically two options: One is to estimate and exploit the non-linear mapping between the BC sensor and the AC microphone [11, 12, 13, 14, 15, 16, 17]. However, an off-line training process is often required. The other option is to take advantage of the special characteristics of the BC sensor for assisting noise reduction of the AC microphone signal. As an aiding sensor, the BC sensor has been proposed to improve speech enhancement of the regular microphone signal in several aspects, e.g., *a priori* SNR estimation [18], direct filtering [19, 20], graphical model based approach [21], low frequency band noise suppression [22], etc.

In this paper, we present the aspect of improving noise PSD estimation for speech enhancement systems with a regular AC microphone by incorporating the BC sensor. To our knowledge, no previous work has utilized BC sensor in the SPP-based noise

10.21437/Interspeech.2018-1046

estimation framework. Therefore, the proposed strategy can potentially be combined with other BC sensor assisted techniques, e.g., the *a priori* SNR estimation in [18]. We show that the special characteristics of the BC sensor allow development of a suitable strategy to mitigate the problem of inaccurate estimation of the noise PSD. It is verified by both objective and subjective tests that the proposed strategy improves the speech quality of the enhanced signal.

2. Signal model

Let n denote the discrete time index. In conventional speech enhancement for a regular AC microphone, we aim to recover the clean speech $x(n)$ from the noisy observation $y(n) = x(n) + v(n)$, where $v(n)$ is an additive noise. Speech enhancement is performed in the T-F domain via the STFT, as depicted in Figure 2. Let $Y(k,m)$, $X(k,m)$, $V(k,m)$ be the STFT coefficients of $y(n)$, $x(n)$, and $v(n)$, respectively, where $k \in \{0, 1, ..., N-1\}$ is the frequency index and m is the frame index. We have $Y(k,m) = X(k,m) + V(k,m)$. Generally, the goal is to come up with a proper real-valued gain function $G(k,m)$ such that an estimate $\hat{X}(k,m) = G(k,m)Y(k,m)$ can approach the clean speech STFT coefficients $X(k,m)$ in some optimal sense. Then the estimated $\hat{X}(k,m)$ is transformed back to the time domain via the inverse STFT (iSTFT) to obtained the enhanced signal $\hat{x}(n)$. For computing the gain function $G(k,m)$, several techniques can be employed, e.g., spectral subtraction [23], Wiener filtering [24], and the Bayesian short-time spectral amplitude estimator [25]. However, all these approaches have one thing in common; they are in general a function of the noise PSD, $\sigma_V^2(k,m) = E[|V(k,m)|^2]$, either directly or indirectly [3]. Therefore, noise PSD estimation is a crucial part of most STFT domain based speech enhancement algorithms, and, in particular, challenging when speech is corrupted by non-stationary noise. In this paper we consider the SPP-based noise estimation.

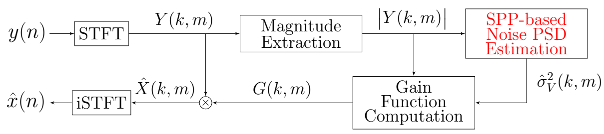

Figure 2: *Block diagram of STFT domain based speech enhancement system.*

3. SPP-based noise estimation

In SPP-based noise estimation, the noise PSD estimate is updated with a soft weighting between the previous estimate $\hat{\sigma}_V^2(k, m-1)$ and the current noisy observation $|Y(k,m)|^2$ as:

$$\begin{aligned}\hat{\sigma}_V^2(k,m) =& \beta(k,m)\hat{\sigma}_V^2(k,m-1) \\ &+ (1-\beta(k,m))|Y(k,m)|^2,\end{aligned} \quad (1)$$

where a T-F dependent smoothing factor

$$\beta(k,m) = \beta_{min} + (1-\beta_{min})p(k,m) \quad (2)$$

is utilized to control the rate of update. In (2), $\beta_{min} \geq 0$ is a constant so that $\beta_{min} \leq \beta(k,m) \leq 1$ and $p(k,m)$ represents the SPP in the (k,m)-th T-F bin that can be estimated differently by several approaches. We can see in (2) that $\beta(k,m)$ is a function of $p(k,m)$: If speech is highly likely to be present in a particular bin (i.e., $p(k,m) \to 1$), then $\beta(k,m) \to 1$ so that the noise PSD estimate is merely updated. On the contrary, when speech is almost absent (i.e., $p(k,m) \to 0$), then we will have $\beta(k,m) \to \beta_{min}$ so that the noise PSD estimate is updated at the highest rate.

3.1. Existing SPP estimation algorithms

Equations (1) and (2) provide a general methodology on how the noise PSD estimate can be updated given the SPP, which can be obtained in a recursive manner [7, 8] or determined from probabilistic modeling [9]. We present two example algorithms.

3.1.1. MCRA-2 [8]

In this approach the SPP is estimated recursively as:

$$p(k,m) = \lambda\, p(k,m-1) + (1-\lambda)\, I(k,m), \quad (3)$$

where λ is a smoothing factor and $I(k,m)$ is a binary indicator for speech presence ($=1$) and absence ($=0$). To determine $I(k,m)$, it first computes the ratio of the smoothed noisy speech PSD to its local minimum which is estimated based on minimum statistics principles. Then speech presence/absence is determined by comparing the ratio to a pre-defined frequency dependent threshold.

3.1.2. MMSE-SPP [9]

In this approach noise estimation is treated as an MMSE optimization problem. Under complex Gaussian distribution assumptions of the noise and speech spectral coefficients, the SPP is derived by probabilistic modeling of the speech presence and absence likelihood functions and priors. The SPP in this approach is given by:

$$p(k,m) = \left(1 + (1+\xi_{\mathcal{H}_1})\, e^{-\frac{|Y(k,m)|^2}{\hat{\sigma}_V^2(k,m-1)}\frac{\xi_{\mathcal{H}_1}}{1+\xi_{\mathcal{H}_1}}}\right)^{-1}, \quad (4)$$

where $\xi_{\mathcal{H}_1}$ is the representative SNR value for speech presence and a fixed value of $10\log_{10}(\xi_{\mathcal{H}_1}) = 15$ dB is suggested. An additional mechanism is further employed to avoid stagnation of the algorithm as suggested in [9].

3.2. Tracking delay and speech leakage

Two major problems with noise estimation have been identified: The noise PSD can be either underestimated or overestimated. Underestimation will lead to an under-suppression of noise content, resulting in an unfavorably large amount of residual noise. This usually happens when the noise level rises abruptly and results in a tracking delay of the noise PSD estimator. On the other hand, overestimation generally leads to an over-suppression of noise content which causes speech distortion. This usually happens when strong speech components are present and mistaken as noise, resulting in speech leakage into the noise PSD estimation. Under a highly non-stationary environment where the noise power changes rapidly, even the state-of-the-art SPP-based methods could fail to properly deal with the two issues. As an example, Figure 3 illustrates these phenomenons observed in the MCRA-2 and MMSE-SPP for a particular frequency bin across different time frames.

4. Proposed method

To overcome the aforementioned two problems in noise estimation, we propose a two-stage strategy in which the BC sensor is

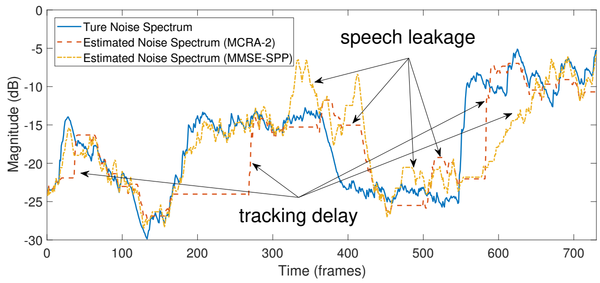

Figure 3: *Illustration of tracking delay and speech leakage.*

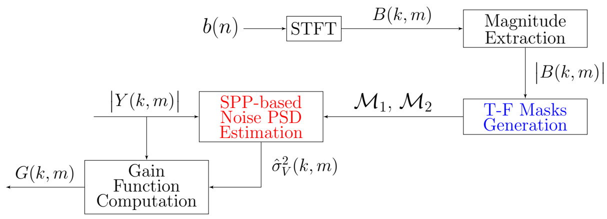

Figure 4: *The proposed BC sensor assisted scheme.*

used as an aiding sensor in assisting the SPP-based techniques by exploiting its special characteristics. In this section we first describe the proposed strategy and then discuss the details of how the BC sensor is incorporated.

4.1. The two-stage strategy

4.1.1. Tracking delay mitigation

In the first stage, the incoming m-th signal frame is analyzed to decide if it contains only noise or not. If it is noise-only, then we force the SPP to be zero for all the N frequency bins, i.e., $p(k,m) = 0$, for $k = 0, 1, ..., N-1$. By doing so, we will have $\beta(k,m) = \beta_{min}$ for all the frequency bins of the frame, ensuring the noise PSD estimate is updated at the highest rate. The motivation behind is that noise is always present but speech is not, which indicates one would like to be more aggressive in terms of updating the noise power estimate when there is only noise within a frame.

4.1.2. Speech leakage alleviation

In the second stage, the SPP is first computed using existing approaches such as the MCRA-2 and MMSE-SPP. Then we determine if a T-F bin contains strong speech content or not. If it does, then we set the corresponding SPP to 1, i.e., $p(k,m) = 1$ for the (k,m)-th bin. This makes the corresponding $\beta(k,m) = 1$ and thus the noise PSD estimate of the particular T-F bin will not be updated. This step acts as a mechanism to prevent strong speech content from leaking into the update of the noise estimate.

4.2. Incorporating the BC sensor

From the above, we see that good detectors for speech presence/absence and for strong speech components are required. This is where the BC sensor comes into play. Figure 4 presents the proposed scheme to utilize the BC sensor characteristics. We use two binary T-F masks, $\mathcal{M}_1$ and $\mathcal{M}_2$, obtained as:

$$\mathcal{M}_1(k,m) = \begin{cases} 1, & \text{if } \left|B(k,m)\right| > t_1 \\ 0, & \text{otherwise} \end{cases}, \tag{5}$$

$$\mathcal{M}_2(k,m) = \begin{cases} 1, & \text{if } \left|B(k,m)\right| > t_2 \\ 0, & \text{otherwise} \end{cases}, \tag{6}$$

where $\left|B(k,m)\right|$ is the spectral magnitude of the BC sensor signal $b(n)$ and t_1 and t_2 are positive threshold values.

Note that $\mathcal{M}_1$ is used to detect noise-only frames for mitigating tracking delay: If in a particular frame the number of T-F bins with a mask value of 1 is smaller than a pre-defined tolerance parameter $\tau > 0$, which is to account for the small amount of noise picked up by the BC sensor, then the frame is declared to be noise-only. Figure 5 (a) presents an example of $\mathcal{M}_1$ obtained with $t_1 = 0.7$. In this example, a frame with fewer then $\tau = 4$ T-F bins with a mask value of 1 is determined as a noise-only frame. On the other hand, $\mathcal{M}_2$ is used to alleviate speech leakage by declaring strong speech content if a T-F bin has a mask value of 1. An example of $\mathcal{M}_2$ obtained with $t_2 = 6.5$ is shown in Figure 5 (b). In this case, any T-F bin with a mask value of 1 is declared to have strong speech content.

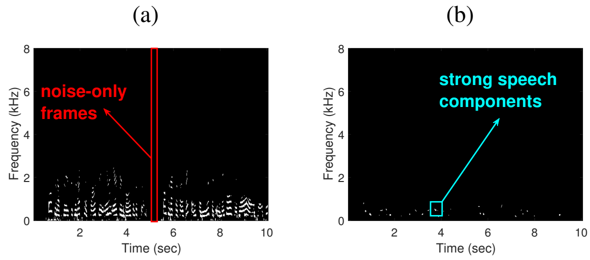

Figure 5: *The two T-F masks: (a) $\mathcal{M}_1$: to detect noise-only frames and (b) $\mathcal{M}_2$: to identify strong speech components.*

We summarize our BC-assisted versions of MCRA-2 and MMSE-SPP in Algorithm 1 and Algorithm 2, respectively. As an example, Figure 6 shows the spectrograms of the enhanced signals of the noisy AC signal in Figure 1 (a) by using the original MCRA-2 and MMSE-SPP, and their corresponding BC-assisted versions utilizing the BC signal in Figure 1 (b).

Algorithm 1: BC-assisted MCRA-2

```
for each time frame m do
    Update the smoothed noisy PSD as [8, eq. 2];
    Minimum tracking using [8, eq. 3];
    Obtain M1(k,m) and M2(k,m) using (5) and (6);
    if sum_{k=0}^{N-1} M1(k,m) < τ then
        p(k,m) ← 0,   for  k = 0, 1, ..., N − 1 ;
    else
        Compute the power ratio as [8, eq. 4];
        Determine if speech is present by [8, eq. 5];
        Update the SPP p(k,m) using (3);
        if M2(k,m) = 1 then
            p(k,m) ← 1;
        end if
    end if
    Compute the smoothing factor β(k,m) as (2);
    Update the noise PSD estimate σ̂²_V(k,m) as (1);
end for
```

Algorithm 2: BC-assisted MMSE-SPP

```
for each time frame m do
    Obtain M_1(k,m) and M_2(k,m) using (5) and (6);
    if Σ_{k=0}^{N-1} M_1(k,m) < τ then
        p(k,m) ← 0,   for   k = 0,1,...,N−1 ;
    else
        Compute the SPP p(k,m) using (4);
        Stagnation avoidance using [9, eq. 23, 24];
        if M_2(k,m) = 1 then
            p(k,m) ← 1;
        end if
    end if
    Compute the smoothing factor β(k,m) as (2);
    Update the noise PSD estimate σ̂_V^2(k,m) as (1);
end for
```

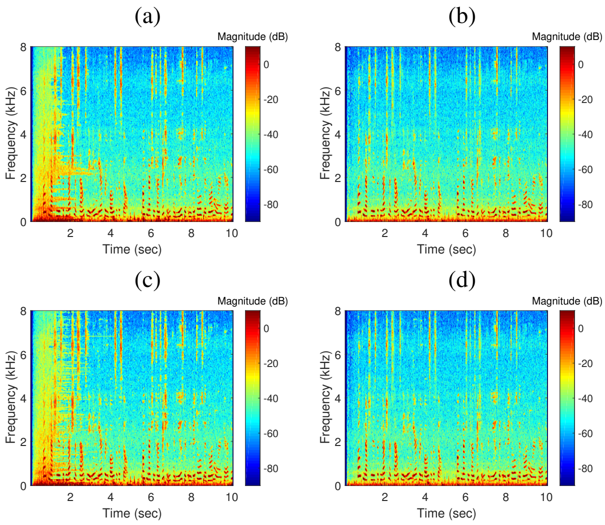

Figure 6: *Spectrograms of enhanced signals using (a) MCRA-2, (b) BC-assisted MCRA-2, (c) MMSE-SPP, and (d) BC-assisted MMSE-SPP.*

5. Simulation results

We compare the proposed BC-assisted versions with the original MCRA-2 and MMSE-SPP and show improvements can be achieved. The experimental setup was as follows: The sampling frequency was 16 kHz. A Hanning window of 512 samples was used with 50% overlap between two consecutive frames. The number of FFT points was $N = 512$. The Wiener filter was employed for computing the gain function with the decision-directed SNR estimator [25]. Parameters of the MCRA-2 and MMSE-SPP were chosen the same as used in their respective original papers. For the T-F masks, we used $t_1 = 0.7, t_2 = 6.5$, and $\tau = 4$. The dataset was provided by Sonion including recordings from regular AC microphones and from their BC sensor – the Voice Pick Up (VPU) sensor. The test signals were sentences recorded in several environments under different SNRs and were spoken by multiple male and female speakers.

An objective test was conducted for evaluation purpose. We considered the speech-to-reverberation modulation energy ratio (SRMR) [26]. We chose to use its extended version, the $\text{SRMR}_{\text{norm}}$ proposed in [27], which has been shown relatively reliable for assessing speech enhancement performance [28]. The obtained numbers are shown in Table 1. Higher values indicate better quality. For comparison with existing BC sensor assisted techniques, we considered the approach proposed in [22] which suggests an intuitive way of utilizing the BC sensor characteristics for noise suppression. The method combines the lower frequency band of the BC sensor signal with the higher frequency band of the AC microphone signal to give the enhanced output. We used the same settings of parameters as in [22] for this approach. From Table 1 we can see that with the aid of the BC sensor, quality improvements can be achieved for the two presented SPP-based methods as well as the existing approach.

Table 1: *Quality in terms of SRMR_{norm} of the objective test.*

Recording Environ.	AC Signal	MCRA-2	MCRA-2 BC-assisted	MMSE-SPP	MMSE-SPP BC-assisted	Method of [22]
Fan & wind	1.94	3.10	**3.92**	2.96	**3.73**	3.12
Cafe 1	2.64	3.42	**3.72**	3.51	**3.69**	3.21
Cafe 2	1.78	2.67	**3.07**	2.66	**3.09**	2.96
Cafe 3	1.64	2.42	**3.00**	2.62	**3.02**	2.98
Car 1	2.36	3.09	**3.62**	3.21	**3.43**	3.18
Car 2	1.44	2.34	**3.05**	2.79	**2.99**	2.82
Cocktail 1	1.81	2.43	**2.82**	2.59	**2.92**	3.10
Cocktail 2	2.24	3.08	**3.99**	3.38	**4.08**	2.78
Cocktail 3	2.82	3.43	**3.79**	3.14	**3.46**	2.86
Average	2.07	2.89	**3.44**	2.98	**3.38**	3.00

To further verify the superiority of the proposed approach over the MCAR-2 and MMSE-SPP, a subjective test was also conducted. Nine subjects were asked to perform an informal preference test. The participants were presented with pairs of sentences via a headphone, one processed with the original method (MCRA-2 or MMSE-SPP) and the other processed with the corresponding BC-assisted version. For each pair the sentences were played to the listener in random order so that they would not know which one was which. They were asked to select one from each pair of the sentences that had a better overall speech quality. A third option of "No Preference" was also available if they did not favor one over the other. The results shown in Table. 2 indicate that the BC-assisted methods can result in more perceptually desirable speech.

Table 2: *Preference scores of the subjective test.*

Techniques	No Preference	Original	BC-assisted
MCRA-2	17.28 %	20.99 %	61.73 %
MMSE-SPP	9.87 %	25.93 %	64.20 %
Total	**13.58** %	**23.46** %	**62.96** %

6. Conclusions

In this paper, the special characteristics of the BC sensor are exploited to assist the SPP-based noise estimation for improved speech enhancement. Two T-F masks are generated from the BC sensor signal to reduce tracking delay and speech leakage caused by underestimation and overestimation of the noise PSD. It has been verified by both objective and subjective tests that the proposed strategy provides significant improvements to the quality of the enhanced signal.

7. Acknowledgements

This work was partially supported by NIH/NIDCD grants and #1R01DC015436 and #1R33DC015046. We are also grateful to Sonion for providing the dataset used in the experiments.

8. References

[1] P. C. Loizou, *Speech enhancement: theory and practice*. 2nd edition, CRC press, 2013.

[2] J. Benesty, J. Chen, and E. A. P. Habets, *Speech enhancement in the STFT domain*. Springer Science & Business Media, 2011.

[3] M. Parchami, W.-P. Zhu, B. Champagne, and E. Plourde, "Recent developments in speech enhancement in the short-time fourier transform domain," *IEEE Circuits Syst. Mag.*, vol. 16, no. 3, pp. 45–77, 2016.

[4] R. C. Hendriks, T. Gerkmann, and J. Jensen, "DFT-domain based single-microphone noise reduction for speech enhancement: A survey of the state of the art," *Synthesis Lectures on Speech and Audio Processing*, vol. 9, no. 1, pp. 1–80, 2013.

[5] S. Van Gerven and F. Xie, "A comparative study of speech detection methods," in *Proc. Europ. Conf. Speech Commun. Techn. (Eurospeech)*, vol. 97, 1997.

[6] R. Martin, "Noise power spectral density estimation based on optimal smoothing and minimum statistics," *IEEE Trans. Speech Audio Process.*, vol. 9, no. 5, pp. 504–512, 2001.

[7] I. Cohen, "Noise spectrum estimation in adverse environments: Improved minima controlled recursive averaging," *IEEE Trans. Speech Audio Process.*, vol. 11, no. 5, pp. 466–475, 2003.

[8] S. Rangachari and P. C. Loizou, "A noise-estimation algorithm for highly non-stationary environments," *Speech Commun.*, vol. 48, no. 2, pp. 220–231, 2006.

[9] T. Gerkmann and R. C. Hendriks, "Unbiased MMSE-based noise power estimation with low complexity and low tracking delay," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 20, no. 4, pp. 1383–1393, 2012.

[10] H. S. Shin, H.-G. Kang, and T. Fingscheidt, "Survey of speech enhancement supported by a bone conduction microphone," in *Proc. ITG Symp. Speech Commun.*, 2012, pp. 1–4.

[11] 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 *Proc. IEEE Work. Autom. Speech Recognit. Underst. (ASRU)*, 2003, pp. 249–254.

[12] Z. Zhang, Z. Liu, M. Sinclair, A. Acero, L. Deng, J. Droppo, X. Huang, and Y. Zheng, "Multi-sensory microphones for robust speech detection, enhancement and recognition," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, vol. 3, 2004, pp. iii–781.

[13] T. Tamiya and T. Shimamura, "Reconstruction filter design for bone-conducted speech," in *Proc. Int. Conf. Spoken Lang. Process. (Interspeech-ICSLP)*, 2004.

[14] S. Srinivasan and P. Kechichian, "Robustness analysis of speech enhancement using a bone conduction microphone - preliminary results," in *Proc. Int. Work. Acoust. Signal Enhancement (IWAENC)*, 2012, pp. 1–4.

[15] P. Kechichian and S. Srinivasan, "Model-based speech enhancement using a bone-conducted signal," *J. Acoust. Soc. Am.*, vol. 131, no. 3, pp. EL262–EL267, 2012.

[16] M. Li, I. Cohen, and S. Mousazadeh, "Multisensory speech enhancement in noisy environments using bone-conducted and air-conducted microphones," in *Proc. IEEE China Summit Int. Conf. Signal Inform. Process. (ChinaSIP)*, 2014, pp. 1–5.

[17] B. Huang, Y. Gong, J. Sun, and Y. Shen, "A wearable bone-conducted speech enhancement system for strong background noises," in *Proc. IEEE Int. Conf. Electronic Packaging Technology (ICEPT)*, 2017, pp. 1682–1684.

[18] H. S. Shin, T. Fingscheidt, and H.-G. Kang, "*A priori* SNR estimation using air- and bone-conduction microphones," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 23, no. 11, 2015.

[19] Z. Liu, Z. Zhang, A. Acero, J. Droppo, and X. Huang, "Direct filtering for air- and bone-conductive microphones," in *IEEE Int. Work. Multimedia Signal Process. (MMSP)*, 2004, pp. 363–366.

[20] 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. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, vol. 1, 2005, pp. I–1093.

[21] A. Subramanya, Z. Zhang, Z. Liu, J. Droppo, and A. Acero, "A graphical model for multi-sensory speech processing in air-and-bone conductive microphones," in *Proc. Europ. Conf. Speech Commun. Techn. (Interspeech-Eurospeech)*, 2005.

[22] M. S. Rahman, A. Saha, and T. Shimamura, "Low-frequency band noise suppression using bone conducted speech," in *Proc. IEEE Pacific Rim Conf. Commun., Computers, Signal Process. (PacRim)*, 2011, pp. 520–525.

[23] S. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 27, no. 2, pp. 113–120, 1979.

[24] R. McAulay and M. Malpass, "Speech enhancement using a soft-decision noise suppression filter," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 28, no. 2, pp. 137–145, 1980.

[25] Y. Ephraim and D. Malah, "Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 32, no. 6, pp. 1109–1121, 1984.

[26] T. H. Falk, C. Zheng, and W.-Y. Chan, "A non-intrusive quality and intelligibility measure of reverberant and dereverberated speech," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 18, no. 7, pp. 1766–1774, 2010.

[27] J. F. Santos, M. Senoussaoui, and T. H. Falk, "An improved non-intrusive intelligibility metric for noisy and reverberant speech," in *Proc. Int. Work. Acoust. Signal Enhancement (IWAENC)*, 2014, pp. 55–59.

[28] A. Avila, B. Cauchi, S. Goetze, S. Doclo, and T. H. Falk, "Performance comparison of intrusive and non-intrusive instrumental quality measures for enhanced speech," in *Proc. Int. Work. Acoust. Signal Enhancement (IWAENC)*, 2016, pp. 1–5.