An Improved Diagonal Loading-Based Minimum Variance Distortionless Response Beamformer

Quan Trong The[a]
Digital Agriculture Cooperative, No. 15 Lane 2, Tho Thap, Dich Vong, Cau Giay, Hanoi, Viet Nam

Keywords: Microphone Array, Speech Enhancement, Minimum Variance Distortionless Response, Diagonal Loading, Desired Target Speaker, Dual-Microphone System, the Signal-to-Noise Ratio.

Abstract: The MVDR beamformer has more prominent solution and much better noise reduction and interference suppression capability than the conventional beamforming method, which required that the associated microphone array steering vector to sound source is accurately known. However, whenever the a priori information about the direction of arrival of the interest signal is not imprecise, microphone mismatch or different microphone sensitivities; the evaluation of MVDR beamformer is often degraded, thus speech distortion, which decreases the speech quality, is unavoidable. For mitigating the drawbacks, diagonal loading has been imposed to enhance MVDR's performance in terms of improving the signal-to-noise ratio (SNR) and removing background noise. So diagonal loading has been a common widely used method to enhance the robustness of MVDR beamformer. The inherent problem of diagonal loading is the choice of optimal parameter λ to increase the effective working of diagonal loading in complex acoustical situation. In this correspondence, the author presented a method for calculating the necessary parameter λ to improve the speech enhancement in dual-microphone system. The illustrated experiment has proven the capability of considered technique via a numerical example.

1 INTRODUCTION

Separation and speech enhancement are the most popular challenging task in digital signal processing. In real environment, target speech signal is often distorted, cause: third-party speaker, noise, transport vehicle, interference. Separation speech refers to the task of saving the target speech speaker and suppressing the unwanted different noisy environment. In this context, speech enhancement is extracting one or more target speakers, and mitigate the effect of annoying noise, interfering environment or reduce some types of speech distortions due to reverberation, the complex surrounding recording scenario. So that, the terms of "signal enhancement" and "source separation" are very necessary in almost industry application. Audio device, hearing aid, teleconference, communication.

Conference has several speakers, which can be considered as target source speech, that requires separating each component from a complicated mixture. Moreover, speech enhancement is the most crucial pre-processing for further speech application, such as:

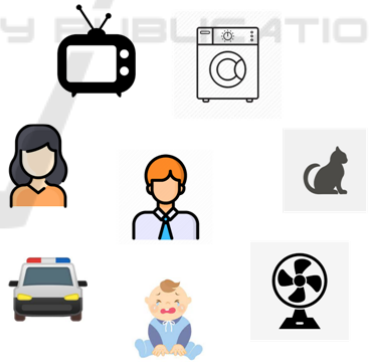

Figure 1: Extracting the desired speech is an essential task in speech enhancement.

dialogue, speech recognition, distant remote, GPS, surveillance device, video game. For dealing these problems, the microphone array (MA) (Benesty et al., 2008, 2016, 2017; Lockwood et al., 2004; Brandstein and Ward, 2001) is used for using the advantage of microphone array geometry, the spatial information of direction-of-arrival (DOA), the coherence between microphones, the characteristics of environment to alleviate the effect of noise while saving the target speaker. MA allows more input signals are multi-

[a] https://orcid.org/0000-0002-2456-9598

The, Q.
An Improved Diagonal Loading-Based Minimum Variance Distortionless Response Beamformer.
DOI: 10.5220/0012009100003561
In *Proceedings of the 5th Workshop for Young Scientists in Computer Science and Software Engineering (CSSE@SW 2022)*, pages 19-26
ISBN: 978-989-758-653-8; ISSN: 2975-9471

channel. The number of microphones has increased in many applications in the last few years. Most of telephone, tablets or hearing-aid require 2-3 microphones.

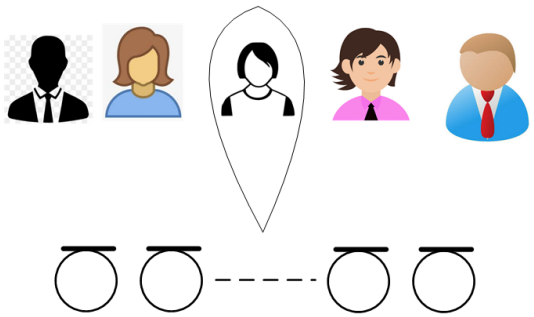

Figure 2: The using of microphone array technology.

The enhancement capabilities of MA are often higher than single-channel method. Because of the designed spatial filter to extract the target directional source speaker while eliminating all interference and noise. MA exploit the spatial a priori information about position, configuration user-defined of MA, difference of phase, more general the different acoustic properties between MA to achieve the better noise reduction while keeping the target speech. In contrast, the single – channel approach doesn't have knowledge of source or noise, so in results, the smaller quality obtained signal.

Minimum Variance Distortionless Response (MVDR) (Pan et al., 2014; Ba et al., 2007; Erdogan et al., 2016; Xiao et al., 2017a,b) has an attractive performance, which is the most widely studied and the basic to some commerce available acoustic devices. MVDR utilizes the information of DOA of target desired speaker for forming beampattern toward this direction while minimizing the output noise power. Based on the precise knowledge of interest signal's DOA, MVDR has ability of extracting the only target directional useful speech component. In practice, the DOA of target speaker if usually is not determined exactly, due to many reasons: position of MA, influence of interference or noisy environment, that degrades seriously performance of MVDR beamformer. A lot of research has been developed to overcome this problem, by extending the region where the target directional sound source can be determined. In this paper, the author proposed the using of diagonal method for solving the problem of imprecise DOA of interest signal to improve speech enhancement.

There are some research directions for enhancing the evaluation of MVDR beamformer. One of the most important parameters is a steering vector, which present the acoustic of sound propagation in environment from the desired source to all element of MA. More generally, a normalized of relative transfer function (RTFs) (Gannot et al., 2001b,a) is used for further signal processing. To improve performance of MVDR beamformer, RTF may be measured a priori or based on knowledge of microphone properties, room acoustic, speaker location, position.

However, in complex situation with presence of microphone mismatches or error of preferred DOA, the diagonal loading (DL) (Wu and Zhang, 1999; Vorobyov et al., 2003; Lorenz and Boyd, 2005; Shahbazpanahi et al., 2003; Chen and Vaidyanathan, 2007) technique is developed to address the problem of degraded performance of MVDR beamformer. DL technology is not only known provides the robustness, which against the DOA mismatch but also to the imprecise steering vector. Several research of DL have been proposed to force the magnitude of final signal in complex recording environment to exceed or equal to the original microphone array signal. The one well-known disadvantage of DL is the way of choosing the exact parameters is still lacking.

In this contribution, the author introduces an improvement of MVDR beamformer (imMVDR) that can be integrated into multi-microphone system for extracting the target directional speaker while eliminating all non-target directional noise or interference.

The rest of this paper is organized as: The next section is the model signal of MVDR beamformer. The proposed method, which use the diagonal technique is presented in section 3. The enhanced evaluation of the suggested method is illustrated in section 4, a comparison the quality output signal between the traditional MVDR beamformer (traMVDR) and imMVDR provides the robustness for separating interested speech source signal. Finally, concluding remarks and the future research of this approach are conducted.

Hundreds of microphone phones have been used for acoustic acquisition sound source from distance. However, dual-microphone array (DMA2) is more popular widely applied in almost speech application, due to it's simplicity, low computational load, compact, and easily installed in almost audio equipment. In experiment, DMA2 is used for verifying and illustrating the effectiveness of suggested method in term of increasing the signal-to-noise ratio (SNR) in real environment.

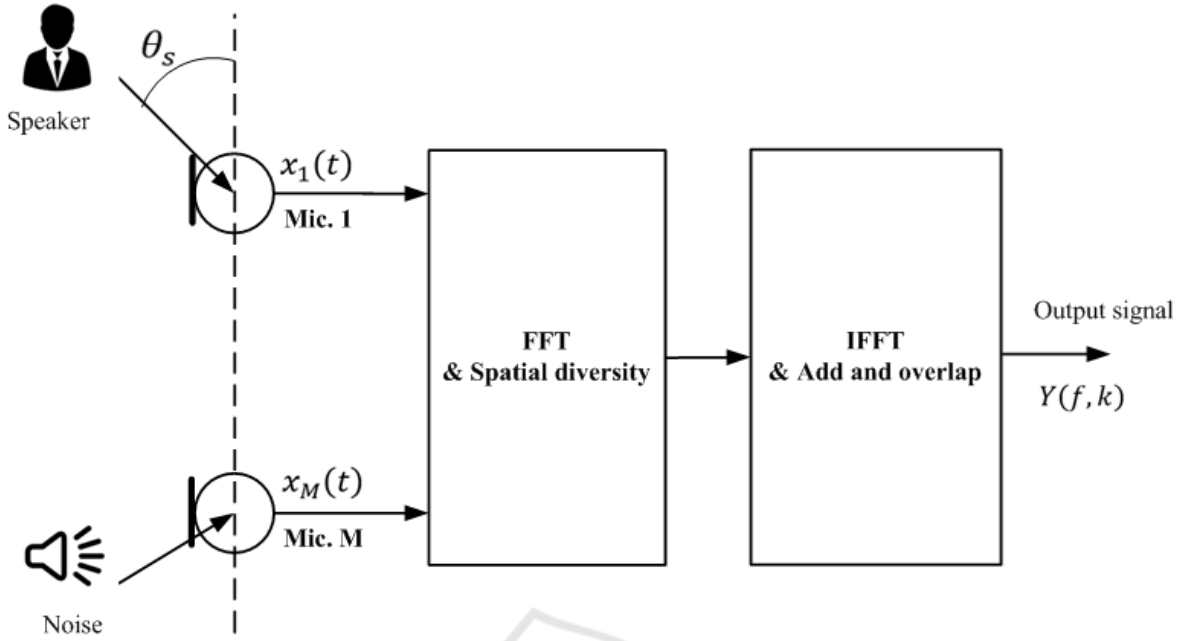

Figure 3: The scheme of beamforming in the frequency domain.

2 THE MVDR BEAMFORMER

In a noisy acoustic recording situation, it is very importance to capture the speech signal from target directional talker, therefore the only capable method is using MA beamforming to acquire the desired signal. It is assumed that a DMA2 is used to record speaker and acoustic environment. With f,k index frequency and frame, a target speaker $S(f,k)$ from a certain direction θ_s, an unwanted noise $V(f,k)$ are captured by DMA2, the observed microphone signals $X_1(f,k), X_2(f,k)$ can be written by in the frequency domain as:

$$X_1(f,k) = S(f,k)e^{j\Phi_s} + V_1(f,k) \quad (1)$$

$$X_2(f,k) = S(f,k)e^{-j\Phi_s} + V_2(f,k) \quad (2)$$

where $e^{j\Phi_s}, e^{-j\Phi_s}$ is the transfer function of target talker relative to microphone 1,2 respectively. $\Phi_s = \pi f \tau_0 cos(\theta_s), \tau_0 = d/c, d$ distance between two microphones, $c = 343(m/s)$ speed of sound propagation in the air, τ_0 is the time delay.

We denote $\boldsymbol{X}(f,k) = [X_1(f,k) \quad X_2(f,k)]^T$, $\boldsymbol{D}(f,\theta_s) = [e^{j\Phi_s} \quad e^{-j\Phi_s}]^T$, $\boldsymbol{V}(f,k) = [V_1(f,k) \quad V_2(f,k)]^T$ with $()^T$ indicates transpose operator, equation (1-2) can be rewritten by:

$$\boldsymbol{X}(f,k) = S(f,k)\boldsymbol{D}(f,\theta_s) + \boldsymbol{V}(f,k) \quad (3)$$

The steering vector $\boldsymbol{D}(f,\theta_s)$ play a major role in all MA algorithm. Due to, $\boldsymbol{D}(f,\theta_s)$ contains the information of DOA desired talker.

The digital signal processing is necessary to find an optimum weight vector $\boldsymbol{W}(f,k)$, which ensures the final output signal $Y(f,k)$ approximate the original signal $S(f,k)$:

$$Y(f,k) = \boldsymbol{W}^H(f,k)\boldsymbol{X}(f,k) \quad (4)$$

where $()^H$ is the symbol of Hermitian conjugation.

MVDR beamformer is aiming to minimizing the power of noise at the output without speech distortion, therefore, the optimum problem is described by the following equation:

$$\begin{aligned} \min_{\boldsymbol{W}(f,k)} \quad & \boldsymbol{W}(f,k)^H \boldsymbol{\Phi}_{VV}(f,k)\boldsymbol{W}(f,k) \\ \text{s.t.} \quad & \boldsymbol{W}(f,k)^H \boldsymbol{D}(f,\theta_s) = 1 \end{aligned} \quad (5)$$

where $\boldsymbol{\Phi}_{VV}(f,k) = E\{\boldsymbol{V}^H(f,k)\boldsymbol{V}(f,k)\}$ is the covariance matrix of noise. The optimum criteria of preserving the target directional speech signal leads to the solution:

$$\boldsymbol{W}(f,k) = \frac{\boldsymbol{\Phi}_{VV}^{-1}(f,k)\boldsymbol{D}(f,\theta_s)}{\boldsymbol{D}^H(f,\theta_s)\boldsymbol{\Phi}_{VV}^{-1}(f,k)\boldsymbol{D}(f,\theta_s)} \quad (6)$$

In realistic speech application, due to not available information about noise, the covariance matrix of observed microphone array signals is used instead of noise $\boldsymbol{\Phi}_{XX}(f,k) = E\{\boldsymbol{X}^H(f,k)\boldsymbol{X}(f,k)\}$. So, the final optimum weight vector is:

$$\boldsymbol{W}(f,k) = \frac{\boldsymbol{\Phi}_{XX}^{-1}(f,k)\boldsymbol{D}_s(f,\theta_s)}{\boldsymbol{D}_s^H(f,\theta_s)\boldsymbol{\Phi}_{XX}^{-1}(f,k)\boldsymbol{D}_s(f,\theta_s)} \quad (7)$$

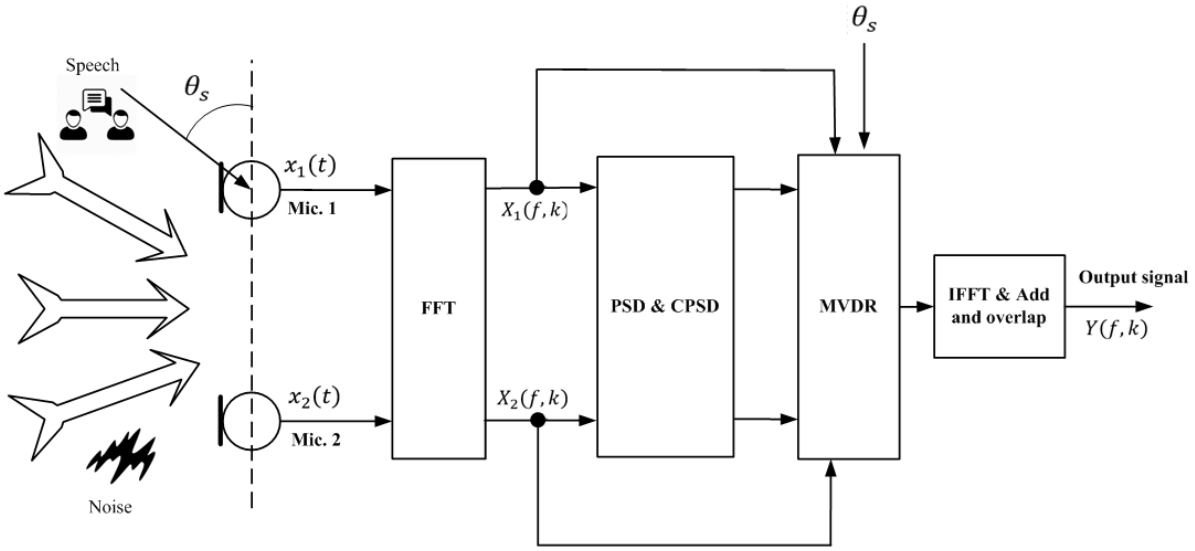

Figure 4: The scheme of MVDR beamformer.

$$\mathbf{\Phi}_{XX}(f,k) = \begin{Bmatrix} E\{X_1{}^*(f,k)X_1(f,k)\} * 1.001 & E\{X_1{}^*(f,k)X_2(f,k)\} \\ E\{X_2{}^*(f,k)X_1(f,k)\} & E\{X_2{}^*(f,k)X_2(f,k)\} * 1.001 \end{Bmatrix} \quad (8)$$

$$P_{X_iX_j}(f,k) = (1-\alpha)P_{X_iX_j}(f,k-1) + \alpha X_i^*(f,k)X_j(f,k) \quad (9)$$

where $\mathbf{\Phi}_{XX}(f,k)$ is denoted by equation (8).

With the power spectral density, $E\{X_i{}^*(f,k)X_j(f,k)\} = P_{X_iX_i}(f,k)$ is calculated as (9), where α is the smoothing parameter.

3 THE DIAGONAL LOADING-BASED PROPOSED METHOD

The matrix covariance $\Phi_{XX}(f,k)$ is one of the most common enhanced for enhancing MVDR beamformer. Diagonal loading technique is an efficient method for increasing the robustness of signal processing and speech quality of the output beamformer, while alleviating all surrounding background noise. Matrix covariance

$$\Phi_{XX}(f,k)$$

is added $\lambda \boldsymbol{I}$, where λ is unknown parameter in range $\{0..1\}$, $\boldsymbol{I}$ is the unity matrix. The problem of determining λ still the most challenging in speech enhancement.

As a result, the speech distortion often occurs in frame, where the signal-to-noise ratio (SNR) high. Due to, the necessary information of noise is more required than target directional speech, the author uses the information the speech presence probability (SPP) (Gerkmann and Hendriks, 2012a,b) and SNR to form an appropriate value of λ.

$$\lambda = SPP(f,k) * \frac{1}{1+SNR(f,k)} \quad (10)$$

where $SPP(f,k)$ was calculated from (Gerkmann and Hendriks, 2012a,b).

In the scenario with these criteria: the speech component of target speaker and noise are uncorrelated, the noise is the same and uncorrelated between two microphones. An estimation of speech covariance $\sigma_s^2(f,k)$ (Zelinski, 1988) can be expressed as:

$$\sigma_s^2(f,k) = \frac{Re\{P_{X_1X_2}(f,k) + P_{X_2X_1}(f,k)\}}{2} \quad (11)$$

where $Re\{.\}$ is the mathematical operator, which gets the real part.

And an estimation of noise covariance:

$$\sigma_n^2(f,k) = \frac{P_{X_1X_1}(f,k) + P_{X_2X_2}(f,k)}{2} - \sigma_s^2(f,k) \quad (12)$$

The temporal $SNR(f,k)$ is computed by:

$$SNR(f,k) = \frac{\sigma_s^2(f,k)}{\sigma_n^2(f,k)} \quad (13)$$

From the equation (7), the denominator plays a role as equalizer for MVDR beamformer. Therefore,

the author proposed the modified MVDR beamformer as the following equation:

$$\boldsymbol{W}(f,k) = \frac{(\boldsymbol{\Phi}_{XX} + \lambda \boldsymbol{I})^{-1}(f,k)\boldsymbol{D}_s(f,\theta_s)}{\boldsymbol{D}_s^H(f,\theta_s)(\boldsymbol{\Phi}_{XX} + \lambda \boldsymbol{I})^{-1}(f,k)\boldsymbol{D}_s(f,\theta_s)} \quad (14)$$

The diagonal loading technique is suitable with complex recording scenarios in presence of diffuse, coherent, incoherent noise field or interference. With an adaptive determined addition to covariance matrix of observed data, the performance of beamformer will rapidly adapt to the change of considered environment.

The next section will analyze the improvement of the proposed technique for reducing speech distortion and enhance speech quality.

4 EXPERIMENTS AND DISCUSSION

In experiment, a DMA2 is used for recording the target directional speech talker in presence of surrounding noise, interference of real situation. The purpose of this experiment is verifying the capability of saving target directional speech in comparison with the conventional MVDR. The distance between two microphones $d = 5(cm)$. The model of experiment is illustrated in figure 5. The desired speaker stand at the direction $\theta_s = 90(deg)$ relative to the axis of DMA2. For further digital signal processing, the author used Hamming window, $\alpha = 0.1$, $FFT = 512$, overlap 50%, the sampling frequency $Fs = 16kHz$. A measurement SNR (Ellis, 2011) is used for estimating the speech quality of obtained signal. The configuration of experiment is shown in figure 5.

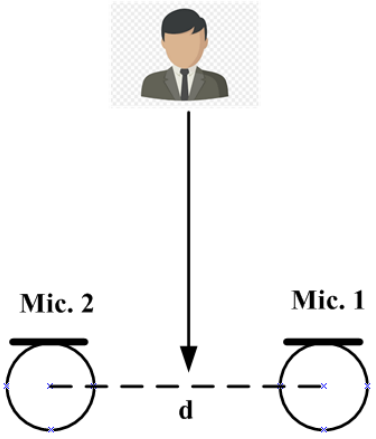

Figure 5: The scheme of experiment.

The author will compare the waveform and energy of microphone array signal and processed signals by traMVDR, imMVDR to realize the effectiveness of the proposed method. The observed microphone signal is shown in figure 6.

The obtained signal by traMVDR and imMVDR are presented in figure 7, 8. The effectiveness of the proposed method is preserving the original speech component while mitigating all background noise. In comparison with the convention MVDR, as we can see, traMVDR removes noise, but the it's weakness is speech distortion due to several reasons. imMVDR has deal it perfectly, and help keeping the original speech signal. With an appropriate addition, which has the information of speech presence probability and the SNR, MVDR beamformer has achieved a better result in extracting the target directional useful speech signal while removing the background noise or coherence noise. MVDR beamformer has the capability of minimizing the noise at the beamformer's output, but because of some reasons, such as the error of direction of arrival (DOA) of target speaker, the microphone mismatches, the different sensitivities of microphones, that degrade the performance of MVDR beamformer. In figure 7, all of surrounding noise are suppressed, but the beamformer has cancelled original signal.

Therefore, as the following of diagonal loading technique, the author has expropriated a small value, which depends on the speech presence probability and temporal SNR. The effectiveness of the proposed has increased the amplitude of received signal. Figure 9 presents the energy of microphone array, traMVDR and imMVDR. imMVDR reduces speech distortion to 3.5 (dB).

The comparison in term of speech quality between two output signals depicted in table 1. The speech quality is increased from 1.8 to 5.4 (dB).

Table 1: The signal-to-noise ratio (dB)

Method Estimation	Microphone array signal	traMVDR	imMVDR
NIST STNR	9.5	24.0	25.8
WADA SNR	6.8	20.4	25.8

So, in the complicated environment, the suggested diagonal loading technique has improved the performance of MVDR beamformer and enhanced the speech quality and intelligibility. The effectiveness of imMVDR was verified and numerical result confirmed the capability of this approach, which uses the information of speech presence probability and instantaneous SNR. The obtained numerical results have satisfied the aim of evaluated experiment.

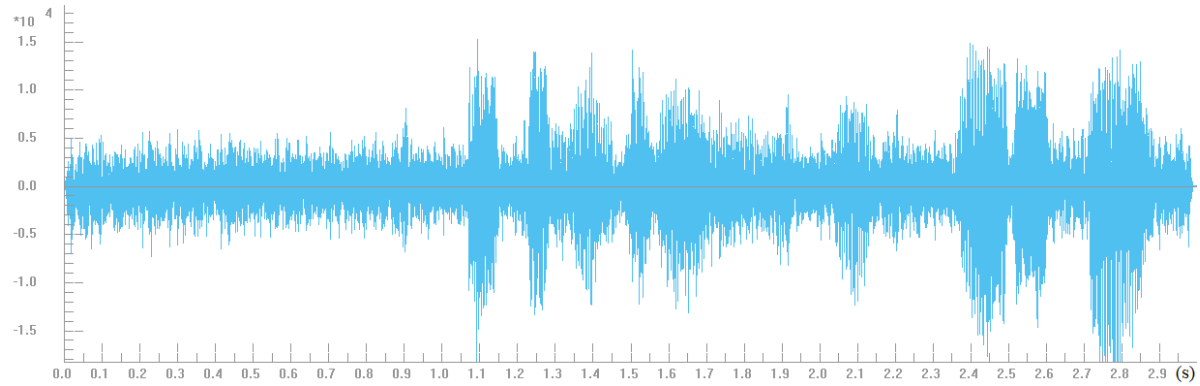

Figure 6: The waveform of the observed microphone array signal.

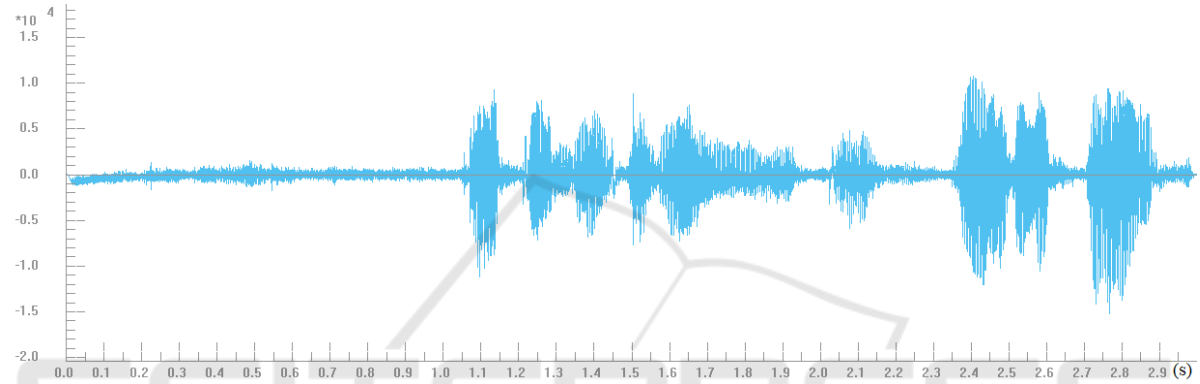

Figure 7: The obtained signal by traMVDR.

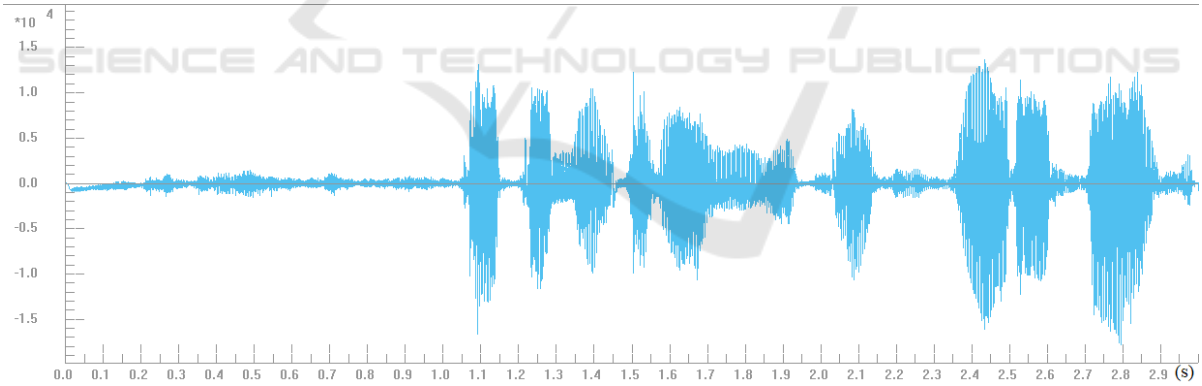

Figure 8: The obtained signal by imMVDR.

5 CONCLUSION

In many speech applications, such as hearing aids, audio devices; extracting of desired speech signal is a challenging problem from a mixture of corrupted signal with surrounding interference and different noise at low SNR. The performance of microphone array signal processing usually significantly deteriorated in the presence of unwanted noise, different speaker or complex recording scenario. Therefore, improvement of diagonal loading is a promising method for enhancing MVDR beamformer to extract useful target signal. This contribution presents an improved of diagonal loading that takes into account the calculation of necessary parameter. Objective experiment was carried out to confirm the ability of suggested technique in increasing of speech quality, noise reduction and the signal-to-noise ratio from 1.8 to 5.4 (dB). The numerical result has ensured that the proposed method can be integrated into multi-microphone system. The es-

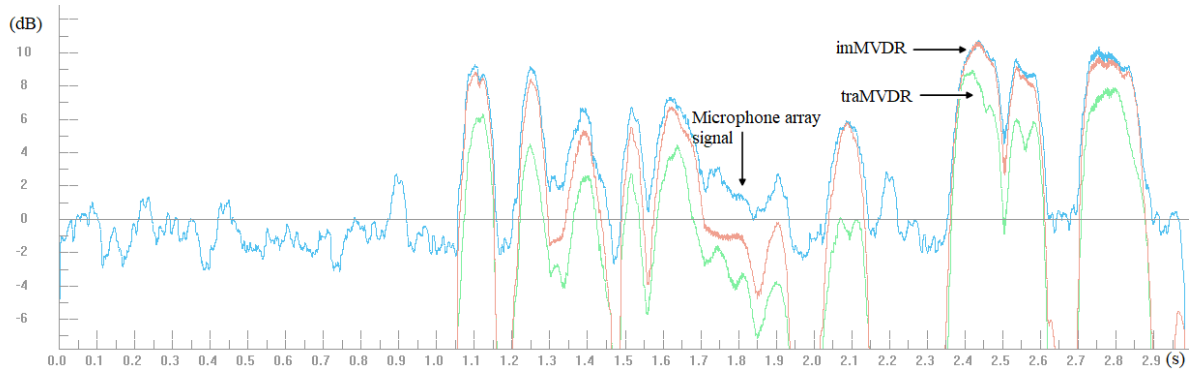

Figure 9: The illustrated energy of microphone array signal and traMVDR, imMVDR.

timation of speech presence probability can be more applied into several approaches to enhance the performance of speech enhancement system.

ACKNOWLEDGEMENTS

This research was supported supported by Digital Agriculture Cooperative. The author thank our colleagues from Digital Agriculture Cooperative, who provided insight and expertise that greatly assisted the research.

REFERENCES

Ba, D. E., Florencio, D., and Zhang, C. (2007). Enhanced MVDR Beamforming for Arrays of Directional Microphones. In *2007 IEEE International Conference on Multimedia and Expo*, pages 1307–1310. https://doi.org/10.1109/ICME.2007.4284898.

Benesty, J., Chen, J., and Huang, Y. (2008). *Microphone Array Signal Processing*, volume 1 of *Springer Topics in Signal Processing*. Springer Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-78612-2.

Benesty, J., Chen, J., and Pan, C. (2016). *Fundamentals of Differential Beamforming*. SpringerBriefs in Electrical and Computer Engineering. Springer Singapore. https://doi.org/10.1007/978-981-10-1046-0.

Benesty, J., Cohen, I., and Chen, J. (2017). *Fundamentals of Signal Enhancement and Array Signal Processing*. John Wiley & Sons Singapore. https://doi.org/10.1002/9781119293132.

Brandstein, M. and Ward, D., editors (2001). *Microphone Arrays: Signal Processing Techniques and Applications*. Digital Signal Processing. Springer Berlin, Heidelberg. https://doi.org/10.1007/978-3-662-04619-7.

Chen, C.-Y. and Vaidyanathan, P. P. (2007). Quadratically Constrained Beamforming Robust Against Direction-of-Arrival Mismatch. *IEEE Transactions on Signal Processing*, 55(8):4139–4150. https://doi.org/10.1109/TSP.2007.894402.

Ellis, D. (2011). Objective measures of speech quality/SNR. https://labrosa.ee.columbia.edu/projects/snreval/.

Erdogan, H., Hershey, J. R., Watanabe, S., Mandel, M. I., and Roux, J. L. (2016). Improved MVDR Beamforming Using Single-Channel Mask Prediction Networks. In *Proc. Interspeech 2016*, pages 1981–1985. https://doi.org/10.21437/Interspeech.2016-552.

Gannot, S., Burshtein, D., and Weinstein, E. (2001a). Signal enhancement using beamforming and nonstationarity with applications to speech. *IEEE Transactions on Signal Processing*, 49(8):1614–1626. https://doi.org/10.1109/78.934132.

Gannot, S., Burshtein, D., and Weinstein, E. (2001b). Theoretical Performance Analysis of the General Transfer Function GSC. In *Proc. Int. Workshop Acoustic Echo Noise Control*. https://www.eng.biu.ac.il/~gannot/articles/Perf.pdf.

Gerkmann, T. and Hendriks, R. C. (2012a). Unbiased MMSE-Based Noise Power Estimation With Low Complexity and Low Tracking Delay. *IEEE Transactions on Audio, Speech, and Language Processing*, 20(4):1383–1393. https://doi.org/10.1109/TASL.2011.2180896.

Gerkmann, T. and Hendriks, R. C. (2012b). Unbiased MMSE-Based Noise Power Estimation With Low Complexity and Low Tracking Delay. *IEEE Transactions on Audio, Speech, and Language Processing*, 20(4):1383–1393. https://doi.org/10.1109/TASL.2011.2180896.

Lockwood, M. E., Jones, D. L., Bilger, R. C., Lansing, C. R., O'Brien, W. D., Wheeler, B. C., and Feng, A. S. (2004). Performance of time- and frequency-domain binaural beamformers based on recorded signals from real rooms. *The Journal of the Acoustical Society of America*, 115(1):379–391. https://doi.org/10.1121/1.1624064.

Lorenz, R. G. and Boyd, S. P. (2005). Robust minimum variance beamforming. *IEEE Transactions on Signal Processing*, 53(5):1684–1696. https://doi.org/10.1109/TSP.2005.845436.

Pan, C., Chen, J., and Benesty, J. (2014). On the noisereduction performance of the MVDR beamformer in noisy and reverberant environments. In *2014 IEEE International Conference on Acoustics, Speech and*

Signal Processing (ICASSP), pages 815–819. https://doi.org/10.1109/ICASSP.2014.6853710.

Shahbazpanahi, S., Gershman, A., Luo, Z.-Q., and Wong, K. M. (2003). Robust adaptive beamforming for general-rank signal models. *IEEE Transactions on Signal Processing*, 51(9):2257–2269. https://doi.org/10.1109/TSP.2003.815395.

Vorobyov, S. A., Gershman, A. B., and Luo, Z.-Q. (2003). Robust adaptive beamforming using worst-case performance optimization: a solution to the signal mismatch problem. *IEEE Transactions on Signal Processing*, 51(2):313–324. https://doi.org/10.1109/TSP.2002.806865.

Wu, S. Q. and Zhang, J. Y. (1999). A new robust beamforming method with antennae calibration errors. In *WCNC. 1999 IEEE Wireless Communications and Networking Conference (Cat. No.99TH8466)*, volume 2, pages 869–872 vol.2. https://doi.org/10.1109/WCNC.1999.796795.

Xiao, X., Zhao, S., Jones, D. L., Chng, E. S., and Li, H. (2017a). On time-frequency mask estimation for MVDR beamforming with application in robust speech recognition. In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3246–3250. https://doi.org/10.1109/ICASSP.2017.7952756.

Xiao, Y., Yin, J., Qi, H., Yin, H., and Hua, G. (2017b). MVDR Algorithm Based on Estimated Diagonal Loading for Beamforming. *Mathematical Problems in Engineering*, 2017:7904356. https://doi.org/10.1155/2017/7904356.

Zelinski, R. (1988). A microphone array with adaptive post-filtering for noise reduction in reverberant rooms. In *ICASSP-88., International Conference on Acoustics, Speech, and Signal Processing*, pages 2578–2581 vol.5. https://doi.org/10.1109/ICASSP.1988.197172.