A Post-Filtering for Performance of Differential Microphone Array

Nguyen Ba Huy[1], Pham Tuan Anh[2], Quan Trong The[3]
[1]Falculty of Control Systems and Robotics, University ITMO St.Petersburg, Russia
[2]Vietnam - Korea University of Information and Communication Technology, The University of Danang, Vietnam
[3]Falculty of Software Engineering and Computer Systems, University ITMO St.Petersburg, Russia
Correspondence: Quản Trọng Thế, Email: quantrongthe1984@gmail.com
Received: 13/03/2023, revised: 21/06/2023, accepted: 25/06/2023
Digital Object Identifier:: 10.32913/mic-ict-research-vn.v2023.n1.1221

Abstract: Microphone array (MA) is a perspective technique for speech enhancement, because of its simplicity and low computation. MA can be decomposed into a concatenation of a spatial filtering, which provides beampattern toward a certain target directional speech source while attenuating noise, interference or third-party speaker from other directions, and a post-filtering, which suppresses the remaining noise at the final output signal. Differential Microphone Array (DIF) is one basic MA, which has many advantages for extracting the desired speech and null-steering to direction of noise. Nevertheless, in real situation, the performance of DIF often deteriorated due to phase error, imprecise the Direction-Of-Arrival (DOA) of interest signal, microphone mismatches or imperfect distribution of MA, which significantly degrade the performance and the speech quality of DIF. In this article, the author proposed a Post-Filtering (PF), which based on the properties of considered recording environment DOA to enhance the speech quality of target speaker and reduce the other non-target speaker at the opposite direction. Experimental results were verified in real scenario with dialogue between two speakers, which stand opposite direction, and show the effectiveness of the suggested PF when compared with the author's previous work.

Keywords: *microphone array signal, noise reduction, post-filtering, the direction of arrival, speech enhancement*

Tiêu đề: **Một phương pháp Post-Filtering khi áp dụng lưới microphone vi sai**

Tóm tắt: Công nghệ lưới microphone là một trong những kỹ thuật tối ưu để nâng cao chất lượng thoại, bởi tính đơn giản, mức độ phép tính thấp. Lưới microphone thường được hiểu là phép lọc không gian, trong đó cung cấp giản đồ đáp ứng xung hướng về một nguồn thoại cụ thể, trên hướng xác định trong khi gây suy hao nhiễu, người nói bên ngoài từ các hướng khác, và kỹ thuật Post- Filtering, cho phép triệt tiêu thành phần nhiễu còn lại tại tín hiệu đầu ra. Lưới microphone vi sai cũng là một dạng cơ bản, có một số ưu điểm về tách lọc tín hiệu thoại gốc và triệt giảm nhiễu xung quanh. Tuy nhiên, trong những tình huống thực tế, kết quả của lưới microphoe vi sai thường bị suy giảm, cho sai số về pha, sự không chính xác về góc tới tín hiệu thoại gốc, mất dữ liệu khi thu ghi hoặc lưới microphone phân bố không đều, điều này có ảnh hưởng rất lớn đến hiệu quả, và chất lượng tín hiệu thoại cuối cùng. Trong bài nghiên cứu này, nhóm tác giả đề xuất một Post-Filtering mới, dựa trên đặc tính hướng góc tới môi trường thu ghi để nâng cao chất lượng thoại mục tiêu trong khi triệt giảm thành phần thoại của người nói hướng đối diện. Kết quả thực nghiệm đã được chứng minh trong môi trường thu ghi thực tế trao đổi thoại giữa hai người nói đối diện nhau, và chỉ rõ hiệu quả của Post-Filtering đề xuất khi so sánh với nghiên cứu trước đây của tác giả.

Từ khóa: *lưới microphone, giảm nhiễu, post-filtering, hướng góc tín hiệu tới, nâng cao chất lượng thoại.*

I. INTRODUCTION

Almost single-channel algorithms are based on the spectral subtraction method, which often leads to speech distortion and decreasing the speech intelligibility and quality of the output signal. Therefore, MA [1-8] is one of the most effective techniques for solving this coherent drawback. Acoustic digital processing devices such as a mobile telephone, cochlear implant, hearing aid, audio device is widely equipped with an MA to capture the acoustic environment, target directional speaker, transport vehicle, interference, and surrounding noise. A mixture of speech plus some complex noisy signals, reverberation, background interference, third-party speaker, imperfect propagation of sound source or microphone mismatch are often occurred with signal processing. Therefore, the task of alleviating background noise and extracting the target speech speaker, known as a critical importance speech enhancement, has been an attractive subject of extensive direction of improvement. Intelligibility and quality of the final output signal is a major factor when estimating a speech system in presence of such environmental noise.

Due to the exploiting the spatial filtering, the knowledge of target speech source, non-target noise, the characteristic of recording situation, MA technology are successfully integrated into various speech enhancement scenarios. However, because of the constraint of physical size or the distance between microphones or geometry of MA, an acceptance degree of extracting the only desired speaker without speech distortion cannot be yielded although there are lot of sufficient spatial beamforming between the observed real microphone arrays data. Consequently, a post-filtering is enough to reduce the remaining noise after a beamformer and increase the speech quality and intelligibility.

DIF is one of the basic elements of MA, which allows blocking a point of noise source by forming a null-beampattern in the direction of noise. In the previous author's paper [26], and equalizer is applied to enhance the DIF's performance in situation of two opposite speaker. However, in many speech applications, the necessary requirement is yielding the only target speaker without noise, interference, or the other different speaker. So, in this article, the author proposed a post-filtering, which is based on the different phase to improve speech enhancement by suppressing the remaining speech component of non-target speakers.

In many decades, a numerous works focused on the designing of compact and small-size DMA [4-8], due to their low computational cost, approximately perfect invariant-frequency beampattern and good directivity index. Spatial diversity of DMAs can be calculated and optimized in any arbitrary order of beampattern, especially first-order are preferred widely commonly utilized in realistic practical applications. In real-life environment, DMAs are affected by many types of noise: white noise, diffuse noise, coherent and incoherent noise, and has many characteristics: phase mismatches, high sensitivity to misplacements, or error of omnidirectional microphones, amplification in low frequency.

The interesting term in signal processing technology is achievement of directional beampattern for compact and small-size MA, which has been an importance require due to the proliferation of embedded system for mobile phone, acoustic surveillance, infotainment. The small-size arrays, as the distance between microphones must be small enough to satisfy the assumption that the directional beampattern can be steered toward target speech. The smaller distance, the wider the spectrum ranges in obtaining beampattern with perfect nearly invariant-frequency characteristic. Moreover, the polar pattern of DMA can be adjusted by varying the value of delay time in microphone array signals. For example, Uniform Linear Arrays (ULAs) and Non-Uniform Linear Arrays (NULAs) allow receiving the symmetric beam with respect to the axis of MA.

The development of strategies for designing steerable beampattern with high order to any arbitrary direction has been attractive to scholars. In this approach, Elko [9-12] proposed a method, which is based on a linear combination of eigen-beam shaped of different order DMA, one direction focused on target talker and other directions toward noise source. Otherwise, many scholars developed designs to receive high robustness, noise reduction, directivity index and flexibility of steering pattern in any arbitrary configuration of MA or distribution of microphones [13-20]. These works have been evaluated and verified the effective performance in extracting target directional speech while suppressing all background noise and increased the robustness of DMA's working.

An essential important issue is the steering flexibility in practical applications with any arbitrary type of MA: linear, circular, planar MA. Numerous efforts have been proposed to enhance the steering flexibility of DMAs [21-25]. With great development has been obtained in the design of DMAs with good steerable pattern, the capability of DMA becomes more importance in microphone arrays signal processing.

Dual-microphone system (DMA2) is the most popular model MA is commonly installed in almost acoustic device, because of it's compact, it's convenient to easy implement almost MA processing algorithm. In experiment, DMA2 is used for recording two speakers and illustrating the proposed post-filtering in comparison with the previous author's work.

This contribution is organized as follows. In the section II, an analysis of principle working of DIF is described, and the model of equalizer, which is used in [26] is presented. The suggested post-filtering is described in Section III. Section IV and V is experiment and Conclusion.

II. DIFFERENTIAL MICROPHONE ARRAY

DMA2 helps achieving good steering flexibility, high noise reduction, high directivity index, a suitable directional pattern, which towards to target speech source. Due to its small size, low computational cost and easy to implement, DMA is the most appropriate structure for extracting desired speech source and eliminating background noise.

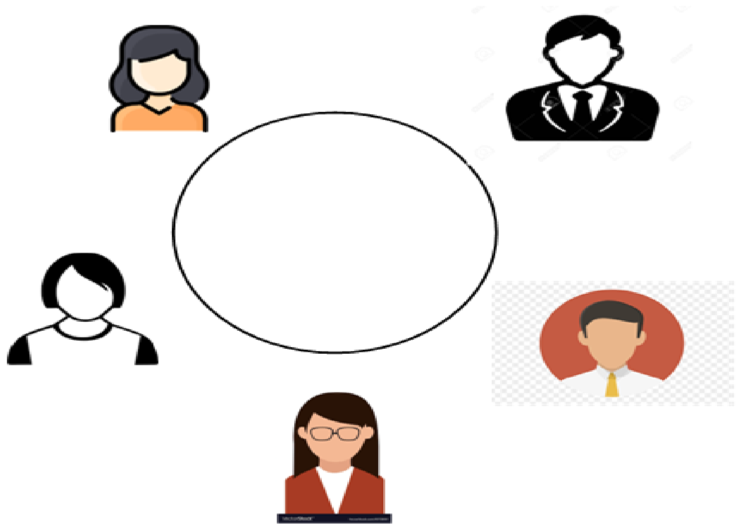

Figure 1. The most difficult task in speech enhancement is speech separation.

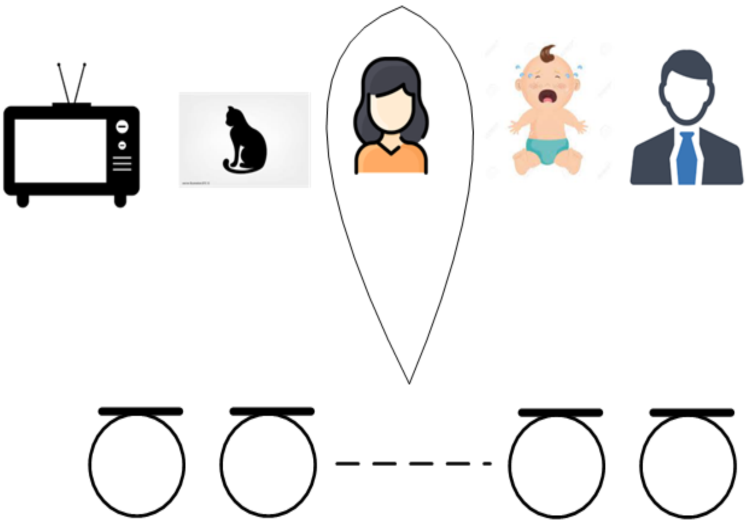

Figure 2. The using of MA to extract the desired target directional speaker.

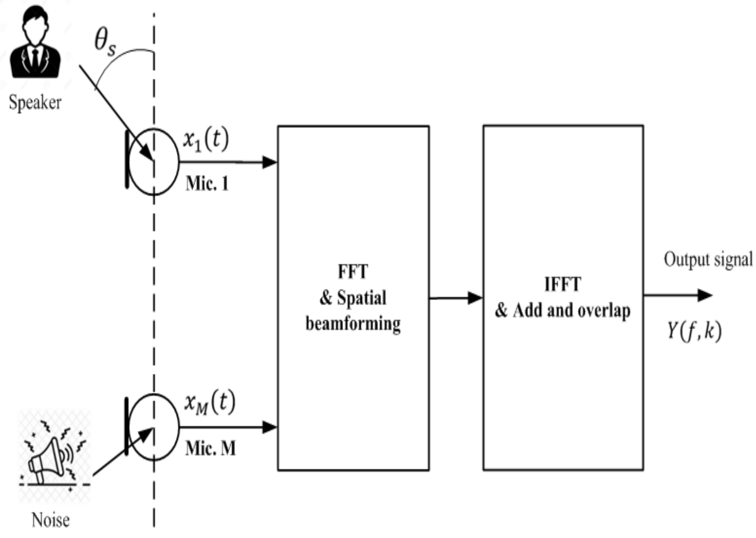

Figure 3. The scheme of MA to process microphone array signals.

We consider the principle DMA2's working in the STFT domain.

Two noisy microphone arrays signals $X_1(f,k), X_2(f,k)$ are represented in STFT domain as:

$$X_1(f,k) = S(f,k)e^{j\Phi_s} \quad (1)$$
$$X_2(f,k) = S(f,k)e^{-j\Phi_s} \quad (2)$$

Where θ is the direction-of-arrival of useful signal, $S(f,k)$ is the target desired speech source, $\Phi_s = \pi f \tau_0 cos\theta$, $\tau_0 = d/c$, d is the inter-distance between two microphones, $c = 343(m/s)$ is the speech of sound propagation; f, k is the index of frequency and frame, respectively.

With τ: an appropriate delay time was added, the high flexible steerable directivity pattern of DMA2 obtained by the described equations:

$$Y_{1DIF}(f,k) = \frac{X_1(f,k) - X_2(f,k)e^{-j\omega\tau}}{2} \quad (3)$$
$$= jS(f,k)sin(\frac{\omega\tau_0}{2}(cos\theta + \frac{\tau}{\tau_0})) \quad (4)$$

$$Y_{2DIF}(f,k) = \frac{X_2(f,k) - X_1(f,k)e^{-j\omega\tau}}{2} \quad (5)$$
$$= -jS(f,k)sin(\frac{\omega\tau_0}{2}(cos\theta - \frac{\tau}{\tau_0})) \quad (6)$$

The signal subtraction operation allows deriving equation of directional beampattern. These patterns can be expressed as:

$$B_1(f,\theta) = |\frac{Y_1(f,k)}{S(f,k)}| = |sin(\frac{\omega\tau_0}{2}(cos\theta + \frac{\tau}{\tau_0}))| \quad (7)$$
$$B_2(f,\theta) = |\frac{Y_2(f,k)}{S(f,k)}| = |sin(\frac{\omega\tau_0}{2}(cos\theta - \frac{\tau}{\tau_0}))| \quad (8)$$

In the previous work, a suggested Equalizer was used for deriving the original speech component of talker. And the Equalizer can be expressed as:

$$H_{eq}(f) = \begin{cases} 6 & 0 \text{ Hz} < \text{f} \le 200 \text{ (Hz)} \\ \frac{1}{\sin(\frac{\pi}{2}\frac{f}{Fc})} & 200 \text{ Hz} < \text{f} \le \text{Fc} \\ 1 & \text{Fc} < \text{f} \le 2\text{*Fc} \\ 0 & 2\text{*Fc} < \text{f} \end{cases} \quad (9)$$

where $Fc = \frac{1}{4*\tau_0}$.

A constrained constant threshold 12 (dB) was set for $H_{eq}(f)$ to avoid gaining the amplitude of original signal.

The obtained final output signals as:

$$Y_1(f,k) = Y_{1DIF}(f,k) * H_{eq}(f) \quad (10)$$
$$Y_2(f,k) = Y_{2DIF}(f,k) * H_{eq}(f) \quad (11)$$

III. THE PROPOSED POST-FILTERING

We can easily determine two additive post-filtering, which use the information about directions. With two preferred directions $\theta_1 = 0(deg), \theta_2 = 180(deg)$, the authors proposed exploiting the different phase $\theta_k(f)$ between two microphone array signals. $\theta_k(f)$ can be calculated as the following way in the frequency domain:

$$\theta_{k1}(f) = \arg(X_1(f,k)) - \arg(X_2(f,k)) - 2\pi f\tau_0 cos\theta_1 \quad (12)$$

$$\theta_{k2}(f) = \arg(X_1(f,k)) - \arg(X_2(f,k)) - 2\pi f\tau_0 cos\theta_2 \quad (13)$$

$\theta_{k1}(f), \theta_{k2}(f)$ is always in range $[-\pi..\pi]$. In the frames, which contain the speech component, $\theta_{k1}(f), \theta_{k2}(f)$ tends to 0 and in the noisy frame, $\theta_{k1}(f), \theta_{k2}(f)$ tends to $-\pi$ or π.

An effective Post-Filtering, which uses $\theta_k(f)$ as:

$$H_{Post-Filtering1}(f) = \cos\frac{\theta_{k1}(f)}{2} \quad (14)$$
$$H_{Post-Filtering2}(f) = \cos\frac{\theta_{k2}(f)}{2} \quad (15)$$

At the frames, in which the desired target speaker exits, $H_{Post-Filtering1}(f), H_{Post-Filtering2}(f)$ equal approximately 1 and will save the speech component. Thus in, at the other frames, $H_{Post-Filtering1}(f), H_{Post-Filtering2}(f)$ close to 0 and remove the different speech component of third-party speaker.

And the final enhanced output signals can be yielded by:

$$Y_{1pro}(f,k) = Y_1(f,k) * H_{Post-Filtering1}(f) \quad (16)$$
$$Y_{2pro}(f,k) = Y_2(f,k) * H_{Post-Filtering2}(f) \quad (17)$$

IV. EXPERIMENTS

In this section, the authors illustrated an perspective experiment with two speakers and a dual-microphone system. The speaker stand at distance $L = 2(m)$, the inter-microphone distance $d = 5(cm)$, two speakers talk with each other at opposite direction. For further signal processing, the sampling rate is $Fs = 16kHz$, overlap 50%, smoothing parameter $\alpha = 0,1$, a Hamming window were used for transforming received arrays signals into STFT domain. The scheme of experiment is described in Figure 6. In this experiment, DMA2 recorded speech component from target directional speaker along the axis of DMA2, $\theta_s = 0(deg)$; while at the opposite direction $\theta_v = 180(deg)$, a noise source corrupted the speech quality. The purpose of experiment is illustrating the capability of suppressing

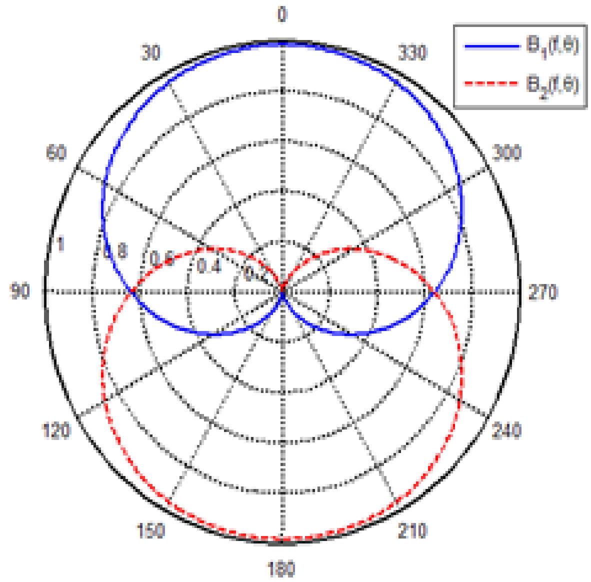

Figure 4. The desired beampattern with DMA2, $d = 5(cm)$, $f = 1500(Hz)$.

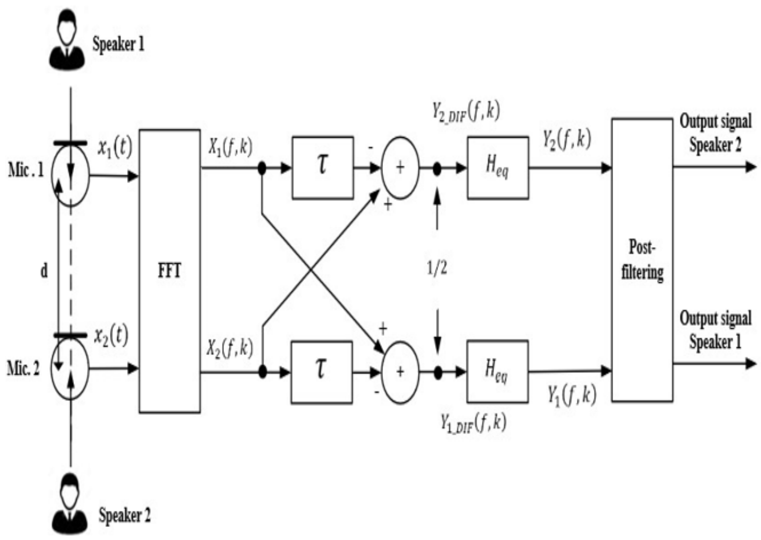

Figure 5. The model of DMA2 for extracting useful speech source.

the remaining second speaker's speech component of second speaker while saving first speaker's target speech in comparison with the previous author's work.

The original microphone array signal is presented in Figure 7.

The processed signal by the proposed method and the previous author's work [26] are depicted in Figure 8. The first speaker's speech component is saved, while the second speaker's component still exist with a substantial amount of speech component.

From figure 9, the effectiveness of the suggested post-filtering was verified and confirmed. The obtained significant benefit is that the remaining of speech component of second talker was suppressed in comparison with the previous work [26]. The degree of second speaker suppression is 10,5 (dB). This direction research can be studied and applied into other speech systems.

DMA2 is the one of the most basic MA configuration, which installed in almost acoustic equipments. Therefore, the author implemented the experiment with two opposite speaker to verify the advantage of Post-Filtering. The different phase between two noisy microphone array signals help extracting the desired talker with the DOA respectively. The proposed Post-Filtering has proven its effectiveness by saving the speaker at $\theta_1 = 0(deg)$ while suppressing the other speaker at $\theta_2 = 180(deg)$. In comparison with the

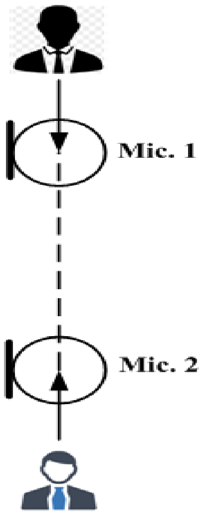

Figure 6. The acoustic recording environment in presence of two speakers.

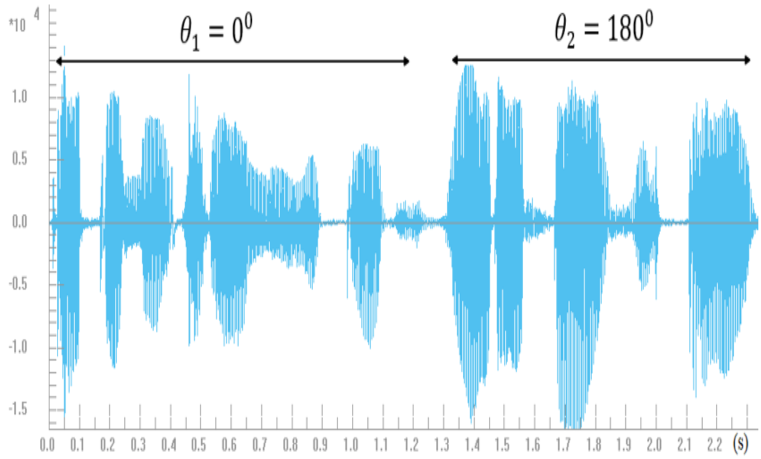

Figure 7. The waveform of microphone array signal.

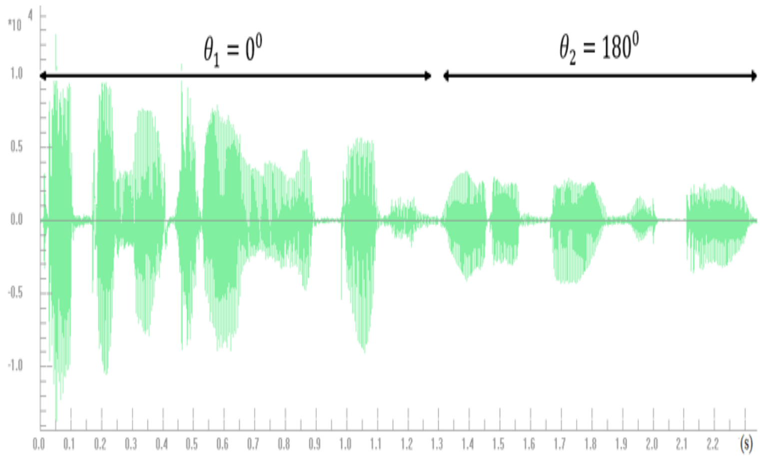

Figure 8. The processed signal by [26].

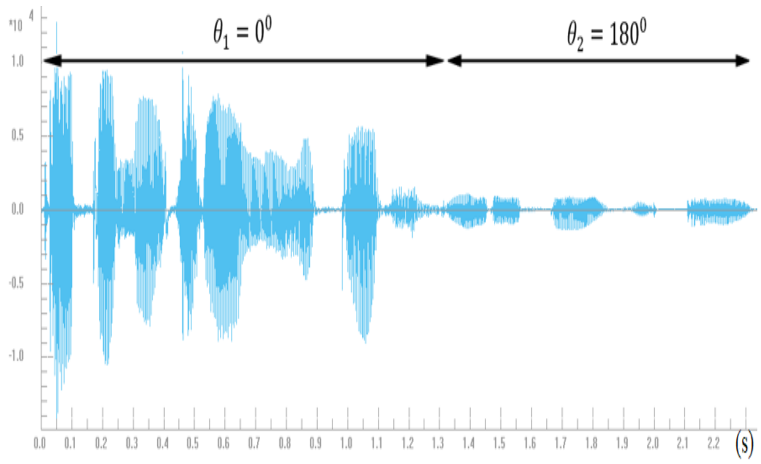

Figure 9. The obtained signal by the proposed Post-Filtering.

previous author's research, the correspondence's achievement is increasing the DIF's performance to the target speaker and removing the third-party talker in the opposite direction.

V. CONCLUSION

A major problem for acoustic processing applications, such as hearing aids, teleconferencing, mobile devices, speech recognition is separating desired speech with good fidelity. In general, it is well-known that microphone arrays and beamforming technique for acquisition of desired target speech has significant advantages compared to single-channel approach. This paper presents a proposed post-filtering, which is suitable for extracting distant speakers while eliminating other different non-target speech. This paper describes the main stage of using DIF and the proposed post-filtering to enhance performance compared to the previous author's work, the obtained degree of suppressing second talker's speech component is to 10.5 (dB). Experimental results showed that the suggested post-filtering can be integrated into multi-microphone system for solving numerous complicated speech enhancement tasks.

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

[1] Brandstein M., Ward D, *Microphone Arrays*, Springer-Verlag, 2001. XVIII, 398 p. doi:10.1007/978-3-66204619-7.

[2] Benesty J., Chen J. Huang Y, *Microphone Array Signal Processing*. Berlin, Germany, Springer-Verlag, 2008.240 p. doi:10.1007/978-3-54078612-2.

[3] Benesty J., Chen J., Pan C, *Fundaments of Differential Beamforming*, Springer, 2016.122 p. doi:10.1007/978-981-10-1046-0.

[4] Elko G. W., Pong A. T. N, "A steerable and variable first order differential microphone array" *Proc IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Apr. 1997, vol. 1. DOI: 10.1109/ICASSP.1997.599609.

[5] Teutsch H., Elko G. W, "First- and second- order adaptive differential microphone arrays", *Proc. IWAENC*, Sep. 2001, pp. 35–38. DOI:10.1121/1.1646972.

[6] Buck M, "Aspects of first-order differential microphone arrays in the presence of sensor imperfections", *European Transactions on Telecommunications*, vol. 13, no. 1, pp. 115–122, Mar.-Apr. 2002. DOI:10.1002/ett.4460130206.

[7] Elko G. W., Huang Y., Benesty J, *Differential microphone arrays, Audio Signal Processing for Next-Generation Multimedia Communication Systems*, MA, Kluwer Norwell, Ed., 2004.

[8] Benesty J., Chen J, "Study and Design of Differential Microphone Arrays", vol. 6, *Springer Science & Business Media*, 2013. ISBN: 978-3-642-33753-6.

[9] Elko G. W., Pong A. T. N, "A steerable and variable first order differential microphone array", *Proc IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, vol. 1, Apr. 1997. DOI: 10.1109/ICASSP.1997.599609.

[10] Elko G. W, Steerable and variable first-order differential microphone array, *Patent* US 006 041 127 A, Mar. 21, 2000.

[11] Elko G. W., Kubli R. A., Meyer J. M, Audio system based on at least second-order eigenbeams, *Patent* WO 2 003 061 336 A1, July 24, 2003.

[12] Elko G. W., Meyer J. M, Polyhedral audio system based on at least second-order eigenbeams, *Patent* US9 197 962 B2, Nov. 24, 2015.

[13] Sena E. D., Hacihabiboglu H., Cvetkovic Z, "On the design and implementation of higher-order differential microphones", *IEEE Trans. Audio, Speech, Lang. Process*. 20, 162–174 (2012). DOI: 10.1109/TASL.2011.2159204.

[14] Mabande E., Schad A., Kellermann K, "Design of robust superdirective beamformers as a convex optimization problem", *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP-09)*, pp. 77-80, 2009-Apr. DOI: 10.1109/ICASSP.2009.4959524.

[15] Huang G., Benesty J., Cohen I., Chen J, "Differen-

tial beamforming on graphs", *IEEE/ACM Trans. Audio, Speech, Lang. Process*. 28(1), 901–913 (2020). DOI: 10.1109/TASLP.2020.2973795.

[16] Bernardini A., Antonacci F., "Sarti A. Wave digital implementation of robust first-order differential microphone arrays", *IEEE Signal Process*. Lett. 25(2), 253–257 (2017). DOI: 10.1109/LSP.2017.2787484

[17] Borra F., Bernardini A., Antonacci F., Sarti A, "Uniform linear arrays of first-order steerable differential microphones", *IEEE/ACM Trans. Audio, Speech, Lang. Process*. 27(12), 1906–1918 (2019). DOI: 10.1109/TASLP.2019.2934567

[18] Borra F., Bernardini A., Antonacci F., Sarti A, "Efficient implementations of first-order steerable differential microphone arrays with arbitrary planar geometry", *IEEE/ACM Trans. Audio, Speech, Lang. Process*. (2020). DOI: 10.1109/TASLP.2020.2998283.

[19] Huang G., Chen J., Benesty J, "On the design of robust steerable frequency-invariant beampatterns with concentric circular microphone arrays", *Proc. IEEE ICASSP - Calgary*, 2018, pp. 506–510. DOI: 10.1109/ICASSP.2018.8461297

[20] Tiana-Roig E., Jacobsen F., Grande E. F,"Beamforming with a circular microphone array for localization of environmental noise sources", *J. Acoust. Soc. Am.*, 128(6), 3535–3542 (2010). https://doi.org/10.1121/1.3500669.

[21] Benesty J., Chen J, *Study and design of differential microphone arrays*, Springer-Verlag, Berlin, Germany, 2012. ISBN: 978-3-642-33753-6.

[22] Huang G., Chen J., Benesty J, "Design of planar differential microphone arrays with fractional orders", *IEEE/ACM Trans. Audio, Speech, Lang. Process*. 28, 116–130 (2019). DOI: 10.1109/TASLP.2019.2949219.

[23] Benesty J., Cohen I., Chen J, *Array processing-Kronecker product beamforming*, Springer-Verlag, Berlin, Germany, 2019. ISBN: 978-3-030-15600-8.

[24] Cohen I., Benesty J., Chen J, "Differential Kronecker product beamforming", *IEEE/ACM Trans. Audio, Speech, Lang. Process*. 27(5), 892–902 (2019). DOI: 10.1109/TASLP.2019.2895241.

[25] Buchris Y., Cohen I., Benesty J., Amar A, "Joint sparse concentric array design for frequency and rotationally invariant beampattern", *IEEE/ACM Trans. Audio, Speech, Lang. Process*. 28, 1143–1158 (2020). DOI: 10.1109/TASLP.2020.2982287.

[26] Stolbov M., Tatarnikova M., The Q.T. (2018), *Using Dual-Element Microphone Arrays for Automatic Keyword Recognition*, Karpov A., Jokisch O., Potapova R. (eds) Speech and Computer. SPECOM 2018. *Lecture Notes in Computer Science*, vol 11096. Springer, Cham. https://doi.org/10.1007/978-3-319-99579-3-68.

Quản Trọng Thế is currently working at Digital Agriculture Cooperative, Cau Giay, Hanoi, Vietnam. He received Master's Degree in Information Systems from Posts and Telecommunications Institute of Technology in 2015, and PhD in Engineering from National Research University ITMO in 2021. His research directions are information systems, signal processing, and speech enhancement.

E-mail: quantrongthe1984@gmail.com.

Nguyễn Bá Huy is currently a PhD student specializing in "System Analysis, Control and Information Processing" at ITMO University, Saint Petersburg, Russia. He received his Bachelor's and Master's degrees in "Control in Technical Systems" from ITMO University in 2019 and 2021 respectively. His research areas include control theory, optimization, analog and digital signal processing.

E-mail: nguyenitmo@gmail.com

Phạm Tuấn Anh is working as a PhD student at Russian Federation from 2017 to present, lecturer University of Information and Communication Technology Vietnamese and Korean communication - University of Danang. Field Research includes: artificial intelligence, machine vision surrounding intelligence system.

E-mail: ptanh@cit.udn.vn