A Different Phase-Based Improved Performance of Differential Microphone Array

Quan Trong The[a]

Digital Agriculture Cooperative, No. 15 Lane 2, Tho Thap, Dich Vong, Cau Giay, Hanoi, Viet Nam

Keywords: Microphone Array, Different Phase, Differential Microphone Array, Dual-System, Speech Enhancement, Noise Reduction, Background Noise.

Abstract: Various speech applications such as speech separation, speech recognition, communication, teleconference, hearing aids common use microphone array (MA) as front-end speech enhancement structure. By using the spatial information, geometry of MA, the characteristic of recording environment, MA help overcome of single-channel algorithm to suppress noise without speech distortion. A critical major component of MA system is the beamforming technique, which forms beampattern to a certain target directional sound source while alleviating different noise. Differential Microphone Array (DIF) is one of the most popular structure is used in several speech application. DIF has a lot of advantages which own the capability of null-steering beampattern to the noise source. In this article, the author proposes a method to enhance performance of DIF for separating speaker in real dialogue conference. Experiment was illustrated in real situation and obtained results show the effectiveness of the suggested method in comparison with the previous author's work.

1 INTRODUCTION

In several speech communication systems, separation target directional speaker of a mixture of interference, noise and third-party speaker is a challenging problem especially in situation with complex surrounding noise and low SNR (signal-to-noise ratio) values. In general, due to the above reason, the speech quality and speech intelligibility are often significant degraded. The necessary of communication is need to improve speech enhancement, even in scenario of low SNR is an importance research. Single-channel approach can't adapt all requirement, because of it uses spectral subtraction method, which lead to speech distortion of the final output. Therefore, MA technology has been developed for dealing this problem. Nowadays, MA is used in almost speech applications, such as speech recognition, mobile device, hearing aid, teleconference, human-machine interface in order to obtain an acceptable degree of speech quality from any algorithms trying to mitigate background noise and extracting desired speech. MA exploits the spatial diversity, the priori of knowledge of the direction-of-arrival (DOA), the properties of acoustic situation, the characteristic of noise field to achieve a beampattern, which toward speech source and remove all of noise.

Figure 1: The complicated task of separating desired speech source.

Because of possibly changing recording scenario and varying position of the noise source relative to MA, DIF beamformer allows null-steering beampattern to noise. In the previous work, the author suggested an additive equalizer for enhancing separation

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

The, Q.
A Different Phase-Based Improved Performance of Differential Microphone Array.
DOI: 10.5220/0012009300003561
In *Proceedings of the 5th Workshop for Young Scientists in Computer Science and Software Engineering (CSSE@SW 2022)*, pages 34-40
ISBN: 978-989-758-653-8; ISSN: 2975-9471

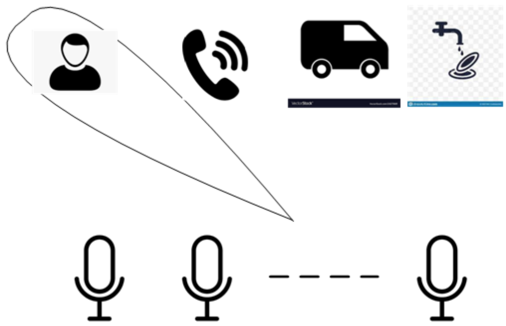

Figure 2: The advantage of utilizing the MA beamforming.

each talker when using DIF.

G. W. Elko (Elko, 1997, 2000; Teutsch and Elko, 2001) formed a cardioid beampattern when combined signal with compact spaced microphones. Therefore, the directivity index of MA was increased.

The designing with small arrays and an appropriate geometry and associated beamforming algorithms, which can pass all through bandpass speech signals still a challenging problem to the scholars (Barfuss et al., 2017a,b). Among different several types of MA, DMA is the most suitable architecture to measure differential of the sound target desired speech , and more appropriate for achieving high diversity, high noise reduction, high directivity index and invariant-frequency beampattern toward the direction of talker (De Sena et al., 2012; Buck, 2002). Recently, many efforts are successful with flexible differential beamforming and robustness of DIF's performance are enhanced (Huang et al., 2020; Benesty et al., 2019; Cohen et al., 2019).

An essential issue in differential beamforming is steering flexibility in any direction of signal or noise. Linear DMAs do not have much flexibility, due to the beampattern varies with steering angle and the optimum directivity factor occurs only in end-fire case DMA. A numerous works are attempted to improving the flexibility of steering DMA's beampattern. In (Elko and Pong, 1997; Derkx and Janse, 2009), a two-dimensions MA are utilized to steer the resulting beampattern in a certain number of directions. First-order steerable DMAs (Wu et al., 2014; Wu and Chen, 2016) are evaluated to construct a combination of monopole and dipoles using 4 microphones square array. In (Benesty et al., 2015), uniform circular DMAs were performed to steering beampattern to some wanted directions. Bernardini et al. (Bernardini et al., 2017) suggested using DMA to steer second-order with monopole and dipole. In (Huang et al., 2017), a beamforming technique was evaluated to design an approximately constant beampattern, and has capable to steer between two directions of the reference beams. In (Parra, 2005, 2006), a designed method was suggested for forming invariant-frequency beampattern for spherical array. In (Lai et al., 2013), a broadband beampattern was presented with arbitrary geometry of microphones for spherical MA.

With a great obtained progress in designing steerable DMA's beampattern with high directivity index, good noise reduction, more robustness, and high capability of steering beampattern.

In this paper, the author continues to deal the problem of separating talker when using the subspace technique to reduce the remaining speech component non-target talker at the output DIF beamformer.

The rest of this paper is organized as follow. The next section presents model signal, advantage of DIF and the previous author's work. Section 3 introduce how to use the information about phase error between two microphones to form a post-filtering, which based on Wiener filter. Section 4 described experiments with two speakers, and a comparison between the proposed method and the previous work. Concluding remark and future work is presented in Section 5.

2 DIFFERENTIAL MICROPHONE ARRAY

We will consider the DMA2's working in STFT domain. DMA2 allows obtaining high diversity, noise reduction, directional beampattern to target speech. DMA2 have super directivity, small size and can null-steering beampattern to the direction of noise, which makes it possible to achieve high performance, even in reverberation, noisy environment, and suitable for almost acoustic equipment. The obtained effectiveness of DMA2 is suppressing complicated background noise and extracting the target speaker.

The representation of two noisy microphone arrays $X_1(f,k), X_2(f,k)$ in the STFT domain as:

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

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

where f,k is the index of frequency and current considered frame, $\Phi_s = \pi f \tau_0 cos(\theta_s), \tau_0 = d/c, d$ is the distance of two microphones, $c = 343(m/s)$ is the speech of sound propagation in the air, Θ is the direction-of-arrival (DOA) of useful speech.

With a certain delay τ is added, we can obtain high directivity beampattern toward two speakers. The output signals, which are based on subtraction signal

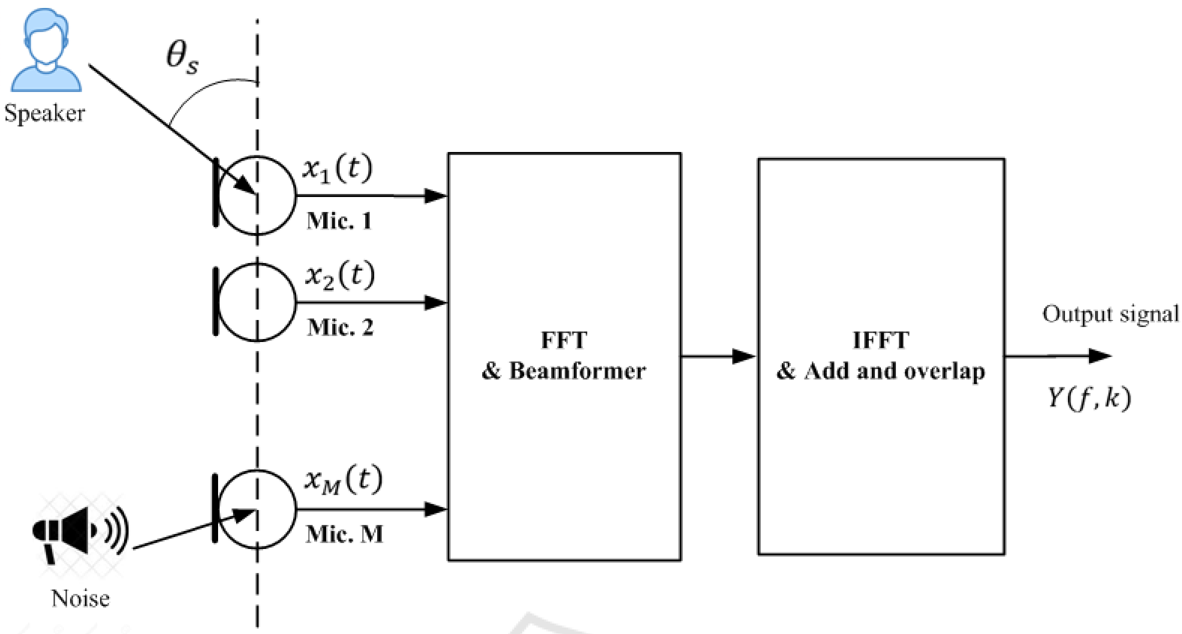

Figure 3: The MA signal processing in the frequency domain.

can be expressed as:

$$Y_{1DIF}(f,k) = \frac{X_1(f,k) - X_2(f,k)e^{-j\omega\tau}}{2} \quad (3)$$
$$= jS(f,k)sin(\frac{\omega\tau_0}{2}(cos\theta + \frac{\tau}{\tau_0})) \quad (4)$$
$$Y_{2DIF}(f,k) = \frac{X_2(f,k) - X_1(f,k)e^{-j\omega\tau}}{2} \quad (5)$$
$$= -jS(f,k)sin(\frac{\omega\tau_0}{2}(cos\theta - \frac{\tau}{\tau_0})) \quad (6)$$

The directional two beampatterns, which received from equation (3) - (6) are derived as:

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

Stolbov et al. (Stolbov et al., 2018) proposed utilizing an equalizer for preserving the useful signal at the low-frequency band. The equalizer can be expressed as the following equation:

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

where $Fc = \frac{1}{4*\tau_0}$. A constant threshold 12(dB) is determined for $H_eq(f)$. The final output signals are:

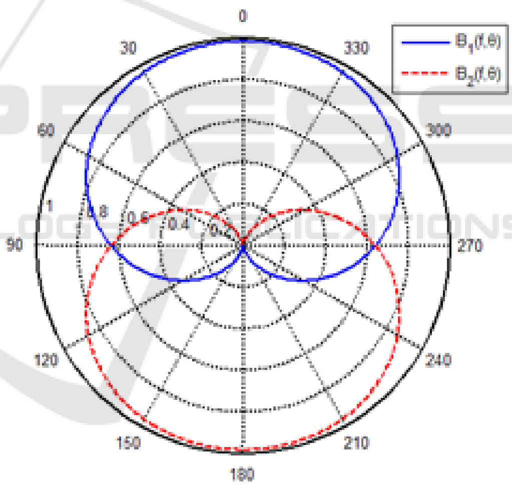

Figure 4: The beampatterns.

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

In many realistic scenario, the remaining noisy component at the output signal often deteriorate the speech quality of DIF beamformer. In the next section, the author show an additive post-filtering, which based on the priori information of DMA2.

3 THE POST-FILTERING

An additive post-filtering, which based on the subspace technique is presented in figure 5.

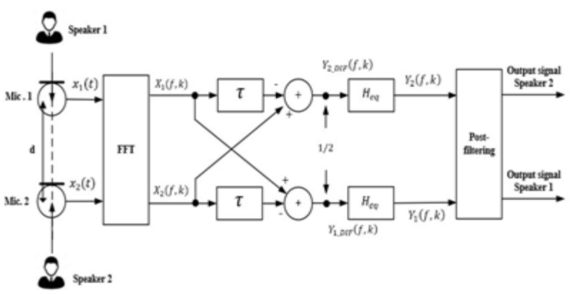

Figure 5: The suggested post-filtering for DMA2 to separate useful signal.

In the STFT-domain, with the assumption as the noise magnitudes are equal for two microphones. The different phase $\theta_k(f)$ is defined as:

$$\theta_k(f) = \arg(X_1(f,k)) - \arg(X_2(f,k)) - 2\pi f \tau_0 \quad (12)$$

A major information that the different phase is always in range $[-\pi..\pi]$. In the frames, which contain the speech component, $\theta_k(f)$ tends to 0 and in the noisy frame, $\theta_k(f)$ tends to π or π.

The author proposed a post-filtering, which use the a priori information $\theta_k(f)$ as:

$$PF(f) = \cos\frac{\theta_k(f)}{2} \quad (13)$$

At the frames, in which the desired target speaker exits, $PF(f)$ equal approximately 1 and will save the speech component. Thus in, at the other noisy frames, $PF(f)$ close to 0 and remove the interference or background noise.

At the final, the received output signal are determined as:

$$Y_{1out}(f,k) = Y_1(f,k) * PF(f) \quad (14)$$

$$Y_{2out}(f,k) = Y_2(f,k) * PF(f) \quad (15)$$

In the next section, the author illustrated and experiment to very the useful of the proposed post-filtering in compared with the previous author's work (Stolbov et al., 2018) by using a DMA2 to extract the desired target talker while suppressing the interference.

4 EXPERIMENTS AND DISCUSSION

The purpose of this experiment is enhancing the performance, which based on the DIF beamformer. The derived problem is alleviating the second talker's component while preserving target directional first talker. A degree of suppression second speaker's speech component is calculated to confirm the promising results. The scheme of experiment is depicted in figure 6. DMA2 with the inter-microphone distance is $d = 5(cm)$. The desired speaker talk while the interference at the other opposite direction.

For calculating the spectral power density, some parameters were used for transformed into STFT domain: Hamming window, NFFT=512, overlap 50%, the sampling frequency $Fs = 16kHz$, smoothing parameter $\alpha = 0.1$. A DMA2 was used for recording speech from target talker at the direction $\theta_s = 0(deg)$ in presence of unwanted interference, which stand at the opposite direction $\theta_v = 180(deg)$.

Figure 6: The evaluated recording scenarios with two speakers.

The waveform of the received signal was illustrated in figure 7.

In figure 8, the obtained processed signal by the previous author's work (Stolbov et al., 2018) and the proposed method are illustrated in figure 9.

From these figures, as we can see that, the remaining of speech component of the second talker is removed to 5(dB) in comparison with (Stolbov et al., 2018). The effectiveness of the proposed technique was confirmed in suppressing the remaining speech component of second talker while ensuring keep the first second's speech. This is the major advantage in comparison with the previous work (Stolbov et al., 2018).

So in this paper, the author exploits the different phase approach to forming an additive post-filtering to enhance the final desired target speaker. As can be observed, the enhancement of suggested method

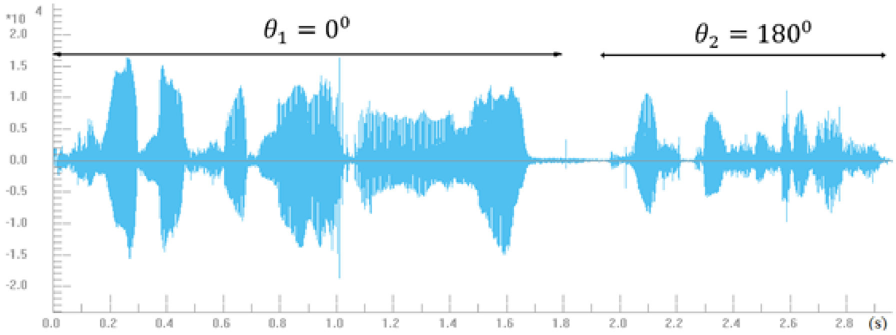

Figure 7: The waveform of the capture microphone array signal.

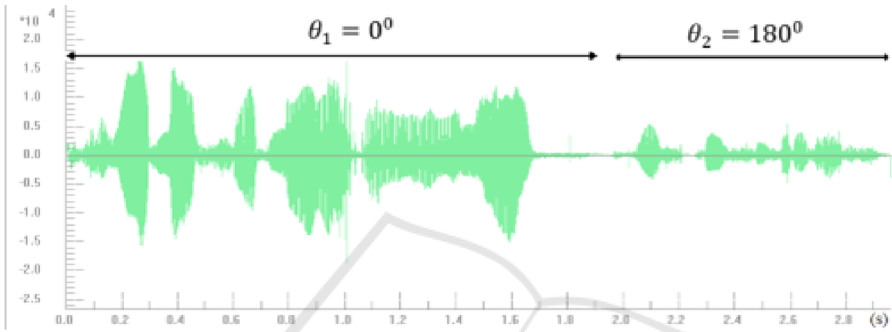

Figure 8: The processed signal by (Stolbov et al., 2018).

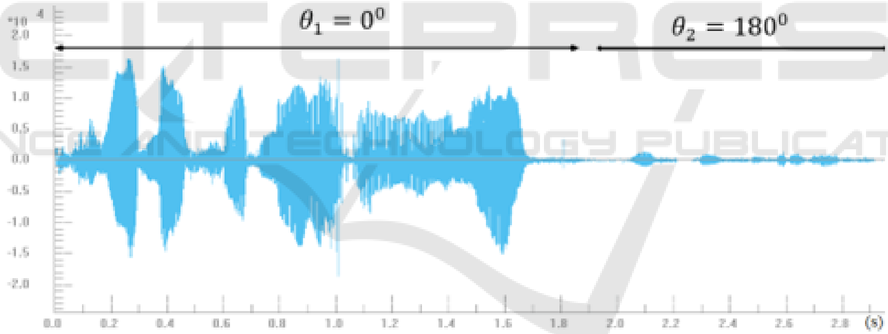

Figure 9: The obtained processed signal by additive post-filtering.

was more significant realized than the work (Stolbov et al., 2018). This technique can be applied into other acoustic system.

DMA2 has the capability of impact, low computation, easy implementation. These advantages make DMA2 is commonly installed in almost acoustic equipments; therefore the demand of enhancing of of high directivity, more robustness still exits. The author use the available information about difference phase to improve the previous work.

Separation of target speaker in complex environment, in which the third-party talker or background noise existed, still a challenge to all scholars. The use of the characteristics between two microphones is the advantage of MA to improve the speech enhancement. Different phase is one of the most interesting object for studying to improve the performance of DMA2.

5 CONCLUSION

MA beamforming are now common installed in various range of acoustic applications such as cellular phones, teleconferencing, speech recognition, robotics, smart device, telephone hands-free. A challenging task is perfectly separating speech each speaker, which requires using model a suitable DMA.

In this paper, the author suggested exploit the characteristic subspace of microphone array signals to improve and increase the robustness of DMA system than the previous work. A knowledge of the power target speech source used for post-filtering to suppress the non-target different speaker without speech distortion of target source. Numerical result was verified in real scenario and confirm the ability of the proposed method. Phase error approach can be further studied in future for installing into multi-microphone system. In the future, the author will investigate the property of environment to enhance the performance of MA algorithms.

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

Barfuss, H., Bachmann, M., Buerger, M., Schneider, M., and Kellermann, W. (2017a). Design of robust two-dimensional polynomial beamformers as a convex optimization problem with application to robot audition. In *2017 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pages 106–110. https://doi.org/10.1109/WASPAA.2017.8170004.

Barfuss, H., Buerger, M., Podschus, J., and Kellermann, W. (2017b). HRTF-based two-dimensional robust least-squares frequency-invariant beamformer design for robot audition. In *2017 Hands-free Speech Communications and Microphone Arrays (HSCMA)*, pages 56–60. https://doi.org/10.1109/HSCMA.2017.7895561.

Benesty, J., Chen, J., and Cohen, I. (2015). *Design of Circular Differential Microphone Arrays*, volume 12 of *Springer Topics in Signal Processing*. Springer Cham. https://doi.org/10.1007/978-3-319-14842-7.

Benesty, J., Cohen, I., and Chen, J. (2019). *Array Processing: Kronecker Product Beamforming*, volume 18 of *Springer Topics in Signal Processing*. Springer Cham. https://doi.org/10.1007/978-3-030-15600-8.

Bernardini, A., D'Aria, M., Sannino, R., and Sarti, A. (2017). Efficient Continuous Beam Steering for Planar Arrays of Differential Microphones. *IEEE Signal Processing Letters*, 24(6):794–798. https://doi.org/10.1109/LSP.2017.2695082.

Buck, M. (2002). Aspects of first-order differential microphone arrays in the presence of sensor imperfections. *European Transactions on Telecommunications*, 13(2):115–122. https://doi.org/10.1002/ett.4460130206.

Cohen, I., Benesty, J., and Chen, J. (2019). Differential Kronecker Product Beamforming. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 27(5):892–902. https://doi.org/10.1109/TASLP.2019.2895241.

De Sena, E., Hacihabiboglu, H., and Cvetkovic, Z. (2012). On the Design and Implementation of Higher Order Differential Microphones. *IEEE Transactions on Audio, Speech, and Language Processing*, 20(1):162–174. https://doi.org/10.1109/TASL.2011.2159204.

Derkx, R. M. M. and Janse, K. (2009). Theoretical Analysis of a First-Order Azimuth-Steerable Superdirective Microphone Array. *IEEE Transactions on Audio, Speech, and Language Processing*, 17(1):150–162. https://doi.org/10.1109/TASL.2008.2006583.

Elko, G. W. (1997). Adaptive noise cancellation with directional microphones. In *Proceedings of 1997 Workshop on Applications of Signal Processing to Audio and Acoustics*, pages 4 pp.–. https://doi.org/10.1109/ASPAA.1997.625628.

Elko, G. W. (2000). Superdirectional Microphone Arrays. In Gay, S. L. and Benesty, J., editors, *Acoustic Signal Processing for Telecommunication*, volume 551 of *The Springer International Series in Engineering and Computer Science*, pages 181–237. Springer US, Boston, MA. https://doi.org/10.1007/978-1-4419-8644-3_10.

Elko, G. W. and Pong, A.-T. N. (1997). A steerable and variable first-order differential microphone array. In *1997 IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 1, pages 223–226. https://doi.org/10.1109/ICASSP.1997.599609.

Huang, G., Benesty, J., and Chen, J. (2017). On the Design of Frequency-Invariant Beampatterns With Uniform Circular Microphone Arrays. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 25(5):1140–1153. https://doi.org/10.1109/TASLP.2017.2689681.

Huang, G., Chen, J., and Benesty, J. (2020). Design of planar differential microphone arrays with fractional orders. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 28:116–130. https://doi.org/10.1109/TASLP.2019.2949219.

Lai, C. C., Nordholm, S., and Leung, Y. H. (2013). Design of Steerable Spherical Broadband Beamformers With Flexible Sensor Configurations. *IEEE Transactions on Audio, Speech, and Language Processing*, 21(2):427–438. https://doi.org/10.1109/TASL.2012.2219527.

Parra, L. C. (2005). Least squares frequency-invariant beamforming. In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, 2005.*, pages 102–105. https://doi.org/10.1109/ASPAA.2005.1540179.

Parra, L. C. (2006). Steerable frequency-invariant beamforming for arbitrary arrays. *The Journal of the Acoustical Society of America*, 119(6):3839–3847. https://doi.org/10.1121/1.2197606.

Stolbov, M., Tatarnikova, M., and The, Q. T. (2018).

Using Dual-Element Microphone Arrays for Automatic Keyword Recognition. In Karpov, A., Jokisch, O., and Potapova, R., editors, *Speech and Computer*, volume 11096 of *Lecture Notes in Computer Science book series*, pages 667–675, Cham. Springer International Publishing. https://doi.org/10.1007/978-3-319-99579-3_68.

Teutsch, H. and Elko, G. W. (2001). First- and second order adaptive differential microphone arrays. In *Seventh International Workshop on Acoustic Echo and Noise Control, IWAENC 2001*, Darmstadt. https://www.iwaenc.org/proceedings/2001/main/data/teutsch.pdf.

Wu, X. and Chen, H. (2016). Directivity Factors of the First-Order Steerable Differential Array With Microphone Mismatches: Deterministic and Worst-Case Analysis. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 24(2):300–315. https://doi.org/10.1109/TASLP.2015.2506269.

Wu, X., Chen, H., Zhou, J., and Guo, T. (2014). Study of the Mainlobe Misorientation of the First-Order Steerable Differential Array in the Presence of Microphone Gain and Phase Errors. *IEEE Signal Processing Letters*, 21(6):667–671. https://doi.org/10.1109/LSP.2014.2312729.