

Chapter 4

Applied Examples and Applications of Localization and Tracking Problem of Multiple Speech Sources

In an enclosed environment, multi-source DOAE is an inspiring due to environmental noise, room reverberation, and the source spectra overlapping. The DOAE of acoustic sources is imperative for several applications including beamforming, automatic camera steering, robotics and surveillance. The incidence of background noise and reverberation should be considered in the real environment. Furthermore, DOAE of multiple and concurrent active sources is considered a challenging problem [1–4]. For these applications, conventional methods often exploit an omnidirectional microphone array with DOAE using phase-delay information between the microphones. Nevertheless, the conventional arrays necessitate a large aperture.

4.1 Simulation of LPA Beamformer

Assume a ULA of ten omnidirectional sensors with the half-wavelength inter-element spacing, and uncorrelated sources with SNR = 0 dB in a single sensor. For three signals $s_1(t)$, $s_2(t)$ and $s_3(t)$ be zero-mean white Gaussian with $\sigma = 1$, have the angles $\theta_i(t)$, $i = 1, 2, 3$. Assume rectangular window has $N = 60$ snapshots and sampling period of $T = 1$ s. Thus, the observation model is represented as follows [5–7]:

$$r(t + kT) = A(t + kT)s(t + kT) + e(t + kT) \quad (4.1)$$

$$A(t + kT) = [a(\theta_1(t + kT)), a(\theta_2(t + kT)), \dots, a(\theta_q(t + kT))] \quad (4.2)$$

$$s(t) = [s_1(t), s_2(t), \dots, s_q(t)]^T \quad (4.3)$$

Afterwards, the time-varying $\theta_i(t)$ should be estimated, where the time-argument is centered around the t of the LPA, where:

$$\theta_i(t + kT) = \theta_i(t) + \theta_i^{(1)}(t)kT \tag{4.4}$$

Thus, $\theta_i(t + kT)$ is linear on time kT with the values of the direction $\theta_i(t)$ and the first derivative $\theta_i^{(1)}(t)$ for the time-instant t . The performance of the LPA beamformer versus the conventional beamformer is studied for different scenarios as follows for multi-speech signals from moving sources.

4.1.1 Case 1 (One Source Case)

Consider a source, moving with three different uniform velocities, namely $\theta^{(1)}(t) = (0, 1, 2)$ deg/sample, where $\theta(t) = 0^\circ$. The LPA function is displayed in Figs. 4.1 through 4.4 as a 2D contour plot and 3D surface plot are illustrated with focusing on the location of interest. Figure 4.1 illustrates the LPA beamformer output of a single stationary source.

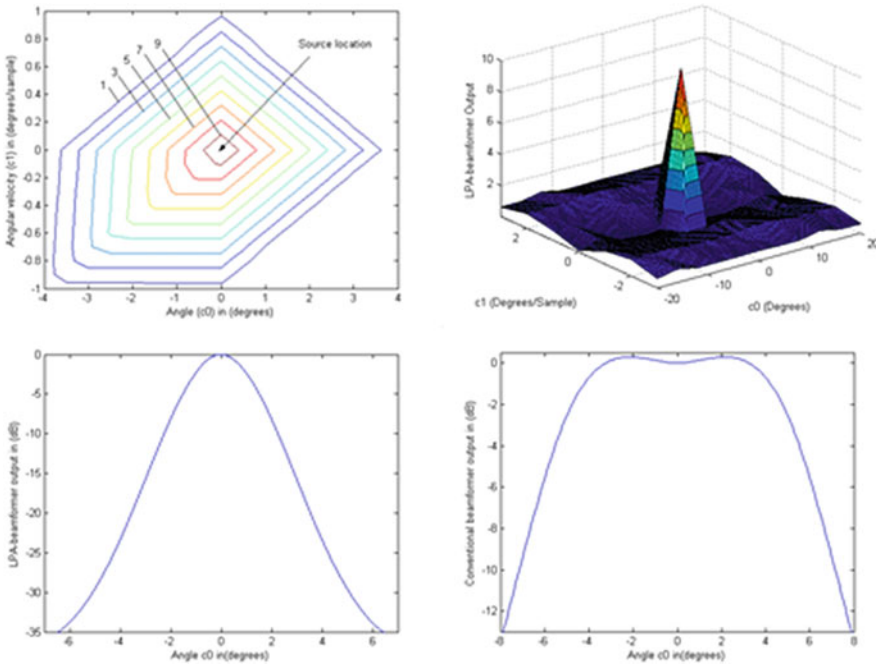


Fig. 4.1 A single stationary source case beamformers output at $\theta = 0$

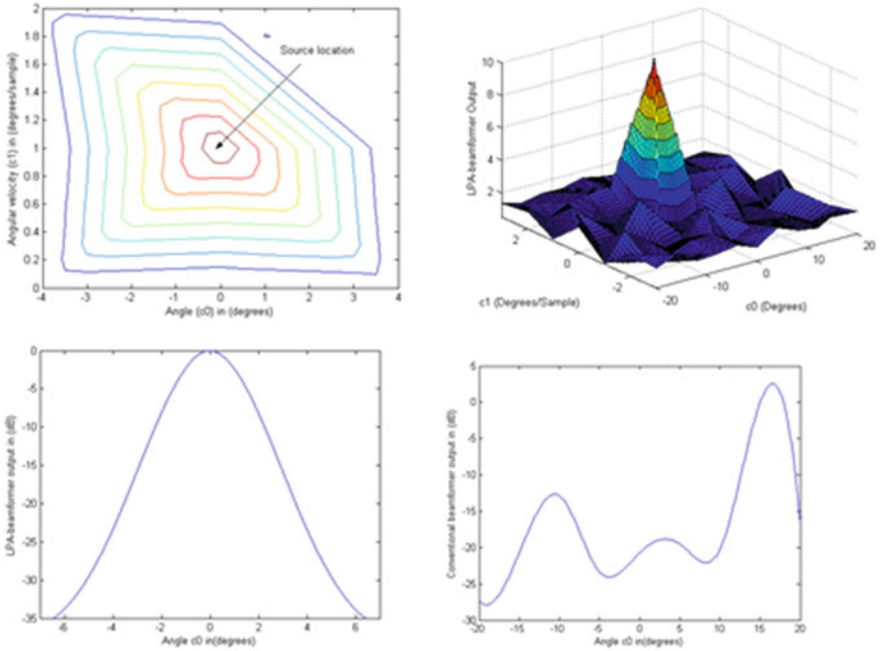


Fig. 4.2 Beamformers output for a single source case at $\theta = 0$ and $\theta^{(1)}(t) = 1$ deg/sample

Figure 4.1 shows the accurate source localization at $\theta = 0$ and $\theta^{(1)}(t) = 0$. Figure 4.2 illustrates the LPA beamformer output compared to the conventional beamformer P_{conv} for a single source case at $\theta = 0$ and $\theta^{(1)}(t) = 1$ deg/sample.

Figure 4.2 establishes that the LPA provided an accurate indication of the source location at $\theta = 0$. Also, P_{LPA} shows $\theta^{(1)}(t) = 1$, while, P_{conv} cannot indicate the source location, where the peak location is shifted. Figure 4.3 illustrates the comparative results of the LPA and the conventional beamformers for a single source case at $\theta = 0$ and $\theta^{(1)}(t) = 2$ deg/sample.

Figure 4.3 illustrates the accurate localization of the source using the LPA beamformer at $\theta = 0$ and $\theta^{(1)}(t) = 2$, while, P_{conv} cannot indicate the source location, where the output is degraded. Figure 4.4 demonstrates the beamformers output for a single source case at $\theta = 0$ and $\theta^{(1)}(t) = -1$ deg/sample.

Figure 4.4 establishes the achieved accurate localization of the single source that located at $\theta = 0$ and $\theta^{(1)}(t) = -1$ using the LPA beamformer, while, P_{conv} cannot indicate the source location. The peaks of all curves indicate the efficient localization of the source position using the LPA beamformer. The true, accurate values of $\theta(t)$ and $\theta^{(1)}(t)$, while the shape of the curves is slightly depending on the angular velocity. The LPA function P_{LPA} is expressed in term of c_0 i.e., θ as well as the conventional beamformer P_{conv} . The peaks of all the LPA curves show the true

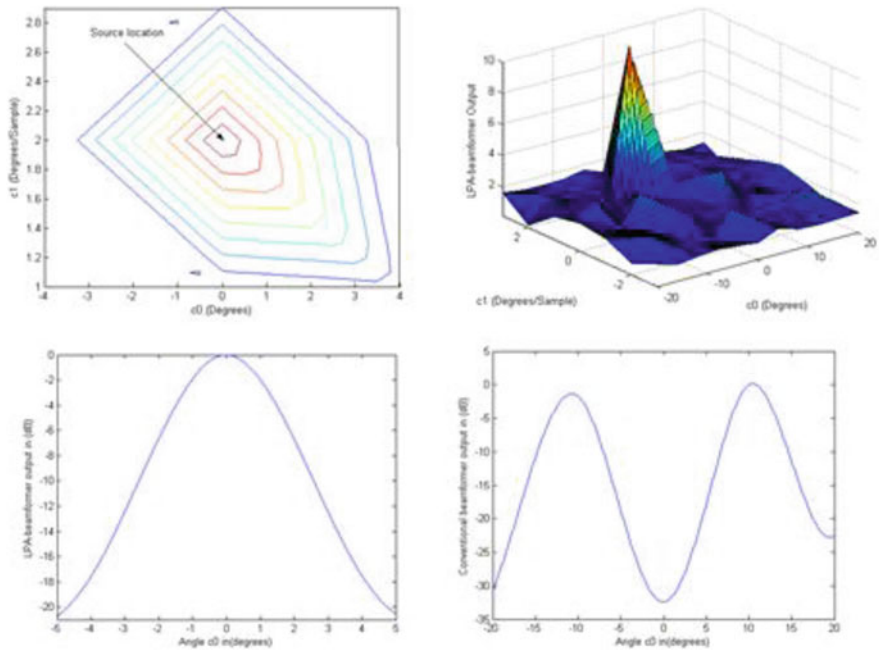


Fig. 4.3 Beamformers output for a single source case at $\theta = 0$ and $\theta^{(1)}(t) = 2$ deg/sample

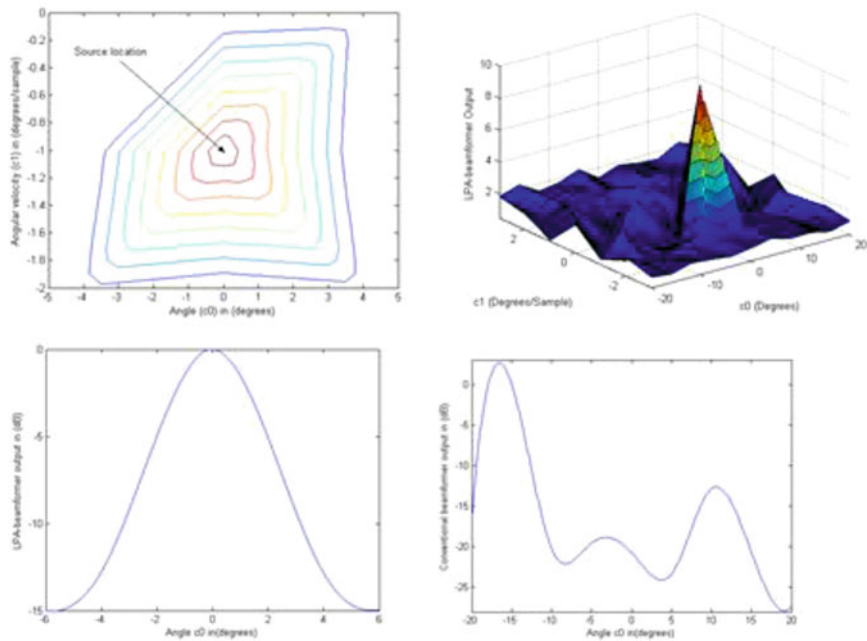


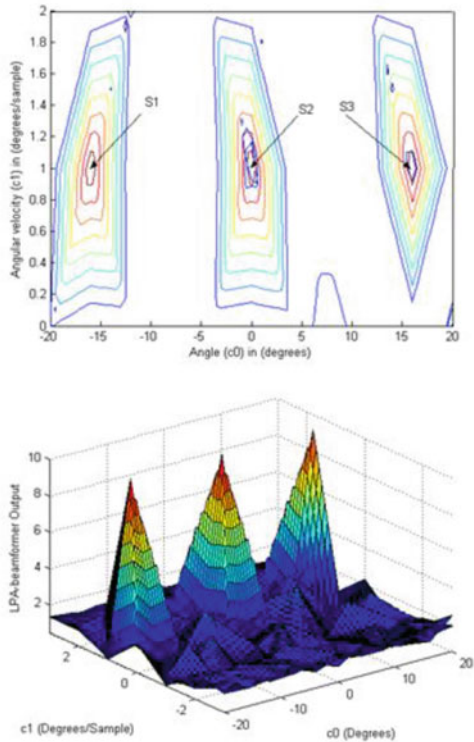
Fig. 4.4 Beamformers output for a single source case at $\theta = 0$ and $\theta^{(1)}(t) = -1$ deg/sample

value of $\theta(t)$ for all velocities. The increasing of the source velocity shifts the P_{conv} peaks from the accurate value of the angle, while a degradation of the one-peak of P_{conv} for larger values of $\theta^{(1)}(t)$ occurs in Fig. 4.3. Comparing the P_{conv} output in Figs. 4.1 and 4.4, it is clear that the peak is shifted to the right when the angular velocity is positive, while it is shifted to left when the angular velocity is negative. Moreover, the P_{LPA} has no side-lobes in a wide range around the peak, which indicates the source location, and its main lobe bandwidth is smaller than that of the conventional beamformer. This can be considered an advantage of the LPA beamformer compared to the conventional beamformer even for the case of unmoving sources. As the source moves with negative uniform angular velocity as in Fig. 4.4, the P_{LPA} beamformer indicates the source location correctly.

4.1.2 Case 2 (Well Separated Multi Sources Case)

Consider the case of three sources (microphones), where the resolution of the sources depends on the direction θ and can be given by:

Fig. 4.5 LPA beamformers output for a well separated sources at $\theta_1(t) = -16^\circ$, $\theta_2(t) = 0$, $\theta_3(t) = 16^\circ$ and they have the same angular velocity $\theta^{(1)}(t) = 1^\circ/sample$



$$\Delta\theta \cong 1/|\Gamma \cos \theta|, \quad \Gamma = (n - 1)d/\lambda \tag{4.5}$$

Assume well separated sources having $\theta_1(t) = -16^\circ$, $\theta_2(t) = 0$, $\theta_3(t) = 16^\circ$ and they have the same angular velocity $\theta^{(1)}(t) = 1^\circ/sample$ as illustrated in Fig. 4.5.

Figure 4.5 establishes that the peaks of the LPA beamformer preserve their position at the actual location of the sources, which provide the right estimate of the source velocities. While the conventional beamformer is degraded and cannot localize the sources correctly where the angular velocity is considered (i.e., the sources are non-stationary).

4.2 Simulation of Frost Beamformers of Microphone Array

Sensor arrays are commonly employed for signal separation from noises based on the DOA information. Frost’s beamformer has a significant role in speech processing for speaker localization. Each sensor in the Frost’s beamformer array is followed by a transversal filter, which has weight as illustrated in Fig. 4.6 [8]. The beamformer output is the filter outputs’ sum, where the weights updated by Frost’s constrained least mean square (CLMS) procedure to minimize the mean square error of the output signal.

The whole Frost’s beamformer system can be supplanted by one transversal FIR filter in the speech signal. The Matlab function in [9] is used to implement a time domain beamformers to recover speech signals by a microphone array of noisy microphone array measurements and to simulate an interference-dominant signal received. The Frost’s beamformer is applied as it has superior interference suppression ability compared to the time-delay approach. The Frost beamformer

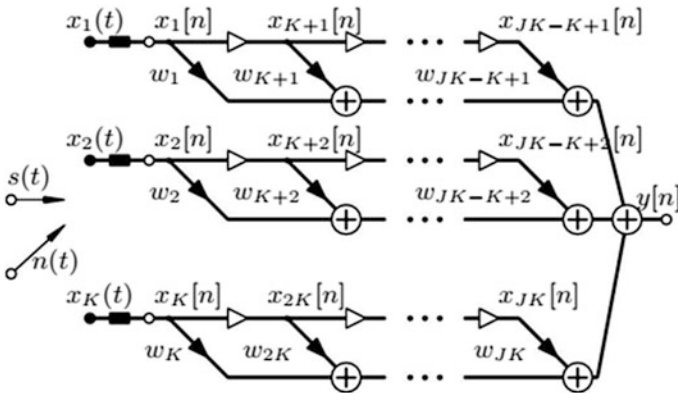


Fig. 4.6 Frost’s beamformer configuration

robustness is achieved using diagonal loading. Several scenarios are applied to show the structure effect of the omnidirectional microphones configuration, the elements' spacing and the number of speech sources.

4.2.1 Case 1 (ULA of Ten Omnidirectional Microphones)

Assume a uniform linear array (ULA) 10 omnidirectional microphones to receive the speech signal. The array elements have 5 cm spacing, where multichannel signals are received by the MA. Two recorded speeches with one laughter recordings have been included, where the laughter audio refers to the interference. The azimuth and elevation directions of the speech signals are $(-30^\circ, 0^\circ)$ and $(-10^\circ, 10^\circ)$ of the first and second speech signals; respectively, while, the interference (laughter) comes from the direction $(20^\circ, 0^\circ)$, which masks the speech signals. In addition, for each sensor, white noise signal of $1e^{-4}$ watts representing the thermal noise is considered. At the array origin, each input single-channel signal is received by a single microphone. Figure 4.7 represents the channel 3 received signal.

In order to compensate the arrival time differences (ATD) across the array, a time delay beamformer is applied with the coming signal from certain direction. A time delay (TD) beamformer is constructed to delineate a steering angle

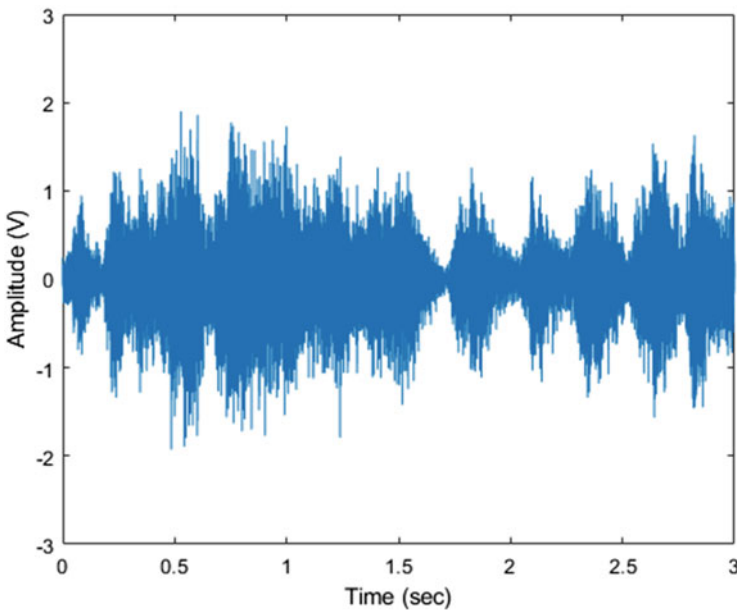


Fig. 4.7 The received speech signal at channel 3

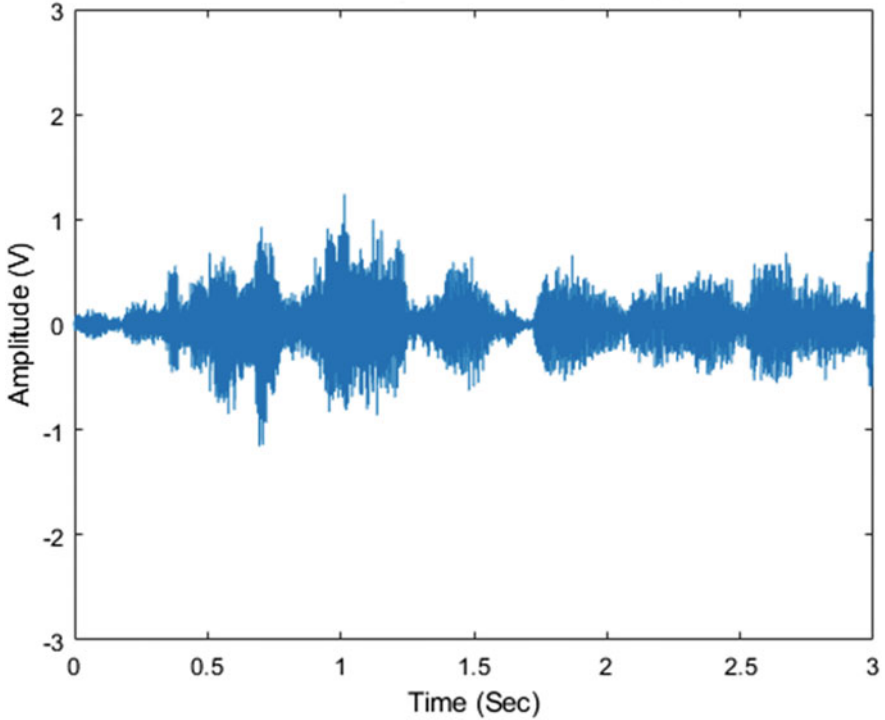


Fig. 4.8 The TD beamformer output

consistent to the first speech signal's incident direction. Figure 4.8 illustrated the TD beamformer output.

The speech enhancement can be reported by measuring the array gain representing the ratio of output to input signal-to-interference-plus-noise ratio (SINR). In addition, a Frost beamformer can be used to acquire superior beamformer performance, where the attached FIR filters to each sensor provided the Frost beamformer with more weights for suppressing the interference. Thus, nulls can be placed at the interference directions for superior interference suppression. The Frost beamformer achieved 14 dB array gain, which is 4.5 dB higher than that of the TD beamformer. Furthermore, the frost's beamformer can be used to steer the array in the direction of the second speech signal. Figure 4.9 illustrates the frost's beamformer output with diagonal loading to improve its performance.

In order to demonstrate the effect of the microphone array configuration and number of elements, the following scenarios are applied.

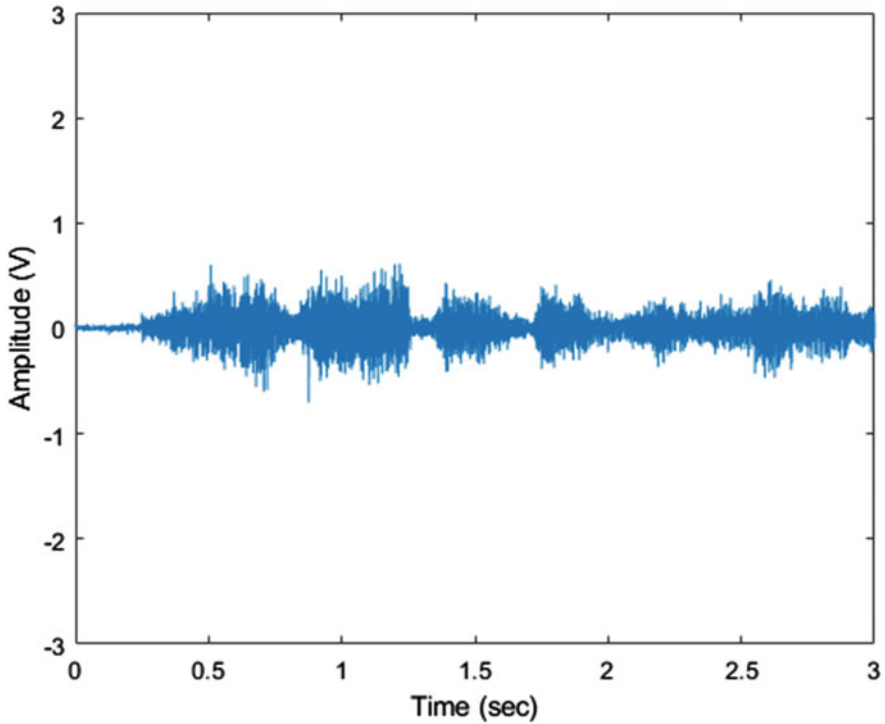


Fig. 4.9 The frost's beamformer output with diagonal loading

4.2.2 Case 2 (ULA of 5 Omnidirectional Microphones)

In this case, the same signals at the same directions are received using a ULA with five omnidirectional microphones to receive the speech signal is used instead of using ten elements. In addition, the spacing between the elements is increased to be 10 cm instead of 5 cm. Figures 4.10, 4.11 and 4.12 represent the channel 3 received signal, the TD beamformer output, and the frost's beamformer output; respectively.

In order to demonstrate the effect of the microphone array configuration, the following scenario is applied.

4.2.3 Case 2 (UCA of 5 Omnidirectional Microphones)

In this case, the same signals at the same directions are received using a uniform circular array (UCA) with five omnidirectional microphones to receive the speech

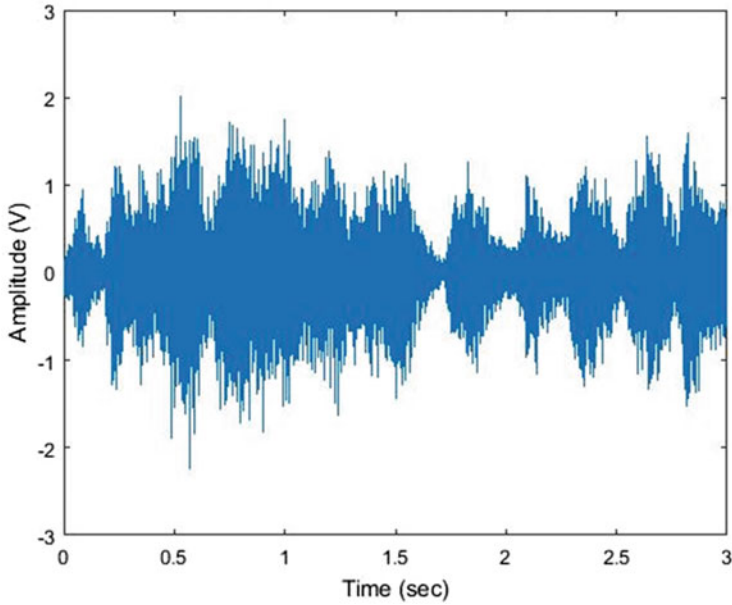
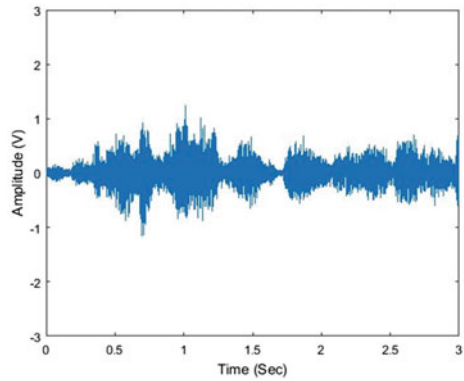


Fig. 4.10 The received speech signal at channel 3

Fig. 4.11 The TD beamformer output



signal. In addition, the radius of the circular array is 1.5 cm. Figures 4.13, 4.14 and 4.15 represent the channel 3 received signal, the TD beamformer output, and the frost's beamformer output; respectively.

A comparative study between the TD beamformer and the frost's beamformer gain values from the preceding results of the different array configurations is illustrated in Fig. 4.16.

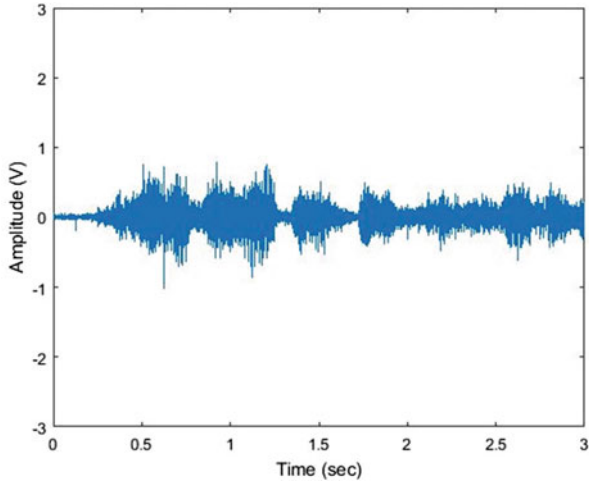


Fig. 4.12 The frost's beamformer output with diagonal loading

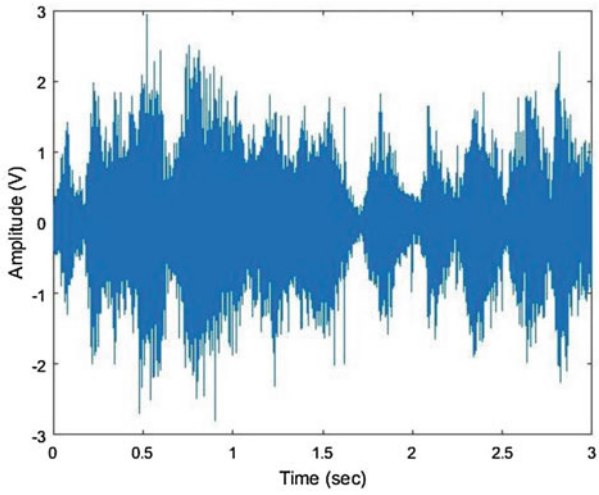


Fig. 4.13 The received speech signal at channel 3

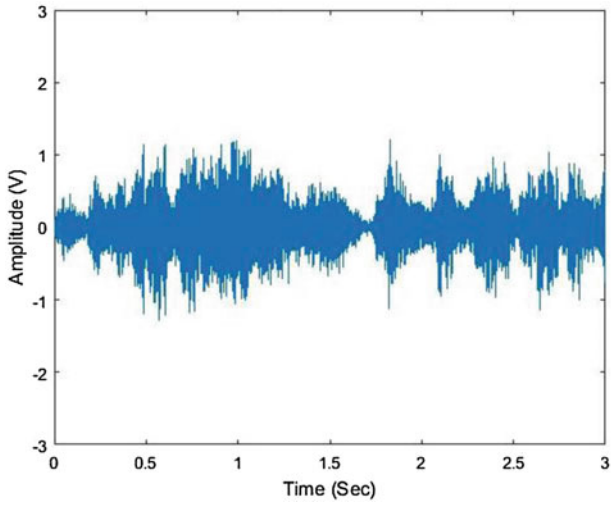


Fig. 4.14 The TD beamformer output

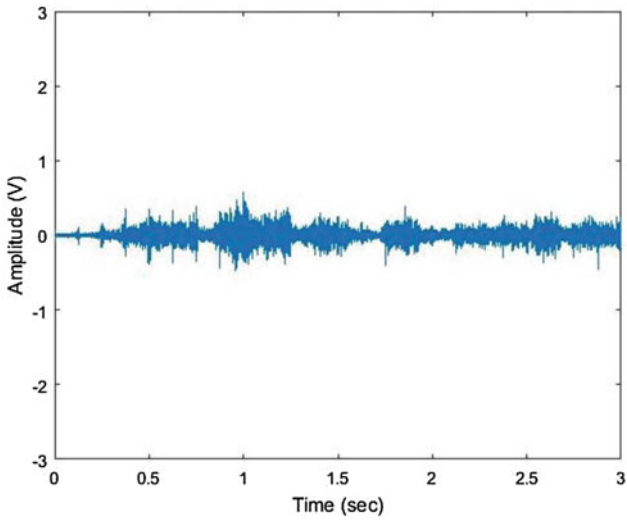


Fig. 4.15 The frost's beamformer output with diagonal loading

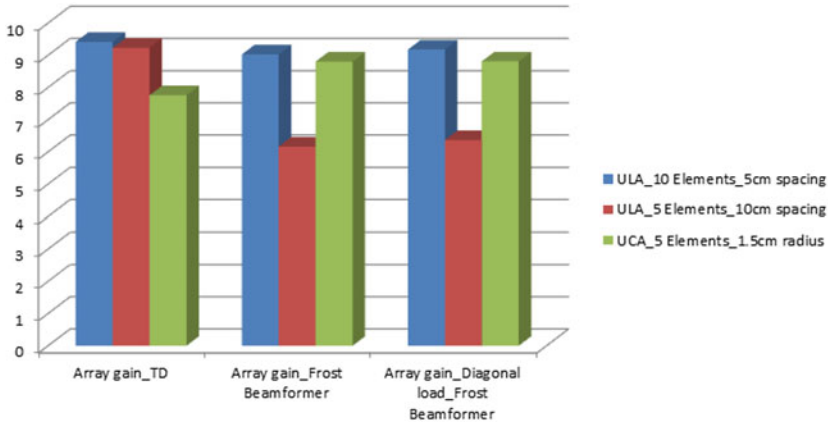


Fig. 4.16 The array gain using different configurations

4.3 Linear Microphone Array for Live Direction of Arrival Estimation

The 4 built-in microphones of the Microsoft Kinect™ are recognized using Matlab 2017 to estimate the linear coordinates [10]. The applied procedure is used independently with pairs of microphones to estimate the DOA, which are then combined to determine a single live DOA output. As the inter-microphone distance increases, DOAE sensitivity increases correspondingly. A bespoke arrow-based polar visualization (Fig. 4.17) is used to illustrate the DOA estimation of the sound source using multiple microphone pairs within the linear array, where the four microphone positions at $[-0.088, 0.042, 0.078, 0.11]$.

Fig. 4.17 The polar plot of the DOAE

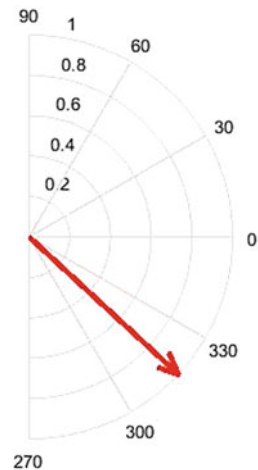


Fig. 4.18 The polar plot of the DOAE of the DOAE

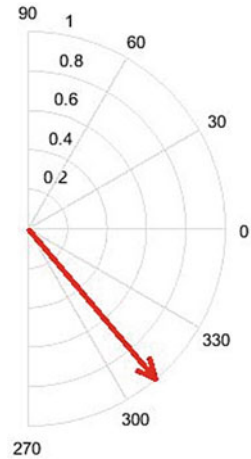


Figure 4.18 illustrates the polar plot of the DOAE when changing the microphone positions to be at $[-0.05, 0.062, 0.098, 0.19]$.

References

1. Di Claudio, E. D., Parisi, R., & Orlandi, G. (2000). Multi-source localization in reverberant environments by ROOT-MUSIC and clustering. In *2000 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2000. ICASSP '00. Proceedings* (Vol. 2, pp. II921–II924). IEEE.
2. Talantzis, F., Constantinides, A. G., & Polymenakos, L. C. (2005). Estimation of direction of arrival using information theory. *IEEE Signal Processing Letters*, *12*(8), 561–564.
3. Zhong, X., & Premkumar, A. B. (2012). Particle filtering approaches for multiple acoustic source detection and 2-D direction of arrival estimation using a single acoustic vector sensor. *IEEE Transactions on Signal Processing*, *60*(9), 4719–4733.
4. Wu, K., Reju, V. G., & Khong, A. W. (2015, April). Multi-source direction-of-arrival estimation in a reverberant environment using single acoustic vector sensor. In *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 444–448). IEEE.
5. Ashour, A. S. (2005). *Smart antenna* (Doctoral dissertation, Ph.D. thesis). Faculty of Engineering, Tanta University, Egypt.
6. Ashour, A. S., Elkamchouchi, H. M., & Nasr, M. E. (2004, June). Planar array for accelerated sources tracking using local polynomial approximation beamformer. In *Antennas and Propagation Society International Symposium, 2004. IEEE* (Vol. 1, pp. 431–434). IEEE.
7. Ashour, A. S. (2014). LPA beamformer for tracking nonstationary accelerated near-field sources. *International Journal of Advanced Computer Science and Applications*, *5*(3), 2–9.
8. Sovka, P., & Strupl, M. (2003). Analysis and simulation of frost's beamformer. *Radio engineering*, *12*(2), 1–9.
9. <https://www.mathworks.com/help/phased/examples/acoustic-beamforming-using-a-microphone-array.html>.
10. <https://www.mathworks.com/help/audio/examples/live-direction-of-arrival-estimation-with-a-linear-microphone-array.html>.