Chapter 2
Microphone Array Principles

Microphone array (MA) involves several microphones positioned at diverse spatial locations. According to the sound propagation fundamentals, the multiple inputs can be handled to attenuate or to enhance the stemming signals from specific directions (desired signal) in the presence of demeaning noise sources [1, 2]. This process is based on the known source location information, and the used MA to guarantee hands-free signal as well as noise sturdiness. Traditional classic microphones should be near the speaker either by wearing the microphone or by using movable microphone. In order to solve such situation, microphone beamforming can be employed by using several microphones to generate directed beam patterns focusing on a specific speaker [3, 4]. These steerable microphones are motivated by the teleconferencing applications. In microphone array speech processing, a microphone array involves multiple microphones located at diverse spatial positions. Such arrays have an excessive impact in several speech processing, real-world applications, owing to their capability to deliver hands-free signal acquisition as well as noise robustness signals. According to the sound propagation fundamentals, the manifold inputs are handled to enrich the desired speech signals and to diminish the unwanted noise signals originating from other directions. This improvement of the input signal is founded on the information about the source location as well as the applied microphone array procedures.

MA is used to communicate acoustic signals, explicitly, speech signals. Array processing handles multiple sensors to transmit/receive the propagated signal waves [5]. These sensors have several applications as in microphone arrays to communicate the acoustic signals, such as speech signals. The sensor array is considered a new research area in speech processing that represents a sampled continuous apertures' version, where the aperture refers to a spatial area which transmits/receives the propagating waves [6]. The transmitting aperture is considered the active aperture, whereas the receiving aperture is a passive one. In acoustics such as speech, the aperture is an electro-acoustic transducer for converting the electrical signals into acoustic signals (loudspeaker) or vice versa (microphone). Inherently, the receiving aperture response is directional, where the detected signal amount by

N. Dey and A. S. Ashour, *Direction of Arrival Estimation and Localization of Multi-Speech Sources*, SpringerBriefs in Speech Technology,
https://doi.org/10.1007/978-3-319-73059-2_2

the aperture differs with the direction of arrival. Thus, the beam pattern/aperture directivity pattern is known as the aperture response, which is a function of the direction of arrival and frequency.

The sensor arrays' ultimate theory is applied in speech signal processing based on the wave propagation theory [7, 8]. In this context, the following array processing principles are considered. Sound waves broadcast as longitudinal waves through fluids, where in the propagation direction, the fluids' molecules move back/forth leading to compression/expansion. For acoustic waves, the complex generalized wave equation depends on the fluid properties as a function of the sound pressure at a point in space and time. In addition, the propagation speed depends on the fluid density and pressure. For array processing algorithms, the spherical wave solution of the wave propagation equation indicates the dependency of the signal amplitude decay rate on the source distance, which has significant associations with sources in the near-field. Typically, sound waves are spherical in nature, however, at an adequate distance from the source, they can be considered as plane waves.

2.1 Models of the Acoustic Signals and Sources

In many applications, there are needs to design antennas with very high gains or directivity to meet the demands of long distance communication. By enlarging the dimensions of the antenna, high gains can be achieved. However, another way to achieve high gain without necessarily increasing the physical size of the individual elements will be to form an assembly of radiating elements arranged in a particular geometrical configuration [9]. This configuration of multi-elements is referred to as an antenna array. In most cases, the multi-elements in an antenna array are identical. Although not necessary, it is often simpler, more convenient and practical when it comes to the design and analysis of an antenna array. However, the primary reason for using antenna arrays is to be able to produce a directive beam that can be repositioned or scanned electronically.

2.1.1 Microphone Array

A phased array consists of multiple stationary radiating elements spaced a distance, d apart. The radiating elements have unity gain in all directions; i.e. they are isotropic radiators [10–12]. The radiating elements can be arranged in a few geometrical configurations, namely linear, circular and planar. Figure 2.1 illustrated the radiating elements different arrangement.

Consider M microphones in a linear microphone array with K incident signals from arrival angles q_k. Assume the incident waves to be plane waves, thus, each microphone in the array will collect each of the signals, s_k with time-delayed forms of each other. A reference point in the system can be considered to calculate each

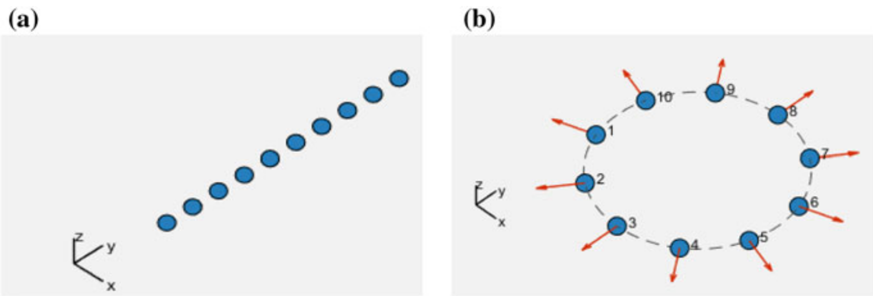
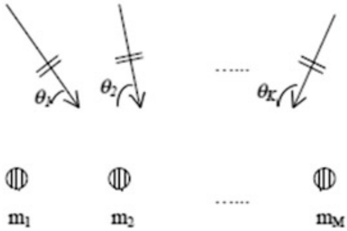


Fig. 2.1 Different microphones array configurations **a** linear array, and **b** circular array

Fig. 2.2 Microphone array structure



signal phase. Figure 2.2 illustrated that microphone 1 is considered the reference point, where d_i represented the distance from microphone m_1 to m_i. The planar distance from source s_k to the microphone m_i that the sound travels relative to the distance to reference microphone is given by [13, 14]:

$$distance_{i,k} = d_i cos\theta_k \tag{2.1}$$

The equivalent time delay to each microphone in the array is represented by:

$$\tau_i = \frac{distance_{i,k}}{v} \tag{2.2}$$

where v represents the sound velocity, which equals 343 m/s. At each microphone, the signal s_k produces an input g_i as follows:

$$g_i(t) = |s_k(t - \tau_i)| \tag{2.3}$$

The same $g_i(t)$ is produced for a source received from a direction between 0° and −180°, where the source reached at its positive angle. For the narrowband cases, this signal can be denoted as a phase shift of the received signal as follows:

$$g_i(t) = s_k(t)e^{-j\omega_0\tau_i} \tag{2.4}$$

At the *ith* sensor, the complete received signal is a collection of the K incoming signals from K sources and the noise, which is given by:

$$g_i(t) = \sum_{k=1}^{K} s_k(t) e^{-j2\pi d_i \cos\theta_k/\lambda} + n_i(t) \tag{2.5}$$

where $\lambda = v/f_0$.

The microphones' spacing is an important issue that affects the microphone array system. Consequently, different microphone configurations can be tested to find the optimal configuration based on the problem under consideration [15]. Moreover, the spatial frequency domain considerations are crucial for avoiding aliasing, where smaller d/l leads to fewer side-lobes with nulls in the directional pattern. An omnidirectional pattern is formed as d/l tends to zero.

2.1.2 Near Field Considerations

For the incoming wavefront, most of the studies consider adaptive arrays having planar wave (far-field) hypothesis. However, in microphone arrays, the near-field hypothesis must be considered leading to spherical wave [16]. The near-field spherical wave has two main parameters, namely θ and r, where r is the radial distance from the source to the array. This complicated the model for solving the near-field problem.

Under the near-field hypothesis, the distance the traveling wave from the source to any point is equal to the line length that connects the two points [17]. Consequently, the distance r_i that the wave takes to each microphone equals the distance from the source to each microphone, where the relative time delay into each microphone can be established by the subtraction of the smallest r from each of r_1, r_2, r_3, r_4 with dividing the result by the sound speed.

2.1.3 Microphones Array Configurations

At different spatial locations, the microphones are placed in the microphone array that act as an omnidirectional acoustic antenna. Linear arrays, circular arrays and spherical microphone arrays are several configurations for designing the microphone array [18, 19]. For a source sound, the spatial location can be determined by the received signal correlation of the separated different microphones [20]. Mainly, the microphone array is used to enhance the signal under concern and to reduce/ suppress the noise reduction using optimal filters for source separation as well as speakers tracking [21, 22]. Recently, the development of efficient speech communication devices has become an active research area with the advancement in the speech processing equipment [23, 24]. Such requirement is realized by microphone array that allows the user with hands free environment deprived of carrying microphone or speaking close to the microphone. Another application of the MA is in the video conferencing as well as for human machine interfaces and speech recognition.

Recently, MAs become popular due their accurate performance even with the limitations of the array processing (high hardware costs and large processing time). Researchers are interested to develop novel speaker localization algorithms to enhance the speech systems. These algorithms can be classified into one phase process, two phase process and spectral estimation [25, 26]. The different environmental conditions affect the performance of each algorithm. For example, room reverberation and various reflections of the received speech signal from a speaker in the acoustic environment lead to corrupt the signal from the background and additive noise. In order to resolve this problem, the speech signal can be recorded using a set of spatially MA, which needs localization as well as tracking of the moving speaker to steer electronically the MA to improve the quality of the speech acquisition. Moreover, speaker localization is essential in the multi-speaker scenarios [27, 28].

In order to perform such localization tasks different configurations of the MA can be used and selected according to the application under concern characteristics and parameters including the signal time variance, the spatial resolution, the acoustical situation, the frequency range, and the data processing [29, 30]. The MA is used to designate a complex sound field. Typically, the MA allows the signal capturing in several points at the same time with the cross spectra between each point to obtain the data presenting the direction of the coming sound at each time. MAs are used for signal recording that can be played back later in another space. Several microphones will help in understanding the sound field complex spatial pattern to renovate this pattern in another space.

2.1.4 Array Geometries

2.1.4.1 Scanning Arrays

As a rule of thumb, more microphones are required for better representation when back-propagating complex radiation pattern or when measuring a complex radiation field. However, increasing the number of microphones requires increasing the number of the recording channels. In order to solve this restriction, a scanning array can be used, where a small set of microphones are located at all desired positions to record the field. The total array configuration depends on the main microphones' mounting. For example, if the microphones are fixed in a linear array; they are appropriate in a regular grid spacing, while, if the microphones are fixed in a half-circle, the circle can be moved around a center point to attain a full sphere covered while the microphone recording process for static signals. A motor is used to automatically move the array around while the recording setup, consequently, it is conceivable to partially record the sound field at the different positions.

2.1.4.2 Planar Microphone Arrays

The planar array is considered the expected postponement to the linear array. Different configurations plane arrays (PAs) that consist of a regular microphones grid can be used for a simplified line array [31]. The Pas can be used in the complex radiating surface reconstruction.

2.1.4.3 Spherical Microphone Arrays

For a simple sound recording nearby a source, the spherical arrays are an expected choice, where the sound source is covered by a sphere in a regular angle, and the radiating body's radiation strength is directly measured. Spherical microphone array arrangements decompose the radiation into spherical harmonics. For the lowest harmonics, the same pressures occur at all microphones. Thus, in such case instead of using the spherical array, a dual-sphere array having two spheres of different radii can be employed. The optimal radius selection is considered a complex problem to be resolved before the recording.

The spherical arrays are considered as compact arrays with a small sphere in space instead of a bulky sphere nearby the radiator [32, 33]. In order to improve the traditional recording, the compact arrays that elaborated microphone recording configurations are considered. In such configuration, different orientation two directional microphones are used for recording.

Furthermore, in order to obtain a response equal that of the non-baffled directions, an omnidirectional microphone is applied. For example, using Matlab function "phased.OmnidirectionalMicrophoneElement" models the omnidirectional microphone that demands the operating frequency range specifications of the microphone [34]. In order to inspect the microphone directionality, namely the polar pattern, an azimuth cut is required by setting the elevation argument as zero single angle. Figure 2.3 both azimuth and elevation are from −90 to +90, which is illustrated the polar plot of a microphone power response at 20 Hz, 1 kHz, and 20 kHz; respectively, where Matlab 2017 is used to construct the omnidirectional microphone element with a response in the audible frequency range of the human, namely 20–20,000 Hz.

In Fig. 2.3, the green color represents the normalized power of '1', where all frequencies have the same response, which means that the microphone has a stable flat frequency response. Generally, in terms of the frequency response, there are two main microphone types, namely (i) flat frequency response, where all audible frequencies ranging from 20 Hz to 20 kHz and have the same output level, and (ii) tailored frequency response at which the microphone may have a peak in the frequency range from 2 to 8 kHz to increase intelligibility for live vocals. The flat

(a)

(b)

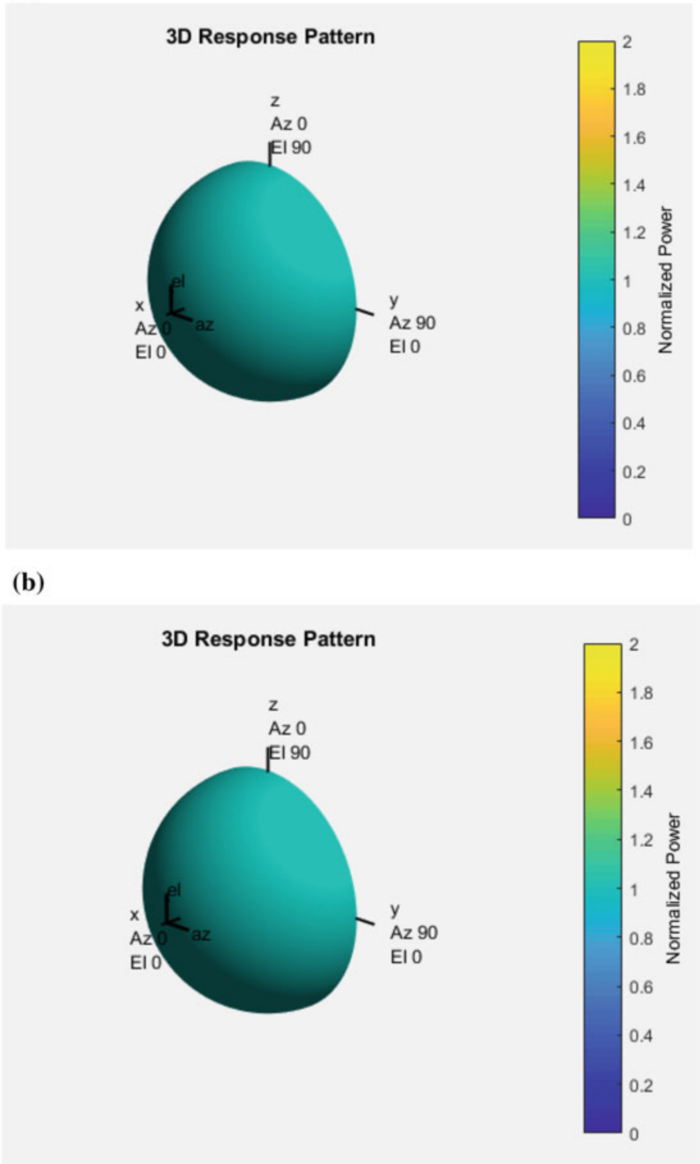


Fig. 2.3 **a** Three-direction (3D) power response pattern of omnidirectional microphone at 20 Hz. **b** 3D power response pattern of omnidirectional microphone at 1 kHz. **c** 3D power response pattern of omnidirectional microphone at 20 kHz

(c)

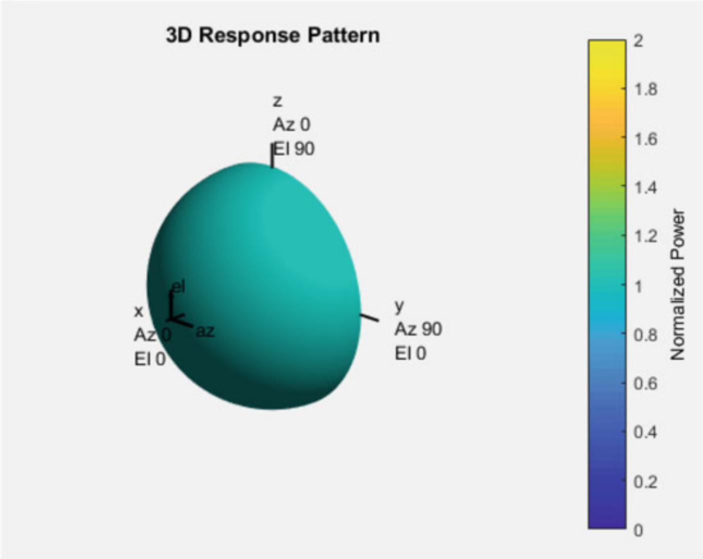


Fig. 2.3 (continued)

frequency response microphones are suitable for applications at which the sound source has to be reproduced (recording) without changing the original sound, while the tailored response microphones are ordinarily designed for the sound source enhancement in a precise application.

2.2 Sensor Arrays

The acoustic (speech) signals broadcast can be measured as functions of time/space variables with conserved signal information through the propagation. Accordingly, over all space and time, in order to renovate the signal it is required to temporally sample the band-limited signal at the specific space location or spatially sample the band-limited signal at the specific time instant. For all sensor/aperture array signal processing processes, the following associations are significant, namely (i) the dependence of the propagation speed on the medium characteristics, (ii) reinstating the signal over all space and time, (iii) the propagated wave decayed amplitude from the source, and (iv) propagating multiple waves prospect without interaction. For DOA estimation and source localization, separating the MAs signals using several algorithms can be engaged to discriminate the different signals inconsistent with their temporal and spatial characteristics, assuming lossless/homogeneous medium without considering the dispersion, propagation speed and diffraction.

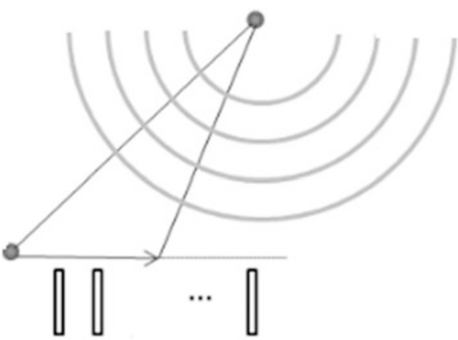
The sampled continuous aperture is a discrete sensor array, where the aperture is excited at a limited number of discrete points. The overall sensor array's response is determined by the individual sensor response superposition that is consistent with the continuous aperture [35]. The equally spaced linear MA directivity pattern of identical sensors is contingent on the array's inter-element spacing, the frequency, and the number of array elements. The effective length of the sensor array is the continuous aperture length, while, the array actual physical length is signified by the distance between the first and last sensors in the array. In order to avoid aliasing and grating lobes in the directivity pattern, the sensor arrays implement spatial sampling with considering the Nyquist frequency. Moreover, the array gain is calculated to measure the sensor array effectiveness that represents the enhancement in signal-to-noise ratio between a reference sensor and the array output.

For a linear aperture, the wave source derives from the aperture far-field as plane waves with neglecting the wavefront curvature [36]. In several applied applications of the sensor arrays, such as speech recognition and DOAE, this criterion is unsatisfied as which the signal source is placed within the array near-field. In the near-field, in order to control a sensor array, the near-field directivity pattern is considered to match the far-field consistent directivity pattern of satisfying the frequency dependent sensor. Commendably, beam steering can be realized by applying time delays to the sensor inputs.

2.3 Speech Processing Requirements

Microphone arrays are the most advanced ways to improve the speech quality. A single microphone can provide the directivity for noise and reverberation reduction without any required post-processing [37]. However, an MA has been efficiently enhanced the speech signal quality by directing the received radiation pattern in the desired signal direction, thereby decreasing the interference and improving the captured sound quality. Figure 2.4 illustrated the MA that can be used for the speech signal enhancement in the direction of interest.

Fig. 2.4 Microphone arrays



Array processing is the main speech processing requirements, where signals can be classified from the statistical point of view into deterministic/random wave-shape [38]. The deterministic wave refers to a known waveform that has unknown parameters, including the delay, amplitude, and/or scale. These waves are encountered in active sonar, data communications, and radar, where the transmitted signal waveform is known to the receiver. Nevertheless, in some applications, unknown signals are transmitted or signals may be affected by the transmission environment leading to changes with space and time. In this case the wave is considered random, where; all the obtained information is enclosed in the probability distribution function. However, in practice, the speech signals are disturbed by interference/noise that mix with the actual signal and modify its properties. Generally, interference has the same signal characteristics and may be produced by a similar source or due to the energy diffusion in a multiplex communication of one channel into the neighboring channel [39], electromagnetic coupling between two end-to-end wires, concurrent pick up of several stations by a single radio receiver, and the multipath transmission. The interference can be confused with the useful signal leading to its destructive nature, while the noise is commonly produced independent of the signal.

In array processing, the wave fields in space are signal-independent. A spherical isotropic noise is designed if the noise-field is produced by uncorrelated random waves propagating in all directions. Generally, the noise-field is assumed to be stationary in space and time. The information may convey by multi-dimensional signals, which are formed using frequency, time, or spatial diversity as in communication over a fading channel. The spatial field is sampled by multiple sensors in array processing. At different channels, the signals are collected from a multi-dimensional signal. The information combination of different dimensions can increase the processing performance. In communication, using a beamformer in array processing diversity diminishes the error reconstruction, and raises the signal-to-noise ratio (SNR).

The array processor goals are classified into field characterization or signal enhancement. Signal enhancement of the array output includes SNR improvement through beam steering in the source direction or through nulls inserting of the beam-pattern in the noise/interference direction [40]. The beam steering conventional methods places delay elements in the sensor output for further computing the delayed outputs' weighted sum. At the delay element's output, the signals having the same specific direction will appear at the same phase in the proper delays selection, which is known as beamforming. The array output signal power is amplified by the square magnitude of the number of sensors if the speech source is positioned at the beam direction. In the beamformer output, the noise power increases linearly with the increase in the sensors' number in an uncorrelated inter-sensor noise case. Consequently, the SNR increased by applying a conventional beamformer.

In order to estimate the source's spatial properties including the range, elevation, azimuth, velocity, and movement direction; the field characterization is required. The field characterization is completed in two stages, namely (i) determining the emitting source number, which is called detection, and (ii) localizing and estimating the signal position in the space. For the far field sources, the incoming wave fields are planar, thus, only azimuth and elevation directions can be estimated. Likewise, if the array and sources are in the same plane, the DOA will be the only spatial parameter of the emitting source.

In wireless communication, the narrow-band array processing is required, where a moving transmitter in indoor or mobile communication produces narrowband signals [41]. The receiver involves an antenna array that receives the original source's signal and its reflections from the nearby objects. In order to reduce the reflected wavefronts effect, the antenna array steered a beam in the source direction to estimate the source location. The antenna array drives the power in the transmission mode in the source direction only by establishing a steered beam. Thus, the energy is preserved and meanwhile the power is transmitted only in a specific direction where smaller interfering influences the other receivers.

Array processing procedures can similarly be realistic with wideband signals, where the frequency bandwidth is comparatively large associated with the center frequency [42]. A microphone array for instance, can be used to localize a speaker in a specific place. At the array, the arriving signal is a wideband speech signal accompanied by reflections from the surroundings. The reflection effect that interferes with the direct signal can be compensated by beam steering towards the speaker direction. Another wideband array processing example is the MA can be used in the hands-free mobile communications inside a car, where the driver voice is collected by microphone array. Forming a beam-pattern with nulls in the noise direction reduces the car environment noise. The foremost noise fields in the microphone array applications are considered according to the correlation degree between the noise signals at dissimilar spatial locations. The coherence is considered the most used measurements of the correlation. The directly propagated signal noise to the microphones is deliberated as coherent noise without any form of dispersion, dissipation, or reflection due to the acoustic environment. Conversely, the measured noise at any certain spatial location in an incoherent noise field is uncorrelated with the measured noise at all other locations. Generally, in the MA applications, the electrical noise in the microphones is distributed randomly, which can be considered an incoherent noise source. A diffuse noise field refers to noise of identical energy that broadcasts simultaneously in all directions. Therefore, in the diffuse noise field, the sensors will receive noise signals, which are poorly correlated, but have the same energy [43]. Numerous practical noise situations can be branded by a diffuse noise field, including the car noise. In a diffuse noise field, the noises coherence at any two points is a function of the distance between the sensors.

2.4 Microphone Array Beamforming

Beamforming techniques can be categorized according to the data type, namely data-dependent/data-independent. In the data-dependent (adaptive beamforming) methods are continuously adapted their parameters based on the received signals

[44]. Nevertheless, the data-independent (fixed beamforming) methods have beamformers with fixed parameters during the operation. For different noise situations, different beamforming methods can be proper to determine the encountered noise fields in the MA applications.

For speech applications, MAs have a small number of elements that should cover only the audible range of the electromagnetic spectrum. Consequently, compared to other applications, such as wireless beamforming, the precise array manifold can be pre-designed. The beamforming MA techniques are generally belongs to signal processing, which an extensive range of applications has extended from source detection/separation to de-noising or de-reverberation that have the common delay and filter methods. These approaches intend to improve the speech signal using a multichannel recording. Hence, these signal reconstruction converges to effective results using four to eight microphones. Various filtering methods are based on the speech processing, including Wiener filter, and Kalman filter [45]. Typically, the extra sources are correlated between the microphones without any correlation with the source under concern. In addition, the reverberation is associated between the microphones as well as with the source of interest. Such unwanted signals should be separately treated.

Direction-of-Arrival (DOA) estimation has an energetic role in several applications, especially for speakers' localization. Beamforming is the most protuberant procedure for estimating the DOA. It is used to accompany by sensors/antenna array to transmit/receive signals to/from certain spatial direction even with the incidence of noise/interference. One of the simplest microphone array beamforming methods is the delay-sum beamforming (DS). In order to steer the directivity pattern's main lobe in the wanted direction, the phase weights apply to the input channels. In the delay-sum beamforming, the sensor inputs in the time domain are mainly delayed by specific seconds, and then a summation process is applied to provide a single array output. In the summation, each channel is prearranged an equal amplitude weighting for unity directivity pattern in the desired direction. The DS is considered one of the filter sum beamformers at which both the amplitude/phase weights are frequency dependent. Several beamformers are considered filter-sum beamformer that use matrix algebra to describe the microphone array methods.

The speech signal is broadband, thus, for frequency invariant beam-pattern; a single linear array design is insufficient [46]. In order to handle the speech broadband signals, an array can be implemented as a series of linear sub-arrays with uniform spacing to provide the desired response characteristics for a specific frequency range. Smaller array length is essential with high frequencies to preserve constant beam-width. Furthermore, in each sub-array, the number of elements must remain the same to guarantee the same side-lobe level across different frequency bands. Thus, the sub-arrays are realized in a nested manner, where any sensor can be used in several sub-arrays.

Super-directive beamforming is another beamforming procedure that achieves the ability for closely spaced end-fire arrays. Its channel filters are expressed to maximize the array gain [47]. Its filters depend only on the array geometry and the source location that are calculated for a specified configuration. The super-directive techniques are efficient as it provides satisfactory array performance at low frequencies for speech processing applications.

For conventional beamforming procedures, low frequency performance is challenging due to the large wavelengths that provide insignificant phase differences between closely spaced sensors that lead to deprived directive discrimination. Near-field super-directivity is an adjustment of the typical super-directive method that considers the amplitude differences and the phase differences [48]. It provides directional sensitivity with greater performance compared to the standard near-field for sources at low frequencies, where the phase differences are insignificant at low frequencies, mainly when the sensors are positioned in an end-fire arrangement, which maximizes the distance difference from the source to each microphone.

2.5 Far-Field and Near-Field Source Location

In several engineering applications, source localization is the vital stage in speech signal processing, array signal processing, and wireless communication. Different DOAE estimation algorithms are developed for far field narrow band sources. However, it becomes intricate in the case of near-field sources in the Fresnel zone of the array aperture. In near-field scenarios, such as the speech enhancement with microphone array in a conference room, seismic exploration; under water source localization; and ultrasonic imaging, the impinging wavefront on the array is spherical, thus, range information is required along with the estimated DOA of the sources for accurate localization. Researchers are interested to develop methods for near-field source localization, such as the two-dimensional (2D) Multiple Signal Classification (MUSIC) method, Maximum Likelihood (ML) method, high-order spectra (HOS) based algorithms, and Estimation of Signal Parameters via Rotational Invariance Technique (ESPRIT) method [49, 50]. In practical situations, especially in the case of closely spaced, multiple sources, these techniques are computationally heavy and needs extra computations to pair the parameters, which causes poor DOAE in low Signal to Noise Ratio (SNR).

2.6 Speech Source Direction of Arrival Estimation and Localization

Microphone arrays are essential for spatial analysis, where numerous microphones perceive more than only one microphone. The speech analysis and reconstruction become easier with more collected information on the space [51]. Microphone array methods are involved in several applications that can be categorized into two core domains according to their problems and mathematical models, namely (i) the beamforming methods are based on signal processing, and (ii) the recorded sound field back-propagation against the radiating body according to the Helmholtz–Kirchhoff integral.

Complex sound/speech environments, such as classrooms/auditoriums, cars, planes, trains, and outdoor places, depend on the listener/speaker position, where any change in the location leads to a corresponding change in the sound perception. Thus, it is essential to determine distance to the source estimation, the source location at a definite angle, separating the different speakers, concentrating with certain speaker; which called cocktail party effect. For example, the position of musical instruments in musical space facilities the separation process of them. The height, depth, or size of a space can be estimated using reverberation information. For speech extraction and speaker selection, some reverberation can be neglected.

The standard beamforming can be considered as a source localization problem that can be enhanced by using several microphones recording from different positions [52]. Such a case occurs in the reverberant situations at which the source is reflected several times, and the source position localization become complex compared to free space. Moreover, this problem becomes more complex in the case of multiple sources, such as many speakers in a room. In conference rooms and in a car, the noise and reverberation are strong, thus, it is essential to localize the source as well as to use the MA to de-noise the recorded signal. In addition, de-reverberate the signal and reconstruct the speech signals are vital. These requirements are achieved through using fixed array to measure the reverberation easily to increase speech intelligibility.

2.6.1 Sound/Speech Source Localization

Localizing single sound source at a distance is considered a near-field problem. In contradiction of the equivalent source procedures that produce a complex radiation pattern, the localization estimation produces a single angle value relative to the normal direction of the MA, which is the preferred single source location [53]. The phase of a signal $s(t)$ is considered the most significant property for its detection. At each of N microphones, the signal phase is different, recording a time series $m_i(t)$, where $i = 1, 2, 3, \ldots N$:

$$m_i(t) = A_i s(t + \Delta t_i) + \xi_i(t) \tag{2.6}$$

where $\xi_i(t)$ is the uncorrelated noise between the microphones and Δt_i are the different time delays between the source point and the i microphones. When a signal traveling from the far field reaches at the array, the arrival time difference of this wave at the microphones [54] can be expressed as:

$$\Delta t_{ij} = \Delta t_i - \Delta t_j = \frac{r_i - r_j}{c} \tag{2.7}$$

where r_i and r_j are the distances between the source and its corresponding microphones as c is the sound speed.

Assume a source located at specific DOA with the angle φ_i between the source and the MA plane at microphone i. Using three microphones of distance d, thus,

$$r_2^2 = r_1^2 + d^2 + 2r_1 dcos(\varphi_1) \tag{2.8}$$

$$r_3^2 = r_1^2 + 4d^2 + 4r_1 dcos(\varphi_1) \tag{2.9}$$

For three microphones, the time delay between a sensor i and its nth neighbour is:

$$\Delta t_i = \frac{(n-1)d}{c} cos(\varphi) \tag{2.10}$$

Consequently, the time delay Δt_i can be estimated from the phase difference between the two sensors and then the DOA φ can be measured. Moreover, once the angle is identified, the distances r_i between the source and sensors can be estimated.

2.6.2 Directional of Arrival Estimation

The angle of arrival or the wave number estimation problem of a plane wave is considered as DOA estimation problem or direction finding. DOAE plays an energetic role in several applications including sonar, radar, electronic surveillance, seismic systems, radio astrology, and medical diagnosis. Beamforming is considered the most protuberant method for estimating the DOA. Over the last several decades, the DOAE has magnetized the researchers' attention due to its widespread applications and the complexity of determining the optimum estimator. Numerous approaches were developed to address the DOAE problem of multiple sources using the received signals at the sensors. Array processing requires either the knowledge of the direction of the desired signal source or a reference signal. Thus, antenna arrays are extensively used to resolve the direction finding problems.

For the sound source, the DOA is imperative information for any beamforming system. The beamformer can direct itself to capture signals coming only from the DOA, while ignoring others [55]. Time delay of arrival is considered one of the most prevalent DOAE approaches, which is called phase differencing. For estimating the DOA delay, consider a sound source located in the far field, two microphones on the same plane will receive the signal with slightly different times. The precise arrival difference value is based on the sound source angle relative to the microphones. Hence, the delay information can be used for estimating the direction at which the source is impinging on the array for further measuring of the propagation delay between the microphones to steer the beamformer toward the captured signals in the desired direction and rejecting all other signals [56].

References

1. Levy, A., Gannot, S., & Habets, E. A. (2011). Multiple-hypothesis extended particle filter for acoustic source localization in reverberant environments. *IEEE Transactions on Audio, Speech and Language Processing, 19*(6), 1540–1555.
2. Doclo, S., Kellermann, W., Makino, S., & Nordholm, S. E. (2015). Multichannel signal enhancement algorithms for assisted listening devices: Exploiting spatial diversity using multiple microphones. *IEEE Signal Processing Magazine, 32*(2), 18–30.
3. Ryan, J. G., & Goubran, R. A. (2000). Array optimization applied in the near field of a microphone array. *IEEE Transactions on Speech and Audio Processing, 8*(2), 173–176.
4. Benesty, J., Chen, J., & Huang, Y. (2008). *Microphone array signal processing* (Vol. 1). Berlin: Springer Science & Business Media.
5. Haimovich, A. M., Blum, R. S., & Cimini, L. J. (2008). MIMO radar with widely separated antennas. *IEEE Signal Processing Magazine, 25*(1), 116–129.
6. Huang, K., & Larsson, E. (2013). Simultaneous information and power transfer for broadband wireless systems. *IEEE Transactions on Signal Processing, 61*(23), 5972–5986.
7. Chan, S. C., & Chen, H. H. (2007). Uniform concentric circular arrays with frequency-invariant characteristics—theory, design, adaptive beamforming and DOA estimation. *IEEE Transactions on Signal Processing, 55*(1), 165–177.
8. Gorodetskaya, E. Y., Malekhanov, A. I., Sazontov, A. G., & Vdovicheva, N. K. (1999). Deep-water acoustic coherence at long ranges: Theoretical prediction and effects on large-array signal processing. *IEEE Journal of Oceanic Engineering, 24*(2), 156–171.
9. Legay, H., & Shafai, L. (1994). New stacked microstrip antenna with large bandwidth and high gain. *IEE Proceedings-Microwaves, Antennas and Propagation, 141*(3), 199–204.
10. Rodenbeck, C. T., Li, M. Y., & Chang, K. (2003). A novel millimeter-wave beam-steering technique using a dielectric-image line-fed grating film. *IEEE Transactions on Antennas and Propagation, 51*(9), 2203–2209.
11. Krnac, B., & Haring, J. (2008, April). Improving of UHF RFID pattern steering by phased antenna array. In *2008 18th International Conference Radioelektronika* (pp. 1–4). IEEE.
12. Cheng, D. K. (1971). Optimization techniques for antenna arrays. *Proceedings of the IEEE, 59*(12), 1664–1674.
13. Brandstein, M., & Ward, D. (Eds.). (2013). *Microphone arrays: Signal processing techniques and applications*. Berlin: Springer Science & Business Media.
14. Brandstein, M. S., Adcock, J. E., & Silverman, H. F. (1995). A practical time-delay estimator for localizing speech sources with a microphone array. *Computer Speech & Language, 9*(2), 153–169.
15. Boone, M. M., Cho, W. H., & Ih, J. G. (2009). Design of a highly directional endfire loudspeaker array. *Journal of the Audio Engineering Society, 57*(5), 309–325.
16. Dmochowski, J. P., Benesty, J., & Affes, S. (2007). A generalized steered response power method for computationally viable source localization. *IEEE Transactions on Audio, Speech and Language Processing, 15*(8), 2510–2526.
17. Rumsey, F., & McCormick, T. (2012). *Sound and recording: An introduction*. Boca Raton, FL: CRC Press.
18. Alexandridis, A., Griffin, A., & Mouchtaris, A. (2017). *U.S. Patent No. 9,549,253*. Washington, DC: U.S. Patent and Trademark Office.
19. Roig, E. T. (2014). *Eigenbeamforming array systems for sound source localization* (Doctoral dissertation, Ph.D. thesis). Technical University of Denmark.
20. Zhu, N. (2011). *Locating and extracting acoustic and neural signals* (Doctoral dissertation). Wayne State University.
21. Brandstein, M., & Ward, D. (Eds.). (2013). *Microphone arrays: Signal processing techniques and applications*. Berlin: Springer Science & Business Media.

22. Valin, J. M., Rouat, J., & Michaud, F. (2004, September). Enhanced robot audition based on microphone array source separation with post-filter. In *2004 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2004 (IROS 2004). Proceedings* (Vol. 3, pp. 2123–2128). IEEE.
23. Perez-Meana, H. (Ed.). (2007). *Advances in audio and speech signal processing: Technologies and applications*. Hershey: Igi Global.
24. Huang, X., Baker, J., & Reddy, R. (2014). A historical perspective of speech recognition. *Communications of the ACM, 57*(1), 94–103.
25. Markovich, S., Gannot, S., & Cohen, I. (2009). Multichannel eigenspace beamforming in a reverberant noisy environment with multiple interfering speech signals. *IEEE Transactions on Audio, Speech and Language Processing, 17*(6), 1071–1086.
26. Kumar, L., Tripathy, A., & Hegde, R. M. (2014). Robust multi-source localization over planar arrays using music-group delay spectrum. *IEEE Transactions on Signal Processing, 62*(17), 4627–4636.
27. Ma, W. K., Vo, B. N., Singh, S. S., & Baddeley, A. (2006). Tracking an unknown time-varying number of speakers using TDOA measurements: A random finite set approach. *IEEE Transactions on Signal Processing, 54*(9), 3291–3304.
28. Minotto, V. P., Jung, C. R., & Lee, B. (2014). Simultaneous-speaker voice activity detection and localization using mid-fusion of SVM and HMMs. *IEEE Transactions on Multimedia, 16*(4), 1032–1044.
29. Brutti, A., & Nesta, F. (2013). Tracking of multidimensional TDOA for multiple sources with distributed microphone pairs. *Computer Speech & Language, 27*(3), 660–682.
30. Plinge, A., Jacob, F., Haeb-Umbach, R., & Fink, G. A. (2016). Acoustic microphone geometry calibration: An overview and experimental evaluation of state-of-the-art algorithms. *IEEE Signal Processing Magazine, 33*(4), 14–29.
31. Rafaely, B. (2005). Analysis and design of spherical microphone arrays. *IEEE Transactions on Speech and Audio Processing, 13*(1), 135–143.
32. Li, Z., & Duraiswami, R. (2007). Flexible and optimal design of spherical microphone arrays for beamforming. *IEEE Transactions on Audio, Speech and Language Processing, 15*(2), 702–714.
33. Balmages, I., & Rafaely, B. (2007). Open-sphere designs for spherical microphone arrays. *IEEE Transactions on Audio, Speech and Language Processing, 15*(2), 727–732.
34. https://www.mathworks.com/help/phased/ug/omnidirectional-microphone.html#btdemjn.
35. Zaunschirm, M. (2012). *Modal beamforming using planar circular microphone arrays* (Doctoral dissertation, MS thesis). University of Music and Performing Arts Graz.
36. Duan, K., & Lü, B. (2003). Nonparaxial analysis of far-field properties of Gaussian beams diffracted at a circular aperture. *Optics Express, 11*(13), 1474–1480.
37. Habets, E. A. P. (2007). Single- and multi-microphone speech dereverberation using spectral enhancement. *Dissertation Abstracts International*, *68*(04).
38. Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience, 8*(5), 393–402.
39. Carrasco, J. M., Franquelo, L. G., Bialasiewicz, J. T., Galván, E., PortilloGuisado, R. C., Prats, M. M., … & Moreno-Alfonso, N. (2006). Power-electronic systems for the grid integration of renewable energy sources: A survey. *IEEE Transactions on industrial electronics*, *53*(4), 1002–1016.
40. Omologo, M., Matassoni, M., & Svaizer, P. (2001). Speech recognition with microphone arrays. In *Microphone arrays* (pp. 331–353). Berlin: Springer.
41. Jensen, M. A., & Wallace, J. W. (2004). A review of antennas and propagation for MIMO wireless communications. *IEEE Transactions on Antennas and Propagation, 52*(11), 2810–2824.
42. Kilfoyle, D. B., & Baggeroer, A. B. (2000). The state of the art in underwater acoustic telemetry. *IEEE Journal of Oceanic Engineering, 25*(1), 4–27.

43. Saligrama, V., Alanyali, M., & Savas, O. (2006). Distributed detection in sensor networks with packet losses and finite capacity links. *IEEE Transactions on Signal Processing, 54*(11), 4118–4132.
44. Li, J., & Stoica, P. (2005). *Robust adaptive beamforming* (Vol. 88). New York: Wiley.
45. Chen, J., Benesty, J., Huang, Y., & Doclo, S. (2006). New insights into the noise reduction Wiener filter. *IEEE Transactions on Audio, Speech and Language Processing, 14*(4), 1218–1234.
46. Brandstein, M., & Ward, D. (Eds.). (2013). *Microphone arrays: Signal processing techniques and applications*. Berlin: Springer Science & Business Media.
47. Benjamin, R. (1983). Signal-space, the key to signal processing. *Radio and Electronic Engineer, 53*(11), 407–422.
48. Tang, M. C., & Ziolkowski, R. W. (2013). Efficient, high directivity, large front-to-back-ratio, electrically small, near-field-resonant-parasitic antenna. *IEEE Access, 1,* 16–28.
49. Sheikh, Y. A., Ullah, R., & Zhonfu, Y. (2016). Range and direction of arrival estimation of near-field sources in sensor arrays using differential evolution algorithm. *International Journal of Computer Applications (IJCA)*, *139*(4), 16–20.
50. Waweru, N. P., Konditi, D. B. O., & Langat, P. K. (2014). Performance analysis of MUSIC, root-MUSIC and ESPRIT DOA estimation algorithm. *International Journal of Electrical, Computer, Energetic, Electronic and Communication Engineering, 8*(1), 209–216.
51. Fristrup, K. M., & Mennitt, D. (2012). Bioacoustical monitoring in terrestrial environments. *Acoustics Today, 8*(3), 16–24.
52. O'Donovan, A., Duraiswami, R., & Neumann, J. (2007, June). Microphone arrays as generalized cameras for integrated audio visual processing. In *IEEE Conference on Computer Vision and Pattern Recognition, 2007. CVPR'07* (pp. 1–8). IEEE.
53. Gotsis, K. A., Siakavara, K., & Sahalos, J. N. (2009). On the direction of arrival (DoA) estimation for a switched-beam antenna system using neural networks. *IEEE Transactions on Antennas and Propagation, 57*(5), 1399–1411.
54. DiBiase, J. H., Silverman, H. F., & Brandstein, M. S. (2001). Robust localization in reverberant rooms. In *Microphone arrays* (pp. 157–180). Berlin: Springer.
55. Zhao, F., & Guibas, L. J. (2004). *Wireless sensor networks: An information processing approach*. Los Altos, CA: Morgan Kaufmann.
56. Valin, J. M., Michaud, F., Rouat, J., & Létourneau, D. (2003, October). Robust sound source localization using a microphone array on a mobile robot. In *2003 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2003 (IROS 2003). Proceedings* (Vol. 2, pp. 1228–1233). IEEE.

Springer

http://www.springer.com/978-3-319-73058-5

Direction of Arrival Estimation and Localization of
Multi-Speech Sources
Dey, N.; Ashour, A.S.
2018, XIV, 53 p. 22 illus., 19 illus. in color., Softcover
ISBN: 978-3-319-73058-5