Chapter 1
Introduction

1.1 Background and Context

The motivation behind this book lies in the rapidly growing interest in spherical microphone arrays over the last decade. Important applications for these arrays include human-human and human-machine speech communication systems and spatial sound recording. While human-human speech communication systems have a long history, speech also plays an ever-growing part in human-machine communication. Indeed, while speech-based interfaces were once confined to the realms of science fiction, they are now becoming an increasingly popular way of interacting with devices such as smartphones, desktop and tablet computers, robots or televisions. This trend has been fuelled by advances in speech recognition technology, as well as the explosion in available computing power, particularly on mobile devices. With the widespread availability of 3D sound cinema systems and virtual reality gear with 3D binaural sound reproduction, the need to capture spatial sound is rapidly growing. Spherical microphone arrays are particularly suitable for capturing all three dimensions of the sound field, including both ambient sounds and sounds from particular directions.

The field of acoustic signal processing seeks to solve a number of problems relating to these systems, which can broadly be divided into three categories: acoustic parameter estimation, acoustic signal enhancement, and spatial audio recording. Acoustic parameter estimation, addressed in Chap. 5, involves the estimation of parameters such as the location or direction of arrival (DOA) of one or more acoustic sources [20, 27, 30, 34, 51–53], the signal-to-diffuse energy ratio or diffuseness of the sound field at a particular position [31, 32, 43, 54, 55], the number of sources present in a sound field [53, 56, 57], or the reverberation time of an acoustic environment [14, 39, 40, 46, 49, 58].

Portions of this chapter were first published in [25], and are reproduced here with the author's permission.

In the aforementioned applications, the signal to be acquired originates from a *distant* source, located at some significant distance from the microphone(s). While in some applications, such as teleconferencing systems, a microphone located close to the source may be available, this is not always a practical option. As a result, the acquired signal is corrupted by the surrounding environment. One major cause for this degradation is the presence of noise, where by *noise* we mean any acoustic signal which is undesired, such as interfering speech signals or background noise [6, 9]. The other is the presence of reflectors and obstacles to the propagation of sound waves, in particular room boundaries (walls, floors and ceiling), which cause *reverberation* [35, 42]. As the distance between the source and microphone(s) increases, the degradative effects of noise and reverberation become increasingly significant.

In the case of speech signals, these effects not only degrade the quality of the acquired signal, but in some cases also its intelligibility, making communication difficult or even impossible [3]. The cognitive effort required to understand highly noisy and reverberant speech can also contribute to listener fatigue. Acoustic signal enhancement or speech enhancement techniques (considered in Chaps. 6–9) seek to mitigate these effects, and extract the desired signal. The main problems of interest within this field are noise reduction [4, 6, 22, 26, 28], echo cancellation [7, 33, 47] and dereverberation [11, 18, 21, 23, 24, 29, 37, 38, 42]. Although the release of the first speakerphone dates back to 1954 [15], these remain open problems and areas of active research.

1.2 Microphone Array Signal Processing

Acoustic signal processing problems are commonly approached with microphone arrays [5, 10, 17], which is an arrangement of microphones in a specific configuration, thereby taking advantage of the spatial properties of the sound field (or *spatial diversity*) in order to improve performance. Owing to the similarity of the problems involved, many microphone array processing techniques are based on narrowband antenna array processing techniques [12]; however, microphone array processing faces its own unique challenges [5]. These include the broadband nature of speech (which covers several octaves), the non-stationarity of speech, and the fact that the desired and noise signals often have very similar spectral characteristics [5]. In addition, the placement and number of microphones is restricted, primarily by cost, aesthetics and available space. Considerations of space limit both the inter-microphone spacing and total microphone array size, and are of particular importance for portable devices, such as hearing aids [13].

A typical application scenario in microphone array signal processing is illustrated in Fig. 1.1. A microphone array captures a mixture of signals with different spatial characteristics, some of which may be desired, and others undesired. Acoustic parameter estimation algorithms seek to accurately estimate the parameters of interest even in the presence of undesired signals that may adversely affect the estimation
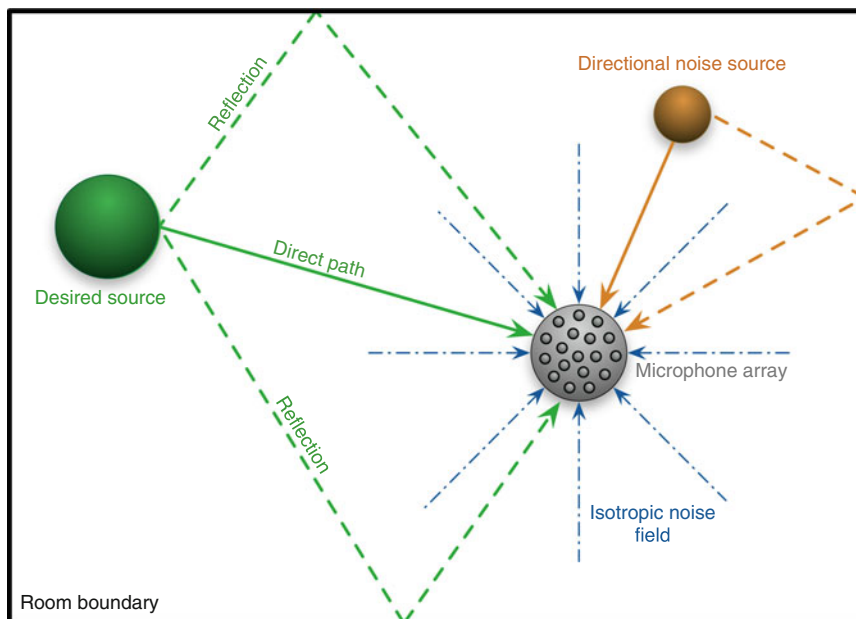
Fig. 1.1 Schematic illustration of a typical application scenario in microphone array signal processing. A microphone array captures a mixture of signals with different spatial characteristics in a reverberant environment

process. Acoustic signal enhancement algorithms aim to extract only the desired signals from the received mixture.

The spatial characteristics of the various captured signals are typically modeled based on their *spatial coherence*. The microphone signals are corrupted by sensor noise, which is spatially *incoherent* (or *spatially white*), that is, the sensor noise signals at each microphone are mutually uncorrelated. The desired signals, originating from one or more desired sources, as well as any directional noise signals, originating from interfering speakers or air-conditioning units, for example, are spatially *coherent*. Finally, *partially coherent* signals can be observed in spherically or cylindrically *isotropic* (or diffuse) sound fields, which can be used to model babble noise or reverberation. The desired signal is normally chosen as either the anechoic signal arriving from the desired source via the *direct path*, or the reverberant signal arriving via the direct path and a number of reflected paths.

In theory, any microphone array configuration is possible; in practice, most microphone arrays are linear or planar, and the microphones respectively lie on a straight line or a flat, two-dimensional surface. Real sound fields are three-dimensional, however, and can only be fully analyzed with a three-dimensional array. The spherical configuration is convenient due to its symmetry giving equal performance in all directions. In addition, the captured sound field can be efficiently described in the

spherical harmonic domain [41, 44], based on a formulation of the wave equation
in spherical coordinates (in Chap. 2). Spherical microphone arrays [1, 16, 19, 36,
45, 48, 50] are usually either *open* or *rigid*, that is, the microphones are either sus-
pended in free space or mounted on a rigid baffle (as discussed in Chap. 3). They
have recently started to become commercially available, in the form of products such
as the *acoustic camera* by GFal, the *Eigenmike* by mh acoustics (Fig. 1.2), Brüel and
Kjær's spherical array (Fig. 1.3), or the *RealSpace Panoramic Audio Camera* by
VisiSonics, yet to date there have been few signal processing algorithms designed
for these arrays. This motivates the work presented in this book.

Fig. 1.2 The em32 Eigenmike spherical microphone array. This rigid array of radius 4.2 cm is
comprised of 32 omnidirectional microphones. Copyright © Emanuël Habets. Used with permission

Fig. 1.3 The Brüel and Kjær spherical microphone array. This rigid array is comprised of 36 or 50 microphones and 12 video cameras. Copyright © Brüel & Kjær. Used with permission

1.3 Organization of the Book

The content of this book is structured as follows:

- In **Chap.** 2, the fundamentals of acoustics are reviewed. We introduce the spherical harmonics, which form a complete set of orthonormal functions. Their importance rests in the fact that any arbitrary function on a sphere can be expanded in terms of a these functions and a set of expansion coefficients.
- **Chapter** 3 examines issues relating to spatial signal acquisition and transformation. We present the short-time Fourier transform and spherical harmonic framework that allow us to efficiently process the signals captured by a spherical microphone array. Common spatial sampling schemes are presented, which determine the placement of microphones on the sphere such that spatial aliasing is

minimized. In addition, we discuss the advantages and disadvantages of two common array types: the open and rigid arrays with omnidirectional microphones.

- In order to comprehensively evaluate spherical array processing algorithms under many different acoustic conditions, it is indispensable to use simulated acoustic impulse responses (AIRs). The image method proposed by Allen and Berkley [2] is a well-established way of doing this for point-to-point AIRs with sensors in free space, however it does not account for the scattering introduced by a rigid sphere. In **Chap.** 4, we present a method for simulating the AIRs between a sound source and microphones positioned on a rigid spherical array. In addition, three examples are presented based on this method: an analysis of a diffuse reverberant sound field, a study of binaural cues in the presence of reverberation, and an illustration of the algorithm's use as a mouth simulator.
- **Chapter** 5 introduces methods for the estimation of two important acoustic parameters: the DOA of a sound source, and the signal-to-diffuse energy ratio at a particular position in a sound field. Later in the book, it will be seen that these quantities can be used for signal enhancement purposes.
- The process of combining signals acquired by a microphone array in order to isolate a signal of interest is known as beamforming or spatial filtering. **Chapter** 6 considers the simplest type of beamformer: the signal-independent (fixed) beamformer, whose weights only depend on the DOA of the source to be extracted, and do not otherwise depend on the desired signal.
- In **Chap.** 7, we derive signal-dependent beamformers, whose weights depend on the second-order statistics of the desired signal and/or of the noise to be suppressed. These beamformers adaptively seek to achieve optimal performance in terms of noise reduction and speech distortion.
- **Chapter** 8 takes a different approach to signal enhancement: a physically-motivated parametric representation of the sound field is introduced. It is shown that the sound field can be manipulated to achieve noise reduction or dereverberation by applying a time- and frequency-dependent gain to a reference signal. The gain is a simple function of the sound field parameters, which can be estimated using the methods presented in Chap. 5.
- The concept of informed array processing is introduced in **Chap.** 9. It involves incorporating relevant spatial information about the specific problem into the design of spatial filters, and into the estimation of the second-order statistics that is required to implement the beamformers in Chap. 7. Informed array processing techniques are developed for two signal enhancement problems: noise reduction and dereverberation.

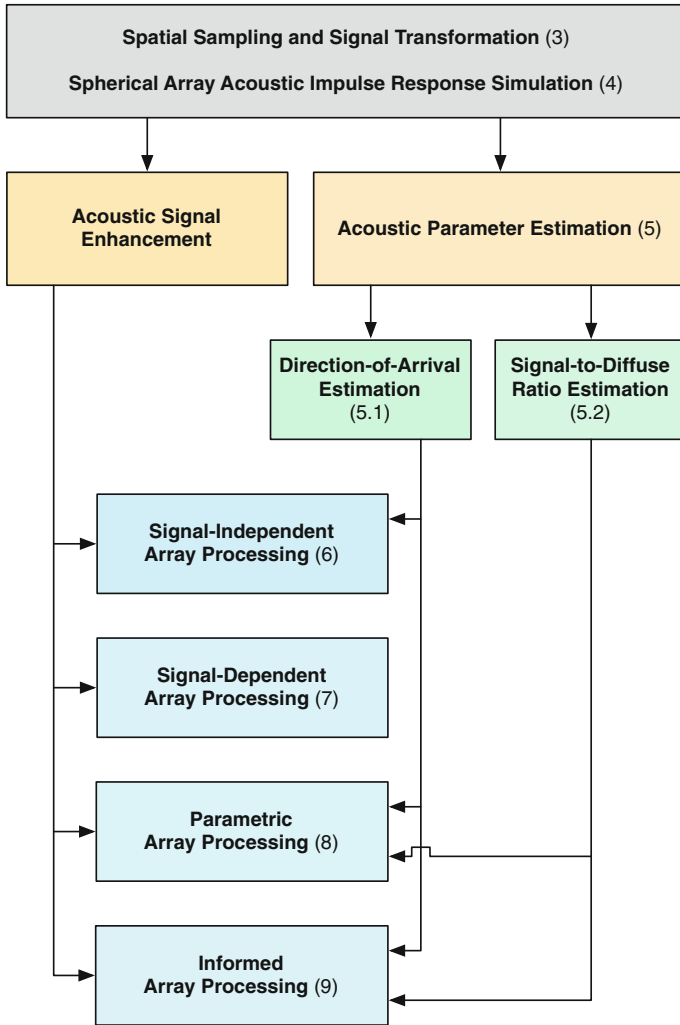The structure of the book, and the relationship between each of the topics it addresses, is illustrated in Fig. 1.4.

Fig. 1.4 Structure of the book. The chapter/section relating to each topic is indicated in *parentheses*

References

1. Abhayapala, T.D., Ward, D.B.: Theory and design of high order sound field microphones using spherical microphone array. In: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), vol. 2, pp. 1949–1952 (2002). doi:10.1109/ICASSP.2002.1006151

2. Allen, J.B., Berkley, D.A.: Image method for efficiently simulating small-room acoustics. J. Acoust. Soc. Am. **65**(4), 943–950 (1979)

3. Assmann, P., Summerfield, Q.: The perception of speech under adverse conditions. In: Greenberg, S., Ainsworth, W.A., Popper, A.N., Fay, R.R. (eds.) Speech Processing in the Auditory System, Chap. 5, pp. 231–308. Springer, Berlin, Germany (2004)

4. Benesty, J., Chen, J., Habets, E.A.P.: Speech Enhancement in the STFT Domain. SpringerBriefs in Electrical and Computer Engineering. Springer, Berlin (2011)

5. Benesty, J., Chen, J., Huang, Y.: Microphone Array Signal Processing. Springer, Berlin, Germany (2008)

6. Benesty, J., Chen, J., Huang, Y., Cohen, I.: Noise Reduction in Speech Processing. Springer, Berlin (2009)

7. Benesty, J., Gänsler, T., Morgan, D.R., Sondhi, M.M., Gay, S.L.: Advances in Network and Acoustic Echo Cancellation. Springer, Berlin (2001)

8. Benesty, J., Sondhi, M.M., Huang, Y. (eds.): Springer Handbook of Speech Processing. Springer, Berlin (2008)

9. Berouti, M., Schwartz, R., Makhoul, J.: Enhancement of speech corrupted by acoustic noise. In: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), vol. 4, pp. 208–211 (1979)

10. Brandstein, M.S., Ward, D.B. (eds.): Microphone Arrays: Signal Processing Techniques and Applications. Springer, Berlin (2001)

11. Braun, S., Jarrett, D.P., Fischer, J., Habets, E.A.P.: An informed spatial filter for dereverberation in the spherical harmonic domain. In: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 669–673. Vancouver, Canada (2013)

12. Compton, Jr., R.: Adaptive Antennas, 1st edn. Prentice-Hall, Upper Saddle River (1988)

13. Doclo, S., Gannot, S., Moonen, M., Spriet, A.: Acoustic beamforming for hearing aid applications. In: Haykin, S., Liu, K.R. (eds.) Handbook on Array Processing and Sensor Networks, chap. 9. Wiley, New York (2008)

14. Eaton, J., Gaubitch, N.D., Naylor, P.A.: Noise-robust reverberation time estimation using spectral decay distributions with reduced computational cost. In: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Vancouver, Canada (2013)

15. Elko, G.W.: Future directions for microphone arrays. In: Brandstein and Ward [10], chap. 17, pp. 383–387

16. Elko, G.W., Meyer, J.: Spherical microphone arrays for 3D sound recordings. In: Huang, Y., Benesty, J. (eds.) Audio Signal Processing for Next-Generation Multimedia Communication Systems, chap. 3, pp. 67–89 (2004)

17. Elko, G.W., Meyer, J.: Microphone arrays. In: Benesty et al. [8], chap. 50

18. Gaubitch, N.D.: Blind identification of acoustic systems and enhancement of reverberant speech. Ph.D. thesis, Imperial College London (2006)

19. Gover, B.N., Ryan, J.G., Stinson, M.R.: Microphone array measurement system for analysis of directional and spatial variations of sound fields. J. Acoust. Soc. Am. **112**(5), 1980–1991 (2002). doi:10.1121/1.1508782

20. Gustafsson, T., Rao, B., Trivedi, M.: Source localization in reverberant environments: modeling and statistical analysis. IEEE Trans. Speech Audio Process. **11**(6), 791–803 (2003)

21. Habets, E.A.P.: Single- and multi-microphone speech dereverberation using spectral enhancement. Ph.D. thesis, Technische Universiteit Eindhoven (2007). http://alexandria.tue.nl/extra2/200710970.pdf

22. Habets, E.A.P., Benesty, J.: A perspective on frequency-domain beamformers in room acoustics. IEEE Trans. Audio, Speech, Lang. Process. **20**(3), 947–960 (2012)
23. Habets, E.A.P., Cohen, I., Gannot, S.: Generating nonstationary multisensor signals under a spatial coherence constraint. J. Acoust. Soc. Am. **124**(5), 2911–2917 (2008). doi:10.1121/1. 2987429
24. Huang, Y., Benesty, J., Chen, J.: Dereverberation. In: Benesty et al. [8], chap. 5
25. Jarrett, D.P.: Spherical microphone array processing for acoustic parameter estimation and signal enhancement. Ph.D. thesis, Imperial College London (2013)
26. Jarrett, D.P., Habets, E.A.P., Benesty, J., Naylor, P.A.: A tradeoff beamformer for noise reduction in the spherical harmonic domain. In: Proceedings of the International Workshop on Acoust. Signal Enhancement (IWAENC). Aachen, Germany (2012)
27. Jarrett, D.P., Habets, E.A.P., Naylor, P.A.: 3D source localization in the spherical harmonic domain using a pseudointensity vector. In: Proceedings of the European Signal Processing Conference (EUSIPCO), pp. 442–446. Aalborg, Denmark (2010)
28. Jarrett, D.P., Habets, E.A.P., Naylor, P.A.: Spherical harmonic domain noise reduction using an MVDR beamformer and DOA-based second-order statistics estimation. In: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 654–658. Vancouver, Canada (2013)
29. Jarrett, D.P., Habets, E.A.P., Thomas, M.R.P., Gaubitch, N.D., Naylor, P.A.: Dereverberation performance of rigid and open spherical microphone arrays: Theory & simulation. In: Proceedings of the Joint Workshop on Hands-Free Speech Communication and Microphone Arrays (HSCMA), pp. 145–150. Edinburgh, UK (2011)
30. Jarrett, D.P., Habets, E.A.P., Thomas, M.R.P., Naylor, P.A.: Simulating room impulse responses for spherical microphone arrays. In: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 129–132. Prague, Czech Republic (2011)
31. Jarrett, D.P., Thiergart, O., Habets, E.A.P., Naylor, P.A.: Coherence-based diffuseness estimation in the spherical harmonic domain. In: Proceedings of the IEEE Convention of Electrical & Electronics Engineers in Israel (IEEEI). Eilat, Israel (2012)
32. Jeub, M., Nelke, C., Beaugeant, C., Vary, P.: Blind estimation of the coherent-to-diffuse energy ratio from noisy speech signals. In: Proceedings of the European Signal Processing Conf. (EUSIPCO). Barcelona, Spain (2011)
33. Kellermann, W.: Acoustic echo cancellation for beamforming microphone arrays. In: Brandstein, M.S., Ward, D.B. (eds.) Microphone Arrays: Signal Processing Techniques and Applications, pp. 281–306. Springer, Berlin, Germany (2001)
34. Khaykin, D., Rafaely, B.: Coherent signals direction-of-arrival estimation using a spherical microphone array: Frequency smoothing approach. In: Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, pp. 221–224 (2009). doi:10.1109/ ASPAA.2009.5346492
35. Kuttruff, H.: Room Acoustics, 4th edn. Taylor & Francis, London (2000)
36. Li, Z., Duraiswami, R.: Flexible and optimal design of spherical microphone arrays for beamforming. IEEE Trans. Audio, Speech, Lang. Process. **15**(2), 702–714 (2007). doi:10.1109/ TASL.2006.876764
37. Lim, F., Naylor, P.A.: Robust low-complexity multichannel equalization for dereverberation. In: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Vancouver, Canada (2013)
38. Lim, F., Thomas, M., Naylor, P.: Mintformer: A spatially aware channel equalizer. In: Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics. New Paltz, USA (2013)
39. Löllmann, H., Vary, P.: Estimation of the frequency dependent reverberation time by means of warped filter-banks. In: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 309 –312 (2011). doi:10.1109/ICASSP.2011. 5946402

40. de M. Prego, T., de Lima, A.A., Netto, S.L., Lee, B., Said, A., Schafer, R.W., Kalker, T.: A blind algorithm for reverberation-time estimation using subband decomposition of speech signals. J. Acoust. Soc. Am. **131**(4), 2811–2816 (2012)
41. Meyer, J., Elko, G.: A highly scalable spherical microphone array based on an orthonormal decomposition of the soundfield. In: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), vol. 2, pp. 1781–1784 (2002)
42. Naylor, P.A., Gaubitch, N.D. (eds.): Speech Dereverberation. Springer, Berlin (2010)
43. Pulkki, V.: Spatial sound reproduction with directional audio coding. J. Audio Eng. Soc. **55**(6), 503–516 (2007)
44. Rafaely, B.: Analysis and design of spherical microphone arrays. IEEE Trans. Speech Audio Process. **13**(1), 135–143 (2005). doi:10.1109/TSA.2004.839244
45. Rafaely, B., Peled, Y., Agmon, M., Khaykin, D., Fisher, E.: Spherical microphone array beamforming. In: I. Cohen, J. Benesty, S. Gannot (eds.) Speech Processing in Modern Communication: Challenges and Perspectives, chap. 11. Springer (2010)
46. Ratnam, R., Jones, D.L., Wheeler, B.C., O'Brien Jr., W.D., Lansing, C.R., Feng, A.S.: Blind estimation of reverberation time. J. Acoust. Soc. Am. **114**(5), 2877–2892 (2003)
47. Sondhi, M.: Adaptive echo cancelation for voice signals. In: Benesty et al. [8], chap. 45. Part H
48. Sun, H., Yan, S., Svensson, U.P.: Robust minimum sidelobe beamforming for spherical microphone arrays. IEEE Trans. Audio, Speech, Lang. Process. **19**(4), 1045–1051 (2011). doi:10.1109/TASL.2010.2076393
49. Talmon, R., Habets, E.A.P.: Blind reverberation time estimation by intrinsic modeling of reverberant speech. In: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Vancouver, Canada (2013)
50. Teutsch, H.: Wavefield decomposition using microphone arrays and its application to acoustic scene analysis. Ph.D. thesis, Friedrich-Alexander Universität Erlangen-Nürnberg (2005)
51. Teutsch, H., Kellermann, W.: EB-ESPRIT: 2D localization of multiple wideband acoustic sources using eigen-beams. In: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), vol. 3, pp. iii/89–iii/92 (2005). doi:10.1109/ICASSP.2005.1415653
52. Teutsch, H., Kellermann, W.: Eigen-beam processing for direction-of-arrival estimation using spherical apertures. In: Proceedings of the Joint Workshop on Hands-Free Speech Communication and Microphone Arrays. Piscataway, New Jersey, USA (2005)
53. Teutsch, H., Kellermann, W.: Detection and localization of multiple wideband acoustic sources based on wavefield decomposition using spherical apertures. In: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 5276–5279 (2008). doi:10.1109/ICASSP.2008.4518850
54. Thiergart, O., Del Galdo, G., Habets, E.A.P.: On the spatial coherence in mixed sound fields and its application to signal-to-diffuse ratio estimation. J. Acoust. Soc. Am. **132**(4), 2337–2346 (2012)
55. Thiergart, O., Del Galdo, G., Habets, E.A.P.: Signal-to-reverberant ratio estimation based on the complex spatial coherence between omnidirectional microphones. In: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 309–312 (2012)
56. Wang, H., Kaveh, M.: Coherent signal-subspace processing for the detection and estimation of angles of arrival of multiple wide-band sources. IEEE Trans. Acoust., Speech, Signal Process. **33**(4), 823–831 (1985)
57. Wax, M.: Detection and localization of multiple sources via the stochastic signals model. IEEE Trans. Signal Process. **39**(11), 2450–2456 (1991)
58. Wen, J.Y.C., Habets, E.A.P., Naylor, P.A.: Blind estimation of reverberation time based on the distribution of signal decay rates. In: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Las Vegas, USA (2008)