Introduction

The term ‘acoustic scene analysis’ (ASA) describes the task of extracting information contained in an acoustic wavefield, such as the waveform itself or parameter describing the source of the wavefield. Since acoustic wavefields are processes spread-out in space and time it follows quite naturally that ASA is predominantly performed by evaluating the signals captured by a number of spatially distinct microphones, i.e. microphone arrays. A standard and widely applied vehicle for evaluating the microphone array signals is built upon classical array signal processing techniques [JD93, Tre02]. In this context, the term ‘classical’ is used to denote signal processing algorithms, to be introduced below, that are applied directly to the individual microphones comprising the array. In contrast, the algorithms to be derived in this book are applied to signals that are obtained by transforming the microphone signals into a domain defined by the eigen-solutions of the acoustic wave equation in two- and three spatial dimensions.

ASA, as considered in this book, can be grouped into three distinct classes.

1. Supervised ASA or waveform estimation. Here, ASA is concerned with the extraction of the desired source signal from the observed wavefield by means of beamforming and beamsteering techniques. Note that although the signal processing steps necessary for the extraction of the desired source involves an estimation of filter coefficients, the primary task is to extract the signal itself. Therefore, the term ‘waveform estimation’ is chosen in this context. Beamforming is an attempt to add the desired signal coherently at the output of the beamformer while the noise is added incoherently. Beamforming methods include, delay-and-sum beamforming [Dol46, Fla85, FJZE85], filter-and-sum beamforming [DB88, Vsd96, WKW01], as well as beamforming using differential [Elk04, TE01a] and superdirective arrays [CZK86, CZO87, BS01]. Both signal-independent as well as signal-dependent implementations are used in practice, see e.g. [Her05]. All of the above mentioned beamformers will be discussed in Chapter 4. For performing the signal alignment, i.e. the
beamsteering step, a prerequisite for this type of ASA is the knowledge of the array’s geometry and the location of the desired acoustic source, hence the term ’supervised’. Typical scenarios set for supervised ASA are teleconferencing applications and front-ends to distant-talker speech recognition systems. Note that supervised ASA techniques typically assume a single desired source signal to be present in a wavefield.

2. Semi-supervised ASA or parameter estimation. In this context, ASA deals with the number and location of possibly multiple simultaneously active acoustic sources using microphone arrays. While the localization of a single desired acoustic source has been the subject of research for the last several decades – mainly by applying the notion of the time-difference-of-arrival between microphone pairs [MS69, KC76, Ben00] – the problem of estimating the number and positions of multiple acoustic sources using non-heuristic methods has not been under investigation only until very recently [BAS+05, TK05a, TK05b, TK05c]. The discussion and an attempt to build a framework for this problem is the main focus of this book. A typical application of semi-supervised ASA is acoustic surveillance. Parameter estimation also serves as a preprocessing step for the techniques based on supervised ASA.

3. Unsupervised ASA or blind system identification / blind source separation. This rather new and emerging area of research does, in principle, not require any knowledge of the parameters present in the wavefield, nor does it rely on the geometry of the microphone array. The desired information is extracted by adaptively estimating the system comprising the acoustic source(s) and the acoustic environment observed by the microphones. This class of ASA is not treated further in this book. Blind system identification and its application to parameter estimation, using classical array signal processing techniques, is described in [BAK05b, BAS+05]. Blind source separation and its application to waveform estimation by extracting the desired acoustic signals from a mixture of signals is described in detail, e.g., in [BAK04, BAK05a].

Note that the ASA considered here is not to be confused with auditory scene analysis which is a technique for encoding Structured Audio bitstreams from acoustic data, as described within the ISO/IEC 14496 MPEG-4 context [MPE].

The work presented in this book is, in contrast to the algorithms as considered in the above mentioned references, not solely based on classical array signal processing but also on the principles of wave propagation and wave scattering offered by the science of classical acoustics. Algorithms for parameter estimation solely based on the pillar ’classical array signal processing’, often suffer from the problem that they rely on a narrowband assumption underlying the signal model, which limit their usability when broadband signals, such as speech, are present in the wavefield under observation. It will be subsequently shown that by erecting a second pillar ’classical acoustics’,