Chapter 1
Introduction

Patrick A. Naylor and Nikolay D. Gaubitch

Abstract Acoustic reverberation will be introduced in this chapter in the context of telecommunication. The adverse effects on speech caused by reverberation are problematic, in particular, in hands-free terminals operating typically at arms-length from the talker's lips. This introductory chapter will provide a system description of room reverberation and will formulate mathematically the dereverberation problem in its most direct form so as to introduce and underpin the more detailed presentation in subsequent chapters. Elements of room acoustics will also be introduced where needed, though detailed study of acoustics is not the aim of this text.

At the time of writing this, dereverberation is a topic of study with many important research questions remaining as yet unanswered. Whilst reviewing the relevant literature later in this chapter, it is intended both to describe the state-of-the-art and to highlight some of the significant open issues. Whereas the former aims to consolidate, perhaps for the first time, the known achievements to date of the research community, the latter aims to highlight potential avenues of future research.

1.1 Background

One can confidently deduce that the phenomenon of reverberation has been known to mankind since the time of prehistoric cave dwellers. Sound reflection effects are believed to have influenced prehistoric cave art [90, 91]. Reverberation is also used in several well known cases by other species, such as bats for navigation during flight.

There is evidence of comprehension of the notion of reflected speech occurring in Plato's *Republic* [76]: *"And what if sound echoed off the prison wall opposite them? When any of the passers-by spoke, don't you think they'd be bound to assume that the sound came from a passing shadow?"*. Pioneering scientific work on sound and acoustics in the 19th century was undertaken by, for example, Rayleigh [80] and

Imperial College London, UK

Sabine [81]. In the 20th century initial efforts in the understanding of reverberation of speech were provided by Bolt [8] and the effects of single echoes by Haas [37].

The level to which humans employ reverberation during everyday life is unclear. There is some evidence to suggest that, through the use of two ears, spatial processing is used to enhance speech intelligibility and enables a useful degree of source separation to be achieved in human speech perception [12].

In music audio processing, the sense of 'space' that can be created by stereo or surround sound reproduction adds greatly to realism and often makes recorded music more attractive and enjoyable. We might then ask ourselves the question: *since reverberation is present in everyday life experience of sound and in some important cases is effective in aiding speech communication, why should we be interested in removing reverberation from speech using dereverberation processing?* The answer to this question is dependent on the application context.

There is a continuously growing demand for high quality hands-free speech input for various telecommunication applications [71, 82]. One driving force behind this development is the rapidly increasing use of portable devices such as mobile telephones, Personal Digital Assistant (PDA) devices and laptop computers equipped for Voice Over Internet Protocol (VoIP) [63]. Furthermore, there is a continuous worldwide expansion of broadband internet access [5]. These factors have paved the way for several advanced speech applications such as wideband teleconferencing with automatic camera steering, automatic speech-to-text conversion, speaker identification, voice-controlled device operation and car interior communication systems [82]. Another important application where speech obtained from a distant talker is of interest is that of hearing aids [82].

1.2 Effects of Reverberation

When speech signals are obtained in an enclosed space by one or more microphones positioned at a distance from the talker, the observed signal consists of a superposition of many delayed and attenuated copies of the speech signal due to multiple reflections from the surrounding walls and other objects, as illustrated in Fig. 1.1. We here define the direct-path as the acoustic propagation path from the talker to the microphone without reflections. We also note that a delay of the superimposed copies arises because all other propagation paths are longer than the direct-path and that additional attenuation occurs at each reflection due to frequency dependent absorption. The perceptual effects of reverberation can be summarized as:

1. *The box effect* – the reverberated speech signal can be viewed as the same source signal coming from several different sources positioned at different locations in the room and thus arriving at different times and with different intensities [3]. This adds spaciousness to the sound [56] and makes the talker sound as if positioned "inside a box".
2. *The distant talker effect* – the perceived spaciousness explained in the previous point makes the talker sound far away from the microphone.

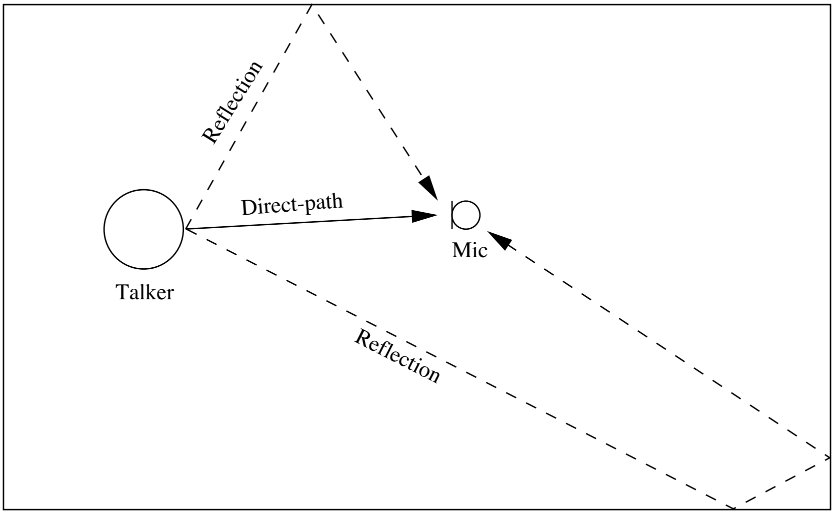

Fig. 1.1 Schematic illustration of room reverberation

When these effects are carefully controlled and moderately applied, the reverberation can add a pleasant sense of the acoustic space in which the sound resides. This is valuable and important in audio rendering but almost always unhelpful in voice communication. When the reverberation effects are severe, intelligibility of speech is degraded. Reverberation alters the characteristics of the speech signal, which is problematic for signal processing applications including speech recognition, source localization and speaker verification, and significantly reduces the performance of algorithms developed without taking room effects into consideration. The deleterious effects are magnified as the distance between the talker and the microphones is increased.

1.3 Speech Acquisition

The problems associated with reverberation can sometimes be overcome in practice by utilizing a headset by means of which the microphone is held close to the mouth. Alternatively, a microphone with a fixed directional sensitivity characteristic positioned in front of the talker can be used. The advent of bluetooth technology has made high quality, low cost wireless headsets feasible. Nevertheless, these solutions impose restrictions on the flexibility and comfort of the talker, which are the main desired features in the use of hands-free equipment [71]. In some applications, such as teleconferencing with multiple talkers on one end, these headset based solutions may not be practical. Therefore, a signal processing approach independent of the relative talker-microphone configuration is certainly preferable.

In hands-free speech acquisition, the talker's lips are typically located at a distance of 0.3-3 m from the microphone. In such a scenario, the speech signal is affected by the user's surrounding environment, which results in the following three distinct effects [71, 82]:

(i) *Additive measurement noise* due to, for example, other audible talkers or passing traffic. When the noise level is comparable to or greater than the speech level, it is difficult for a listener to distinguish the desired speech signal from the noise, and thus intelligibility and listening comfort are reduced.
(ii) *Acoustic echoes* due to speech from a far-end talker which is picked up by the near-end microphones and retransmitted back to the far-end talker with delay. This results in the talker hearing an echo of their own voice, which greatly disturbs the communication.
(iii) *Reverberation* that arises whenever sound is produced in enclosed spaces, such as offices and other rooms, due to reflections from walls and surrounding objects.

These components jointly contribute to an overall degradation in the quality of the observed speech signals, which significantly reduce the perceived speech quality for the listener and the performance of applications such as speech recognizers [71]. Speech enhancement and acoustic echo cancellation are two widely researched fields that address problems (i) and (ii) respectively. Several significant contributions have been made in these areas [6, 7, 11, 13, 30, 41, 60] and many algorithms have been implemented and are in use in commercial applications [77]. The problem of reverberation on the other hand, received much less attention in the literature until recently. Nevertheless, finding solutions to this problem is essential for the future development of applications with hands-free speech acquisition. This indeed motivates the focus of the forthcoming chapters of this book.

1.4 System Description

A generic system diagram for multichannel dereverberation is shown in Fig. 1.2. The speech signal, $s(n)$, from the talker propagates through acoustic channels, $H_m(z)$ for $m = 1$ to M. The output of each channel is observed using M microphones to give signals $x_m(n)$. All noise in the system is assumed additive and is represented by $v_m(n)$.

The observed signal, $x_m(n)$, at microphone m can be described as the superposition of (i) the direct-path signal, which propagates by line-of-sight from the talker to the microphone with corresponding attenuation and propagation delay and (ii) a theoretically infinite set of reflections of the talker signal arriving at the microphone at later time instances [56] with attenuation dependent on the properties of the reflecting surfaces. This can be expressed as

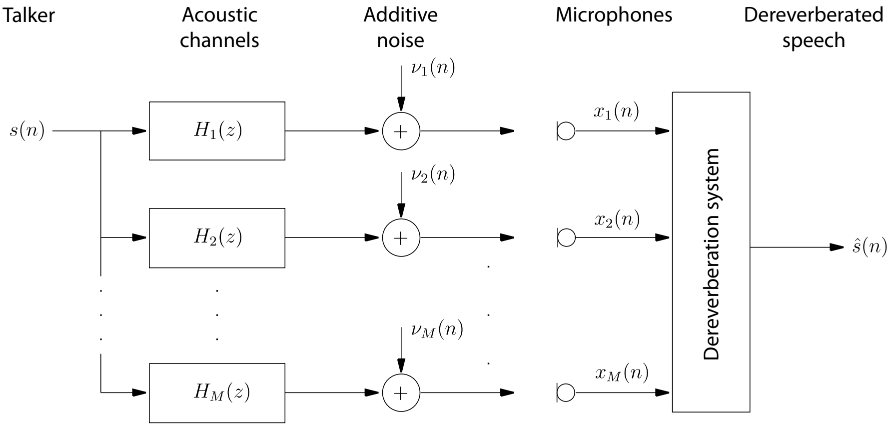

Fig. 1.2 Generic multichannel reverberation-dereverberation system model

$$x_m(n) = \sum_{i=0}^{\infty} h_{m,i}(n)s(n-i), \tag{1.1}$$

where the acoustic channel impulse responses $h_{m,i}(n)$ represent the attenuation and the propagation delay corresponding to the direct signal and all the reflected components.

The aim of speech dereverberation is to find a system with input $x_m(n)$, $m = 1,\ldots,M$ and output $\hat{s}(n)$, which is a 'good' estimate of $s(n)$. The definition of 'good' is this context is application dependent. It may, for example, be desired to estimate $s(n)$ with minimum Mean Square Error (MSE). Alternatively, other criteria may be relevant, such as those related to perceptual quality. This is a blind problem since the acoustic channels $H_m(z)$ are unknown.

Recent efforts in acoustic signal processing have produced several algorithms for speech dereverberation and reverberant speech enhancement. These methods can be divided broadly into three main categories:

1. *Beamforming* – the signals received at the different microphones are filtered and weighted so as to form a beam of enhanced sensitivity in the direction of the desired source and to attenuate sounds from other directions. Beamforming is dependent on the availability of multi-microphone inputs. Beamforming is a multiple input single output process.
2. *Speech enhancement* – the speech signals are modified so as to represent better some features of the clean speech signal according to an *a priori* defined model of the speech waveform or spectrum. Speech enhancement is often a single input single output process, though many speech enhancement techniques benefit from the use of multiple inputs.
3. *Blind deconvolution* – the acoustic impulse responses are identified blindly, using only the observed microphone signals, and then used to design an inverse filter that compensates for the effect of the acoustic channels.

1.5 Acoustic Impulse Responses

The Acoustic Impulse Response (AIR) characterizes the acoustics of a given enclosure and therefore study of the AIR is a natural approach to dereverberation. This section will introduce some of the characteristics of AIRs. The focus is on the AIRs of rooms where reverberation has a significant effect on telecommunication applications. Further relevant details of room acoustics are given in Chap. 2. Whereas AIR is used to refer to acoustic impulse responses in general, there are some cases where it is more appropriate to limit the acoustic context to be within a room, in which case, the impulse response is referred to as a Room Impulse Response (RIR). In this book we will use AIR and also RIR, depending on the acoustic scenario being considered.

Several models of room impulse responses have been considered in the literature, including both Finite Impulse Response (FIR) and Infinite Impulse Response (IIR) structures [40, 47, 48, 65, 66, 74]. The choice of AIR model will generally influence the algorithmic development.

An often-used quantification of the impulse response of a room is the reverberation time, originally introduced by Sabine [56]. The reverberation time, T_{60}, is defined as the time taken for the reverberant energy to decay by 60 dB once the sound source has been abruptly shut off. The reverberation time for a room is governed by the room geometry and the reflectivity of the reflecting surfaces.

The reverberation time is approximately constant when measured at any location in a given room. However, the impulse response is spatially variant and will vary as the talker, the microphones or other objects in the room change location [56]. A particular characteristic that varies with the talker-microphone separation is the relation between the energy of the direct-path component and the energy of the reflected components of the AIR. The critical distance is the distance such that these two energies are equal.

Figure 1.3 shows an example room impulse response. Direct-path propagation from the sound source to the microphone gives rise to an initial short period of near-zero amplitude, sometimes referred to as the direct-path propagation delay, followed by a peak. The amplitude of this peak due to direct-path propagation may be greater or less than the amplitude of the later reflections depending on the source-microphone distance and the reflectivity of the surfaces in the room. The example of Fig. 1.3 shows a relatively strong direct-path component, indicating that the source-microphone distance is relatively short.

The early and the late reflections are indicated in the figure as two distinct regions of the AIR. The early reflections are often taken as the first 50 ms of the impulse response [56], and constitute well defined impulses of large magnitude relative to the smaller magnitude and diffuse nature of the late reflections. The propagation from the talker's lips to the microphone is represented by the convolution of the speech signal with the AIR. The AIR early reflections cause spectral changes and lead to a perceptual effect referred to as coloration [56]. In general, closely spaced echoes are not distinguished by human hearing due to masking properties of the ear, and it has been shown that early reflections can have a positive impact on the intelligibility

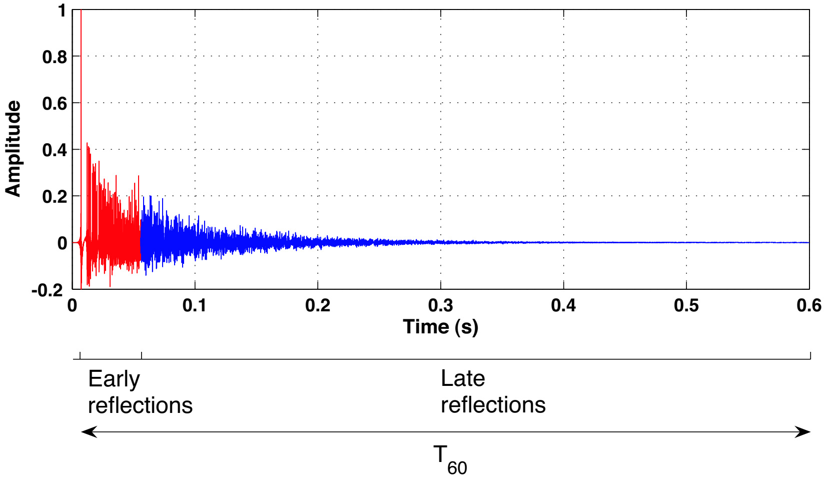

Fig. 1.3 An example room impulse response

of speech with an effect similar to increasing the strength of the direct-path sound [9, 56, 71]. However, coloration can degrade the quality of recorded speech [56]. The late reflections are referred to as the tail of the impulse response and constitute closely spaced, decaying impulses, which are seemingly randomly distributed. The late reflections cause a ‘distant’ and ‘echo-ey’ sound quality, we refer to as the reverberation tail and provides the major contribution to what is generally perceived of as reverberation in everyday experience.

In terms of spectral characteristics, the room transfer function is proportional to the sound pressure [56, 92] and has been studied extensively in the room acoustics literature, where many properties have been established [56]. One property of interest in the context of dereverberation is the average magnitude difference between minimum and maximum spectral points, which has been shown to extend beyond 10 dB [56, 83]. Since the room transfer function changes depending on the location of the source and the microphone, it can been described as a random process [56, 79, 92]. Neely and Allen [68] concluded that the AIRs in most real rooms posses non-minimum phase characteristics.

Rooms are generally stable systems with the coefficients $h_{m,i}(n)$ tending to zero with increasing index i and therefore, it is sufficient to consider only the first L_h coefficients in (1.1). The choice of L_h is often linked to the reverberation time of the room. Taking into account any additive noise sources, the observed signal at the m^{th} microphone can be written in a vector form

$$x_m(n) = \mathbf{h}_m^T(n)\mathbf{s}(n) + \nu_m(n), \tag{1.2}$$

where $\mathbf{h}_m(n) = [h_{m,0}(n)\ h_{m,1}(n)\ \ldots\ h_{m,L_h-1}(n)]^T$ is the L_h-tap impulse response of the acoustic channel from the source to microphone m, $\mathbf{s}(n) = [s(n)\ s(n-1)\ \ldots\ s(n-L_h+1)]^T$ is the speech signal vector and $\nu(n)$ is observation noise.

In the frequency domain this can be expressed, equivalently, as

$$X_m(e^{j\omega}) = H_m(e^{j\omega})S(e^{j\omega}) + \mathcal{N}_m(e^{j\omega}), \tag{1.3}$$

where $X_m(e^{j\omega})$, $H_m(e^{j\omega})$, $S(e^{j\omega})$ and $\mathcal{N}_m(e^{j\omega})$ are the Fourier transforms of $x_m(n)$, $\mathbf{h}_m(n)$, $s(n)$ and $\nu_m(n)$, respectively.

Having introduced these properties of the room impulse response, the following can be deduced regarding the processing of reverberant speech:

1. Hand-free telephony users can be expected to move around their acoustic environment and so the AIRs will vary with time.
2. The use of measured impulse responses is not feasible for dereverberation due to the dependence on talker-microphone position and on the room geometry.
3. If the talker-microphone separation is much smaller than the critical distance, the effects of reverberation are likely to be negligible. Thus, dereverberation is of greatest importance when the source-microphone distance (D) is larger than the critical distance (D_c), $D \geq D_\mathrm{c}$.
4. The reverberation time in typical office-sized rooms can be expected to vary in the range 0.1-1 s. Consequently, this involves FIR filters of several thousand taps for typical sampling frequencies.
5. Although undermodelling of the channel is possible, late reflections are important in dereverberation, in particular in the case when $D \geq D_\mathrm{c}$.
6. The non-minimum phase property and the large spectral dynamic range of room transfer functions will raise challenges in designing AIR equalization filters.

1.6 Literature Overview

This section presents an overview of the existing literature, which deals explicitly with enhancement of reverberant speech with the aim to serve as an introduction to the topic and to provide an annotated bibliography. A more thorough treatment with additional bibliographic records of several methods mentioned here is provided in the relevant chapters.

1.6.1 Beamforming Using Microphone Arrays

Beamforming techniques are fundamentally important and among the first multi-channel processing approaches for enhancement of speech acquisition in noisy and reverberant environments [11]. The most direct and straightforward technique is the Delay-and-sum Beamformer (DSB) in which the microphone signals are delayed, to compensate for different times of arrival, and then weighted and summed [15, 89] as a convex combination. The output of the DSB can be written as

$$\bar{x}(n) = \sum_{m=1}^{M} w_m x_m(n - \tau_m), \tag{1.4}$$

where τ_m is the propagation delay in samples from the source to the m^{th} sensor and w_m is the weighting applied to the m^{th} sensor. In this way, the coherent components across channels, due to the direct-paths, are added constructively, while incoherent components, due to reverberation or noise, are attenuated [15].

The DSB can also be interpreted as forming a beam of sensitivity in the chosen direction. From this spatial filtering interpretation it can be seen intuitively that the beamformer approach works best for strongly localized sources and is less effective when the sound field is diffuse. The design of the weights is the spatial equivalent to the design of temporal FIR filters; the number of microphones is analogous to the number of taps and the spacing between sensors is analogous to the sampling frequency [15, 89]. Consequently, there is a spatial sampling criterion analogous to the time domain Nyquist sampling criterion, which relates the distance between microphones to the frequency components in the signal such that spatial aliasing can be avoided. This is defined as [15]

$$\|\mathbf{q}_{\text{mic},m} - \mathbf{q}_{\text{mic},m+1}\|_2 < \frac{c}{2f}, \tag{1.5}$$

where $\|\cdot\|_2$ denotes the Euclidean norm, $\mathbf{q}_{\text{mic},m}$ is the three-dimensional position vector of the m^{th} microphone and c the speed of sound. Talantzis and Ward studied an alternative design of optimal weights in [85]. It can be seen from the expression in (1.5) that, for broadband signals such as speech, a linear array may not be the optimal solution. Consequently, several designs have been proposed with three or four subarrays and with different microphone spacing such that each of these subarrays covers a different bandwidth [11, 15].

Several variants of the DSB exist. For example, the DSB can be extended into the filter-and-sum beamformer in which the scalar weights are replaced each by an FIR filter [15]. Alternatively, in an approach employing frequency subbands by Allen *et al.* [4], the signals are co-phased in each frequency band and the gain is adjusted based on the cross-correlation between the channels to remove incoherent components before the summation. A two-dimensional microphone array was proposed by Flanagan *et al.* [19], which uses a DSB with a 'track-while-scan' approach where the area under consideration is quantized into overlapping regions that are scanned sequentially and speech characteristics are incorporated to distinguish a speech source from noise. The extension to three-dimensional arrays has also been considered [20] and also the use of spherical microphone arrays [58, 61]. Adaptive beamforming approaches have been studied, which automatically adjust the weights of the beamformer [15, 43] and which may also include constraints in the adaptation rule [22]. Generally, beamformers have been found to be efficient in applications to suppress localized additive noise sources [11]. Reverberation can be partially reduced, as will be shown in Chap. 2. However, since diffuse reverberant sound comes from all possible directions in a room [56], it will always enter the look-direction of the beam and hence will be only partially suppressed.

Improvements to beamforming applied in reverberant environments can be achieved using multiple beamformers where, instead of only forming a single beam in the direction of the desired source, a three-dimensional array can be used to form additional beams that are steered in the direction of the strong initial reflections [20, 70]. The additional reflections are treated as virtual sources in a similar way to the source-image method for simulation of room acoustics [3] described in Chap. 2. Another approach is the matched filter beamformer where the microphone signals are convolved with the time-reversed room impulse responses [1, 20, 32, 53, 54]. However, both these methods require at least partial knowledge of the room impulse response and can rather be treated as an alternative to inverse filtering.

1.6.2 Speech Enhancement Approaches to Dereverberation

An early technique in the class of speech enhancement dereverberation was proposed by Oppenheim and Schafer [72, 73]. The authors first introduce the observation that simple echoes are observed as distinct peaks in the cepstrum of the speech signal. Consequently, they use a peak picking algorithm to identify these peaks and attenuate them with, for example, a comb filter. An alternative to this was also considered, where a lowpass weighting function was applied to the cepstrum assuming that most of the energy of speech is in the lower quefrencies. However, this approach was not found suitable for more complex reverberation models [73].

A class of techniques emerged from the observation that the linear prediction residual signal contains the effects of reverberation, comprising peaks corresponding to excitation events in voiced speech together with additional peaks due to the reverberant channel [10, 98]. These techniques aim to suppress the effects of reverberation without degrading the original characteristics of the residual such that dereverberated speech can be synthesized using the processed residual and the all-pole filter resulting from prediction analysis of the reverberant speech. It is assumed in these methods that the effect of reverberation on the Autoregressive (AR) coefficients is insignificant [10]. It was shown in [29] that the AR coefficients of the clean speech can be estimated accurately from multichannel observations.

An early idea based on linear prediction processing was proposed in a patent by Allen [2] where the author suggested that synthetic clean speech could be generated from reverberant speech by identifying the Linear Predictive Coding (LPC) parameters from one or more reverberant observations. Griebel and Brandstein *et al.* [33, 34] used wavelet extrema clustering to reconstruct an enhanced prediction residual. In [35] the authors employ coarse room impulse response estimates and apply a matched filter type operation to obtain weighting functions for the reverberant residuals. Yegnanarayana *et al.* [97] used multichannel time-aligned Hilbert envelopes to represent the strength of the peaks in the prediction residuals. The Hilbert envelopes are then summed and the result used as a weight vector, which is applied to the prediction residual of one of the microphones. In [98] the authors derive a weighting function based on the signal-to-reverberant ratio in different regions of the prediction residual. Gillespie *et al.* [31] demonstrate the kurtosis of the residual to be a useful reverberation metric, which they then maximize using an adaptive filter. This method was extended by Wu and DeLiang [94], who added a spectral subtraction stage to further suppress the remaining reverberation. Although these methods do attenuate the impulses due to reverberation in the prediction residual, they also significantly reduce naturalness in the dereverberated speech. This problem was ameliorated using a spatiotemporal averaging approach, where the speech signals are first spatially averaged and the prediction residual is further enhanced using temporal averaging of neighbouring larynx cycles [24, 27, 28, 86]. A further discussion on the processing of the linear prediction residual and the spatiotemporal averaging method will be given in Chap. 4.

A related method was proposed by Nakatani *et al.* [67]. This assumes a sinusoidal speech model. First the fundamental frequency of the speech signal is identified from the reverberant observations, then the remaining sinusoidal components are identified. Using the identified magnitude and phases of these sinusoids, an enhanced speech signal is synthesized. Subsequently, the reverberant and the dereverberated speech signals are used to derive an equivalent equalization filter. The processing is performed in short frames and the inverse filter is updated in each frame. It is shown that this inverse filter tends to the AIR equalization filter. However, this method may be computationally demanding [67].

Spectral subtraction has been widely applied, with some success, in noise reduction [6, 13]. Spectral subtraction was applied to dereverberation by Lebart *et al.* [57] and extended to the multichannel case by Habets [38, 39]. The authors assume a statistical model of the room impulse response comprising Gaussian noise modulated by a decaying exponential function. The decay rate of this exponential function is governed by the reverberation time. It is then shown that, if the reverberation time can be blindly estimated and in combination with multichannel spatial averaging, the power spectral density of the impulse response can be identified and subsequently removed by spectral subtraction. This method has shown promising results [38, 93], provided that the assumed unknowns are available, and will be elaborated in Chap. 3.

In summary, several speech enhancement approaches to dereverberation have appeared in the literature. These do not assume explicit knowledge of the room impulse response. However, blind identification of other features is often required. Nevertheless, many of these methods are computationally efficient and suitable for real-time implementation.

1.6.3 Blind System Identification and Inversion

The effects of reverberation can be removed if the AIR from the talker to at least one microphone can be identified and inverted so as to give a perfect equalizer for the acoustic channel. This approach presents several technical challenges that are the subject of much current research. Significant progress has been made towards addressing these difficulties though, at the time of writing this, many issues related to algorithm design and implementation remain open.

1.6.3.1 Blind System Identification

Blind multichannel system identification using second order statistics is usually based on the cross-relation between two observations x_1 and x_2 and the corresponding two AIRs h_1 and h_2, where the time index is temporarily omitted for brevity. The cross-relation is given by [95]: $x_1 * h_2 = (s * h_1) * h_2 = x_2 * h_1$, which leads to the system of equations $\mathbf{R}\mathbf{h} = \mathbf{0}$, where in general for M channels $\mathbf{R}$ is a correlation-like matrix [50] and $\mathbf{h} = [\mathbf{h}_1^T \ \mathbf{h}_2^T \ \ldots \ \mathbf{h}_M^T]^T$ is a vector of the concatenated AIRs. It can be seen from this system of equations that the desired solution is the eigenvector corresponding to the zeroth eigenvalue in $\mathbf{R}$ or, in the presence of noise, the smallest eigenvalue. Several alternative solutions have been proposed. A Least Squares (LS) approach for solving this problem is given in [95]. An eigendecomposition method was proposed by Gürelli and Nikias [36]. Gannot and Moonen [23] use eigendecomposition methods for blind system identification both in the full-band and in frequency subbands. Huang and Benesty proposed the use of adaptive filters and derived multichannel LMS and Newton adaptive filters both in the time domain [49, 51, 52] and in the frequency domain [50].

This type of blind system identification requires that the following identifiability conditions are satisfied [95]:

1. The unknown channels must not include common zeros.
2. The correlation matrix of the source signal must be full rank.

Blind acoustic system identification algorithms additionally have to overcome the following challenges:

1. Acoustic channels are normally time-varying and therefore system identification must be performed adaptively.
2. AIRs have a duration typically corresponding to thousands of coefficients, and estimation of systems with such high order requires robust algorithms with high numerical precision and that typically present high computational requirements.
3. Noise in the observations can cause the adaptive algorithms to misconverge. Some approaches have been developed to improve robustness [25, 26, 42, 52];
4. Many approaches assume knowledge of the order of the unknown system. This issue has been addressed, for example, in [23] and [21];
5. Solutions for $\mathbf{h}$ are normally found only to within a scale factor [23, 52, 95].

Other approaches include Subramaniam's [84] proposed use of the cepstrum for blind system identification between two channels. It is shown that the channels can be reconstructed from their phases using an iterative approach, where the phases are identified from the cepstra of the observed data [75, 84] but that the method is sensitive to zeros close to the unit circle – a situation which often arises in acoustic systems as was shown in [59]. A method introduced by Triki and Slock [88] comprises multichannel Linear Prediction (LP) to whiten the input signal and subsequent multichannel linear prediction which is used to identify the channels. A different approach to multichannel LP for dereverberation was taken in [14]. Recent developments of this class of methods will be discussed in more detail in Chap. 9. Finally, in [48] it is proposed to use an autoregressive model of channel impulse response, which is assumed to be stationary, in contrast to the FIR model employed in all the above methods. Furthermore, it is assumed that the source signal is a locally stationary AR process but that it is globally nonstationary. In this way, the parameters of the all-pole channel filter can be identified by observing several frames of the input signal and collecting information regarding the poles either by using a histogram approach or a more robust Bayesian probabilistic framework. Over several frames, the poles due to the stationary channel become apparent and the channel can thus be identified. One major advantage of this method is that, by using an AR model of the channel, the order of the channel is reduced compared to the FIR channel models. Further extensions based on this idea have been developed in [16–18]. This approach will be discussed in more depth in Chap. 8. Nevertheless, problems of sensitivity to noise and channel order estimation are common to all approaches and the subject of much current research, which will be discussed in Chaps. 5, 6 and 8.

1.6.3.2 Inverse Filtering

If the acoustic impulse responses from the talker to the microphones, $\mathbf{h}_m(n)$, are available, for example, from a blind system identification algorithm, dereverberation can be achieved in principle by an inverse system, $\mathbf{g}_m$, satisfying $\mathbf{h}_m^T(n)\mathbf{g}_m = \kappa\delta(n-\tau)$, where κ and τ are, respectively, arbitrary scale and delay factors. However, direct inversion of an acoustic channel presents several significant technical challenges.

1. AIRs have duration typically corresponding to thousands of coefficients and inversion of systems with such high order requires robust algorithms with high numerical precision and that typically present high computational requirements.
2. Acoustic channels typically exhibit non-minimum phase characteristics [68].
3. Acoustic channels may contain spectral nulls, which after inversion give strong peaks in the spectrum causing narrow band noise amplification.

Several alternative approaches have been studied for single channel inversion. For example, single channel LS inverse filters can be designed by minimizing the error $\hat{\mathbf{g}}_m = \min_{\mathbf{g}_m} ||\mathbf{h}_m^T(n)\mathbf{g}_m - \delta(n-\tau)||_2^2$ [64, 66]. Homomorphic inverse filtering has also been investigated [3, 64, 78, 87], where the impulse response is decomposed into a minimum phase component, $\mathbf{h}_{\mathrm{mp},m}(n)$ and an all-pass component, $\mathbf{h}_{\mathrm{ap},m}(n)$, such that $\mathbf{h}_m(n) = \mathbf{h}_{\mathrm{ap},m}^T(n)\mathbf{h}_{\mathrm{mp},m}(n)$. Consequently, magnitude and phase are equalized separately, where an exact inverse can be found for the magnitude, while the phase can be equalized, e.g., using matched filtering [55, 78]. An important result is that equalization of only the magnitude results in audible distortion [68, 78].

In the multichannel case, an exact inverse can be found by application of multichannel least squares design [51, 62]. The Multiple-input/output INverse Theorem (MINT) approach was the first such multichannel inversion method proposed by Miyoshi and Kaneda [62], which was also implemented in a subband version [96]. Adaptive versions have also been considered in [69]. If there are no common zeros between the two channel transfer functions, a pair of inverse filters, $\mathbf{g}_1$ and $\mathbf{g}_2$ can be found such that: $\mathbf{h}_1^T(n)\mathbf{g}_1 + \mathbf{h}_2^T(n)\mathbf{g}_2 = \delta(n)$. Thus, exact inverse filtering can be performed, with inverse filters of length similar to the channel length [51, 62]. Undermodelled estimates of $\mathbf{h}_m(n)$ are problematic for this type of inversion, and it has been observed that true channel inverses are of limited value for practical dereverberation when the channel estimate contains even moderate estimation errors. Regularized multichannel equalization was shown to increase the equalization robustness to noise and estimation errors [44–46, 99]. Acoustic channel equalization will be discussed in Chaps. 7 and 9.

1.7 Outline of the Book

The remainder of this book is organized as follows:

Chapter 2 reviews the acoustic characteristics of typical rooms and discusses measurement and simulation of acoustic impulse responses. Furthermore, subjective and objective measures of reverberation in speech are discussed.

Chapter 3 introduces a statistical model of the room impulse response and uses that to develop a multichannel spectral subtraction based algorithm for speech dereverberation.

Chapter 4 reviews the use of processing of the linear prediction residual for dereverberation of speech. A spatiotemporal averaging method for linear prediction residual processing is introduced and its application to speech dereverberation is demonstrated.

Chapter 5 develops a multichannel eigendecomposition method for blind identification of room impulse responses in the presence of coloured noise. The identified impulse responses are then used to design equalization filters for speech dereverberation.

Chapter 6 introduces a class of adaptive blind system identification methods with implementations both in the time and frequency domain. The adverse effects of noise on these algorithms are explored and several approaches to added noise robustness are presented and discussed.

Chapter 7 presents a multichannel Acoustic Transfer Function (ATF) equalizer design framework using oversampled and decimated subbands. The method is shown to allow for approximate equalization of long, non-minimum phase ATFs at low computational cost.

Chapter 8 considers blind dereverberation in time-varying acoustic environments. The source and the AIRs are presented by parametric models that are employed in combination with Bayesian inference to estimate the room acoustic parameters.

Chapter 9 uses multichannel linear prediction to derive an equalization filter without necessarily estimating the acoustic impulse responses first. This equalization filter results in excessive whitening of the speech and, consequently, four different methods to overcome this problem are presented.

Chapter 10 presents TRINICON – a generic framework for Multi-Input Multi-Output (MIMO) signal processing. It is applied here to derive two dereverberation algorithms: one where the AIRs are first identified blindly and used to design equalization filters, and the second where the equalization filters are identified directly from the reverberant observations.

References

1. Affes, S., Grenier, Y.: A signal subspace tracking algorithm for microphone array processing of speech. IEEE Trans. Speech Audio Process. **5**(5), 425–437 (1997)
2. Allen, J.B.: Synthesis of pure speech from a reverberant signal. U.S. Patent No. 3786188 (1974)
3. Allen, J.B., Berkley, D.A.: Image method for efficiently simulating small-room acoustics. J. Acoust. Soc. Am. **65**(4), 943–950 (1979)
4. Allen, J.B., Berkley, D.A., Blauert, J.: Multimicrophone signal-processing technique to remove room reverberation from speech signals. J. Acoust. Soc. Am. **62**(4), 912–915 (1977)
5. BBC, U.: Iceland comes first in broadband. [Online] (2006). URL `http://news.bbc.co.uk/1/hi/technology/4903776.stm`
6. Benesty, J., Makino, S., Chen, J. (eds.): Speech enhancement. Springer (2005)
7. Benesty, J., Sondhi, M.M., Huang, Y. (eds.): Springer handbook of speech processing. Springer (2007)
8. Bolt, R.H., MacDonald, A.D.: Theory of speech masking by reverberation. J. Acoust. Soc. Am. **21**(6), 577–580 (1949)
9. Bradley, J.S., Sato, H., Picard, M.: On the importance of early reflections for speech in rooms. J. Acoust. Soc. Am. **113**(6), 3233–3244 (2003)
10. Brandstein, M.S., Griebel, S.M.: Nonlinear, model-based microphone array speech enhancement. In: S.L. Gay, J. Benesty (eds.) Acoustic Signal Processing For Telecommunication, pp. 261–279. Kluwer Academic Publishers (2000)
11. Brandstein, M.S., Ward, D.B. (eds.): Microphone arrays: Signal processing techniques and applications, 1 edn. Springer (2001)
12. Cherry, C.: On human communications, third edn. MIT Press (1980)
13. Davis, G.M. (ed.): Noise reduction in speech applications. CRC Press (2002)
14. Delcroix, M., Hikichi, T., Miyoshi, M.: Precise dereverberation using multichannel linear prediction. IEEE Trans. Audio, Speech, Lang. Process. **15**(2), 430–440 (2007)
15. Elko, G.W.: Microphone array systems for hands-free telecommunication. Speech Communication **20**(3-4), 229–240 (1996)
16. Evers, C., Hopgood, J.R.: Parametric modelling for single-channel blind dereverberation of speech from a moving speaker. IET Communications **2**, 59–74 (2008)
17. Evers, C., Hopgood, J.R., Bell, J.: Acoustic models for blind source dereverberation using sequential Monte Carlo methods. In: Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP) (2008)

18. Evers, C., Hopgood, J.R., Bell, J.: Blind speech dereverberation using batch and sequential Monte Carlo methods. In: Proc. Int. Symp. on Circuits and Systems (2008)
19. Flanagan, J.L., Johnston, J.D., Zahn, R., Elko, G.W.: Computer-steered microphone arrays for sound transduction in large rooms. J. Acoust. Soc. Am. **78**(5), 1508–1518 (1985)
20. Flanagan, J.L., Surendran, A.C., Jan, E.E.: Spatially selective sound capture for speech and audio processing. Speech Communication **13**(1-2), 207–222 (1993)
21. Furuya, K., Kaneda, Y.: Two-channel blind deconvolution for non-minimum phase impulse responses. In: Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP), pp. 1315–1318 (1997)
22. Gannot, S., Burshtein, D., Weinstein, E.: Signal enhancement using beamforming and nonstationarity with applications to speech. IEEE Trans. Signal Process. **49**(8), 1614–1626 (2001)
23. Gannot, S., Moonen, M.: Subspace methods for multi-microphone speech dereverberation. EURASIP J. on App. Signal Process. **2003**(11), 1074–1090 (2003)
24. Gaubitch, N.D.: Blind identification of acoustic systems and enhancement of reverberant speech. Ph.D. thesis, Imperial College London (2007)
25. Gaubitch, N.D., Hasan, M.K., Naylor, P.A.: Generalized optimal step-size for blind multichannel LMS system identification. IEEE Signal Process. Lett. **13**(10), 624–627 (2006)
26. Gaubitch, N.D., Hasan, M.K., Naylor, P.A.: Noise robust adaptive blind identification using spectral constraints. In: Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP), pp. V–93–V–96. Toulouse, France (2006)
27. Gaubitch, N.D., Naylor, P.A.: Spatiotemporal averaging method for enhancement of reverberant speech. In: Proc. IEEE Int. Conf. Digital Signal Processing (DSP). Cardiff, UK (2007). DOI 10.1109/ICDSP.2007.4288655
28. Gaubitch, N.D., Naylor, P.A., Ward, D.B.: Multi-microphone speech dereverberation using spatio-temporal averaging. In: Proc. European Signal Processing Conf. (EUSIPCO), pp. 809–812. Vienna, Austria (2004)
29. Gaubitch, N.D., Ward, D.B., Naylor, P.A.: Statistical analysis of the autoregressive modeling of reverberant speech. J. Acoust. Soc. Am. **120**(6), 4031–4039 (2006)
30. Gay, S.L., Benesty, J. (eds.): Acoustic signal processing for telecommunication. Kluwer Academic Publishers (2000)
31. Gillespie, B.W., Malvar, H.S., Florêncio, D.A.F.: Speech dereverberation via maximum-kurtosis subband adaptive filtering. In: Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP), vol. 6, pp. 3701–3704 (2001)
32. Grenier, Y., Affes, S.: Microphone array response to speaker movements. In: Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP), vol. 1, pp. 247–250 (1997)
33. Griebel, S.M.: A microphone array system for speech source localization, denoising and dereverberation. Ph.D. thesis, Harvard University, Cambridge, Massachusetts (2002)
34. Griebel, S.M., Brandstein, M.S.: Wavelet transform extrema clustering for multi-channel speech dereverberation. In: Proc. Int. Workshop Acoust. Echo Noise Control (IWAENC). Pocono Manor, Pennsylvania (1999)
35. Griebel, S.M., Brandstein, M.S.: Microphone array speech dereverberation using coarse channel estimation. In: Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP), vol. 1, pp. 201–204 (2001)
36. Gürelli, L., Nikias, C.L.: EVAM: An eigenvector-based algorithm for multichannel blind deconvolution of input colored signals. IEEE Trans. Signal Process. **43**(1), 143–149 (1995)
37. Haas, H.: The influence of a single echo on the audibility of speech. J. Audio Eng. Soc. **20**, 145–159 (1972)
38. Habets, E.A.P.: Multi-channel speech dereverberation based on a statistical model of late reverberation. In: Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP), vol. 4, pp. iv/173–iv/176. Philadelphia (2005)
39. Habets, E.A.P.: Single- and multi-microphone speech dereverberation using spectral enhancement. Ph.D. thesis, Technische Universiteit Eindhoven (2007). URL `http://alexandria.tue.nl/extra2/200710970.pdf`
40. Haneda, Y., Makino, S., Kaneda, Y.: Common acoustical pole and zero modeling of room transfer functions. IEEE Trans. Speech Audio Process. **2**(2), 320–328 (1994)

41. Hänsler, E., Schmidt, G. (eds.): Topics in acoustic echo and noise control. Springer (2006)
42. Hasan, M.K., Benesty, J., Naylor, P.A., Ward, D.B.: Improving robustness of blind adaptive multichannel identification algorithms using constraints. In: Proc. European Signal Processing Conf. (EUSIPCO). Antalya, Turkey (2005)
43. Haykin, S.: Adaptive filter theory, 4 edn. Prentice Hall, Upper Saddle River, N.J. (2001)
44. Hikichi, T., Delcroix, M., Miyoshi, M.: Inverse filtering for speech dereverberation less sensitive to noise. In: Proc. Int. Workshop Acoust. Echo Noise Control (IWAENC), pp. 1–4 (2006)
45. Hikichi, T., Delcroix, M., Miyoshi, M.: On robust inverse filter design for room transfer function fluctuations. In: Proc. European Signal Processing Conf. (EUSIPCO) (2006)
46. Hikichi, T., Delcroix, M., Miyoshi, M.: Inverse filtering for speech dereverberation less sensitive to noise and room transfer function fluctuations. EURASIP J. Advances in Signal Processing **2007**, 1–12 (2007)
47. Hopgood, J.R., Rayner, P.J.W.: A probabilistic framework for subband autoregressive models applied to room acoustics. In: Proc. IEEE Workshop Statistical Signal Processing, pp. 492–495 (2001)
48. Hopgood, J.R., Rayner, P.J.W.: Blind single channel deconvolution using nonstationary signal processing. IEEE Trans. Speech Audio Process. **11**(5), 476–488 (2003)
49. Huang, Y., Benesty, J.: Adaptive multi-channel least mean square and Newton algorithms for blind channel identification. Signal Processing **82**(8), 1127–1138 (2002)
50. Huang, Y., Benesty, J.: A class of frequency-domain adaptive approaches to blind multichannel identification. IEEE Trans. Signal Process. **51**(1), 11–24 (2003)
51. Huang, Y., Benesty, J., Chen, J.: A blind channel identification-based two-stage approach to separation and dereverberation of speech signals in a reverberant environment. IEEE Trans. Speech Audio Process. **13**(5), 882–895 (2005)
52. Huang, Y., Benesty, J., Chen, J.: Optimal step size of the adaptive multichannel LMS algorithm for blind SIMO identification. IEEE Signal Process. Lett. **12**(3), 173–176 (2005)
53. Jan, E., Flanagan, J.L.: Microphone arrays for speech processing. In: Int. Symposium on Signals, Systems, and Electronics, pp. 373–376 (1995)
54. Jan, E., Svazier, P., Flanagan, J.L.: Matched-filter processing of microphone array for spatial volume selectivity. In: Proc. Int. Symp. on Circuits and Systems, vol. 2, pp. 1460–1463 (1995)
55. Kennedy, R.A., Radlović, B.D.: Iterative cepstrum-based approach for speech dereverberation. In: Proc. Int. Symposium on Signal Processing and Its Applications (ISSPA), vol. 1, pp. 55–58 (1999)
56. Kuttruff, H.: Room acoustics, 4 edn. Taylor & Francis (2000)
57. Lebart, K., Boucher, J.M., Denbigh, P.N.: A new method based on spectral subtraction for speech dereverberation. Acta Acoustica **87**, 359–366 (2001)
58. 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)
59. Lin, X., Gaubitch, N.D., Naylor, P.A.: Two-stage blind identification of SIMO systems with common zeros. In: Proc. European Signal Processing Conf. (EUSIPCO). Florence, Italy (2006)
60. Loizou, P.C.: Speech enhancement theory and practice. Taylor & Francis (2007)
61. Meyer, J., Agnello, T.: Spherical microphone array for spatial sound recording. In: Audio Engineering Society, 115th Convention, preprint 5975. New York (2003)
62. Miyoshi, M., Kaneda, Y.: Inverse filtering of room acoustics. IEEE Trans. Acoust., Speech, Signal Process. **36**(2), 145–152 (1988)
63. Mobile Operators Association: History of cellular mobile communications. [Online] (2005). URL `http://www.mobilemastinfo.com/information/history.htm`
64. Mourjopoulos, J., Clarkson, P., Hammond, J.: A comparative study of least-squares and homomorphic techniques for the inversion of mixed phase signals. In: Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP), vol. 7, pp. 1858–1861 (1982)
65. Mourjopoulos, J., Paraskevas, M.A.: Pole and zero modeling of room transfer functions. J. Sound Vib. **146**(2), 281–302 (1991)
66. Mourjopoulos, J.N.: Digital equalization of room acoustics. J. Audio Eng. Soc. **42**(11), 884–900 (1994)

67. Nakatani, T., Miyoshi, M., Kinoshita, K.: Single-microphone blind dereverberation. In: J. Benesty, S. Makino, J. Chen (eds.) Speech Enhancement, 1 edn. Springer Verlag (2005)
68. Neely, S.T., Allen, J.B.: Invertibility of a room impulse response. J. Acoust. Soc. Am. **66**(1), 165–169 (1979)
69. Nelson, P.A., Orduña-Brustamante, F., Hamada, H.: Inverse filter design and equalization zones in multichannel sound reproduction. IEEE Trans. Speech Audio Process. **3**(3), 185–192 (1995)
70. Nishiura, T., Nakanura, S., Shikano, K.: Speech enhancement by multiple beamforming with reflection signal equalization. In: Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP), vol. 1, pp. 189–192 (2001)
71. Omologo, M., Svazier, P., Matassoni, M.: Environmental conditions and acoustic transduction in hands-free speech recognition. Speech Communication **25**(1), 75–95 (1998)
72. Oppenheim, A.V., Schafer, R.W.: Digital signal processing, 1 edn. Prentice Hall (1975)
73. Oppenheim, A.V., Schafer, R.W., Stockham, T.G.: Nonlinear filtering of multiplied and convolved signals. IEEE Trans. Audio Electroacoust. **AU-16**(3), 437–466 (1968)
74. Paatero, T.: Modeling of long and complex responses using Kautz filters and time-domain partitions. In: Proc. European Signal Processing Conf. (EUSIPCO), pp. 313–316. Vienna, Austria (2004)
75. Petropulu, A.P., Nikias, C.L.: Blind deconvolution using signal reconstruction from partial higher order cepstral information. IEEE Trans. Signal Process. **41**(6), 2088–2095 (1993)
76. Plato: The republic. Penguin Books Ltd (2003)
77. Polycom: Polycom communicator. [Online] (2006). URL `http://www.polycom.com/`
78. Radlović, B.D., Kennedy, R.A.: Nonminimum-phase equalization and its subjective importance in room acoustics. IEEE Trans. Speech Audio Process. **8**(6), 728–737 (2000)
79. Radlović, B.D., Williamson, R.C., Kennedy, R.A.: Equalization in an acoustic reverberant environment: Robustness results. IEEE Trans. Acoust., Speech, Signal Process. **8**(3), 311–319 (2000)
80. Rayleigh, J.W.S.: The theory of sound. Dover Publications (1976)
81. Sabine, W.C.: Collected papers on acoustics. Dover Publications (1964)
82. Schmidt, G.: Applications of acoustic echo control – an overview. In: Proc. European Signal Processing Conf. (EUSIPCO), pp. 9–16. Vienna, Austria (2004)
83. Schroeder, M.R.: Statistical parameters of the frequency response curves of large rooms. J. Audio Eng. Soc. **35**(5), 299–305 (1987)
84. Subramaniam, S., Petropulu, A.P., Wendt, C.: Cepstrum-based deconvolution for speech dereverberation. IEEE Trans. Acoust., Speech, Signal Process. **4**(5), 392–396 (1996)
85. Talantzis, F., Ward, D.B.: Robustness of multi-channel equalization in an acoustic reverberant environment. J. Acoust. Soc. Am. **114**(2), 833–841 (2003)
86. Thomas, M.R.P., Gaubitch, N.D., Gudnason, J., Naylor, P.A.: A practical multichannel dereverberation algorithm using multichannel DYPSA and spatiotemporal averaging. In: Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics. New Paltz, NY (2007)
87. Tohyama, M., Lyon, R.H., Koike, T.: Source waveform recovery in a reverberant space by cepstrum dereverberation. In: Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP), vol. 1, pp. 157–160 (1993)
88. Triki, M., Slock, D.T.M.: Delay-and-predict equalization for blind speech dereverberation. In: Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP). Toulouse, France (2006)
89. VanVeen, B.D., Buckley, K.M.: Beamforming: a versatile approach to spatial filtering. IEEE Signal Process. Mag. **5**(2), 4–24 (1988)
90. Waller, S.J.: Sound and rock art. Nature **363** (1993)
91. Waller, S.J.: Psychoacoustic influences of the echoing environments of prehistoric art. J. Acoust. Soc. Am. **112** (2002)
92. Ward, D.B.: On the performance of acoustic crosstalk cancellation in a reverberant environment. J. Acoust. Soc. Am. **110**(2), 1195–1198 (2001)

93. Wen, J.Y.C., Gaubitch, N.D., Habets, E.A.P., Myatt, T., Naylor, P.A.: Evaluation of speech dereverberation algorithms using the MARDY database. In: Proc. Int. Workshop Acoust. Echo Noise Control (IWAENC). Paris, France (2006)
94. Wu, M., Wang, D.: A two-stage algorithm for one-microphone reverberant speech enhancement. IEEE Trans. Audio, Speech, Lang. Process. **14**(3), 774–784 (2006)
95. Xu, G., Liu, H., Tong, L., Kailath, T.: A least-squares approach to blind channel identification. IEEE Trans. Signal Process. **43**(12), 2982–2993 (1995)
96. Yamada, K., Wang, J., Itakura, F.: Recovering of broad band reverberant speech signal by subband MINT method. In: Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP), pp. 969–972 (1991)
97. Yegnanarayana, B., Prasanna, S.R.M., Rao, K.S.: Speech enhancement using excitation source information. In: Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP), vol. 1, pp. 541–544 (2002)
98. Yegnanarayana, B., Satyanarayana, P.: Enhancement of reverberant speech using LP residual signal. IEEE Trans. Acoust., Speech, Signal Process. **8**(3), 267–281 (2000)
99. Zhang, W., Gaubitch, N.D., Naylor, P.A.: Computationally efficient equalization of room impulse responses robust to system estimation errors. In: Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP) (2008)