Binaural Dereverberation

A. Tsilfidis, A. Westermann, J. M. Buchholz, E. Georganti and J. Mourjopoulos

1 Room Reverberation

When a sound is emitted by a source in an enclosed space, a listener will initially receive the direct sound followed by multiple reflections from the walls or objects placed in the room—see Fig. 1.

The energy of the reflected sound will be attenuated according to the frequency-dependent absorption of the reflecting surfaces. Moreover, assuming an omnidirectional source, the sound pressure of both the direct and the reflected sounds will decrease according to the $1/r$ distance law [83]. Although the sound energy decays over time, t, roughly following an exponential function, the reflection density increases with t^3, forming an increasingly diffuse sound field [83]—see also [42]. The required time for the energy in a room to decrease by 60 dB after the sound source has stopped emitting sound is the reverberation time, T_{60}, being the most commonly used parameter for specifying the acoustic properties of a given room [83, 110].

Assuming that the room acoustics are modeled as a *linear, time-invariant*, LTI, system, the *room impulse response*, RIR, provides a complete description of the direct and reflective paths in a room from a sound source to the receiver. In a general multichannel scenario with one source and i receivers, the reverberant signal, $x_i(n)$, for each specific source-receiver position in the room can be expressed as the convolution of the anechoic signal, $s(n)$, with the corresponding RIRs, $h_i(n)$, as follows,

$$x_i(n) = \sum_{j=0}^{J_h-1} h_i(j)s(n-j) \ , \tag{1}$$

A. Tsilfidis · E. Georganti · J. Mourjopoulos (✉)
Audio and Acoustic Technology Group, Electrical and Computer Engineering department,
University of Patras, Patras, Greece
e-mail: mourjop@upatras.gr

A. Westermann · J. M. Buchholz
National Acoustic Laboratories, Australian Hearing, Macquarie University, Sydney, Australia

J. Blauert (ed.), *The Technology of Binaural Listening*, Modern Acoustics
and Signal Processing, DOI: 10.1007/978-3-642-37762-4_14,

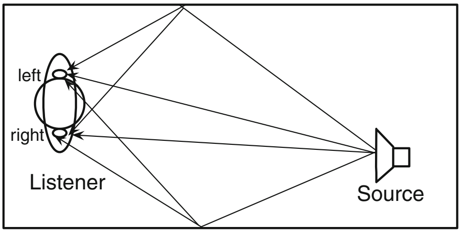

Fig. 1 A human listener and a sound source in a reverberant room

where n represents the discrete time index and J_h is the length of the impulse response.

In the binaural scenario the room response is combined with the related left- and right-ear *head-related impulse responses*, HRIRs. The latter are measured in anechoic conditions. As a consequence, assuming an ideal omni-directional source, a *binaural room impulse response*, BRIR, for the left-ear channel, $h_L(n)$, can be expressed as

$$\begin{aligned} h_L(n) &= g(r_s)\delta(n-n_s) * h_{HRIR,L,\theta_d,\phi_d}(n) \\ &\quad + \sum_{m=0}^{J_{h_m}-1} h_{m,L}(n) * h_{HRIR,L,\theta_m,\phi_m}(n) \ , \end{aligned} \tag{2}$$

where $g(r_s)$ is a gain reduction that depends on the source-receiver distance r_s, $\delta(n)$ refers to a Kronecker-delta function, n_s is the delay mainly depending on the source-receiver distance, r_s, and the physical characteristics of the propagation medium. $h_{HRIR,L,\theta_d,\phi_d}(n)$ is the left HRIR for the direct sound, corresponding to θ_d and ϕ_d, namely, the horizontal and vertical angles between source and receiver. The value $h_m(n)$ denotes the response of the m-th reflection. J_{h_m} is the number of individual reflections. $h_{HRIR,L,\theta_m,\phi_m}$ is the HRIR corresponding to such a reflection. Finally, θ_m and ϕ_m are the horizontal and vertical angles between receiver and m-th reflection.[1] A similar equation also applies for the BRIR, $h_R(n)$.

Hence, the reverberant signal at the left and right ear of a listener, $x_L(n)$ and $x_R(n)$, can be described as a convolution of the anechoic source signal, $s(n)$, with the left- and right-ear binaural room impulse responses, $h_L(n)$ and $h_R(n)$, that is,

$$x_L(n) = \sum_{j=0}^{J_{h_L}-1} h_L(j)s(n-j) \ , \tag{3}$$

$$x_R(n) = \sum_{j=0}^{J_{h_R}-1} h_R(j)s(n-j) \ . \tag{4}$$

Examples of BRIRs measured in a stairway with a reverberation time of approximately 0.86 s are shown in Fig. 2 for the left and right ear [67]. The initial delay before the arrival of the first peak of each RIR depicts the delay, n_s, due to the

[1] In some cases n_s is a fractional delay, and the delta function is not well defined.

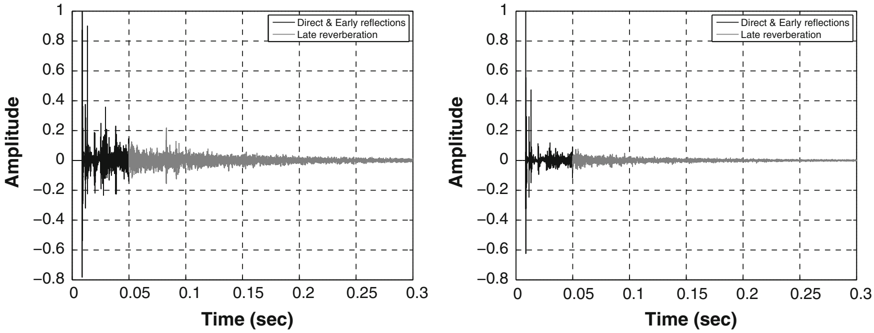

Fig. 2 Binaural room impulse responses in a stairway having a reverberation time of approximately 0.86 s. *Left* left-ear room impulse response. *Right* right-ear room impulse response

source-receiver distance, r_s. After the direct sound, the *early reflections* arrive. The early reflections are considered relatively sparse and span a short time interval of about 50–80 ms after the arrival of the direct sound. The last part of the RIR is called *late reverberation* and results to the reverberant tail of the signal [58, 77]. Due to the interaction of body, head and torso of the listener with the reverberant sound field, the RIRs at the left and right ear exhibit frequency-dependent interaural differences in arrival time and level. These interaural differences are essential parameters for binaural dereverberation algorithms, as is further described in Sects. 3.4 and 4.

In room acoustics and, consequently, in speech- and audio-enhancement applications, RIRs are often modeled as the sum of two components, one denoting the direct path and the early reflections, $h_{i,e}(n)$, and the other one the late reverberation, $h_{i,l}(n)$, as

$$h_i(n) = h_{i,e}(n) + h_{i,l}(n). \tag{5}$$

By combining (1) and (5), each reverberant signal can be written as the sum of a signal part affected by early reflections only, $x_{i,e}(n)$, and a signal part affected by late reverberation, $x_{i,l}(n)$, namely,

$$x_i(n) = \underbrace{\sum_{j=0}^{J_h-1} h_{i,e}(j)s(n-j)}_{x_{i,e}(n)} + \underbrace{\sum_{j=0}^{J_h-1} h_{i,l}(j)s(n-j)}_{x_{l,e}(n)}. \tag{6}$$

As will be further discussed in Sect. 2.2, these two components of room reverberation affect the received signal in a different way and are thus treated separately in most dereverberation applications.

2 Speech Signals in Rooms

2.1 Auditory Perception in Rooms

In an anechoic environment, where only the direct sound is present, a normal hearing listener can accurately localize arbitrary sound sources due to the presence of unambiguous interaural-time and -level cues as well as spectral cues that are provided by interaction of pinnae, head and torso with the sound field—see [13]. In particular due to the presence of interaural binaural cues, the auditory system is also able to suppress interfering sounds that arrive from different locations than a target sound and, thereby, for instance, improve speech intelligibility significantly. This phenomenon is commonly referred to as *spatial release from masking* [20].

In the case that a sound is presented in a reverberant environment, the direct sound is accompanied by early reflections and reverberation—see Sect. 1. This results in distortion of the available auditory cues and, typically, leads to reduced auditory performance, for instance, in localization or speech intelligibility. Whereas the early reflections, which arrive within a time window of about 50–80 ms after the direct sound, improve speech intelligibility [4, 17], late reverberation generally has a negative effect on speech intelligibility [61].

Auditory localization in rooms is aided by auditory mechanisms that are associated with the *precedence effect* [13, 88] and may be linked to a cue-selection mechanism that takes advantage of a measure of *interaural coherence* [36]. In particular, early reflections change the timbre of a sound and introduce perception of *coloration* [11, 23], a phenomenon that is significantly suppressed by the binaural auditory system [22, 130]. Also, late reverberation, which is mainly perceived within the temporal gaps inherent in the source signal, is reduced by the binaural system [26]. In [19] it has been shown that familiarization with a reverberant environment can result in enhanced speech intelligibility. Finally, it should be mentioned that auditory masking renders many reflections to be inaudible [24].

Besides the aforementioned detrimental effects of room reverberation on auditory performance, which are partly compensated by different auditory mechanisms, the room also introduces a number of additional cues that are utilized by the auditory system. The direct-to-reverberant energy ratio, for instance, provides a very reliable cue for distance perception [128]. Such mechanisms are described in detail in this volume [42]. Moreover, early lateral reflections extend the apparent width of a sound source [6] and late lateral reverberation energy makes a listener feel enveloped in an auditory scene [18]. The latter two phenomena are highly appreciated when listening to music [46]. The perceived spaciousness introduced by a room has often been related to interaural coherence, whereby, as can be stated as a rule, the lower the interaural-coherence, the higher is the perceived spaciousness [13].

When considering signal processing methods that aim at reducing room reverberation for applications with human listeners, such as in hearing aids or telecommunication devices, it is important that binaural cues are preserved. When binaural cues are distorted, the listener may not be able to correctly localize sound sources any more, although this is obviously very important for the orientation in an auditory scene as well as for perceiving warnings from potential threats. Moreover, with binaural cues being preserved, the binaural system may provide additional benefit by suppressing coloration, reverberation and interfering sound sources. Finally, successful dereverberation methods will increase the direct-to-reverberant energy ratio and, thus, may modify the perceived distance of a sound source and/or also modify the perception of apparent source width and envelopment.

2.2 Early- and Late-Reverberation Effects

The typical effects of reverberation in speech spectrograms are presented in Fig. 3, obtained with a fast fourier transform (FFT) length of 23.2 ms. Figure 3a shows

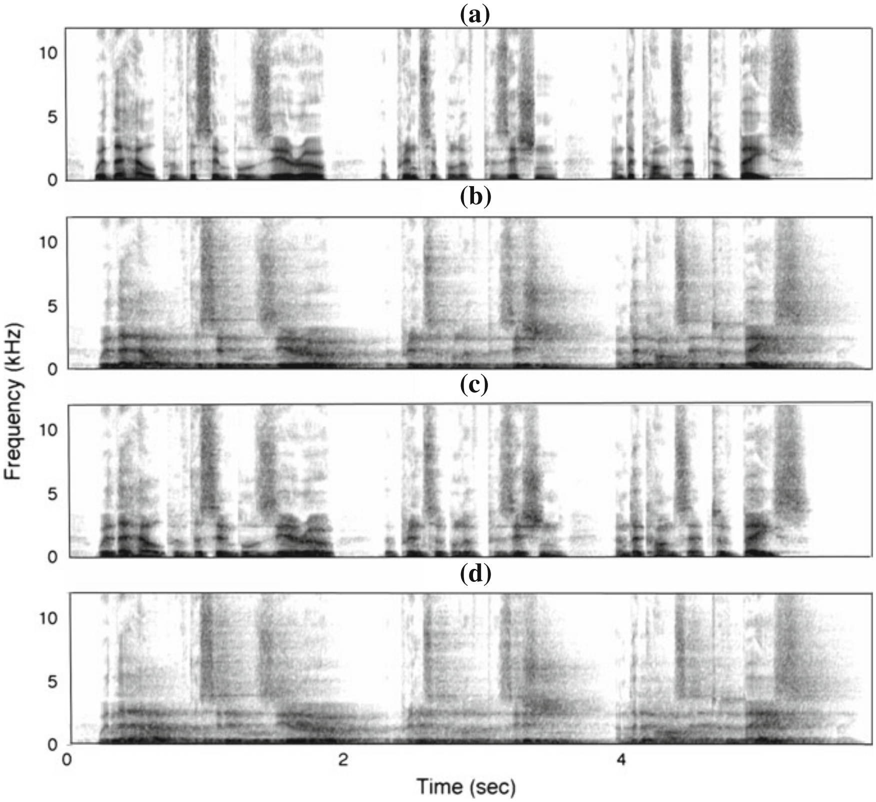

Fig. 3 Spectrograms illustrating the effects of reverberation on speech. **a** Anechoic input signal. **b** Reverberant signal. **c** Reverberant signal due to early reflections only. **d** Reverberant signal due to late reverberation only

an anechoic speech signal of a male speaker, and Fig. 3b shows the corresponding reverberant signal. The reverberant signal is produced via convolution with an RIR recorded in a lecture hall with a reverberation-time value of T_{60} =1 s at a distance of 4 m from the source. A comparison of Fig. 3a,b reveals that a large number of temporal gaps and spectral dips that can be seen in the anechoic speech are now filled due to the reverberation, that is, reverberation leads to a smearing of both temporal and spectral features. Moreover, the reverberation produced by signal components with high energy may mask later components with lower energy. In Fig. 3c the effect of early reflections on the reverberant signal is shown in isolation, obtained by convolving the anechoic signal with only the first 50 ms of the impulse response, since for speech applications this is considered to be the boundary between early reflections and late reverberation. Obviously, the early reflections alone do not significantly alter the anechoic speech spectrogram, but more careful observation reveals a smearing of the spectral speech profile. In Fig. 3d only the late-reverberant speech is shown, produced by convolving the anechoic signal with an artificially-modified impulse response where the first 50 ms were set to 0. It is evident that late reverberation significantly distorts the spectrogram of the anechoic signal and generates a reverberation tail between temporal speech gaps.

The effects of early and late reflections on the long-term speech spectrum (smoothed in 1/6 octave bands) can be observed in Fig. 4. The FFT length for these illustrations was equal to the signal length, namely, 7.8 s. In Fig. 4a the spectrum of the anechoic signal is compared to the spectrum of the same signal, contaminated by early reflections. In Fig. 4b the long-term spectrum of the anechoic signal is presented along with the spectrum of the same signal, contaminated by late reverberation. It is obvious that the early reflections significantly degrade the long-term speech spectrum, especially in the lower frequencies. This distortion is perceived as coloration of the sounds. In contrast, late reverberation introduces a more flat, white-noise like effect for the same frequency band of the signal.

Finally, Fig. 5 depicts speech spectrograms as obtained via convolution of an anechoic speech excerpt with (a) a left-ear impulse response and (b) a room impulse response obtained from an omni-directional microphone at exactly the same position. Both impulse responses were recorded in a lecture hall with $T_{60} = 0.79$ s at a source-receiver distance of 10.2 m [67]. It clearly appears that the spectrogram of the received speech signal is not significantly different for the binaural scenario. Such observation relates especially to the late reverberation signal components which are usually generated by diffuse reflections and, hence, are less susceptible to binaural cues. These late-reverberation effects can be treated by adapting single-channel dereverberation methods to the binaural scenario.

2.3 Interaural Coherence

An often used measure of similarity between two binaural signals or BRIRs is the interaural coherence, IC, defined as

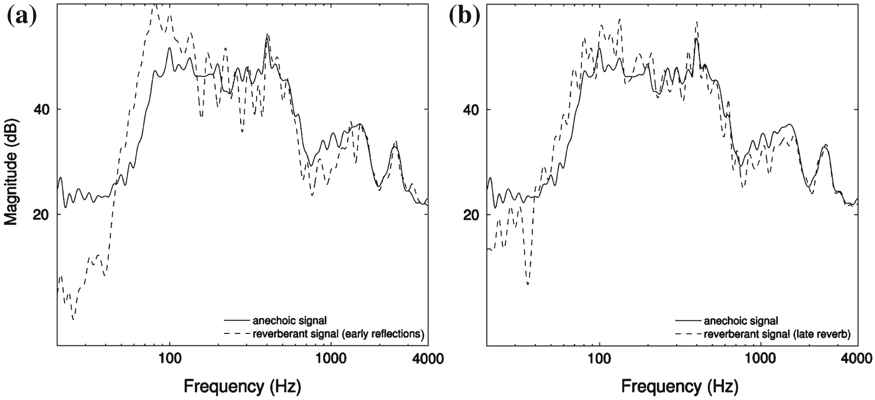

Fig. 4 Effect of early and late reflections on the long-term speech spectrum, smoothed in 1/6 octave bands. **a** spectrum of anechoic speech signal and of the signal contaminated by early reflections. **b** spectrum of anechoic speech signal and of the signal contaminated by late reverberation

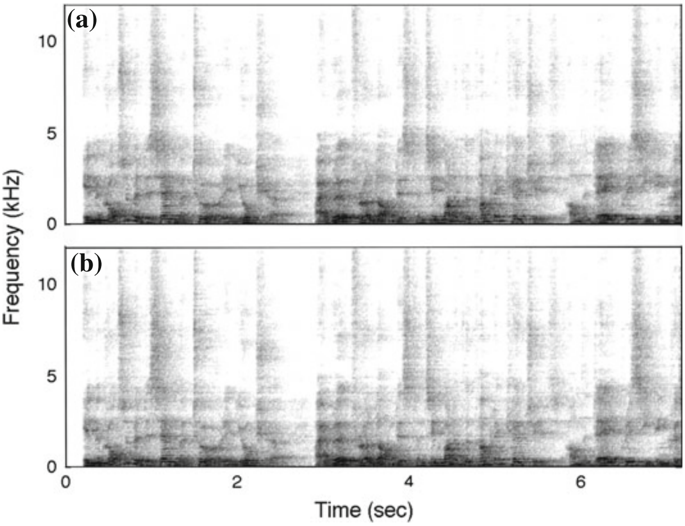

Fig. 5 Speech spectrograms obtained by convolving an anechoic speech sample. **a** *Left-ear* room impulse response measured in a lecture room with $T_{60} = 0.79$ s at 10.2 m from the source. **b** Room impulse response measured with an omni-directional microphone in the same room at exactly the same position [67]

$$IC_{X_L,X_R}(k) = \frac{|\langle X_L(k) \cdot X_R^*(k)\rangle|}{\sqrt{\langle X_L(k) \cdot X_L^*(k)\rangle\langle X_R(k) \cdot X_R^*(k)\rangle}} \, , \tag{7}$$

with k being the frequency band. $X_L(k)$ is the Fourier transform of $x_L(n)$. $X_R(k)$ is the Fourier transform of $x_R(n)$. "$*$" denotes the complex conjugate. $\langle s \rangle$ is the expected value of s. The magnitude-squared coherence, MSC, is referred to the square of (7). The IC behavior of speech in rooms is highlighted here with four examples—taken from [123] and [124]. The IC is estimated using the method described by (11). Figure 6a shows the IC plot for speech presented in a reverberation chamber, dominated by diffuse reflections and calculated from the binaural recordings of [51]. First, the algorithm defined in Sect. 4.1 was applied to obtain a 6.4 m short-term IC of the binaural representation of an entire sentence spoken by a male talker. From the resulting coherence values, the coherence plots were derived. The gray-graduation scale reflects the number of occurrences in a given frequency channel. As expected for the ideally diffuse sound field, an increased coherence is observed below 1 kHz. Above 1 kHz, most coherence values are between 0.1 and 0.3, whereby the minimum coherence that can be derived is limited by the duration of the time window applied in the coherence estimate.

Figure 6b–d shows examples of coherence plots for 0.5, 5 and 10 m source-receiver distances in an auditorium with $T_{60} = 1$ s. The overall coherence decreases with increasing distance between the source and the receiver. This results from the

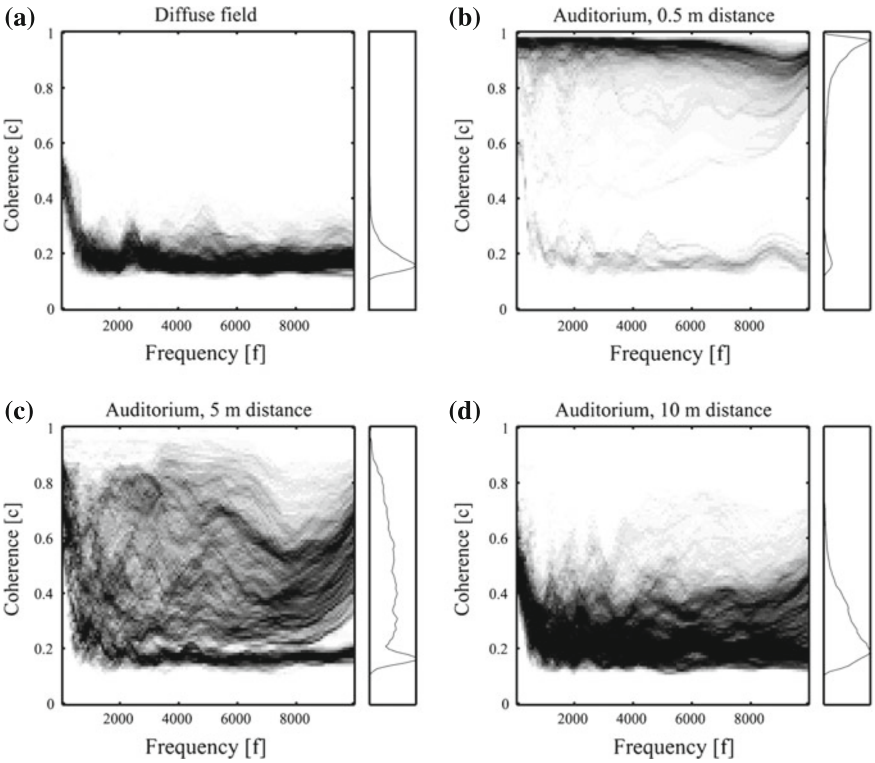

Fig. 6 **a** Interaural-coherence plots of speech presented in a diffuse field as a function of frequency. **b–d** Interaural-coherence plots in an auditorium at different distances from the sound source. The histograms summed across frequency are shown in the side panels

decreased direct-to-reverberant energy ratio at longer source-receiver distances. At very small distances—Fig. 6b—most coherence values are close to one, indicating that mainly direct-sound energy is present. In addition, coherence values arising from the diffuse field, having values between 0.1 and 0.3, are separated from those arising from the direct sound field. For the 5 m distance, frames with high coherence values are no longer observed. This is because frames containing direct-sound information are now affected by reverberation, and there is no clear separability between frames with direct and diffuse energy. At a distance of 10 m, this trend becomes even more profound as the coherence values drop further and the distribution resembles the one as found in the diffuse field, where very little direct sound information is available.

3 Review of Dereverberation-Techniques Literature

Since the early works of Flanagan and Lummis [37], Mitchell and Berkley [95] and Allen, Berkley and Blauert [2], many *blind*- or *non-blind*-dereverberation techniques have been developed, utilizing single or multiple input channels. As was shown in Sect. 2.2, early and late reverberation have different effects on anechoic signals. Hence, most of the research efforts handle early and late-reverberant-signal components separately, enhancing either $x_{i,e}(n)$ or $x_{i,l}(n)$. In the following sections, a summary of the existing literature on devererberation is presented.

3.1 Suppression of Early Reflections and Decoloration

Inverse Filtering

Inverse filtering of the RIR [97–99, 104] is used to minimize the coloration effect produced by the early reflections. In theory, an ideal RIR inversion will completely remove the effect of reverberation—both early and late reflections. However, the RIR is known to have non-minimum phase characteristics [104] and the non-causal nature of the inverse filter may introduce significant artifacts. In addition, exact measurements of the RIR must be available for the specific source-receiver room position, even if the RIRs are known to present common features in different room positions, as in [50]. The above limitations can be avoided by compensating exclusively for the broad spectral-coloration effect. For this, many single or multichannel techniques have been proposed, such as based on minimum-phase inverse [104], least-squares [100], frequency warping [52, 53, 111], complex smoothing [54, 55], Kautz filters [71, 107], frequency-dependent regularization [75] and filter clustering [10, 99]. Many of them are already incorporated in commercial room-correction systems, which rely on in-situ RIR measurements. However, results from perceptual tests show that some of these techniques do not always achieve the desired perceptual effect [56, 105].

Cepstral Techniques

In 1975, Stockham restored old Caruso recordings through cepstral blind deconvolution [106, 113]. The technique was based on homomorphic signal processing, exploring the fact that deconvolution may be represented as a subtraction in the log-frequency domain. Similar dereverberation techniques based on the same principle were later proposed in [8, 100, 108].

LP-Residual Enhancement

Using the source-filter production model, the speech can be represented as a convolutive mixture of the *linear-prediction*, LP, coefficients and the LP residual [29]. The fundamental assumption of the LP-residual dereverberation techniques is that the excitation signal is distorted by the room reflections, while the LP coefficients are not significantly affected from reverberation. Hence, the above techniques enhance the LP residual and recover the speech by applying the reverberant LP coefficients [40, 43, 45, 81, 102, 127].

3.2 Late-Reverberation Suppression

Temporal-Envelope Filtering

A class of techniques mostly aiming at compensating for late reverberation is based on temporal envelope filtering [5]. They are motivated by the concept of *modulation index* [62]. The modulation index is reduced when the late-reverberation tails fill the low-energy regions of a signal [84]. Mourjopoulos and Hammond [101] have shown that dereverberation of speech can be achieved by temporal envelope deconvolution in frequency sub-bands. Furthermore, the temporal envelope-filtering principle has been found to be advantageous when used in combination with other techniques such as LP-residual enhancement [127] and spectral subtraction [81]. Further, in [119], a sub-band temporal envelope-filtering technique, based on a computational auditory-masking model [21], has been proposed.

Spectral Enhancement

A number of dereverberation techniques based on spectral-enhancement techniques have been developed inspired by a multi-microphone reverberation-reducing method proposed by Flanagan and Lummis [37]. The same concept was later explored in the dereverberation method proposed by Allen et al. [2]. Spectral subtraction was mainly explored for denoising applications [9, 16, 29, 33, 89]. The classical technique is implemented in the STFT domain. Its main principle is to subtract an estimate of the noise-power spectrum from the power spectrum of the noisy signal. Usually, a speech-activity detector is involved in order to update the estimation of noise characteristics during the non-speech frames.

The most common processing artifact introduced by spectral enhancement is the so-called *musical noise*. It is generated when spectral bins of the noisy signal are strongly attenuated, because they are close to or below the estimated noise spectrum. As a result, the residual noise contains annoying pure-tone components at random frequencies. Most spectral-enhancement methods are trying to accurately estimate the noise spectra and avoid or reduce the musical noise [25, 89, 118, 120].

As indicated by (3), reverberation is a convolutive distortion. However, late reverberation can be considered as an additive degradation with noise-like characteristics—see (6). Hence, in the dereverberation context spectral subtraction has been adapted for the suppression of late reverberation. The basic principle of spectral-subtraction dereverberation for single-channel signals, originally presented in [86], is estimating the short-time spectrum of the clean signal, $S_e(m, k)$, by subtracting an estimation of the short-time spectrum of late reverberation, $R(m, k)$, from the short-time spectrum of the reverberant signal, $X(m, k)$, that is,

$$S_e(m, k) = X(m, k) - R(m, k) \ , \tag{8}$$

where k and m are the frequency bin and time frame index respectively. Following an alternative formulation, the estimation of the short-time spectrum of the clean signal can be derived by applying appropriate weighting gains, $G(m, k)$, to the short-time spectrum of the reverberant signal, such as

$$S_e(m, k) = G(m, k)X(m, k) \ , \tag{9}$$

where

$$G(m, k) = \frac{X(m, k) - R(m, k)}{X(m, k)} \ . \tag{10}$$

Further examples of spectral enhancement dereverberation methods can be found in [34, 38, 47, 48, 118, 126].

3.3 Dereverberation Methods Based on Multiple Inputs

Multichannel dereverberation may be considered as a somewhat easier task than the single-channel dereverberation, since the spatial diversity of the received signals can be further exploited. A set of such multichannel techniques is based on beamforming [121]. They explore the directivity properties of microphone arrays and require some a-priori knowledge of the array configuration. For a given system, the improvement depends on the microphone arrangement and the source-receiver positions, but it is independent of the reverberation time, T_{60}, of the room [41]. In simple implementations, the beamforming microphone arrays may present fixed-directivity characteristics such as in fixed-beamforming techniques, however adaptive beamforming setups where the processing parameters are adjusted to the environment also exist. Most beamforming algorithms assume that the noise and the source signal are statistically independent. This assumption does not stand for reverberation, which is a convolutive distortion. Therefore, the performance of such algorithms is poor in the dereverberation context [12].

Some early methods for multichannel dereverberation were presented in [2, 15, 37]. Miyoshi et al. [96] have shown that in non-blind multichannel systems perfect inverse filtering can be achieved when the captured RIRs do not share any common zeros. A technique that performs multiple-point room equalization using adaptive filters has been presented in [31]. Complete reverberation reduction may be theoretically achieved by applying blind deconvolution [59]. However, in order to perform blind deconvolution, the signal and the RIR must be irreducible, that is, they cannot be expressed as the convolution of two other signals [82]. The LTI systems are usually reducible and hence in principle blind deconvolution cannot be applied. In order to overcome the above limitation, single or multichannel blind-deconvolution implementations often involve a very low channel order and the number of reflections in the tested RIRs is unrealistically low—being based on simulations, as in [32, 39, 59, 60]. A set of room impulse response shortening techniques has been also proposed [70, 94, 129].

Further, multichannel blind-deconvolution methods for speech based on the LP analysis have been developed, based on the following principle. When the input of a system is white it can be equalized through multichannel LP. For speech dereverberation, the reverberant speech signal is pre-whitened in order to estimate a dereverberation filter. Then this filter is applied to the reverberant signal [28, 38, 74, 114]. A multichannel combined noise- and reverberation-suppression technique based in matched filtering has been presented in [30].

3.4 Binaural Techniques

Dereverberation is particularly important for binaural applications, for example, in digital hearing aids, binaural telephony, hands-free devices, and immersive audio [49, 91, 125]. However, developing models for binaural dereverberation and/or adapting single or multichannel techniques for binaural processing is not a trivial task. Binaural dereverberation cannot be considered as just a subset of the multichannel dereverberation techniques. Apart from the challenging task of reducing reverberation without introducing audible artifacts, binaural dereverberation methods should preserve the interaural-time-difference, ITD, and interaural-level-difference, ILD, cues, because it has been shown that bilateral signal processing can otherwise adversely affect source localization [49].

As was already discussed earlier in this section, such methods can be historically related to Allen et al. [2], who proposed a binaural approach where gain factors are determined by the diffuseness of the sound field between two spatially-separated microphones—see also [14, 15]. The technique involves two methods for calculating gain factors, one of which representing the coherence function of the two channels. However, because of a cophase-and-add stage that combines the binaural channels, only a monaural output was provided by this early method. Kollmeier et al. extended the original approach by applying the original coherence-gain factor separately to both channels, thus providing a binaural output [80]. A binaural variant of the original Allen et al. algorithm was also presented in [85].

In [112], a coherence-based Wiener filter was suggested that estimates the reverberation noise from a model of coherence between two points in a diffuse field. The method was further refined in [93] and [69] where acoustic shadow effects from a listener's head and torso were included. Jeub et al. [68] proposed a two-stage dereverberation algorithm that explicitly preserves binaural cues. They demonstrated that synchronized spectral weighting across binaural channels is important for preserving binaural cues. In [91] and also in [68], a binaural version of the single-channel spectra-subtraction technique presented in [86] is employed. In [115], a unified framework for binaural spectral subtraction dereverberation has been discussed. Lee et al. [87] presented a semi-blind method where they estimated a dereverberation filter from a pre-trained whitening filter and a whitened signal. Note that despite the great importance of binaural dereverberation, only few studies have been published up to now in the existing literature.

4 Examples of Dereverberation Algorithms

4.1 Method Based on Interaural Coherence

Historically, coherence-based methods (see 3.4) directly apply the coherence estimates as a gain to both binaural channels. Considering the processing as a mapping between coherence and gain, these methods apply a frequency-independent linear coherence-to-gain mapping. However, the strong source-receiver distance dependency observed in Fig. 6 and the inherent variations of the coherence across frequency highlights the necessity for applying acoustic scenario specific coherence-to-gain mapping functions. While for close source-receiver distances—Fig. 6b—a rather shallow mapping function is already able to suppress reverberant components and to preserve direct sound components, a much steeper function is required for larger distances—Fig. 6c,d. Therefore, a method is proposed here, taken from [123], which applies a parameterized mapping function that is controlled by an estimate of the present coherence statistics. The signal-processing steps for this dereverberation method are illustrated in Fig. 7. Two reverberant time signals, recorded at the left and right ear of a person or a dummy head, $x_L(n)$ and $x_R(n)$, are transformed to the time-frequency domain using short-time fourier transform, STFT [3]. This results in complex-valued short-term spectra, $\underline{X}_L(m,k)$ and $\underline{X}_R(m,k)$, where m denotes the

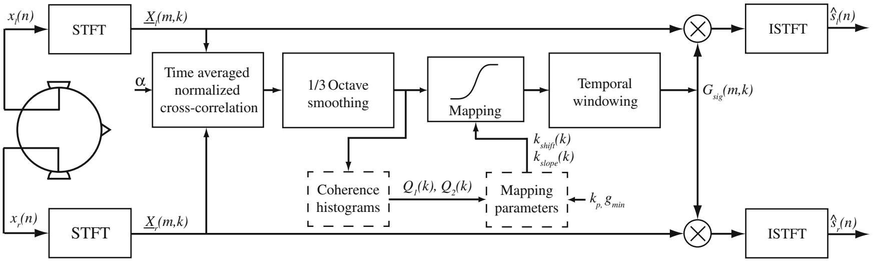

Fig. 7 Block diagram of a dereverberation method utilizing IC. The signals recorded at the ears, $x_L(n)$ and $x_R(n)$, are transformed via the STFT to the time-frequency domain, resulting in $\underline{X}_L(m,k)$ and $\underline{X}_R(m,k)$. The IC is calculated for each time-frequency bin and third-octave smoothing is applied. Statistical long-term properties of the IC are used to derive parameters of a sigmoidal mapping stage. The mapping is applied to the IC to realize a coherence-to-gain mapping, and subsequent temporal windowing is performed. The derived gains or weights are applied to both channels, $\underline{X}_L(m,k)$ and $\underline{X}_R(m,k)$. The dereverberated signals, $\hat{s_L}(n)$ and $\hat{s_R}(n)$, are reconstructed by applying an inverse SFTF

time frame and k the frequency band. For the STFT, a Hanning window of length L, including zero-padding of length $L/2$, and a 75 % overlap between successive windows are used. For each time-frequency bin, the absolute value of the IC, referred also as coherence, is calculated according to (7), which is implemented as follows

$$C_{LR}(m,k) = \frac{|\Phi_{LR}(m,k)|}{\sqrt{\Phi_{LL}(m,k)\Phi_{RR}(m,k)}} \ , \tag{11}$$

with $\Phi_{LL}(m,k)$, $\Phi_{RR}(m,k)$ and $\Phi_{LR}(m,k)$ representing the exponentially-weighted short-term cross-correlation and auto-correlation functions, namely,

$$\Phi_{ll}(m,k) = \alpha\Phi_{ll}(m,k-1) + \left|\underline{X}_l(m,k)\right|^2 \tag{12}$$

$$\Phi_{rr}(m,k) = \alpha\Phi_{rr}(m,k-1) + \left|\underline{X}_r(m,k)\right|^2 \tag{13}$$

$$\Phi_{lr}(m,k) = \alpha\Phi_{lr}(m,k-1) + \underline{X}_r(m,k)\underline{X}_l^*(m,k) \tag{14}$$

where α is the recursion constant.

The resulting IC estimates are spectrally smoothed using third-octave smoothing [57]. From the long-term statistical properties of the smoothed IC estimates, the parameters of a sigmoidal mapping function are derived. This mapping is subsequently used to transform the coherence estimates to the gain function, $G_{\text{sig}(m,k)}$. In order to suppress potential aliasing artifacts that may be introduced, temporal windowing is applied [72]. This is realized by applying an inverse STFT to the derived filter gains and then truncating the resulting time-domain representation to a length of $L/2+1$. The filter response is then zero-padded to a length of L and another STFT is performed. The resulting filter gain is applied to both channels, $\underline{X}_L(m,k)$, and, $\underline{X}_R(m,k)$. The dereverberated signals, $\hat{s_L}(n)$, and, $\hat{s_R}(n)$, are finally reconstructed by applying the inverse STFT and then adding the resulting overlapping signal segments [3].

Coherence-to-Gain Mapping

In order to cope with the different frequency-dependent distributions of the IC observed in different acoustic scenarios—see Sect. 2.3—a coherence-distribution dependent gain-to-coherence mapping is introduced. This is realized by a sigmoid function which is controlled by an online estimate of the statistical properties of the IC in each frequency channel. The function is derived from a normal sigmoid and given by

$$G_{\text{sig}}(m,k) = \frac{(1-G_{\min})}{1+e^{-k_{\text{slope}}(k)(C_{LR}(m,k)-k_{\text{shift}}(k))}} + G_{\min}\ , \tag{15}$$

where k_{slope} and k_{shift} control the sigmoidal slope and the position. The minimum gain, $G_{\min}$, is introduced to limit signal-processing artifacts.

In order to calculate the frequency-dependent parameters of the sigmoidal mapping function, coherence samples for a duration defined by t_{sig} are gathered in a histogram. The method yields best performance with a t_{sig} in the range of several seconds, assuming that the source-receiver locations are kept constant. For moving sources and varying acoustic environments, the method for updating the sigmoidal mapping function might need revision. The mapping functions are determined as two predefined points, Q_1 and Q_2, corresponding to the 1st and 2nd quartiles of the estimated IC-histogram distributions. A coherence histogram shown as a Gaussian distribution for illustrative purposes is exemplified in Fig. 8a by a gray curve together with the corresponding 1st and 2nd quartiles. An example sigmoidal coherence-to-gain mapping function is represented by a black solid curve. The linear mapping function as applied by [2] is indicated by the black dashed curve.

The degree of processing is determined by k_p, which directly controls the slope of the sigmoidal mapping function. The parameters k_{slope} and k_{shift} of the mapping function can be derived from $\varsigma(Q_1) = G_{\min} + k_p$ and $\varsigma(Q_2) = 1 - k_p$ as follows,

$$k_{\text{shift}} = \left(\frac{ln(\varsigma(Q_1)^{-1})}{ln(\varsigma(Q_2)^{-1})}Q_2 + Q_1\right)\cdot\left(1 - \frac{ln(\varsigma(Q_1)^{-1})}{ln(\varsigma(Q_2)^{-1})}\right)^{-1} \tag{16}$$

$$k_{\text{slope}} = \frac{ln(\varsigma(Q_1)) - 1}{Q_1 - k_{\text{shift}}}\ , \tag{17}$$

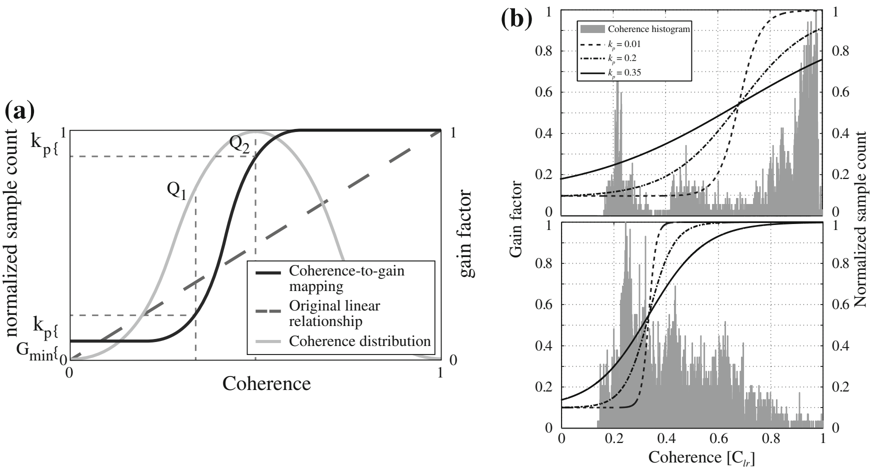

Fig. 8 **a** Idealized IC histogram distribution in one frequency-channel (*gray curve*). The coherence-to-gain relationship in the specific channel is calculated to intersect $\varsigma(Q_1) = G_{\min} + k_p$ and $\varsigma(Q_2) = 1 - k_p$. Thereby, $G_{\min}$ denotes the maximum attenuation and k_p determines the processing degree. **b** IC histogram distribution of speech presented in an auditorium with 0.5 m source-receiver distance (*top panel*) and 5 m source-receiver distance (*bottom panel*). Sigmoidal coherence-to-gain relationship for three different processing degrees of k_p are shown

whereby Q_1 and Q_2 are estimated in each frequency channel from the measured coherence histograms and k_p a predetermined parameter—see Fig. 8a. In addition, $G_{\min}$ is introduced to avoid signal artifacts related to applying infinite attenuation.

For speech presented in an auditorium with source-receiver distances of 0.5 m and 5 m—see Sect. 2.3—examples of sigmoidal mapping functions are shown in Fig. 8b for different values of k_p in the 751.7 Hz frequency channel. It can be seen that the coherence-to-gain mapping steepens as k_p increases. In addition, with the distribution broadening, that is, from 5 m to 0.5 m, the slope of the coherence-to-gain mapping decreases. Hence, in contrast to the original coherence-based-dereverberation approach in [2], which considered a linear coherence-to-gain mapping—plotted with dashed line in Fig. 8—the approach presented here provides a mapping function with added flexibility that can be adjusted by the parameter k_p and to any given acoustic condition.

4.2 Spectral-Subtraction Framework

As shown in (8), in the spectral-subtraction framework, the dereverberation problem is deduced to an estimation of the late-reverberation short-time spectrum. Several single-channel techniques have been introduced to blindly provide such estimates. Under specific conditions, as discussed later in this section, such blind dereverberation methods may be also adapted for binaural processing. For instance, Lebart et al. [86] proposed a method, referred to in the following as LB, that is based on exponential-decay modeling of the RIR, $h(n)$, as

$$h(n) = b(n)\exp(-3\ln 10/T_{60,\,n})\ , \tag{18}$$

where $b(n)$ is a zero-mean Gaussian stationary noise. T_{60} is the reverberation time in seconds [83].[2] The short-time spectral magnitude of the reverberation is estimated as

$$|R(m,k)| = \frac{1}{\sqrt{SNR_{pri}(m,k)+1}}|X(m,k)|\ , \tag{19}$$

where $SNR_{pri}(m,k)$ is the a-priori signal-to-noise ratio that can be approximated by a moving average relating to the a-posteriori signal-to-noise ratio, $SNR_{post}(m,k)$, in each frame,

$$\begin{aligned} SNR_{pri}(m,k) = \beta SNR_{pri}(m-1,k) + \\ (1-\beta)\max(0, (SNR_{post}(m,k)-1))\ , \end{aligned} \tag{20}$$

where β is a constant taking values close to one. The a-posteriori SNR is defined as

$$SNR_{post}(m,k) = \frac{|X(m,k)|^2}{E[|X(m,k)|]}\ . \tag{21}$$

Thus, $S_e(m,k)$ is estimated by subtraction and is combined with the phase of the reverberant signal, so that the dereverberated signal in the time domain is finally obtained through an overlap-add process.

An alternative method, as proposed by Wu and Wang [126], referred to hereinafter as WW, is motivated by the observation that the smearing effect of late reflections produces a smoothing of the signal spectrum in the time domain. Hence, similarly to the approach of [86], the late-reverberation power spectrum is considered a smoothed and shifted version of the power spectrum of the reverberant speech, namely,

$$|R(m,k)|^2 = \gamma w(m-\rho) * |X(m,k)|^2\ , \tag{22}$$

where ρ is a frame delay. γ is a scaling factor and "*" denotes convolution. The term $w(j)$ represents an assymetrical smoothing function given by the Rayleigh distribution

$$w(m) = \begin{cases} \dfrac{m+\alpha}{\alpha^2}\exp\left(\dfrac{-(m+\alpha)^2}{2\alpha^2}\right) & \text{if } j < -\alpha\ , \\ 0 & \text{otherwise}\ , \end{cases} \tag{23}$$

[2] Note that this model holds when the direct-to-reverberant ratio is smaller than 0 dB [48].

where α represents a constant number of frames. The phase of the reverberant speech is combined with the spectrum of the estimated clean signal and overlap-add is used to extract the time domain estimation.

Alternatively, Furuya and Kataoka [38] proposed a method, referred to hereinafter as FK, where the short-time power spectrum of late reverberation in each frame can be estimated as the sum of filtered versions of the previous frames of the reverberant signal's short time power spectrum, that is,

$$|R(m,k)|^2 = \sum_{l=1}^{M} |a_l(m,k)|^2 |X(m-l,k)|^2 \ , \tag{24}$$

where M is the number of frames that corresponds to an estimation of the T_{60}. $a_l(m,k)$ are the coefficients of late reverberation. The FK method assumes that an inverse filtering step, which reduces spectral degradation produced by the early reflections, precedes the spectral subtraction. Hence, in such a case the short-time power spectrum of the reverberant signal is considered to roughly approximate the short-time power spectrum of the anechoic signal. The coefficients of late reverberation are derived from

$$a_l(m,k) = E\left\{ \frac{X(m,k)X^*(m-l,k)}{|X(m-l,k)|^2} \right\} . \tag{25}$$

With these coefficients an estimation of the clean signal in the time domain can be derived through overlap-add from the short-time spectrum of the dereverberated signal, $S_e(m,k)$, as follows,

$$S_e(m,k) = \left\{ \frac{|X(m,k)|^2 - |R(m,k)|^2}{|X(m,k)|^2} \right\} X(m,k) \ . \tag{26}$$

Overlap-add is finally applied in order to estimate the time-domain dereverberated signal.

Although the above methods were originally employed for single-channel dereverberation, they can be adapted for binaural processing. For such case, as discussed in Sect. 3.4, in order to preserve the binaural ITD and ILD cues identical processing should be applied to the left and right signal channels. Similar principles apply to the binaural noise reduction, as in [72]. An effective approach for extending the LB method to a binaural context is to derive a reference signal using a delay-and-sum beamformer, DSB [68], where the time delays are estimated utilizing a method based on the generalized cross-correlation with phase transform as proposed in [76]. The reference signal is then calculated as the average of the time aligned left and right reverberant signals. Using the reference, appropriate weighting gains are derived, and identical processing is applied to both left and right channels. In [115], the DSB approach is also implemented for both the WW and FK methods in order to evaluate the efficiency of different late-reverberation-estimation techniques in a binaural scenario. However, in binaural applications, the time delay between the left and right channels of the speech signal is limited by the width of the human head. Therefore, it can be assumed to be shorter than the length of a typical analysis window used in spectral-subtraction techniques. Hence, in [115], it was shown that the time alignment stage can be omitted.

A different approach in order to adapt single channel spectral subtraction dereverberation in the binaural scenario is to process the left and right-ear channel signals independently. This results in the corresponding weighting gains, $G_L(m,k)$ and $G_R(m,k)$. These two gains can be combined, and different adaptation strategies have been investigated for each algorithm, namely,

(a) The binaural gain can be derived as the maximum of the left and right-channel weighting gains,

$$G(m,k) = \max(G_L(m,k), G_R(m,k)) \; . \tag{27}$$

This approach, maxGain, achieves moderate late-reverberation suppression, but it is also less likely to produce overestimation artifacts.

(b) The binaural gain can be derived as the average of the left and right channel weighting gains,

$$G(m,k) = \frac{(G_L(m,k) + G_R(m,k))}{2} \; . \tag{28}$$

This gain-adaptation strategy, avgGain, compensates equally for the contribution of the left and right channels.

(c) The binaural gain can be derived as the minimum of the left and right channel weighting gains,

$$G(m,k) = \min(G_L(m,k), G_R(m,k)) \; . \tag{29}$$

This adaptation technique, minGain, results in maximum reverberation attenuation, but the final estimation may be susceptible to overestimation artifacts.

After the derivation of the adapted gain, a *gain-magnitude regularization*, GMR, technique can be applied. The purpose of such as step is twofold. Firstly, the GMR has been proved to be a low-complexity approach reducing annoying musical-noise artifacts [79, 117]. Furthermore, the GMR is utilized in order to constrain the suppression and thus, to prevent from overestimation errors. An overestimation of the late reverberation is less likely to happen in spectral regions with a high signal-to-reverberation-ratio, SRR, such as signal steady states [118]. Yet, such problem is more likely to affect the low SRR regions. Therefore a low SRR detector is employed [68] and GMR is applied only on the lower-gain parts. Consequently, the new constrained gain, $G'(m,k)$, is derived as

$$G'(m,k) = \begin{cases} \dfrac{G(m,k)-\theta}{r} + \theta & \text{when } \zeta < \zeta_{th} \text{ and } G(m,k) < \theta \ , \\ G(m,k) & \text{otherwise} \ , \end{cases} \tag{30}$$

and

$$\zeta = \frac{\sum_{k=1}^{K} G(m,k)|Y(m,k)|^2}{\sum_{k=1}^{K} |Y(m,k)|^2} \ , \tag{31}$$

where θ being the threshold for applying the gain constraints, r is the regularization ratio, ζ is the power ratio between the enhanced and the reference signal, ζ_{th} the threshold of the low-SRR detector. K is the total number of frequency bins.

The effect of the GMR is further explained in Fig. 9. In Fig. 9a, a typical illustration of a frequency domain binaural gain is shown. In Fig. 9b–d the effect of the GMR step on the binaural gain is presented for $\theta = 0.2$, $r = 4$, for $\theta = 0.2$, $r = 8$ and for $\theta = 0.6$, $r = 4$ respectively. It can be observed that larger regularization ratios, r, result in larger gain values—in other words, the method suppresses less reverberation. Moreover, a constraint of the reverberation reduction is observed for larger thresholds, θ. By comparing Fig. 9b–d, it becomes clear that the effect of the

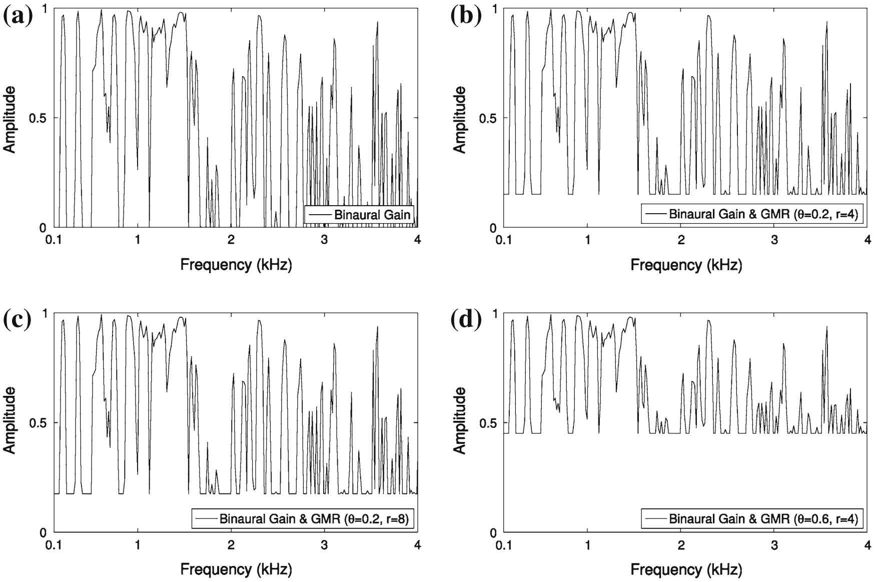

Fig. 9 **a** Typical binaural gain in the frequency domain. **b** Effect of the GMR step for $\theta = 0.2$ and $r = 4$. **c** Effect of the GMR step for $\theta = 0.2$ and $r = 8$. **d** Effect of the GMR step for $\theta = 0.6$ and $r = 4$

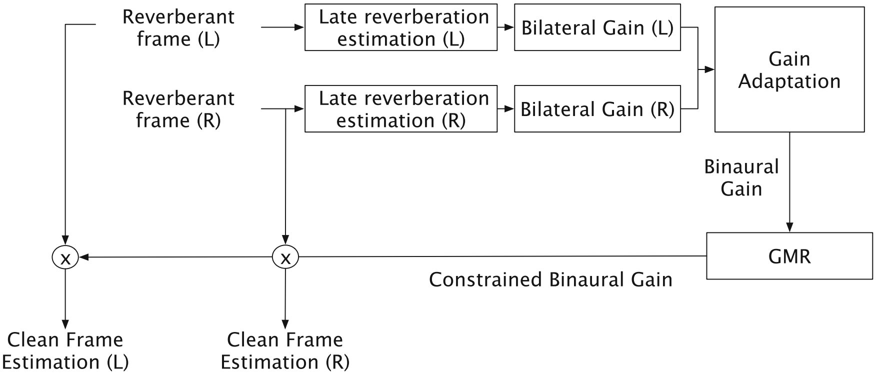
Reverberant frame (L)
Reverberant frame (R)
Late reverberation estimation (L)
Late reverberation estimation (R)
Bilateral Gain (L)
Bilateral Gain (R)
Gain Adaptation
Binaural Gain
GMR
Constrained Binaural Gain
x
x
Clean Frame Estimation (L)
Clean Frame Estimation (R)

Fig. 10 Block diagram of the spectral-subtraction binaural-dereverberation approach that preserves the interaural cues. The late-reverberation estimation can be based on either of the techniques described in Sect. 4.2

regularization ratio, r, of (30) is more subtle than the effect of the threshold, θ. Therefore, the parameter r can be used for fine-tuning purposes.

To conclude this section, Fig. 10 presents a block diagram of the framework applied for binaural spectral-substraction, as described above.

5 Evaluation Methods

5.1 *Objective Measures of Dereverberation*

The evaluation of the potential improvement of speech or audio enhancement techniques has proven to be a rather difficult task. Many objective measures have been developed and often they can predict the perceived quality of the enhanced signals accurately enough [89]. However, there are cases where such objective measures fail to correctly evaluate the performance of a reference algorithm [90]. The evaluation of the performance of dereverberation algorithms has proven to be more difficult. This happens for the same reason that dereverberation is generally more demanding than denoising: the reverberation noise is correlated with the anechoic signal. For the above reasons, denoise measures are not always appropriate for dereverberation evaluation.

This difficulty increases further for the evaluation of binaural dereverberation methods. In this case, apart from the assessment of the output signal's temporal and spectral characteristics, the binaural qualities of the processed output must be also taken into account. Unfortunately, the development of objective or perceptual dereverberation evaluation metrics that explicitly take into account the binaural conditions and cues is still an open research issue. Therefore, single-channel measures are adapted in the binaural scenario by combining through simple addition the left and right-channel results. Hence, one must be very careful when interpreting the values of such metrics.

Most dereverberation-evaluation measures require a-priori knowledge of the anechoic signal. In principle they calculate some type of distance between the dereverberated and the anechoic signal. Such metrics are, for example, the *signal-to-reverberation-ratio*, SRR, the *frequency-weighted signal-to-reverberation-ratio*, fwSRR, the *weighted-slope spectral distance*, WSS, the *Itakura-Saito distance*, IS, the *Bark spectral distortion*, BSD, the *cepstral distance*, CD, and the *log-spectral distortion*, LSD [44, 63, 103, 122]. Moreover, metrics based on auditory modeling have been also used for the evaluation of dereverberation algorithms, such as the *perceptual evaluation of speech quality*, PESQ, the *noise-to-mask ratio*, NMR, the *perceptual-similarity measure*, PSM, and the non-intrusive *speech-to-modulation-energy ratio*, SRMR [35]. In order to improve the overall evaluation performance, some researchers have modified and/or combined subsets of the above metrics, as in [27, 78].

5.2 Perceptual Measures of Dereverberations

When listening inside reverberant spaces, the auditory system applies several mechanisms assisting both intelligibility and localization. These include monaural/binaural decoloration, binaural auditory dereverberation [13, 22, 130] and the precedence effect [88]. Objective measures of dereverberation processing often do not incorporate or take account of these mechanisms or other features of the auditory system. In addition, these objective measures have shown varying correlation with perceptual measures [78, 122]. If signals processed via dereverberation algorithms are intended for human listeners, such a discrepancy needs to be taken into account.

Until now, the literature has only sporadically used perceptual evaluation for evaluation of dereverberation algorithms. In [38] and [118] a mean opinion score, MOS, of signal quality was used. In [68] a preference comparison task was implemented for the different processed signals. A modified version of the ITU P.835 test has been also employed for perceptual evaluation of dereverberation [66]. The test evaluates (i) the speech signal naturalness, (ii) the reverberation intrusiveness and (iii) the overall signal quality [35, 64, 116]. Similar tests have been extended to cover multiple attributes, such as the amount of reverberation, source width and sound envelopment [92].

The *multiple-stimuli-with-hidden-reference-and-anchor test*, MUSHRA, [109] has been also applied for dereverberation [105, 119]. This test is especially successful at detecting small signal impairments, since stimuli are presented simultaneously and evaluated on a scale. For future evaluation of dereverberation algorithms this test can be extended in order to include attributes such as *amount of reverberation* and *overall quality*. One dilemma faced when designing a MUSHRA test are the anchors. Anchors are an inherent trait of MUSHRA experiments to increase the reproducibility of the results and to prevent contraction bias—see [7]. These are normally made by low-pass filtering the reference signal. To evaluate the quality of speech, the anchor should be implemented by introducing distortions similar to those resulting from the dereverberation processing, for example, by using an *adaptive multi-rate*, AMR, speech coder, available from [1], but other distortion types could be also applied. Anchors for judging the amount of reverberation can be created by applying a temporal half cosine to the BRIRs and thereby artificially reducing the resulting reverberation while keeping direct sound and early reflections. Pilot studies have shown that presenting the unprocessed reference stimulus as a hidden anchor resulted in significant compression bias of the listeners—for further details, see [7]. Therefore, this hidden anchor can be omitted and replaced by a separate *reference-button* which allows listeners to hear the unprocessed signal. This test could be employed and combined with reference-processing methods for more reliable results, as shown later in this chapter.

6 Tests and Results

6.1 Results for the Coherence-Based Algorithm

Signal-to-Reverberation Ratio

In this section, the objective results of the coherence-based algorithm (using the processing parameters in Table 1) , described in Sect. 4.1, are presented. The method was compared with the method of Allen et al. [2] and a binaural version of the Lebart et al. [86] spectral subtraction method. Hereinafter the IC-based algorithm of Sect. 4.1 will be referred to as WB, to the Allen et al. method as AB and to the Lebart et al. method as LB. Figure 11 shows the signal-to-reverberation ratio, $\varDelta$segSRR, for the different processing schemes. All algorithms show a significant reduction of the amount of reverberation, as all exhibit positive values.

For the 0.5 m distance—left panel—the WB algorithm for $k_p = 0.2$ provides the best performance. For the lowest degrees of processing, $k_p = 0.35$, the performance is slightly below the one attained for the LB algorithm. For the 5 m distance—right panel—the WB method shows a performance that is comparable to the LB method

Table 1 Processing parameter values for the coherence-based method at $fs = 44.1$ kHz

Parameter	Value
Frame length (L)	6.4 ms
Recursion constant (α)	{0.01; 0.2; 0.35}
Recursion constant (α)	0.97
Gain threshold ($G_{\min}$)	0.1
Mapping updating time (t_{sig})	3 s

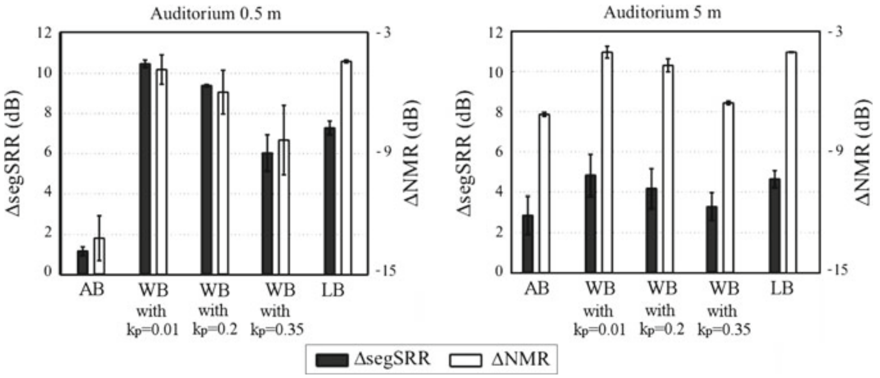

Fig. 11 Estimates of reverberation suppression, *Δ*segSRR, and loss of quality, *Δ*NMR, between the clean signal and the processed reverberant signal for different methods. *Left panel* 0.5 m source-receiver distance. *Right panel* 5 m source-receiver distance

for the highest processing degree, $k_p = 0.01$. As expected, the performance of the WB method generally drops with decreasing processing degree—that is, increasing k_p value. The AB method shows the poorest performance in general and provides essentially no reverberation suppressions in the 0.5 m condition.

Noise-to-Mask Ratio

The *noise-to-mask ratio*, NMR, is an objective measure that determines the audible non-masked noise components. Lower NMR values denote better signal quality [119]. In Fig. 11, also *Δ*NMR is shown, whereby smaller values correspond to less audible noise. For the different processing conditions, the AB approach shows the best quality overall for both source-receiver distances. Considering the very small amount of dereverberation provided by this algorithm—see Fig. 11—this observation is not surprising since the algorithm only has a minimal effect on the signal. The NMR performance of the WB method for high degrees of processing, that is, $k_p = 0.01$, is similar or slightly better than that obtained with the LB approach. The sound quality of the WB method increases with decreasing degree of processing, namely, $k_p = 0.2$ and 0.3. However, at the same time, the strength of dereverberation, as indicated by segSRR, also decreases—see the gray bars in Fig. 11. Considering both measures, segSRR and the NMR, the WB method is superior for close sound sources, in our case the 0.5 m condition with $k_p = 0.2$, and exhibits performance similar to the LB method for the 5 m condition.

Perceptual Evaluation

For the perceptual evaluation of the different dereverberation methods, binaural signals were presented to 10 listeners via headphones. The signals were generated by convolving anechoic sentences with BRIRs, measured in an auditorium with $T_{30} = 1.9$ s. A MUSHRA test was applied to measure (i) strength of dereverberation and (ii) overall loss of quality. As described in Sect. 5.2, an AMR speech coder at 7.95 kbits/sec was used as anchor in the quality measure and a 600 ms long cosine window was applied to the measured BRIRs to generate the anchor for the dereverberation measure. Further details are described in [123]. The results from the perceptual evaluation for each processing method are shown in Fig. 12. For better comparison with the objective results, the measured data were inverted, that is, 100—original score. Considering the strength of dereverberation—indicated by the gray bars—the WB approach exhibited the best performance for $k_p = 0.01$ at both distances. As the degree of processing decreases, that is, for increasing values of k_p, the strength of dereverberation decreases. The improvement relative to the LB approach is considerably higher for the 0.5 m distance—left panel—than for the 5 m distance—right panel. The AB approach of [2] produced the lowest strength of dereverberation for both source-receiver distances. The differences in scores between the AB approach and the others were noticeably larger for the 0.5 m distance than for 5 m. This indicates that, for very close sound sources, the other methods are more efficient than the AB approach.

The overall quality loss of the signals processed with the WB method were found to be substantially higher for the 0.5 m condition compared to the 5 m condition. This difference is not as large with the AB approach as well as the LB method,

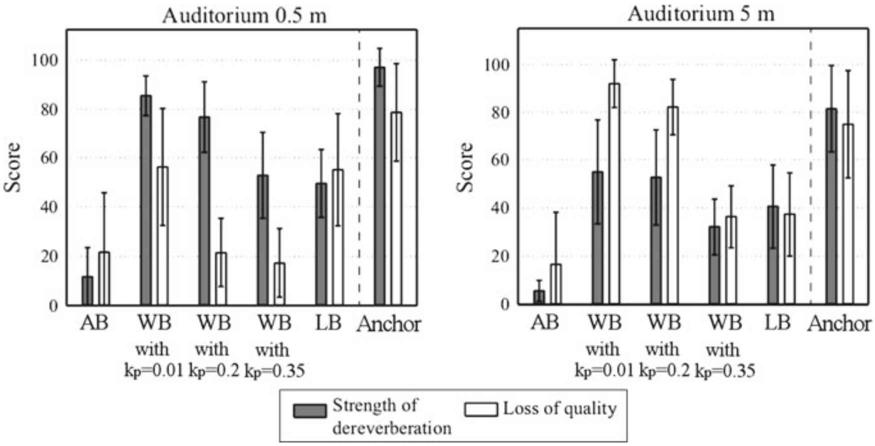

Fig. 12 The mean and standard deviation of perceptual results judging *Strength of dereverberation* and *Overall loss of quality* for the 0.5 m source-receiver distance (*left panel*) and 5 m source-receiver distance (*right panel*)

indicating that the WB is particularly successful for very close sound sources. As in the objective quality evaluation, increasing the degree of dereverberation processing, that is, by decreasing k_p, results in a drop of the overall quality. However, this effect is not as prominent when decreasing k_p from 0.35 to 0.2 at the 0.5 m distance.

Considering the combination of reverberation reduction and overall quality, the WB method with $k_p = 0.2$ exhibits a clearly superior performance at the 0.5 m distance. Even when applying the highest degree of processing, namely, $k_p = 0.01$, the quality is similar to that obtained with LB, but the strength of dereverberation is substantially higher. For the 5 m distance, increasing the degree of processing has a negligible effect on the strength of dereverberation but is detrimental for the quality. However, for $k_p = 0.35$, the performance of the WB method is comparable to that obtained with the LB approach. An analysis of variance, ANOVA, showed significance for the sample effect at source-receiver distances of 0.5 m, namely, $(F = 97.65,\ p < 0.001)$ and 5 m, $(F = 41.31,\ p < 0.001)$. No significant effect of listeners was found.

6.2 Results for the Spectral-Subtraction Framework

In this section, the results of the binaural spectral-subtraction framework are shown—for details see Sect. 4.2. The presented dereverberation methods LB [86], WW [126] and FK [38] are binaural extensions of the original single-channel methods.

A first evaluation has been made for 16 kHz signals [115]. Eight anechoic phrases uttered by both male and female speakers of the TIMIT database were convolved with real BRIRs. Four BRIRs measured in a Stairway Hall with a reverberation time of T_{60}=0.69 s at a source-receiver distance of 3 m and azimuth angles of 0, 30, 60 and 90° were chosen from the Aachen database [68]. In addition, three BRIRs measured in a Cafeteria with a $T_{60} = 1.29$ s at source-receiver distances of 1.18, 1 and 1.62 m and azimuth angles of approximately −30, 0 and 90° were chosen from the Oldenburg database [73]. The authors made informal tests to select optimal values for the analysis parameters, that is, 16 kHz, 16 bit—see Table 2. The θ and ζ_{th} values of the GMR step, described by (30) in Sect. 4.2, were set to 0.15, the regularization ratio, r, was 4. The T_{60} was calculated from the impulse responses. All parameter values that are not detailed here were set according to the values proposed by the authors of the

Table 2 Processing parameter values for comparing the spectral subtraction based methods at fs =16 kHz

Parameter	LB	WW	FK
Total frame length	1024	1024	2048
Zero padding	512	128	128
Frame overlap	0.125	0.25	0.25

original works. In addition, for the FK and LB techniques, two additional relaxation criteria were imposed [118] as they were previously found by the authors to have advantageous effects on the performance. The WW and FK methods assume that an inverse-filtering stage precedes the spectral subtraction implementation. Here, however, the implementation of an 1/3-octave RIR minimum-phase inverse filtering was not found to notably alter the relative improvement achieved by the tested methods. Therefore, a generalized case where the spectral subtraction is applied directly to the reverberant signals is presented.

The produced signals were evaluated by means of the PESQ variation [65], compared to the reverberant signals. PESQ was not originally developed to assess the dereverberation performance—see Sect. 5—and it implements a perceptual model in order to assess the quality of a processed speech signal. Rating is performed according to the five-grade *mean-opinion-score*, MOS, scale. The results are presented in Table 3 with the bold values denoting optimum performance. For the case of the Stairway Hall the bigger PESQ improvement is achieved utilizing the WW method with the minGain adaptation technique. The same gain adaptation technique seems to be also the optimal choice when used in conjunction with the LB method. It can be assumed that in a scenario where bilateral late-reverberation estimations are successful this technique presents superior performance. However, it is not beneficial when used with the FK method where probably the bilateral processing resulted to inferior results. The FK method produces better results when used with the avgGain technique. In general, the WW method shows a significant PESQ improvement for all tested adaptation techniques. For the Cafeteria, the LB method produces a relatively stable PESQ improvement independent of the employed binaural adaptation. On the other hand, better results are derived with the WW method for all binaural-adaptation schemes—although the best results are achieved with the avgGain approach. The FK method seems to produce processing artifacts despite the utilized binaural-adaptation scheme and decreases the PESQ values in every case. Finally, note that the DSB implementation has the advantage of lower computational cost as it involves calculations in a single channel for the estimation of the weighting-gain functions. This is in contrary to the binaural-adaptation schemes that

Table 3 PESQ improvement for the binaural spectral subtraction framework, showing results for various binaural gain adaptation options—see Sect. 4.2

Method	BSD	maxGain	avgGain	minGain
Stairway hall				
LB	0.153	0.142	0.147	0.158
WW	0.206	0.160	0.208	**0.258**
FK	0.160	0.180	0.186	−0.029
Cafeteria				
LB	0.133	0.136	0.135	0.133
WW	0.205	0.208	**0.216**	0.198
FK	−0.235	−0.141	−0.228	−0.428

require dual-channel calculations. On the other hand, the gain-adaptation techniques discussed in Sect. 4.2 involve bilateral processing but do not necessitate the initial time-delay estimation.

For the perceptual test, the methods were applied for broadband signals, sampled at 44100 Hz [116]. A modified version of the ITU P.835 test was used for the perceptual evaluation, as explained in Sect. 5.2. Note that the listeners were not guided to directly rate the binaural qualities of the output signals. However, some listeners reported that they were inherently taken them into account in their assessments. Four phrases from two male and two female speakers along with three BRIRs measured in a Stairway Hall with $T_{60} = 0.69$ s, at a source-receiver distance of 3 m and azimuth angles of $0, 45$ and 90° were used [68]. The original single-channel dereverberation methods, LB, WW and FK, were optimized for lower signal resolutions. Here, the optimal values for their application in broadband signals were extracted through informal listening tests. The STFT analysis parameters, that is, total frame length, zero padding and frame overlap, for each tested method are detailed in Table 4, the θ and ζ_{th} values of the GMR step according to (30) were set at 0.15 and 0.8, respectively, while the regularization ratio r was 4. For the FK and LB techniques, the two additional relaxation criteria proposed in [118] were also implemented. In order to reduce the experimental conditions the authors conducted informal listening tests to choose the optimum gain-adaptation scheme for each dereverberation method. Hence, the avgGain adaptation has been chosen for the LB and WW methods while the maxGain has been used for the FK method. Twenty self-reported normal-hearing listeners participated in the tests and a training session preceded the formal experiment.

Figure 13 presents the perceptual scores in terms of speech naturalness, reverberation reduction and overall signal quality for the proposed binaural dereverberation techniques. The results were subjected to an ANOVA analysis of variance and a highly significant effect for the tested method was revealed for the speech naturalness, namely, $F(3, 228) = 112.7$, $p < 0.001$, for the reverberation reduction, $F(3, 228) = 62.1$, $p < 0.001$ and for the overall quality, $F(3, 228) = 38.8$, $p < 0.001$. No significant effect was found for the tested azimuth angles. Following the ANOVA multiple Tukey's, HSD tests were made to reveal significant differences between algorithms.

In all cases, listeners rated that the unprocessed reverberant signals were significantly more natural than the dereverberated signals—$p < 0.001$. This was due to the artifacts introduced from the dereverberation processing. On the other hand, the FK method performed significantly worse than the other two methods in terms of

Table 4 Analysis parameter values for the employed methods at $fs = 44.1$ kHz

Parameter	LB	WW	FK
Total Frame Length	2048	8192	8192
Zero padding	1024	4096	4096
Frame Overlap	0.5	0.25	0.25

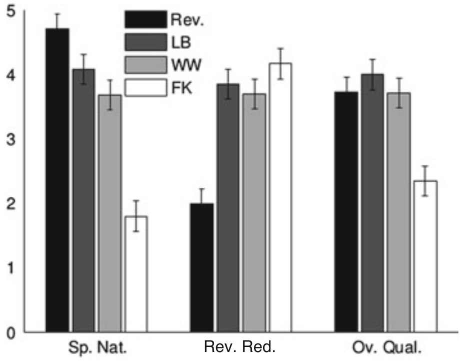

Fig. 13 Perceptual results for speech naturalness, reverberation reduction and overall signal quality for unprocessed signals and the three dereverberation methods tested

speech naturalness. No significant difference was noticed between the LB and WW methods—$p > 0.05$.

Furthermore, the three dereverberation methods have significantly reduced the reverberation—$p < 0.001$. The FK method performed significantly better than the WW method in terms of perceived reverberation suppression—$p < 0.05$. However, no significant difference between the FK and the LB or the LB and WW methods was found—$p > 0.05$. Finally, the LB, the WW methods and the reverberant signals were rated significantly better in terms of overall quality than the FK method—$p < 0.001$—, but no significant quality difference was found between the LB method, the WW method and the reverberant signals—$p > 0.05$.

From the objective and perceptual results it appears that all methods suppress reverberation significantly, but the introduced processing artifacts reduce the naturalness of the speech signals. The FK method achieves greater reverberation suppression than the LB and WW methods; however, it appears that the produced signals are more degraded. On the other hand, the LB and WW dereverberation methods moderately reduce the reverberation, but they preserve the perceptual signal quality.

7 Conclusions

Binaural dereverberation is rapidly evolving as a significant and unique research field having assimilated signal-processing methods and algorithms introduced previously into the broader areas of acoustic signal enhancement and noise suppression but also from more specialized areas such as speech dereverberation, correction of room acoustics, auditory-scene analysis, and from a wealth of perceptual studies and models related to auditory mechanisms. In contrast to some of the mostly single-input-channel methods as were introduced in the signal-enhancement-research fields and were targeted towards machine listening and automatic speech recognition, or were geared towards speech-intelligibility improvements, binaural dereverberation attempts to utilize concepts and optimize processing specifically tailored to binaurally-received signals by human listeners. As is well known, during everyday life, the human auditory system has an impressive ability to analyze, process and select individual source signals from complex acoustic environments, significantly so from signals contaminated by room reflections and reverberation. It has been well established by earlier research that this ability is to a large extend due to auditory and cognitive mechanisms which rely on the binaural signals as these allow the listeners to analyse auditory scenes and suppress unwanted signal components.

Recently, technological and other developments dictate the ever expanding use of portable devices for receiving auditory information via headphones or earpieces and in many such applications binaural dereverberation is an essential preprocessing step in order to ensure reception comparable or better to that of normal listening. Such applications are currently mostly driven by the digital-hearing-aid sector where the problem of reverberation intrusion is prohibitive to speech intelligibility and auditory-scene interpretation. At the same time, the processing capabilities of the commercially available hearing aids allow the real-time implementation of the emerging methods. However, other applications such as binaural telephony, teleconferencing, hands-free devices and interfaces, immersive-audio rendering, and so on, seem to rapidly adopt such binaural dereverberation processing.

Given the extensive capabilities of the hearing system, any binaural dereverberation method is facing significant challenges. Traditionally, most acoustic-signal enhancement methods aim at suppressing some unwanted interference and potentially improve speech intelligibility, and/or audio signal quality. However, the perception of room reverberation is inherently linked to many cues useful to the listener that are generated by room reflections and relate to source position, listener orientation, room size, and further properties that must be retained after processing. Hence, binaural dereverberation methods appear to aim at retaining some useful auditory cues and signal qualitative features, thus compromising their ability to suppress reverberation. From this discussion it is becoming clear that the prominent aims of the binaural methods appear as follows: improving the ratio of direct-to-reverberant signal energy, removing unwanted timbral coloration due to room reflections, removing late-reverberant effects and energy during temporal signal gaps, improving source localization and separation, and suppressing unwanted sound sources and noise. However, after processing, it is imperative that binaural cues necessary for source localization and auditory-scene analysis must be fully preserved or even enhanced. Further, it is sometimes desirable that important cues such as the precedence effect, spectral qualities, as well as room size and envelopment, are retained.

In the previous sections of this chapter it is shown that binaural dereverberation is largely relying on a linear-system model of the room, described via the binaural impulse responses between source(s) and the listener's ears and that often this function is separated into the direct path, early reflection and late-reverberation components. Given that different physical and perceptual effects can be identified due to each of these response components, many dereverberation methods attempt to compensate specifically for effects due to those parts. Another common theme during the evolution of the binaural-dereverberation methods is that some of those methods were based on dual-input-channel processing and others on bilateral adaptation of single channel dereverberation methods, those typically relying on spectral subtraction, a technique which has been widely used for noise suppression. Significantly, all binaural dereverberation must be implemented without any prior measurement of the room response, thus being *blind*, or at least semi-blind, when some broad parameters related to the acoustic environment have to be known.

Starting from those earlier dual-channel dereverberation methods—for example, [2], the significance of the interchannel/interaural coherence function as an indicator of the direct-path-signal contribution in the received signals has been established, leading to techniques that can effectively enhance this signal component in the short-term spectral domain, improving thus the direct-to-reverberant ratio in the processed signal. In this chapter, a detailed presentation of a recent IC based method is given [123, 124]. In contrast to the earlier method [2], where a linear coherence-to-gain function for the spectral modification was proposed, this recent method has introduced a flexible parametric sigmoidal function that can easily be adjusted to the desired filtering-gain form appropriate for specific room-acoustical and source-receiver configurations. Furthermore, the form of the IC estimates and hence the parameters of the filter can be directly obtained from long-term spectral analysis of the received signals. The performance of this IC method was evaluated by testing and comparing the change in the ΔsegSRR, the change in the ΔNMR, as well as via perceptual-evaluation tests. It was found that the method achieved superior overall quality compared to the original IC-based method, as well as to a spectral-subtraction based method. The method was also found to perform better for dereverberating signals from closely located sound sources, for instance, at 0.5 m, instead of those derived from distant sources, for instance, at 5 m. This illustrates that dereverberation gains via IC-based methods depend largely on the degree of the direct-signal energy within the received signals, typically this happening for shorter source-listener distances. For such cases, the overall quality of the processed signal is superior to the results obtained via other processing methods. When this condition is not satisfied, then processing may result in signal-quality loss.

Apart from enhancing the direct signal components, many further single channel dereverberation methods have been developed, based on the concept of spectral subtraction. Although reverberation is a convolutive distortion, late reverberation has usually exponentially decaying white-noise-like properties and, hence, it may be effectively modeled as additive-noise distortion. Traditionally, spectral-subtraction methods suppress such additive noise by subtracting in the short-term spectral domain its estimate and, following this line of thinking, they were also adopted for late-reverberation suppression. For the blind estimation of the late reverberation, essential for deriving the short-term spectral filter gains, a number of established methods exist, originally proposed for speech applications. These methods have either employed an exponentially-decaying noise-like estimate function—compare the LB method, [86]—or an estimate based on smoothed-shifted version of the reverberant speech spectrum—WW method, [126]—or an estimate based on a sum of filtered versions of earlier reverberant speech spectra—FK method, [38].

For adopting such methods for the binaural case, their bilateral application to each signal channel may be followed. Given that the relevant ITD and ILD cues must be preserved in the processed dereverberated signals, identical processing must be applied on the left- and right-ear signals by using appropriately adapted common gains derived from the received signals in each channel, since it is likely that filter-gain estimates may vary for the two different paths to the ears. Such alternative gain-adaptation strategies were studied in [115] and were presented in detail in Sect. 4.2 of this chapter—along with an optimized GMR stage [79, 116] that restricts via a parametric function the extend of spectral modifications during processing. For the perceptual performance tests, described in detail in [115, 116], the PESQ was employed along with perceptual tests based on the MOS and a modified version of the ITU P.835 test. The results indicate that a trade-off between the degree of dereverberation and the perceptual quality of the processed signal, with the estimators derived via the LB and WW methods, achieve the best performance. For these two methods it was also found that average-gain weighting of the individually estimated gains for the left- and right-ear paths, was the best way for adapting these functions to the binaural processing scenario.

In this chapter, an analysis of the concepts involved in human sound reception and perception inside reverberant rooms has been presented, along with a literature review concerning past attempts on the open problems of signal dereverberation. A focus was put on more detailed presentation of two recent blind binaural-dereverberation methods, the first one based on interaural coherence to enhance the direct-to-reverberant ratio and the second one geared towards optimal adaptation of single channel, spectral subtraction based methods for suppressing late reverberation. In both cases, as is the case with most other signal enhancement methods, it was found that reverberation suppression and processed signal quality are two mutually exclusive items that restrict the overall performance of the methods. In comparison with the performance achieved by the human auditory system in similar tasks, such as for the precedence effect, signal decoloration and reverberation suppression, the performance of the current dereverberation methods is clearly inferior, in particular, when being considered under all possible acoustic and source-listener configurations. Furthermore, a unique problem facing all these dereverberation methods is the definition of the desirable results that should be aimed at by such processing. The difficulty of adapting existing objective and perceptual-performance measures and methodologies into the binaural-dereverberation case, especially with respect to retaining the complex cues associated by auditory-scene analysis performed by human listeners inside enclosed spaces, as well as the challenge of dealing with both noise and reverberation likewise, illustrates the complexity and the open issues facing this promising and relatively young research field.

Acknowledgments The authors would like to thank their two anomynous reviewers for substantial and constructive suggestions. This research has been co-financed by the European Union through the European Social Fund, ESF, and the Greek national funds through its Operational Program *Education and Lifelong Learning* of the National Strategic Reference Framework, NSRF. Research Funding Programs: *Thalis-University of Athens-Erasitechnis, grant number: MIS 375435*.

References

1. 3GPP TS26.073. ANSI-C code for the Adaptive Multi Rate (AMR) speech codec. Technical report, 3rd Generation Partnership Project, Valbonne, France, 2008.
2. J. B. Allen, D. A. Berkley, and J. Blauert. Multimicrophone signal-processing technique to remove room reverberation from speech signals. *J. Acoust. Soc. Amer.*, 62:912–915, 1977.
3. J. B. Allen and L. R. Rabiner. A unified approach to short-time fourier analysis and synthesis. In *Proc. of the IEEE*, volume 65, pages 1558–1564, 1977.
4. I. Arweiler and J. M. Buchholz. The influence of spectral and spatial characteristics of early reflections on speech intelligibility. *J. Acoust. Soc. Am.*, 130:996–1005, 2011.
5. C. Avendano and H. Hermansky. Study on the dereverberation of speech based on temporal processing. In *Proc. ICSLP*, 1996.
6. M. Barron and A. Marshall. Spatial impression due to early lateral reflections in concert halls: The derivation of a physical measure. *J. Sound Vibr.*, 77:211–232, 1981.
7. S. Bech and N. Zacharov. *Perceptual Audio Evaluation: Theory, method and application,* pages 39–96. Wiley and Sons Ltd., West Sussex, Great Britain, 2006.
8. D. Bees, M. Blostein, and P. Kabal. Reverberant speech enhancement using cepstral processing. In *Proc. IEEE-ICASSP*, volume 2, pages 977–980, 1991.
9. M. Berouti, R. Schwartz, and J. Makhoul. Enhancement of speech corrupted by acoustic noise. In Proc. *IEEE-ICASSP*, volume 4, pages 208–211, 1979.
10. S. Bharitkar, P. Hilmes, and C. Kyriakakis. Robustness of spatial average equalization: A statistical reverberation model approach. *J. Acoust. Soc. Am.*, 116:3491–3497, 2004.
11. F. A. Bilsen and R. J. Ritsma. Repetition pitch and its implications for hearing theory. *Acustica*, 22:63–73, 1969/70.
12. J. Bitzer, K. U. Simmer, and K.-D. Kammeyer. Theoretical noise reduction limits of the generalized sidelobe canceller (GSC) for speech enhancement. In *Proc. IEEE-ICASSP*, volume 5, pages 2965–2968, 1999.
13. J. Blauert. *Spatial Hearing*. MIT Press, 1997.
14. J. P. Bloom. Evaluation of a dereverberation process by normal and impaired listeners. In *Proc. IEEE-ICASSP*, volume 5, pages 500–503, 1980.
15. P. Bloom and G. Cain. Evaluation of two-input speech dereverberation techniques. In *Proc. IEEE-ICASSP*, volume 7, pages 164–167, 1982.
16. S. Boll. Suppression of acoustic noise in speech using spectral subtraction. *IEEE Trans. Acoust. Speech, Signal Process.*, 27:113–120, 1979.
17. J. S. Bradley, H. Sato, and M. Picard. On the importance of early reflections for speech in rooms. *J. Acoust. Soc. Amer.*, 113:3233–3244, 2003.
18. J. S. Bradley and G. A. Soulodre. The influence of late arriving energy on spatial impression. *J. Acoust. Soc. Am.*, 97:263–271, 1995.
19. E. Brandewie and P. Zahorik. Prior listening in rooms improves speech intelligibility. *J. Acoust. Soc. Am.*, 128:291–299, 2010.
20. A. Bronkhorst and R. Plomp. The effect of head-induced interaural time and level differences on speech intelligility in noise. *J. Acoust. Soc. Am.*, 83:1508–1516, 1988.
21. J. Buchholz and J. Mourjopoulos. A computational auditory masking model based on signal-dependent compression. I. Model description and performance analysis. *Acta Acust./Acustica*, 90:873–886, 2004.
22. J. M. Buchholz. Characterizing the monaural and binaural processes underlying reflection masking. *Hearing Research*, 232:52–66, 2007.
23. J. M. Buchholz. A quantitative evaluation of spectral mechanisms involved in auditory detection of coloration by a single wall reflection. *Hearing Research*, 277:192–203, 2011.
24. J. M. Buchholz, J. Mourjopoulos, and J. Blauert. Room masking: understanding and modelling the masking of room reflections. In *Proc. 110th Conv. Audio Eng. Soc.*, Amsterdam, NL, 2001.
25. O. Cappe. Elimination of the musical noise phenomenon with the Ephraim and Malah noise suppressor. *IEEE Trans. Speech and Audio Process.*, 2:345–349, 1994.

26. L. Danilenko. Binaurales Hören im nichtstationären diffusen Schallfeld (Binaural hearing in a nonstationary, diffuse sound field). *Kybernetik*, 6:50–57, 1969.
27. A. A. de Lima, T. de M. Prego, S. L. Netto, B. Lee, A. Said, R. W. Schafer, T. Kalker, and M. Fozunbal. On the quality-assessment of reverberated speech. *Speech Communication*, 54:393–401, 2012.
28. M. Delcroix, T. Hikichi, and M. Miyoshi. Precise Dereverberation Using Multichannel Linear Prediction. *IEEE Trans. Audio, Speech and Language Process.*, 15:430–440, 2007.
29. J. Deller, J. Hansen, and J. Proakis. *Discrete-time processing of speech signals*. Wiley-IEEE Press, 1999.
30. S. Doclo and M. Moonen. Combined frequency-domain dereverberation and noise reduction technique for multi-microphone speech enhancement. In *Proc. IEEE IWAENC*, pages 31–34, Darmstadt, Germany, 2001.
31. S. Elliott and P. Nelson. Multiple-point equalization in a room using adaptive digital filters. *J. Audio Eng. Soc.*, 37:899–907, 1989.
32. K. Eneman and M. Moonen. Multimicrophone Speech Dereverberation: Experimental Validation. EURASIP *J. Audio Speech and Music Process.*, pages 1–20, 2007.
33. Y. Ephraim and D. Malah. Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator. *IEEE Trans. Acoust., Speech and, Signal Process.*, 32:1109–1121, 1984.
34. J. S. Erkelens and R. Heusdens. Correlation-Based and Model-Based Blind Single-Channel Late-Reverberation Suppression in Noisy Time-Varying Acoustical Environments. *IEEE Trans. Audio, Speech and Language Process.*, 18:1746–1765, 2010.
35. T. Falk, C. Zheng, and W.-Y. Chan. A non-intrusive quality and intelligibility measure of reverberant and dereverberated speech. *IEEE Trans. Audio, Speech and Language Process.*, 18:1766–1774, 2010.
36. C. Faller and J. Merimaa. Source localization in complex listening situations: Selection of binaural cues based on interaural coherence. *J. Acoust. Soc. Am.*, 116:3075–3089, 2004.
37. J. L. Flanagan and R. Lummis. Signal processing to reduce multipath distortion in small rooms. *J. Acoust. Soc. Am.*, 47:1475–1481, 1970.
38. K. Furuya and A. Kataoka. Robust speech dereverberation using multichannel blind deconvolution with spectral subtraction. *IEEE Trans. Audio, Speech and Language Process.*, 15:1571–1579, 2007.
39. S. Gannot and M. Moonen. Subspace Methods for Multimicrophone Speech Dereverberation. EURASIP *J. Advances in, Signal Process.*, pp. 1074–1090, 2003.
40. N. Gaubitch, P. Naylor, and D. Ward. On the use of linear prediction for dereverberation of speech. In *Proc. of the IEEE IWAENC*, pages 99–102, 2003.
41. N. D. Gaubitch and P. A. Naylor. Analysis of the Dereverberation Performance of Microphone Arrays. In *Proc. IEEE-IWAENC*, 2005.
42. E. Georganti, T. May, S. van de Par, and J. Mourjopoulos. Extracting sound-source-distance information from binaural signals. In J. Blauert, editor, *The technology of binaural listening*, chapter 7. Springer, Berlin-Heidelberg-New York NY, 2013.
43. B. W. Gillespie, H. S. Malvar, and D. A. F. Florencio. Speech dereverberation via maximum-kurtosis subband adaptive filtering. In *Proc. IEEE-ICASSP*, volume 6, pages 3701–3704, 2001.
44. S. Goetze, E. Albertin, M. Kallinger, A. Mertins, and K.-D. Kammeyer. Quality assessment for listening-room compensation algorithms. In *Proc. IEEE-ICASSP*, pages 2450–2453, 2010.
45. S. Griebel and M. Brandstein. Wavelet Transform Extrema Clustering For Multi-Channel Speech Dereverberation. In *Proc. of the IEEE-IWAENC*, pages 27–30, 1999.
46. D. Griesinger. The psychoacoustics of apparent source width, spaciousness and envelopment in performance spaces. *Acta Acust./Acustica*, 83:721–731, 1997.
47. E. Habets. *Single- and multi-microphone speech dereverberation using spectral enhancement*. PhD thesis, Technische Univ. Eindhoven, 2007.
48. E. Habets, S. Gannot, and I. Cohen. Late reverberant spectral variance estimation based on a statistical model. *Signal Process. Letters, IEEE*, 16(9):770–773, 2009.

49. V. Hamacher, J. Chalupper, J. Eggers, E. Fischer, U. Kornagel, H. Puder, U. Rass. Signal Processing in High-End Hearing Aids: State of the Art, Challenges, and Future Trends. EURASIP *J. Applied, Signal Process.*, pp. 2915–2929, 2005.
50. Y. Haneda, S. Makino, and Y. Kaneda. Common acoustical pole and zero modeling of Room Transfer Functions. *IEEE Trans. Speech and Audio Process.*, 2:320–328, 1994.
51. V. Hansen and G. Munch. Making recordings for simulation tests in the Archimedes project. *J. Audio Eng. Soc.*, 39:768–774, 1991.
52. A. Härmä, M. Karjalainen, L. Savioja, V. Valimaki, U. Laine, and J. Huopaniemi. Frequency-warped signal processing for audio applications. *J. Audio Eng. Soc.*, 48:1011–1031, 2000.
53. A. Härmä and U. K. Laine. A comparison of warped and conventional linear predictive coding. *IEEE Trans. Speech and Audio Process.*, 9:579–588, 2001.
54. P. Hatziantoniou and J. Mourjopoulos. Generalized fractional-octave smoothing of audio and acoustic responses. *J. Audio Eng. Soc.*, 48:259–280, 2000.
55. P. Hatziantoniou and J. Mourjopoulos. Errors in real-time room acoustics dereverberation. *J. Audio Eng. Soc.*, 52:883–899, 2004.
56. P. Hatziantoniou, J. Mourjopoulos, and J. Worley. Subjective assessments of real-time room dereverberation and loudspeaker equalization. In *Proc. 118th Conv. Audio Eng. Soc.*, 2005.
57. P. D. Hatziantoniou and J. N. Mourjopoulos. Generalized fractional-octave smoothing of audio and acoustic responses. *J. Audio Eng. Soc.*, 48:259–280, 2000.
58. T. Hidaka, Y. Yamada, and T. Nakagawa. A new definition of boundary point between early reflections and late reverberation in room impulse responses. *J. Acoust. Soc. Am.*, 122(1):326–332, 2007.
59. J. Hopgood. *Nonstationary Signal Processing with Application to Reverberation Cancellation in Acoustic Environments*. PhD thesis, University of Cambridge, 2001.
60. J. Hopgood, C. Evers, and J. Bell. Bayesian single channel blind speech dereverberation using Monte Carlo methods. *J. Acoust. Soc. Am.*, 123:3586, 2008.
61. T. Houtgast, H. Steeneken, and R. Plomp. Predicting speech intelligibility in rooms from the modulation transfer function. i. general room acoustics. *Acustica*, 61:60–72, 1980.
62. T. Houtgast and H. J. M. Steeneken. A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria. *J. Acoust. Soc. Am.*, 77:1069–1077, 1984.
63. Y. Hu and P. Loizou. Evaluation of objective quality measures for speech enhancement. *IEEE Trans. Audio, Speech and Language Processing*, 16, 2008.
64. Y. Hu and P. C. Loizou. Subjective comparison and evaluation of speech enhancement algorithms. *Speech Communication*, 49:588–601, 2007.
65. International Telecommunications Union, Geneva, Switzerland. *Perceptual evaluation of speech quality (PESQ), and objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs*, 2000.
66. International Telecommunications Union (ITU-T, P.835), Geneva, Switzerland. *Subjective test methodology for evaluating speech communication systems that include noise suppression algorithm*, 2003.
67. M. Jeub, M. Schaefer, and P. Vary. A Binaural Room Impulse Response Database for the Evaluation of Dereverberation Algorithms. In *Proc. 17th Digital Signal Process. Conf.*, DSP, Santorini, Greece, 2009.
68. M. Jeub, M. Schafer, T. Esch, and P. Vary. Model-based dereverberation preserving binaural cues. *IEEE Trans. Audio, Speech, and Lang. Process.*, 18(7):1732–1745, 2010.
69. M. Jeub and P. Vary. Binaural dereverberation based on a dual-channel Wiener filter with optimized noise field coherence. In *Proc. IEEE-ICASSP*, pages 4710–4713, 2010.
70. M. Kallinger and A. Mertins. Impulse response shortening for acoustic listening room compensation. In *Proc. IEEE-IWAENC*, 2005.
71. M. Karjalainen and T. Paatero. Equalization of loudspeaker and room responses using Kautz filters: direct least squares design. EURASIP *J. Appl. Signal Process.*, 2007.
72. J. M. Kates. *Digital Hearing Aids*, pages 221–262. Plural Publishing, San Diego, CA, USA, 2008.

73. H. Kayser, S. D. Ewert, J. Anemuller, T. Rohdenburg, V. Hohmann, and B. Kollmeier. Database of Multichannel In-Ear and Behind-the-Ear Head-Related and Binaural Room Impulse Responses. EURASIP *J. Appl. Signal Process.*, 2009:1–10, 2009.
74. K. Kinoshita, M. Delcroix, T. Nakatani, and M. Miyoshi. Suppression of Late Reverberation Effect on Speech Signal Using Long-Term Multiple-step Linear Prediction. *IEEE Trans. Audio, Speech and Language Process.*, 17:534–545, 2009.
75. O. Kirkeby and P. Nelson. Digital filter design for inversion problems in sound reproduction. *J. Audio Eng. Soc.*, 47:583–595, 1999.
76. C. Knapp and G. Carter. The generalized correlation method for estimation of time delay. *IEEE Trans. Acoust., Speech and, Signal Process.*, ASSP-24:320–327, 1976.
77. A. Koening, J. Allen, D. Berkley, and C. T. Determination of masking level differences in a reverberant environment. *J. Acoust. Soc. Am.*, 61:1374–1376, 1977.
78. K. Kokkinakis and P. C. Loizou. Evaluation of objective measures for quality assessment of reverberant speech. In *Proc. IEEE-ICASSP*, pages 2420–2423, 2011.
79. E. Kokkinis, A. Tsilfidis, E. Georganti, and J. Mourjopoulos. Joint noise and reverberation suppression for speech applications. In *Proc. 130th Conv. Audio Eng. Soc.*, London, UK, 2011.
80. B. Köllmeier, J. Peissig, and V. Hohmann. Binaural noise-reduction hearing aid scheme with real-time processing in the frequency domain. *Scandinavian Audiol. Suppl.*, 38:28–38, 1993.
81. P. Krishnamoorthy and S. R. Mahadeva Prasanna. Reverberant speech enhancement by temporal and spectral processing. *IEEE Trans. Audio, Speech and Language Process.*, 17:253–266, 2009.
82. D. Kundur and D. Hatzinakos. Blind image deconvolution. *IEEE Signal Process. Mag.*, 13:43–64, 1996.
83. H. Kuttruff. *Room acoustics, 4th Edition*. Taylor & Francis, 2000.
84. T. Langhans and H. Strube. Speech enhancement by nonlinear multiband envelope filtering. In *Proc. IEEE-ICASSP*, 1982.
85. K. Lebart, J. Boucher, and P. Denbigh. A binaural system for the suppression of late reverberation. In *Proc. European Signal Process. Conf.*, pages 1–4, 1998.
86. K. Lebart, J. Boucher, and P. Denbigh. A new method based on spectral subtraction for speech dereverberation. *Acta Acust./Acustica*, 87:359–366, 2001.
87. J. H. Lee, S. H. Oh, and S. Y. Lee. Binaural semi-blind dereverberation of noisy convoluted speech signals. *Neurocomputing*, 72:636–642, 2008.
88. R. Y. Litovsky, S. H. Colburn, W. A. Yost, and S. J. Guzman. The precedence effect. *J. Acoust. Soc. Amer.*, 106:1633–1654, 1999.
89. P. Loizou. *Speech enhancement: theory and practice*. CRC Press, 1st edition, 2007.
90. P. Loizou and G. Kim. Reasons why current speech-enhancement algorithms do not improve speech intelligibility and suggested solutions. *IEEE Trans. Audio, Speech, and Lang. Process.*, 19:47–56, 2011.
91. H. W. Löllmann and P. Vary. Low delay noise reduction and dereverberation for hearing aids. EURASIP *J. Advances in Signal Process.*, 2009:1–9, 2009.
92. G. Lorho. *Perceived Quality Evaluation - An application to Sound reproduction over headphones*. PhD thesis, Aalto University, School of Science and Technology, Department of Signal Processing and Acoustics, Finland, 2010.
93. I. McCowan and H. Bourlard. Microphone array post-filter based on noise field coherence. *IEEE Trans. Speech and Audio Process.*, 11:709–716, 2003.
94. A. Mertins, T. Mei, M. Kallinger. Room impulse response shortening/reshaping with infinity- and p-norm optimization. *IEEE Trans. in Acoust., Speech and, Signal Process.*, 18:249–259, 2010.
95. O. M. M. Mitchell and D. A. Berkley. Reduction of long time reverberation by a center clipping process. *J. Acoust. Soc. Am.*, 47:84, 1970.
96. M. Miyoshi and Y. Kaneda. Inverse filtering of room acoustics. *IEEE Trans. Acoustics, Speech and, Signal Processing*, 36:145–152, 1988.

97. J. Mourjopoulos. On the variation and invertibility of room impulse response functions. *J. Sound and Vibr.*, 102:217–228, 1985.
98. J. Mourjopoulos. Digital equalization methods for audio systems. In *Proc. 84th Conv. Audio Eng. Soc.*, 1988.
99. J. Mourjopoulos. Digital equalization of room acoustics. *J. Audio Eng. Soc.*, 42:884–900, 1994.
100. J. Mourjopoulos, P. Clarkson, and J. Hammond. A comparative study of least-squares and homomorphic techniques for the inversion of mixed phase signals. In *Proc. IEEE-ICASSP*, 1982.
101. J. Mourjopoulos and J. K. Hammond. Modelling and enhancement of reverberant speech using an envelope convolution method. In *Proc. IEEE-ICASSP*, 1983.
102. P. A. Naylor and N. D. Gaubitch. *Speech Dereverberation*, pages 57–387. Springer, London, Great Britain, 2010.
103. P. A. Naylor, N. D. Gaubitch, and E. A. P. Habets. Signal-based performance evaluation of dereverberation algorithms. *J. Electrical and Computer Eng.*, 2010:1–5, 2010.
104. S. Neely and J. B. Allen. Invertibility of room impulse response. *J. Acoust. Soc. Am.*, 66:165–169, 1979.
105. S. Norcross, G. Soulodre, and M. Lavoie. Subjective investigations of inverse filtering. *J. Audio Eng. Soc.*, 52:1003–1028, 2004.
106. A. V. Oppenheim. *Applications of digital signal processing*. Prentice-Hall, 1978.
107. T. Paatero and M. Karjalainen. Kautz Filters and Generalized Frequency Resolution: Theory and Audio Applications. *J. Audio Eng. Soc.*, 51:27–44, 2003.
108. A. P. Petropulu and S. Subramaniam. Cepstrum based deconvolution for speech dereverberation. In *Proc. IEEE-ICASSP*, pages I/9-I12, 1994.
109. I. RBS.1534-2001: Method for the subjective assessment of intermediate quality levels of coding systems. 2003.
110. W. C. Sabine. *Collected Papers on Acoustics*. Peninsula Publishing, Los Altos, 1993.
111. L. Savioja and V. Valimaki. Multiwarping for enhancing the frequency accuracy of digital waveguide mesh simulations. *IEEE Signal Process. Letters*, 8:134–136, 2001.
112. K. Simmer, S. Fischer, and A. Wasiljeff. Suppression of coherent and incoherent noise using a microphone array. Ann. *Telecommunications*, 49:439–446, 1994.
113. T. Stockham, T. M. Cannon, and R. B. Ingebretsen. Blind deconvolution through digital signal processing. In *Proc. IEEE*, volume 63, pages 678–692, 1975.
114. M. Triki and D. T. M. Slock. Delay and Predict Equalization for Blind Speech Dereverberation. In *Proc. IEEE-ICASSP*, volume 5, 2006.
115. A. Tsilfidis, E. Georganti, and J. Mourjopoulos. Binaural extension and performance of single-channel spectral subtraction dereverberation algorithms. In *Proc. IEEE-ICASSP*, Prague, Czech Republic, 2011.
116. A. Tsilfidis, E. Georganti, and J. Mourjopoulos. A binaural framework for spectral subtraction dereverberation. In *Forum Acusticum 2011*, Aalborg, Denmark, 2011.
117. A. Tsilfidis, K. E. Kokkinis, and J. Mourjopoulos. Suppression of late reverberation at multiple speaker positions utilizing a single impulse response measurement. In *Forum Acusticum 2011*, Aalborg, Denmark, 2011.
118. A. Tsilfidis and J. Mourjopoulos. Signal-dependent constraints for perceptually motivated suppression of late reverberation. *Signal Processing*, 90:959–965, 2010.
119. A. Tsilfidis and J. Mourjopoulos. Blind single-channel suppression of late reverberation based on perceptual reverberation modeling. *J. Acoust. Soc. Am.*, 129:1439–1451, 2011.
120. D. Tsoukalas, J. Mourjopoulos, and G. Kokkinakis. Speech enhancement based on audible noise suppression. *IEEE Trans. Speech and Audio Process.*, 5:497–513, 1997.
121. H. L. Van Trees. *Optimum array processing*, volume 4. Wiley-Interscience. New York, NY, USA, 2002.
122. J. Wen, N. Gaubitch, E. Habets, T. Myatt, and P. Naylor. Evaluation of speech dereverberation algorithms using the MARDY database. In *Proc. IEEE-IWAENC*, 2006.

123. A. Westermann, J. Buchholz, and T. Dau. Using long-term coherence estimates for binaural dereverberation. In *Forum Acusticum 2011*, Aalborg, Denmark, 2011.
124. A. Westermann, J. Buchholz, and T. Dau. Binaural dereverberation based on interaural coherence histograms. *J. Acoust. Soc. Am.*, xxx:in press, 2013.
125. T. Wittkop and V. Hohmann. Strategy-selective noise reduction for binaural digital hearing aids. *Speech Communication*, 39:111–138, 2003.
126. M. Wu and D. Wang. A two-stage algorithm for one-microphone reverberant speech enhancement. *IEEE Trans. Audio, Speech and Language Process.*, 14:774–784, 2006.
127. B. Yegnanarayana and P. S. Murthy. Enhancement of reverberant speech using LP residual signal. *IEEE Trans. Audio, Speech and Language Process.*, 8:267–281, 2000.
128. P. Zahorik. Auditory distance perception in humans: A summary of past and present research. *Acta Acust./Acustica*, 91:409–420, 2005.
129. W. Zhang, E. Habets, and P. Naylor. On the use of channel shortening in multichannel acoustic system equalization. In *Proc. of the IEEE IWAENC*, 2010.
130. P. M. Zurek. Measurements of binaural echo suppression. *J. Acoust. Soc. Amer.*, 66:1750–1757, 1979.