On the Combination of Systems for Listening-Room Compensation and Acoustic Echo Cancellation in Hands-Free Teleconference Systems

PhD-Thesis

to obtain the academic degree

Doktor der Ingenieurwissenschaften (Dr.-Ing.)

University of Bremen

(Dept. of Communications Engineering)

by

Dipl.-Ing. Stefan Goetze

Day of public colloquium:	16. Januar 2013
Reviewer of this PhD-thesis	Prof. Dr.-Ing. K.D. Kammeyer
	Prof. Dr.-Ing. A. Mertins
Further examinors:	Prof. Dr. phil. nat. D. Silber
	Prof. Dr.-Ing. M. Schneider

Bremen, July 2013

Preface

This thesis is the result of my activity as a research assistant at the Dept. of Communications Engineering, Institute for Telecommunications and High-Frequency Technology at University of Bremen, Germany.
It always has been a pleasure - even during numerous long nights - to work for and within this research group. For this always positive environment I would like to thank all my colleagues and particularly Prof. Dr.-Ing. Karl-Dirk Kammeyer who provided a maximum of freedom and support which allowed me to develop my knowledge and skills within the research community. Furthermore, I am very grateful to Prof. Dr.-Ing. Alfred Mertins for acting as my second supervisor and for all the fruitful discussions which always provided helpful ideas, especially when I struggled with the mathematics. I furthermore thank Prof. Dr. phil. nat. Dieter Silber and Prof. Dr.-Ing. Martin Schneider very much for reviewing this thesis.
I would like to express special thanks for an excellent time at the University of Bremen to my former office mates Dr.-Ing. Volker Mildner, Dr.-Ing. Mark Petermann and Dr.-Ing. Peter Klenner for sharing large portions of time with me and making my working and free time in Bremen very enjoyable.
I would like to thank Dr.-Ing. Mark Petermann, Prof. Dr.-Ing. Markus Kallinger, M.Sc. Feifei Xiong, Dr. rer. nat. Anna Warzybok and M.Sc. Ina Kodrasi for various fruitful, interesting or diverting discussions as well as for final proof reading of this thesis.
Likewise, I want to thank Deutsche Forschungsgemeinschaft (DFG) for financially supporting large parts of this work.
Special thanks go to my parents who always helped and supported me on all my ways!

Bremen, July 2013

Stefan Goetze

Contents

Chapter 1

Introduction

1.1 Overview

The most natural form of human communication is speech. During the last decades an increasing demand for natural and comfortable speech communication over long distances can be observed and hands-free telecommunication setups are widely used nowadays. Examples for such systems (without claim for completeness) are video-conferencing systems, hands-free front-ends in cars, information terminals, e.g. at railway stations, airports, or public places, smart homes, messaging-software like ICQ© or Skype©, or computer software and computer games with sound output and speech input. Hands-free systems can be used to increase security while driving a vehicle or to increase communication comfort in teleconference situations. By this, the user of a hands-free telecommunication system may use both hands for other tasks and/or can move freely in a room. Furthermore, the use of a hands-free setup for communication while driving a car is required mandatorily, e.g. by German law since 2001.

If the usual handset of a telephone is replaced by one or more loudspeakers and microphones, several problems occur for the digital signal processing unit of a hands-free system, that will be described in the following. A scheme of a hands-free telephony situation is shown in **Figure 1.1**. Here, $s_n[k]$ is the near-end speaker's signal that has to be processed and transmitted to the far-end side unaffectedly. It is the desired signal for beamforming noise reduction schemes but also a disturbance, e.g. for adaptive algorithms for acoustic echo cancellation. Ambient noise is denoted by $n[k]$ and $\psi[k]$ is the acoustic echo due to the acoustic coupling between microphones and loudspeakers. The acoustic coupling can be described by the so-called room

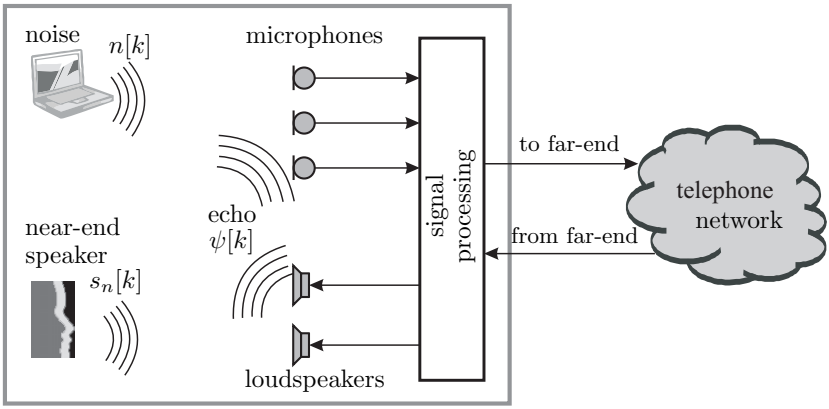

Figure 1.1: Multi-channel hands-free telephony setup.

impulse responses (RIRs) (cf. Section 2.1).

Figure 1.2 shows a more detailed schematic of a common setup for hands-free tele-communication. It contains several sub-systems for the specific problems such systems have to tackle. The signal of the far-end speaker $s_f[k]$ is picked up by one or more microphones in the far-end room, transmitted to the near-end room, radiated by the loudspeakers, and picked up again by the microphones due to the acoustic coupling, expressed here by the room impulse response(s) $h_{\mathrm{AEC}}[k]$. The desired sound source for the microphone array in the near-end room is the near-end speaker $s_n[k]$. Its signal is superimposed by the noise disturbances $n[k]$ and the acoustic echo $\psi[k]$.

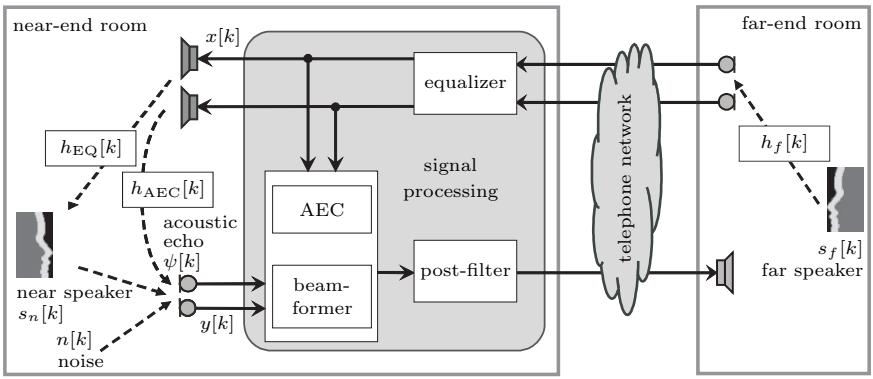

Figure 1.2: Signal processing in a hands-free setup.

The setup in Figure 1.2 leads to several problems that will be described in the following. They have to be solved in combination by the signal processing system, containing an equalizer for listening-room compensation (dereverberation) of the loudspeaker signal, an acoustic echo canceller (AEC), a noise reduction subsystem (beamformer), and a post-filter which may perform dereverberation of the near-end speaker's signal, suppression of residual echoes, and residual noise, bandwidth reduction, and signal coding. Please note, that this thesis focuses on the problems of dereverberation by means of pre-equalization of the signal and on the system identification by means of an AEC which is needed for that purpose. This section tries to briefly provide a general overview about the problems in hands-free communication, without any claim for completeness, to embed the topics discussed into a general framework (cf. also Section 1.2).

- **Suppression of ambient noise:** Hands-free communication leads to a drastically reduced signal-to-noise ratio (SNR) at the microphone compared to the use of a hand-set. The desired signal which is the near-end speaker $s_n[k]$ is superimposed by disturbances in the near-end room, such as ambient noise, PC-fans, air conditioning, thermal noise, etc. All ambient noise sources which may occur in a hands-free setup besides the echo signal will be denoted by $n[k]$. A too low volume control of the hands-free system further decreases the SNR. High noise levels lead to disturbances for adaptive algorithms which are used for the acoustic echo canceller. Furthermore, noisy signals that are presented to the user of the hands-free system result in a lower speech intelligibility and cause tiredness of the system user. Since speech and noise signals may overlap in time as well as in frequency the separation of the desired signal from the disturbance without distorting the desired signal may be difficult, especially if only one microphone is available.

- **Cancellation of acoustic echoes:** The speech signal of the far-end user, which is radiated by the loudspeakers in the near-end room, is picked up again by the microphones and is transmitted back to the far-end user. He or she perceives his or her own voice as an echo delayed by the round-trip delay of the system which may be up to several hundred milliseconds. This is very annoying and drastically disturbs natural communication.
  Although noise reduction schemes are able to reduce acoustic echoes since they treat echoes as disturbance, acoustic echo cancellation is preferable whenever a reference signal (in this case the loudspeaker signal) is available, since the echo canceller allows for maximum echo

suppression of this kind of *interferers*. Theoretically, AECs are able to cancel the acoustic echo without introducing distortions to the near-end speaker's signal. Since the first proposal by Sondhi [Son67] for compensation of acoustic echoes based on modelling the room impulse response by a digital linear filter much research has been done on single-channel acoustic echo cancellation as well as the extension to the multi-channel case. A detailed discussion of the challenges of AEC and some possible solutions will be given in Chapter 3.

- **Dereverberation:** In common hands-free environments signals not only travel directly from source to microphone but are reflected at the room boundaries numerous times. This causes reverberation of the sound signal which reduces speech intelligibility. Reverberation effects get more and more perceivable in larger rooms which are characterized by high room reverberation times as it is experienced while listening to speech in churches or large halls. Dereverberation schemes can target two different signals in Figure 1.2. Either reverberation can be removed from the microphone signal that contains a reverberant version of the near-end speaker's signal $s_n[k]$ before it is transmitted to the far-end user or reverberation of the loudspeaker signal in the near-end room caused by the RIR $h_{\mathrm{EQ}}[k]$ between loudspeaker and near-end listener can be reduced. The latter method is known as listening-room compensation (LRC) and will be one of the major foci of this thesis. A brief discussion of the influence of reverberation is given in Chapter 2. Chapters 4 and 5 address the problem of dereverberation.

- **Sound source coding:** Especially for high-quality multi-channel hands-free systems an enormous amount of data has to be transmitted to the far-end user. This problem gets even more severe if video is sent in addition to the audio signal or if wireless networks have to be used. Since the required bandwidth for a direct transmission may not be available or too costly, the amount of data to be transmitted should be reduced by an appropriate coding scheme. Fortunately, powerful methods exist for coding of speech signals as well as arbitrary audio signals such as linear predictive coding (LPC) for speech signals or the widely used MP3 and AAC coding schemes which were developed for coding of music signals. Please note that source coding will not be a topic of this thesis.

1.2 Focus, Outline and Main Contributions

As described above, numerous problems have to be solved for hands-free communication systems addressing the subsystems individually or the combination of one or more subsystems. Although research results could also be published as well in the fields of noise reduction [GMK06b, MGK06c, MGK06b, GMK06a, MGK06a, MGKM07, RGH+08a, RGA09, RGA11, RAGA10, GXR+10, GRA10, GMA+10, MGA11, CGD12, RG12, RGB+12, SCS+13, MSA+13, SMS+13], sound position estimation [GRH+08, RGH+08b, GGBD11, GGD12, RGB+12, GSG+12], or voice activity detection [HSGA10, WGH+11, RGH+11], the remainder of this thesis will focus on listening-room compensation, acoustic echo cancellation and the mutual influences of LRC and AEC only. The following chapters are organized as follows:

In Chapter 2 some fundamentals on room acoustics are introduced. Section 2.1 particularly describes the properties of RIRs which are needed for the following discussions about identification and equalization of RIRs and the respective problems for AEC and LRC algorithms. Furthermore, the multi-delay filtering structure known from literature [MAG95] is introduced in Section 2.2 in vector/matrix notation which will be used throughout this thesis.

Chapter 3 introduces the basic principles of acoustic echo cancellation that are mainly used for the system identification needed by the LRC sub-systems in this thesis [GKK05, GKMK06b, RGA09, RGA11, GRA10]. Main contributions in this chapter are the system identification for equalized systems by means of proportionate update schemes (cf. Section 3.2.2) [GXJ+11], conventional AECs [GKMK08d] and post-filters (cf. Section 3.3) [GKK05, GKMK06b, XAG12] for system identification.

Chapter 4 introduces the signal processing strategies for LRC. Main contributions in this chapter are the analysis of various objective quality measures for LRC which was lacking in the literature [GAR+10b, ARG10, GAK+10, SGR+11, BRX+12, GAR+14] (cf. Section 4.2 and Appendix A to C), analysis of LRC robustness for different algorithms [GKKM07, GKMK08d, GKMK09, JGM11, JMGM11, KGD12a, KGD12b, XGM13, KGD13a, KGD13b] and the development of a new type of gradient algorithm for LRC [GKMK08a, GKMK08b] (cf. Section 4.5) which converges quickly and is computationally efficient.

Chapter 5 discusses different possibilities for combinations of subsystems for AEC and LRC and the respective mutual influences of these subsystems [GKMK06a, GKMK07]. Main contributions in this chapter are the system identification and the influences on the LRC approaches

[GKMK08c, GKMK08d] and the identification of equalized impulse responses [GXJ+11], as well as a method to increase LRC robustness based on the knowledge of the AEC convergence state [GKMK08b] (cf. Section 5.1).

1.3 Notation

To distinguish between scalar values, vectors, and matrices in time and in frequency-domain the following notation is determined, which is summarized in **Table 1.1**.

Vectors are written in bold letters to distinguish them from scalars. Matrices are written in bold uppercase letters. Time-domain variables are written *italic* while frequency-domain variables are written in $\mathsf{sans-serif}$ letters, to allow the reader to distinguish between time-domain and frequency-domain even if the dependence on time or frequency is omitted, e.g. for readability reasons in long formulas. Thus $y = \boldsymbol{x}^T \boldsymbol{h}$ clearly indicates a multiplication of two vectors in time-domain while $\mathsf{y} = \mathbf{x}^T \mathbf{h}$ indicates the multiplication of the corresponding vectors in frequency-domain.

Furthermore, the discrete time and frequency-domain can be distinguished from the continuous domains by the use of squared brackets, e.g. $x[k], \mathsf{x}[n]$, instead of round parentheses, e.g. $x(t)$ and $\mathsf{x}(e^{j\Omega})$. Here k, n, t and $e^{j\Omega}$ are the arguments for the discrete time, the discrete frequency, the continuous time, and the continuous frequency, respectively. The index ℓ will be used for the block-time throughout the work, if block processing is used.

Table 1.1 summarizes the conventions given above.

	Continuous time-domain	Discrete time-domain	Continuous freq.-domain	Discrete freq.-domain
Scalar	$x(t), \xi(t),$	$x[k], \xi[k],$	$\mathsf{x}(e^{j\Omega}), \mathsf{\xi}(e^{j\Omega}),$	$\mathsf{x}[n], \mathsf{\xi}[n],$
Vector	$\boldsymbol{x}(t), \boldsymbol{\xi}(t),$	$\boldsymbol{x}[k], \boldsymbol{\xi}[k],$	$\mathbf{x}(e^{j\Omega}), \mathbf{\xi}(e^{j\Omega}),$	$\mathbf{x}[n], \mathbf{\xi}[n],$
Matrix	$\boldsymbol{X}(t), \boldsymbol{\Xi}(t),$	$\boldsymbol{X}[k], \boldsymbol{\Xi}[k],$	$\mathbf{X}(e^{j\Omega}), \mathbf{\Xi}(e^{j\Omega}),$	$\mathbf{X}[n], \mathbf{\Xi}[n],$

Table 1.1: Definitions for scalars, vectors and matrices.

Typically, the length of a vector is denoted by L with a sub-index indicating the vector, e.g. L_x denotes the length of the vector $\boldsymbol{x}$. If vector notation is chosen, vectors having time-varying coefficients (like adaptive filters) can be distinguished from vectors with time-constant coefficients by a successive

$[k]$ as for

$$\boldsymbol{h}[k] = [h_0[k], h_1[k], h_2[k], \cdots, h_{L_h-1}[k]]^T \tag{1.3.1}$$

instead of

$$\boldsymbol{h} = [h_0, h_1, h_2, \cdots, h_{L_h-1}]^T . \tag{1.3.2}$$

The superscripts $(\cdot)^T$, $(\cdot)^*$, $(\cdot)^H$, and $(\cdot)^+$ denote the transposition, the conjugate, the Hermitian transpose, and the Moore-Penrose pseudoinverse, respectively. The operator $*$ denotes the convolution of two sequences, $\mathrm{E}\{\cdot\}$ is the expectation operator, and the operator $\mathrm{convmtx}\{\boldsymbol{h}, L_{\mathrm{EQ}}\}$ generates a convolution matrix of size $(L_{\mathrm{EQ}} + L_h - 1) \times L_{\mathrm{EQ}}$ (cf. (4.2.7)). The operator $\mathrm{diag}\{\cdot\}$ builds up a matrix of size $L \times L$ from a vector of size $L \times 1$ that has the vector's elements on its main diagonal and zeros elsewhere, and the operator $\mathrm{bdiag}\{\cdot\}$ generates a matrix of size $L'L \times L'L$ having matrices of size $L \times L$ on its block diagonal and zeros else (cf. (2.2.12)).

Chapter 2

Basics

2.1 Fundamentals of Room Acoustics

In this chapter some background information about room acoustics will be given that will be basis of the following chapters. Particularly, room impulse responses are discussed briefly and some of their properties are described that are needed for a deeper understanding of the problems to be solved in successive chapters. More detailed discussion about room acoustics can be found in the literature, cf. e.g. [Kut00, Bor89] and references therein.
A sound signal that is played back via headphones sounds differently than the same signal played back in an acoustic environment, such as a car, a concert hall, or a church. Characterizing for the sound is, amongst other properties, room volume and shape, reflection at the surfaces, or the distance between sound source and listener. A sound wave hitting a surface can be reflected, absorbed or transmitted as it is depicted in **Figure 2.1(a)**. Of course, partly reflections are possible and most common. The ratio between reflection and transmission/absorption is defined by the so-called reflection coefficient.
A further possibility is the so-called diffusion as depicted in the bottom of Figure 2.1 (b). Here, the sound energy of the impinging wave is not reflected to one certain direction but scattered to numerous directions. This can be archived, e.g., by book shelves in a common living room.

2.1.1 Room Impulse Responses

In room acoustics the room impulse response plays an important role, since it characterizes the sound propagation from one spatial position

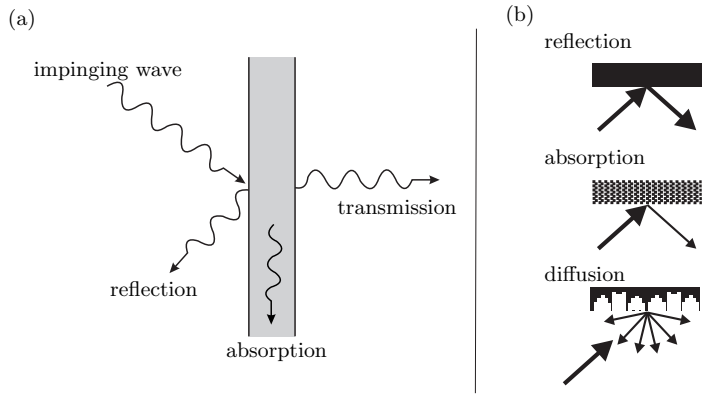

Figure 2.1: Reflection, absorption and diffusion of sound sources.

$\boldsymbol{s} = [s_x, s_y, s_z]^T$ to another position $\boldsymbol{p} = [p_x, p_y, p_z]^T$. Here, $\boldsymbol{s}$ and $\boldsymbol{p}$ may be the 3-dimensional coordinates of a sound source and a microphone. If a room impulse response has to be measured [RV89a, Van94, MM01, PS01] or identified by means of an adaptive filter [HBC06, GKMK08d], the influence of loudspeaker and microphone is always contained in the measurement result. This so-called loudspeaker room microphone (LRM) system is depicted in **Figure 2.2 (a)**.

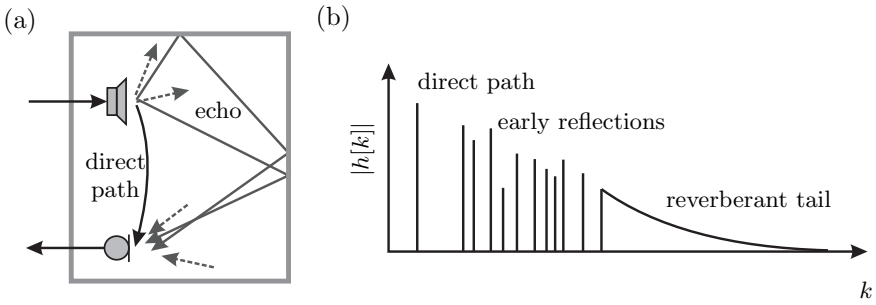

Figure 2.2: (a) LRM system, acoustic coupling between loudspeaker and microphone. (b) Room impulse response (schematically).

Please note, that the influence of the loudspeaker and the microphone can often be neglected compared to the influence of the room. Thus, the term RIR or acoustic environment will always describe such a LRM system and these terms are used as synonyms in the following.
Typical room impulse responses have a similar structure as schematically

shown in Figure 2.2 (b). The first peak of the RIR corresponds to the direct path between source and microphone. The next peaks correspond to early reflections of the sound wave at the room boundaries. The direct path and the early reflections are important for sound source localizations and the spectral characteristic of the room [SL61, Hal01, WN07]. Since more and more reflections overlap, arriving at the microphone with decaying energy, the later part of the RIR is called reverberant tail. This tail has a stochastic nature and is linked to the perception of reverberation. Thus, the longer the late tail, the more reverberation is perceived in a specific room.
The following subsections briefly introduce some basic properties of RIRs that will be needed in the successive chapters of this thesis to understand the problems of AEC, LRC and quality assessment of dereverberation algorithms.

2.1.2 Time- and Frequency-Domain Properties of Room Impulse Responses

The left panels of **Figure 2.3** show two typical RIRs for different room reverberation times ($\tau_{60} = 50$ ms and $\tau_{60} = 500$ ms, cf. also Section 2.1.4 for a discussion of the reverberation time τ_{60}) scaled (a) linearly and (b) logarithmically. Panels (c) and (d) of Figure 2.3 show the corresponding frequency-domain representations (room transfer functions) linearly in panel (c) and logarithmically in panel (d). It can be seen from the room transfer function $|\mathsf{h}(f)|$ in the right panels of Figure 2.3 that, in general, all frequencies are transmitted by the room (as one could expect). The mean transfer function equals one (or 0 dB). However, strong fluctuations are visible especially for higher room reverberation times. This is due to the effect of destructive interference of different sound propagation paths. As more clearly visible in Figure 2.3 (d) the room transfer function is characterized by numerous notches that are caused by numerous zeros very close to the unit circle in the z-domain (cf. Section 2.1.6). The higher the room reverberation time is, the higher is the density of the notches.
Since listening-room compensation aims at equalization of room impulse responses, every notch in frequency-domain has to be compensated for by a peak in the corresponding transfer function of an equalization filter. It is obvious that an inaccurate compensation of the notches of a room transfer function (RTF), e.g. caused by a frequency shift, may lead to severe distortions (cf. Section 4.4.2 for a more detailed discussion of robustness issues).
If a speech signal $s(t)$ is transmitted from one position of a room to another in a hands-free situation, reverberation is added to the signal due

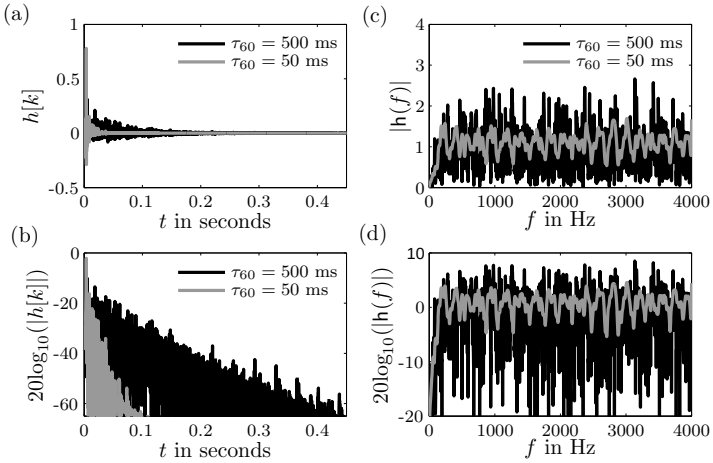

Figure 2.3: Typical room impulse responses in time-domain in linear scale (a) and logarithmic scale (b) and corresponding transfer functions (c) and (d).

to the convolution with the corresponding RIR $h(t)$. **Figure 2.4** shows a clean speech signal $s(t)$ in (a) time and (b) time-frequency-representation (spectrogram). Panels (c) and (d) of Figure 2.4 show the reverberant signal

$$x(t) = s(t) * h(t), \tag{2.1.1}$$

which is obtained by a convolution with an RIR $h(t)$ having a room reverberation time of $\tau_{60} = 500$ ms.

For the anechoic signal the so-called formants and the pitch, which correspond to the harmonic structure of speech due to the speech production in humans and the resonances of the human speech production system [VM06], respectively, can clearly be observed in the spectrogram. The phonemes[1] in Figure 2.4 are well separated in time. In panels (c) and (d), blurring of the formants and smearing of the phonemes are visible. Speech pauses are partly filled up by reverberant signal parts due to the reverberant tail of the RIR and phonemes overlap which may lead to a decreased speech intelligibility [SH02].

[1]The phoneme is the smallest segmental unit of sound that leads to meaningful contrasts between utterances [VM06].

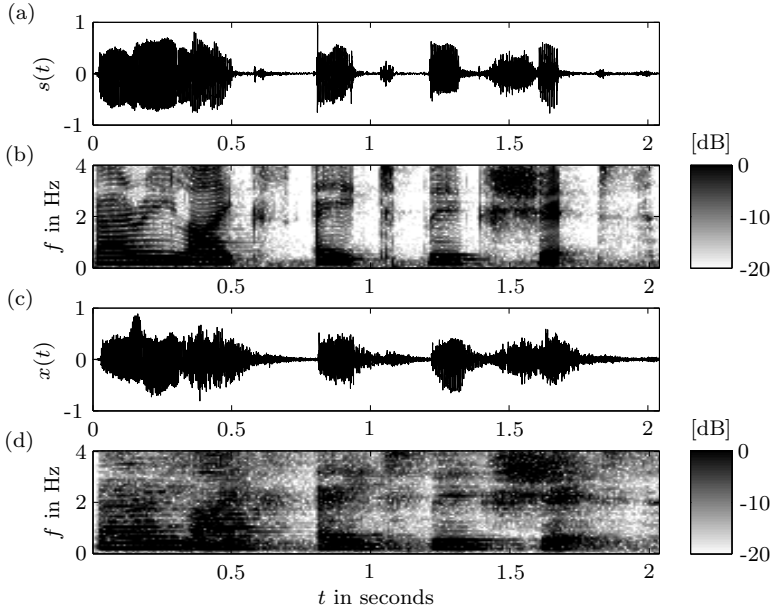

Figure 2.4: Influence of a room impulse responses on a speech signal. (a) anechoic speech signal $s(t)$ in time-domain, (b) spectrogram of anechoic signal $s(t)$, (c) reverberant speech signal $x(t)$ after convolution with an RIR ($\tau_{60} = 500$ ms), (d) spectrogram of reverberant signal $x(t)$.

2.1.3 Stochastic RIR Modelling

Since at least the reverberant tail of an RIR can be considered as a stochastic process and RIRs typically are characterised by an exponential decay in time-domain, it is common to model an RIR as an exponentially damped Gaussian process [Moo79, Hab07, GKMK09]

$$h_M[k] = b[k] \exp\left(-\frac{(k - k_{\text{init}})}{\eta}\right) u[k - k_{\text{init}}] \tag{2.1.2}$$

with k_{init} being the initial delay of the room impulse response model, $b[k]$ a white Gaussian random process, $u[k - k_{\text{init}}]$ the time-shifted Heaviside step function, f_s the sampling frequency and

$$\eta = \frac{2\tau_{60} f_s}{\ln(10^{-6})} \tag{2.1.3}$$

a damping constant that depends on the room reverberation time τ_{60}. **Figure 2.5** shows an RIR obtained by (2.1.2) in grey colour and the corresponding power delay profile (PDP) aka. power delay spectrum (PDS) [PRLN92] in black.

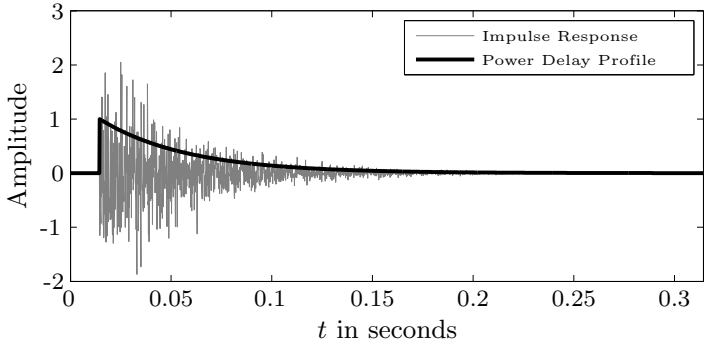

Figure 2.5: RIR and power delay profile (PDP).

2.1.4 Room Reverberation Time and Energy Decay Curve

Room impulse responses and room transfer functions can be described by several properties and measures. Some of these measures are sometimes also used for quality assessment of dereverberation algorithms and, thus, will be described in Chapter 4.2 and Appendix A.1 because they may slightly differ for common RIRs and other impulse responses, e.g. those of equalized systems.

One important measure characterizing an RIR is the so-called room reverberation time τ_{60} which is also known as RT60 in the literature. It is defined as the time of an 60 dB decay of the energy of the RIR. The room reverberation time is characteristic for a specific room. It depends on the room properties, such as sound absorption of the room boundaries (walls, floor and ceiling), interior and room volume. In general, the room reverberation time is frequency dependent, since typical wall materials have different sound absorption properties at different frequencies. Common reflection coefficients range from 0.99 (concrete) to 0.3 (sound absorbing slabs) [Bor89]. In contrast to the RIR, the room reverberation time does not depend on the exact positions of source and receiver.

A frequency independent approximation of the room reverberation time is

given by the so-called Sabine[2] reverberation formula [Joy75],

$$\tau_{60} = \frac{24\ln(10)V}{c\frac{1}{I}\sum_i \beta_i S_i} = 0.163\ [\text{sec/m}] \cdot \frac{V}{\frac{1}{I}\sum_i \beta_i S_i}. \tag{2.1.4}$$

Here, V is the room volume in m^3, $c \approx 340$ m/s is the speed of sound, S_i is the wall surface area in m^2 of a specific room boundary i and I is the total number of room surfaces. Accurate results can be obtained by using (2.1.4) as long as the absorption coefficient β is less than about 0.3, thus, for most realistic scenarios [Bor89].
To determine the room reverberation time, the so-called energy decay curve (EDC) can be calculated from a given RIR by [Sch65]

$$\text{EDC}(t) = C \int_{t'=t}^{\infty} h^2(t'). \tag{2.1.5}$$

In (2.1.5), the constant C is fixed by normalizing $\text{EDC}(t = 0)$ to 0 dB as depicted in **Figure 2.6 (b)**. Figure 2.6 (a) shows 4 RIRs having different room reverberation times and Figure 2.6 (b) the corresponding energy decay curves (EDCs) according to (2.1.5) and an estimate for the room reverberation time obtained from the EDC curves. For this purpose, a line is fitted to the linear part of an EDC and the intersection point with the -60 dB line indicates the corresponding room reverberation time τ_{60} [Sch65]. Usually the line is fitted to the values at -5 dB and -35 dB. However, due to measurement noise that leads to a deviation of the strictly linear decay in logarithmic domain, as it can be seen at the later parts of EDC curves e.g. for $\tau_{60} = \{50, 150\}$ ms in Figure 2.6 (b), particularly for shorter room reverberation times different points may be more appropriate such as -5 dB and -15 dB that were used for generating Figure 2.6 (b).

2.1.5 Critical Distance

The critical distance is the distance at which the sound pressure of the direct sound and all reflected parts are equal. It is defined [Hab07] depending on the directivity of the source compared to a sphere Q and the room constant $R = \bar{\alpha}S/(1 - \bar{\alpha})$ in m^2, i.e.

$$D_c = \sqrt{\frac{QR}{16\pi}} \tag{2.1.6}$$

[2]Wallace Clement Sabine, (June 13, 1868 - January 10, 1919) was an American physicist who was a pioneer in the field of architectural acoustics.

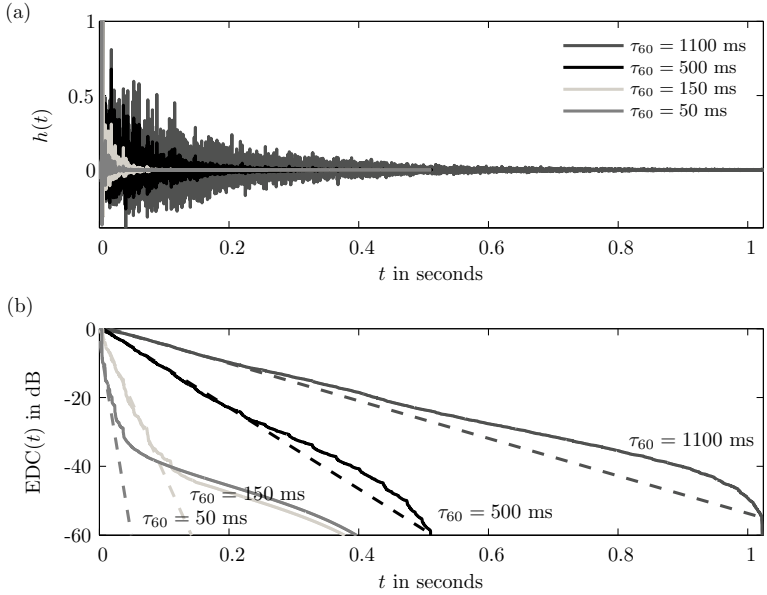

Figure 2.6: (a) RIRs and (b) energy decay curves (EDCs) for different room reverberation times τ_{60}. Panel (b) shows the EDCs according to (2.1.5) (solid lines) and the corresponding (dashed) lines fitted to the points at -5 dB and -15 dB.

For an omnidirectional sound source and assuming that the speed of sound in air is 344 m/s it can be approximated to [Kut00, Hab07]

$$D_c \approx 0.1 \text{ [sec/m]} \sqrt{\frac{V}{\pi \tau_{60}}} \tag{2.1.7}$$

only depending on the room volume V and the room reverberation time τ_{60}. If a human listener or a microphone is located closer to a sound source than the critical distance, the direct sound is dominant while the reverberant part is dominant if the distance is greater than the critical distance. For speech in highly reverberant rooms like churches a high speech intelligibility can be obtained if the listener is very close to the sound source despite high reverberation times.

2.1.6 z-Domain Properties of Room Impulse Responses

The distribution of the zeros of a common RIR in the z-domain is shown exemplarily in **Figure 2.7**.

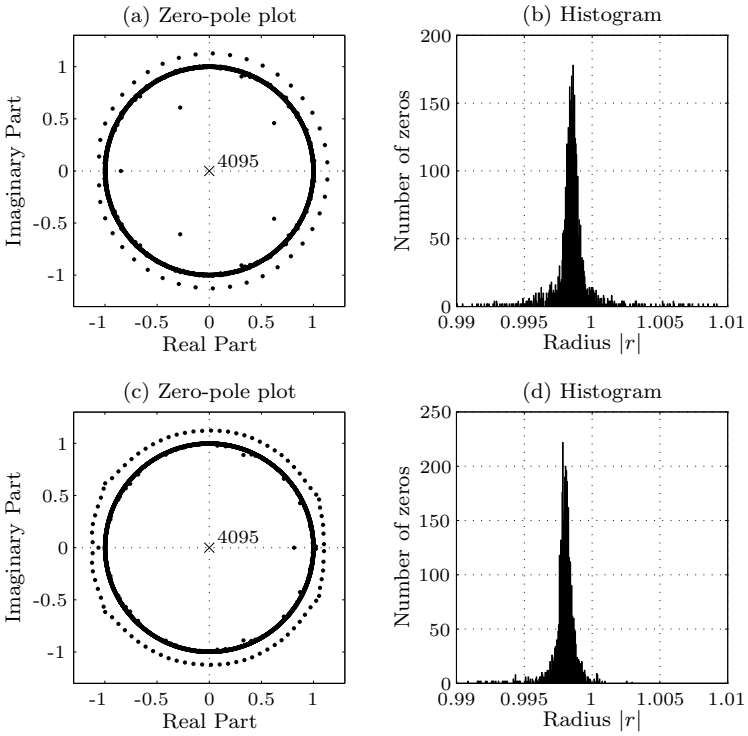

Figure 2.7: (a) Zero-pole plot and (b) histogram of radii $|r|$ of zeros in z-plane of a common simulated [AB79] RIR ($\tau_{60} \approx 500$ ms at $f_s = 8$ kHz). (c) Zero-pole plot and (d) histogram of radii $|r|$ of zeros in z-plane of a common measured [MM01] RIR ($\tau_{60} \approx 500$ ms at $f_s = 8$ kHz). 98.6% and 97.7% of zeros of the RIRs are within the depicted area of $0.99 \leq |r| \leq 1.01$ in panel (b) and (d), respectively.

The zero-pole plots in panels (a) and (c) of Figure 2.7 show the distribution of zeros in the z-domain for a simulated RIR [AB79] (panel (a)) and a measured RIR [MM01] (panel (c)) while panels (b) and (d) of Figure 2.7

show the corresponding distributions of the absolute values of the radii $|r|$ of the complex zeros close to the unit circle in the range of $0.99 \leq |r| \leq 1.01$. It can be seen that numerous zeros are located very close to the unit circle which corresponds to the notches in frequency-domain depicted in Figure 2.3. Furthermore, some zeros are located outside the unit circle. An RIR is, thus, a mixed phase system[3] and only its minimum phase part can be inverted by a causal stable infinite impulse response (IIR) filter [NA79]. However, since the maximum phase part of an RIR considerably contributes to the perceived reverberation a compensation of the minimum phase part only is not sufficient for dereverberation. Furthermore, as it will be illustrated in Section 4.3, equalization of zeros that are close to the unit circle is a hard task for an equalizer. And finally, since spectral notches in an RTF have to be compensated by spectral peaks of a corresponding equalizer, such an equalization cannot be very robust w.r.t. changes in the RTF (cf. Section 4.4.2).

2.2 Multi-Delay Filtering

The filtering result of an input signal $x[k]$ and an impulse response $h[k]$ can be obtained directly in time-domain by the convolution $y[k] = x[k] * h[k]$ or by a multiplication of the corresponding signal spectrum[4] $\mathsf{x}[n]$ and the transfer function $\mathsf{h}[n]$ in frequency-domain [KK09, VM06]. A large number of adaptive filter algorithms have been developed as well in time-domain [WS85, Hay02] as in frequency-domain [DMW78, Fer80, MG82]. If these filters have to deal with very long impulse responses as for the tasks of LRC or AEC, problems arise as well in time-domain as in frequency-domain. Since typical room impulse responses have lengths up to several thousand coefficients even for low sampling rates, time-domain algorithms normally suffer from slow convergence [BDH+99]. Highly correlated input signals, such as speech, further slow down convergence of time-domain algorithms since the maximum convergence speed depends on the ratio of minimum and maximum eigenvalue of the input correlation matrix [Hay02, Kam08]. Frequency-domain algorithms are an attractive solution for these problems. The computational effort for the convolution can be heavily reduced by using the fast Fourier transform (FFT) algorithm [Shy92, BM00]. Furthermore, the discrete Fourier transform (DFT) approximately produces uncorrelated

[3] Systems having all zeros inside the unit circle in the z-domain are called minimum phase systems. Systems having all zeros outside the unit circle are called maximum phase systems. Mixed phase systems have zeros as well inside as outside the unit circle.

[4] Please note that according to the notation declarations in Section 1.3 sans serif letters like $\mathsf{x}[n]$ indicate the frequency-domain representation of $x[k]$.

signals which gives the opportunity to choose the step-size of an adaptive algorithm independently for every frequency bin. This allows for a nearly uniform convergence even for large variations of the input power spectrum. Since frequency-domain algorithms rely on block processing a delay of at least one block is introduced. If the block size L is chosen to the length of the adaptive filter L_c, the processing delay of the algorithm may be too large, since the filter length L_c may be several thousand. More sophisticated algorithms, so-called multi-delay filters (MDFs), which are based on the classical overlap-save (OLS) method, overcome this drawback by partitioning the impulse response into smaller blocks [SP90, Som90]. A generalization to the case of arbitrary block sizes and weighted overlap-add (WOLA) structures is described in [MAG95]. By this, the block size L can be chosen as small as desired leading to a trade-off between computational efficiency and processing delay.
In the following, a brief overview on multi-delay filtering will be given and the notation that will be used throughout this thesis for block-frequency-domain adaptive algorithms will be introduced.

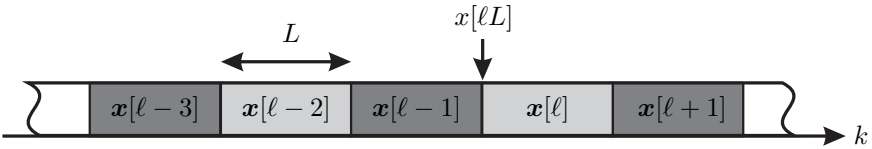

Figure 2.8: Input signal blocks $\boldsymbol{x}[\ell + i]$ of length L.

Each block of the input signal can be described as a vector

$$\boldsymbol{x}[\ell] = [x\,[\ell L]\,,\, x\,[\ell L + 1]\,,\, ...,\, x\,[(\ell + 1)L - 1]]^T \tag{2.2.1}$$

containing L input samples as depicted in **Figure 2.8**. Here, ℓ is the block-time index.
Accordingly, a vector $\boldsymbol{h}$ of length $L_h = L'_h L$ is defined containing the impulse response as depicted in **Figure 2.9** (a). Here, L'_h is the number of partitions needed to cover a vector of length L_h, given a certain block length L.

$$\boldsymbol{h} = \left[\boldsymbol{h}_0^T,\, \boldsymbol{h}_1^T,\, ...,\, \boldsymbol{h}_{L'_h-1}^T\right]^T \tag{2.2.2a}$$

$$\boldsymbol{h}_i = \left[h_{iL},\, h_{iL+1},\, ...,\, h_{(i+1)L-1}\right]^T \tag{2.2.2b}$$

The number of blocks needed for covering the partitioned room impulse response is denoted as $L'_h \in \mathbb{N}^+$. Please note that the impulse response

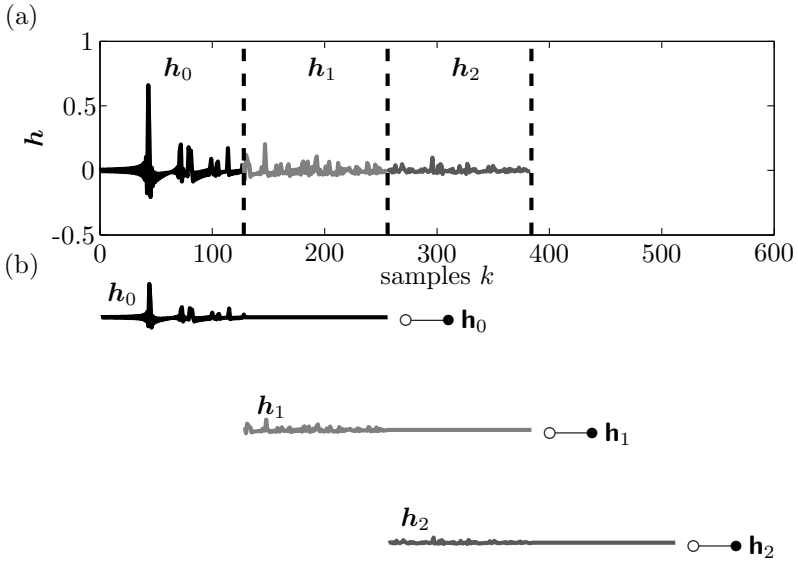

Figure 2.9: (a) Partitioning of room impulse response $\boldsymbol{h}$ of length $L_h = 384$ samples into $L'_h = 3$ partitions of length $L = 128$ and (b) transformation of each zero-padded partition to the frequency-domain.

vector $\boldsymbol{h}$ is defined here being time-invariant for simplicity reasons although, in general, it is time-variant and, thus, depends on the block index ℓ.

The DFT of a time series of length L is defined as

$$\mathrm{DFT}\{h[k]\} = \mathsf{h}[n] = \sum_{k=0}^{L-1} h[k] e^{-j2\pi kn/L} \tag{2.2.3}$$

and a 50% zero-padded DFT is given by

$$\mathrm{DFT}\{h[k]\} = \mathsf{h}[n] = \sum_{k=0}^{L-1} h[k] e^{-j2\pi kn/(2L)}. \tag{2.2.4}$$

In vector notation the DFT of a vector of length L can be written by a

multiplication with the DFT matrix

$$\boldsymbol{F}_{L\times L} = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & e^{-j2\pi\frac{1(L-1)}{L}} & e^{-j2\pi\frac{2(L-1)}{L}} & \dots & e^{-j2\pi\frac{(L-1)(L-1)}{L}} \\ 1 & e^{-j2\pi\frac{1(L-2)}{L}} & e^{-j2\pi\frac{2(L-2)}{L}} & \dots & e^{-j2\pi\frac{(L-1)(L-2)}{L}} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & e^{-j2\pi\frac{1(L-(L-1))}{L}} & e^{-j2\pi\frac{2(L-(L-1))}{L}} & \dots & e^{-j2\pi\frac{(L-1)(L-(L-1))}{L}} \end{bmatrix} \tag{2.2.5}$$

of size $L \times L$ which contains the DFT rotation factors[5].

$$\mathbf{h} = \boldsymbol{F}_{L\times L}\boldsymbol{h} \tag{2.2.6}$$

In the following, the DFT length L_{DFT} is chosen to equal twice the block length ($L_{\mathrm{DFT}} = 2L$) to obtain a 50% zero-padded spectrum. The zero-padded DFT can be calculated by

$$\mathbf{h} = \boldsymbol{F}_{2L\times 2L}\boldsymbol{W}^{10}_{2L\times L}\boldsymbol{h} \tag{2.2.7}$$

with the window matrix

$$\boldsymbol{W}^{10}_{2L\times L} = [\boldsymbol{I}_{L\times L}, \mathbf{0}_{L\times L}]^T. \tag{2.2.8}$$

In (2.2.8), $\boldsymbol{I}_{L\times L}$ is the identity matrix and $\mathbf{0}_{L\times L}$ a matrix containing zeros only, both of size $L \times L$. In the following, the truncated DFT matrix corresponding to a zero-padded FFT will be defined as

$$\boldsymbol{F}_{2L\times L} = \boldsymbol{F}_{2L\times 2L}\boldsymbol{W}^{10}_{2L\times L}. \tag{2.2.9}$$

The frequency-domain vector

$$\mathbf{h} = \left[\mathbf{h}_0^T, \mathbf{h}_1^T, \dots, \mathbf{h}_{L'_h-1}^T\right]^T \tag{2.2.10}$$

of size $2LL'_h \times 1$ containing the transfer functions

$$\mathbf{h}_i = \boldsymbol{F}_{2L\times 2L}\boldsymbol{W}^{10}_{2L\times L}\boldsymbol{h}_i \tag{2.2.11}$$

of all zero-padded partitions of the impulse response $\boldsymbol{h}_i$ as defined in (2.2.2b) can be obtained by

$$\mathbf{h} = \mathrm{bdiag}_{L'_h}\{\underbrace{\boldsymbol{F}_{2L\times L}, \boldsymbol{F}_{2L\times L}, \dots, \boldsymbol{F}_{2L\times L}}_{L'_h}\}\boldsymbol{h}. \tag{2.2.12}$$

[5]The inverse discrete Fourier transform of a vector is defined by $\boldsymbol{F}^{-1}_{L\times L} = \boldsymbol{F}^{*}_{L\times L}/L$.

The operator $\text{bdiag}_{L'_h}\{\cdot\}$ in (2.2.12) generates a block diagonal matrix of size $L'_h 2L \times L'_h L$ from L'_h DFT matrices of size $2L \times L$ each.

$$\text{bdiag}_{L'_h}\{\boldsymbol{F}_{2L\times L},\, \boldsymbol{F}_{2L\times L},\, ...,\, \boldsymbol{F}_{2L\times L}\} = \begin{bmatrix} \boldsymbol{F}_{2L\times L} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \boldsymbol{F}_{2L\times L} \end{bmatrix} \tag{2.2.13}$$

The result $\tilde{\mathbf{y}}[\ell]$ of the fast convolution corresponding to the well known overlap-save method [VM06] is obtained by multiplication of the zero-padded block transfer function $\mathbf{h}$ according to (2.2.12) and a matrix $\mathbf{X}[\ell]$ of size $2L \times 2LL'_h$ which contains the spectra of the signal obtained from $L'_h + 1$ subsequent time-domain data blocks:

$$\tilde{\mathbf{y}}[\ell] = \mathbf{X}[\ell]\mathbf{h} \tag{2.2.14}$$

with

$$\mathbf{X}[\ell] = \left[\breve{\mathbf{X}}[\ell],\, ...,\, \breve{\mathbf{X}}[\ell - L'_h + 1]\right] \tag{2.2.15}$$

$$\breve{\mathbf{X}}[\ell] = \text{diag}\{\boldsymbol{F}_{2L\times L}\boldsymbol{x}[\ell] + \tilde{\boldsymbol{I}}_{2L\times 2L}\, \boldsymbol{F}_{2L\times L}\boldsymbol{x}[\ell - 1]\} \tag{2.2.16}$$

$$\begin{aligned} \tilde{\boldsymbol{I}}_{2L\times 2L} &= \text{diag}\left\{\left[e^{-j\pi 0}, e^{-j\pi 1}, ..., e^{-j\pi(2L-1)}\right]\right\} \\ &= \text{diag}\left\{[1, -1, ..., 1, -1,]\right\} \end{aligned} \tag{2.2.17}$$

Equations (2.2.15) to (2.2.17) that define the block-frequency-domain input data matrix $\mathbf{X}[\ell]$ will be explained in the following. For the MDF, as an extension of the OLS method, two consecutive blocks of time-domain input data are transformed to the frequency-domain ($\boldsymbol{F}_{2L\times L}\boldsymbol{x}[\ell]$ and $\boldsymbol{F}_{2L\times L}\boldsymbol{x}[\ell - 1]$) and are combined to result in one block in the frequency-domain. Before adding them up in the frequency-domain the shifting property of the DFT [Rad79, KK09] is used to delay one block.

$$\begin{aligned} \text{DFT}\{x_0[k] + x_1[k - L]\} &= \text{DFT}\{x_0[k]\} + \text{DFT}\{x_1[k]\}e^{j\frac{2\pi}{L_{\text{DFT}}}nL} \\ &= \mathsf{x}_0[n] + \mathsf{x}_1[n]e^{j\pi n} \end{aligned} \tag{2.2.18}$$

For $L_{\text{DFT}} = 2L$ the rotation factors $e^{j\frac{2\pi}{L_{\text{DFT}}}nL} = e^{j\pi n}$ result in an alternating series of 1 and -1 for $0 \leq n \leq 2L - 1$ (cf. the definition of the shifting matrix in (2.2.17)). Equation (2.2.18) is expressed in vector/matrix form in (2.2.16) generating the block-frequency-domain matrix $\breve{\mathbf{X}}[\ell]$ depending on two consecutive time-domain input blocks. As a last step the block input signal matrix $\mathbf{X}[\ell]$ needed in (2.2.14) is generated by L'_h sub matrices $\breve{\mathbf{X}}[\ell]$.

As obvious from (2.2.18) and (2.2.16), only one FFT is needed for each block of time-domain input data.
The first L samples of the OLS output

$$\tilde{\boldsymbol{y}}[\ell] = \boldsymbol{F}_{2L\times 2L}^{-1}\tilde{\mathbf{y}}[\ell] \tag{2.2.19}$$

contain cyclic convolution products which have to be removed by a constraining matrix $\boldsymbol{G}$. The constrained frequency-domain block-filter result which does not contain cyclic convolution products can be expressed by

$$\mathbf{y}[\ell] = \boldsymbol{G}\tilde{\mathbf{y}}[\ell] = \boldsymbol{G}\mathbf{X}[\ell]\mathbf{h} \tag{2.2.20}$$

with

$$\boldsymbol{G} = \boldsymbol{F}_{2L\times 2L}\boldsymbol{W}_{2L\times L}^{01}\boldsymbol{W}_{L\times 2L}^{01}\boldsymbol{F}_{2L\times 2L}^{-1}, \tag{2.2.21}$$

$$\boldsymbol{W}_{2L\times L}^{01} = [\mathbf{0}_{L\times L}, \boldsymbol{I}_{L\times L}]^T, \tag{2.2.22}$$

$$\boldsymbol{W}_{L\times 2L}^{01} = [\mathbf{0}_{L\times L}, \boldsymbol{I}_{L\times L}]. \tag{2.2.23}$$

The matrix $\boldsymbol{G}$ removes cyclic convolution products by transforming one data block of length $2L$ to the time-domain, setting the first L samples to zero and transforming the constrained data back to the frequency-domain. The time-domain vector $\boldsymbol{y}[\ell]$ containing the filtering result of the current block is obviously obtained by

$$\boldsymbol{y}[\ell] = \boldsymbol{W}_{L\times 2L}^{01}\boldsymbol{F}_{2L\times 2L}^{-1}\mathbf{y}[\ell]. \tag{2.2.24}$$

With $\boldsymbol{W}_{L\times 2L}^{01}\boldsymbol{F}_{2L\times 2L}^{-1}\boldsymbol{F}_{2L\times 2L}\boldsymbol{W}_{2L\times L}^{01} = \boldsymbol{I}_{L\times L}$, (2.2.20) can be written as

$$\boldsymbol{y}[\ell] = \boldsymbol{W}_{L\times 2L}^{01}\boldsymbol{F}_{2L\times 2L}^{-1}\mathbf{X}[\ell]\mathbf{h}. \tag{2.2.25}$$

The MDF approach is visualized exemplarily in **Figure 2.10** for a filter impulse response $\boldsymbol{h}$ of length $L_h' = 2$ blocks which implies that 3 blocks of input data $\boldsymbol{x}[\ell]$ have to be used for calculating the current block $\boldsymbol{y}[\ell]$. Each block of input data is transformed to the frequency-domain by a DFT of length $2L$ (50% zero padding). According to (2.2.16), each data block $\check{\mathbf{X}}[\ell]$ in frequency-domain is calculated by two successive data blocks $\mathbf{x}[\ell]$ exploiting the shifting property of the DFT (2.2.18). In Figure 2.10, this is visualized by the ruled areas, i.e. the current frequency-domain signal block $\check{\mathbf{x}}[\ell]$ is calculated from the current block of input data $\mathbf{x}[\ell] = \boldsymbol{F}_{2L\times L}\boldsymbol{x}[\ell]$ and the previous block of input data $\mathbf{x}[\ell-1] = \boldsymbol{F}_{2L\times L}\boldsymbol{x}[\ell-1]$. The previous block of input data $\check{\mathbf{x}}[\ell-1]$ which is calculated by $\mathbf{x}[\ell-1] = \boldsymbol{F}_{2L\times L}\boldsymbol{x}[\ell-1]$ and $\mathbf{x}[\ell-2] = \boldsymbol{F}_{2L\times L}\boldsymbol{x}[\ell-2]$ has already been calculated previously and

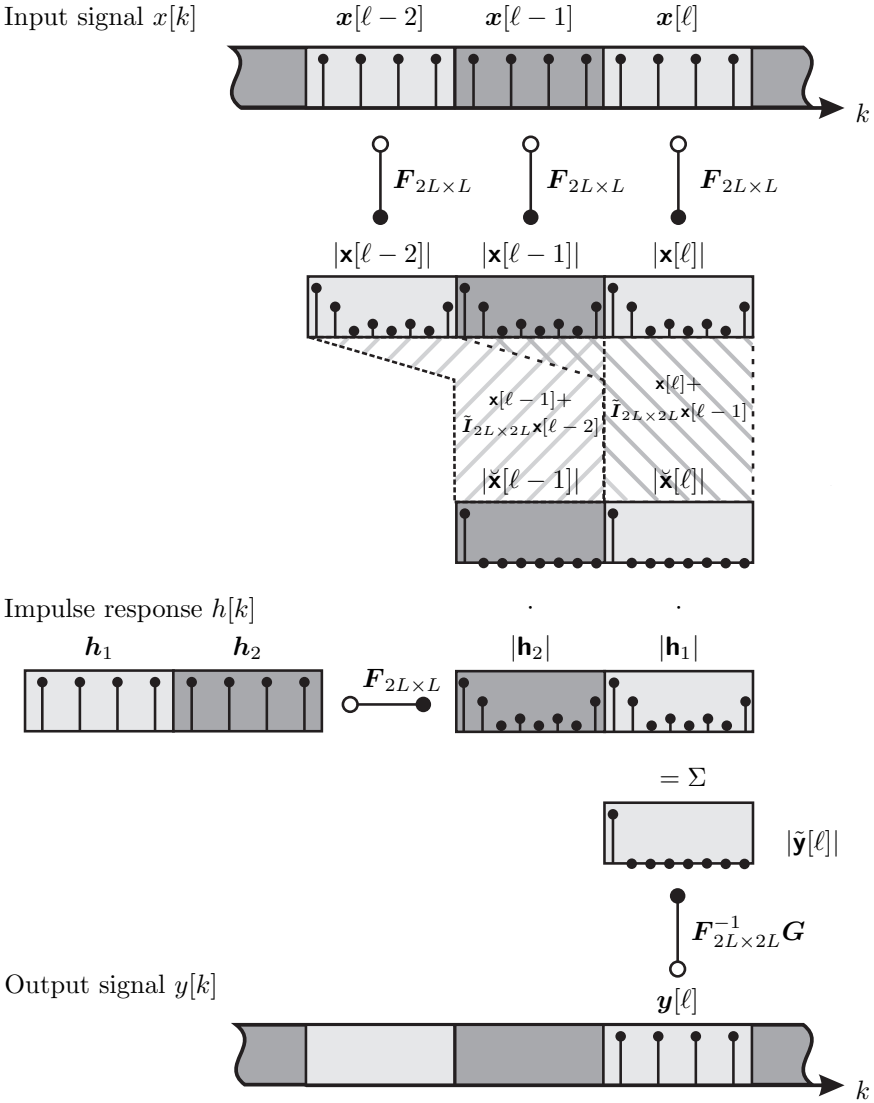

Figure 2.10: Illustration of the multi-delay filter (MDF) method expressed by vector/matrix notation for a block length of $L = 4$.

can be taken from memory. The frequency-domain data blocks $\check{\mathbf{X}}[\ell] = \operatorname{diag}\{\check{\mathbf{x}}[\ell]\}$ and $\check{\mathbf{X}}[\ell-1] = \operatorname{diag}\{\check{\mathbf{x}}[\ell-1]\}$ are then multiplied by the block

transfer functions $\mathbf{h}_1$ and $\mathbf{h}_1$ according to (2.2.14) and summed up to result in $\tilde{\mathbf{y}}[l]$. Cyclic parts of the convolution are removed by multiplication with the constraining matrix $\boldsymbol{G}$ resulting in one block output data $\boldsymbol{y}[\ell]$.

A schematic for an MDF implementation is given in **Figure 2.11**. The current block of input data $\boldsymbol{x}[\ell]$ of length L is transformed to the frequency-domain using the FFT with 50 % zero-padding. The resulting frequency-domain vector $\mathbf{x}[\ell]$ is summed up with the spectrum calculated in the previous block $\mathbf{x}[\ell-1]$ which was delayed by the block delay unit indicated by z^{-L} in Figure 2.11 and multiplied by the shifting vector $\tilde{\boldsymbol{I}}_{2L\times 2L}$. Each block of the RTF $\mathbf{h}_i$ is then multiplied by the corresponding part of the input signal, all block results are summed up and transformed back to the time-domain by the inverse fast Fourier transform (IFFT) to obtain the time-domain signal $\tilde{\boldsymbol{y}}[\ell]$. Cyclic convolution products are removed from $\tilde{\boldsymbol{y}}[\ell]$ by setting the first half to zero and taking only the second half of $\tilde{\boldsymbol{y}}[\ell]$.

$$\tilde{\boldsymbol{y}}[\ell] = [\tilde{y}_0[\ell],\, \tilde{y}_1[\ell] \ldots,\, \tilde{y}_{2L-1}[\ell]]^T \tag{2.2.26}$$

$$\boldsymbol{y}[\ell] = [\tilde{y}_L[\ell],\, \tilde{y}_{L+1}[\ell] \ldots,\, \tilde{y}_{2L-1}[\ell]]^T \tag{2.2.27}$$

If a frequency-domain signal $\tilde{\mathbf{y}}[\ell]$ is needed for further processing e.g. for adaptive filters, it can be obtained by

$$\mathbf{y}[\ell] = \boldsymbol{F}_{2L\times 2L} \left[\mathbf{0}_{1\times L},\, \boldsymbol{y}^T[\ell]\right]^T \tag{2.2.28}$$

which is equivalent to (2.2.20).

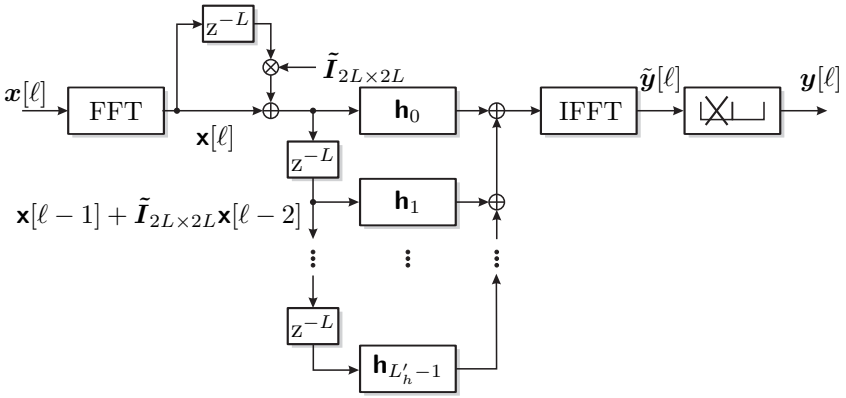

Figure 2.11: Schematic of a multi-delay filter (MDF).

2.3 Chapter Summary

This chapter introduced some fundamental properties of RIRs in Section 2.1, which will be needed for the following discussions about identification and equalization of RIRs and the respective problems for the AEC and LRC algorithms. It was shown that RIRs are characterized by thousands of zeros very close to the unit circle in z-domain, resulting in strong fluctuations, i.e. notches, in frequency-domain and an impulse response length of thousands of taps in time-domain, even for moderate sampling frequencies.
In Section 2.2 the multi-delay filtering structure known from the literature [MAG95] was introduced in the vector/matrix notation which will be used throughout the remainder of this thesis, i.e. for the derivation of an acoustic echo suppression filter in Section 3.3 and a quickly converging frequency-domain LRC filter that will be derived in Section 4.5.3.

Chapter 3

Acoustic Echo Cancellation

Acoustic echoes arise from the acoustic coupling between loudspeaker and microphone in an enclosure as previously depicted in Figure 2.2 on page 10. Thus, they occur in all modern voice communication systems with hands-free transducers. The acoustic coupling which is caused by numerous reflections at the room boundaries can be described by the room impulse response $h[k]$ as depicted in **Figure 3.1**.

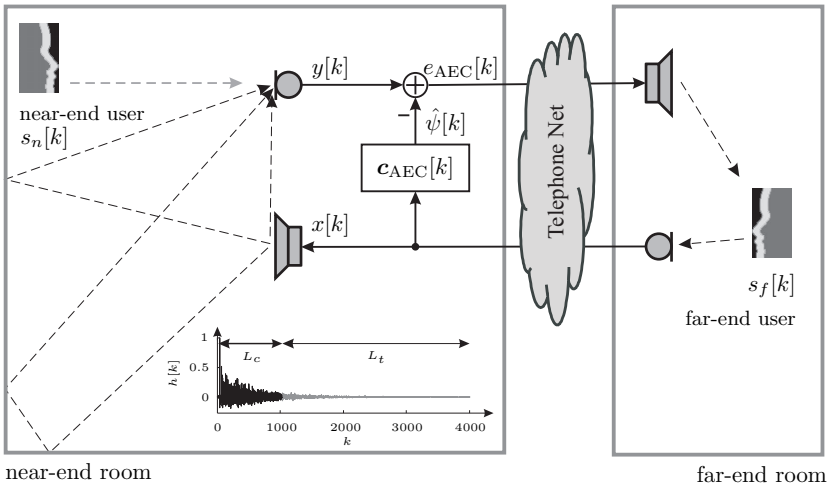

Figure 3.1: Hands-free system for acoustic echo cancellation.

As obvious from Figure 3.1, the acoustic echo has to be cancelled to prevent it from being transmitted back to the far-end user. Hearing his or her own voice delayed by the round trip delay of the system is annoying and does not allow for a natural communication. Furthermore, the risk of instability of the involved systems is given due to the closed electro-acoustic loop. The first linear filtering approach for estimating acoustic echoes was proposed by Sondhi in 1967 [Son67]. His approach to model the room impulse response $\boldsymbol{h}[k]$ by a linear filter $\boldsymbol{c}_{\mathrm{AEC}}[k]$ which is updated by a gradient algorithm based on the well-known least-mean-squares (LMS) algorithm [WS85] is still used in most of the state-of-the-art echo cancellation systems. The filter output is an estimate of the echo $\hat{\psi}[k]$ which can be subtracted from the microphone signal $y[k]$ as depicted in Figure 3.1. Much work has been carried out during the last decades in the field of echo cancellation, e.g. [YK82, Pol88, Kel88, Hän92, Hän94, Hän95, BMS98b, BDH+99, GB01b, BGM+01, KBK03a, HS04, GKMK06b, VM06, Enz08, XAG12]. In 1997, Hänsler coined the statement *From algorithms to systems - it's a rocky road* [Hän97] since the problem which may have appeared easy at a first glance turned out to cause difficulties, such as described in the following:

- **Very high filter orders are required in highly reverberant environments:** Typical RIRs have lengths of several thousand taps which have to be considered even at low sampling rates (cf. also Section 2.1). The identification of such RIRs requires filters of high orders [BDH+99]. While about 250 filter coefficients may be sufficient for echo cancellation in a car environment, AECs for office environments usually have lengths of several thousand filter coefficients to reduce the echo energy sufficiently, since the required filter length depends directly on the reverberation time τ_{60} of the environment. Common RIRs are of infinite length while the filter order of the AEC is limited. The unmodelled tail of the RIR, which is depicted in Figure 3.1 in gray for a filter order of 1024 exemplarily, always leads to a residual echo at the output of the AEC. For correlated input signals this so-called *tail-effect of acoustic echo cancellation*, furthermore, leads to a biased system identification which gets more severe for multi-loudspeaker hands-free systems [BMS98b, Kal07, GKMK07].

  Furthermore, the transmission path between loudspeaker and microphone is time-varying, in general, mainly due to movement of loudspeakers, microphones, or the user of the hands-free system. Therefore, the AEC filter has to track the time-varying RIRs continuously and quickly converging gradient algorithms are necessary to track changes of the impulse response between loudspeaker and microphone.

Recommendations given by the International Telecommunication Union (ITU) specify the requirements which have to be fulfilled by systems for acoustic echo cancellation [ITU93a, ITU93b, ITU88]. According to these recommendations, AEC systems must provide an echo suppression of 45 dB echo return loss enhancement (ERLE) (cf. Section 3.1 for the definition of the ERLE measure). If both speakers use hands-free systems a suppression of 40 dB is sufficient and in case of double-talk (both users speak at the same time) a suppression of 25-30 dB is required. In case of quickly changing RIRs which have to be tracked by the AEC system at least 10 dB echo suppression has to be achieved. If gradient algorithms such as the common normalized least-mean-squares (NLMS) algorithm [Hay02] are used for echo cancellation, high filter orders lead to slow convergence [BDH$^+$99]. The ITU recommendations furthermore require 20 dB echo suppression after 1 second convergence time. Especially for non-stationary and highly correlated input signals like speech this requires quickly converging algorithms like affine projection (AP) algorithms [OU84, GT95, Dou96, HS04] or the recursive least-squares (RLS) algorithm [Hay02, SK91]. Although the RLS algorithm leads to the fastest convergence speed it is not applicable directly since it suffers from heavy computational load. The so-called frequency-domain adaptive filters (FDAFs) [DMW78, MG82, SP90, Shy92, Nit00, BM00, BBK03] which exploit the properties of the fast Fourier transform (FFT) and their extension, the multi-delay filters (MDFs) [MAG95, BG03, GKMK06b, GKMK08b], lead to computationally efficient filtering algorithms with fast convergence and a small processing delay (cf. Section 2.2). Such a filter structure, which is already designed to be integrated seamlessly into the LRC systems introduced in the following chapters, will be described in Section 3.3.

Imperfect system identification of acoustic echo cancellers or acoustic echo suppression filters will have a strong influence on the LRC systems discussed in Chapter 4. Thus, quickly converging algorithms leading to a reliable system identification will be necessary in the following.

- **Ambient Disturbances:** If an AEC has to work in a car environment the previously described problems become less important since reverberation times in cars are very small ($\tau_{60} \approx 50$ ms). This allows for sufficient echo reduction by short and, thus, quickly converging AECs. However, high-level ambient noise inside a car, originating from engine, wind, rain, or tires, acts as a disturbance for adaptive

gradient algorithms which prevents fast convergence [HS04]. Hence, much research effort has been carried out on joint reduction of acoustic echoes and local interferences, e.g. [MV93, FB95a, Kel97, Gus99, Kel01, DMDC00, HS00, BSFB01, HKN04, HS04, KMK05, Her05, Kal07]. An additional *disturbance* which is always present in a scenario of hands-free communication is the local speaker. The adaptation of gradient algorithms has to be slowed down or even stopped [MPS00] for an active near-end speaker.

- **Residual Echoes:** In theory, AEC filters are capable to completely remove the acoutic echo from the microphone signal. However, due to the previously described tail effect, slow filter convergence and inevitable estimation errors, the acoustic echo cancellation is not sufficient most of the time. Residual echoes remain in the microphone signal which are clearly perceivable and disturbing.

  Post-filters, as described in Section 3.3, are quickly converging acoustic echo suppression (AES) filters in the microphone path which contribute to the echo reduction especially in periods of AEC convergence. These post-filters converge much faster than gradient-based AEC algorithms. However, post-filters inevitably lead to signal distortions since they are based on the principle of short time spectral attenuation (STSA). Post-filters can be designed to suppress noise [Wie49, Cap94, SBM01, EM85, GMK06b, GMK06a, MGK06b], residual echoes [GS99, EMV02, EV03, KBK03b, GKMK06b, Enz08], reverberation [Hab07] or combinations of these interferences [SB96, MV96, TGS97a, GMV98, Hab07]. Thus, they increase the signal-to-interference ratio (SIR) but always affect both, the residual echo and the signal of the near-end speaker which should be transmitted unaffectedly.

  Post-filters are not restricted to identification of the transmission path [KBK03a, Kal07, GKMK06b] since they only have to estimate the residual echo component [KBK03a, Fal03, FT05, FC05, KKS$^+$08, FFK$^+$08a] within the microphone signal. A reliable estimate of the power spectral density (PSD) of the residual echo is, thus, crucial for the design of echo suppression by post-filtering [EMV02, KBK03a, GKK05, Enz08]. For noise suppression the interference is assumed to be approximately stationary [Gie88, BF95, Mar01]. The residual echo signal, however, generally is highly non-stationary.

  Psychoacoustically motivated weighting rules [Gus99, Fal03, GMK06a, MGK06b] which incorporate knowledge about the masking effects of the human auditory system [Int92, WG00] lead

to perceptually *better sounding* post-filters since they produce less disturbing artefacts such as the well-known *musical noise*. Such psychoacoustically motivated weighting rules can be found in [TPM93, TGS97b, Vir99] for the single-channel case, in [GT98] for the multi-loudspeaker case, and in [BTSG98, Gus99] in combination with noise reduction.

Since the LRC sub-system may introduce severe distortions if the RIR estimate is not of sufficient quality, AES post-filter approaches (cf. Section 3.3) will be introduced and used to increase the performance of AEC filters in this thesis.

- **Demand for robust adaptation control:** If convenient and natural communication is provided by the hands-free system, the AEC as well as the post-filter has to be able to cope with so-called double-talk, which means that both speakers may talk at the same time. Although the concept of a linear filter lying in parallel to the RIR and subtracting the estimated echo signal according to [Son67] theoretically provides the possibility for perfect echo cancellation while unaffected transmission of the near-end speaker at the same time, the adaptation of the filter has to be stopped during periods of an active near-end speaker [MPS00, Nit00, Her05, KMK05, IG07, SMZ08]. First simple approaches for control of AEC adaptation in case of double-talk were based on simple suppression of the speaker with lower energy (*automatic gain control (AGC)*) [Hie95, IK97]. State-of-the-art double-talk detection can be achieved by analysis of the coherence between loudspeaker and microphone [SKW92, AG97, BMC00, GB01b, GB01a, BG02a, BG03], introducing artificial delays in the microphone path [AGQ97, VM06] or so-called shadow filtering [OAO97, IG07]. Besides the aforementioned approaches that were explicitly designed for AEC control, also general voice activity detection (VAD) or speech activity detection (SAD) strategies which are predominantly developed in the field of noise reduction or speech/music discrimination can be applied and adapted. Such algorithms are based on estimation of the SIR [HBSH98, GZ92, HS99] or evaluation of features such as energy [LZTZ02, MK02, MV08, TRB$^+$06, LCC10, EM06, RS75, RSB$^+$04], entropy [PMV$^+$06], higher statistical moments [NGM01], long-term observations [RSB$^+$04, RSB$^+$05, RGS07, HSGA10], modulation energy [SA02, AS03, SAP04, MMGP07], pitch [LE06, KK05b], auto-correlation statistics [SAVJ09], signals pre-processed by noise reduction algorithms [SS98, CKM06, TBA10], vocal detection [YY09], zero-crossing rate [RS75, KK05b], and combination of different features

[Pee04, SS97, WGH+11, RGH+11]. Most approaches compare internal variables to a certain threshold that can be fixed or adaptive [HTK03, HNK04, Her05, HBNK05a, HBNK05b].

Although obviously important, adaptation control of AEC and AES systems will not be within the scope of this thesis. Various references for robust adaptation control are given above, e.g. [MPS00]. Thus, in the following, the near-end speaker can be assumed to be inactive and no adaptation control is needed. However, in this thesis the ideas of adaptation control are used in Section 5.1.2 to incorporate knowledge about the AEC convergence state and, by this, increase the robustness of the LRC system.

- **Nonlinearities:** As already stated in Section 2.1 the AEC has to model the LRM system which contains the chain of loudspeaker, RIR, and microphone. Common linear approaches for acoustic echo cancellation are not able to model nonlinear effects which occur if low-cost loudspeakers, microphones, amplifiers, or analogue/digital (AD) converters are used in a hands-free system. Nonlinear effects on acoustic echo cancellation were studied in e.g. [SK00, KK05a, KK06] but will not be considered in the following since they are out of the scope of this thesis and, furthermore, introduce higher complexity and computational load. It will be assumed that all LRM systems can be modelled by linear approaches with sufficiently small errors.

- **AEC stereo problem:** If multi-channel systems are considered, common AEC algorithms can be extended easily to the multi-microphone case [Kel97] while the extension to a multi-loudspeaker system in general leads to a heavily reduced performance [BK01]. Since multiple-loudspeaker systems allow for the transmission of spatial information about the acoustic scene in the far-end room, such systems are highly demanded. They allow to distinguish the direction of different far-end speakers by the near-end listener, which enhances speech intelligibility. Whenever the terms multi-channel acoustic echo cancellation or stereo acoustic echo cancellation are used, generally hands-free systems with two or more loudspeakers are considered if not stated otherwise.

  In principle, simple and straightforward extensions of well-known single-channel gradient algorithms to the multi-channel case are possible [BAGG95, SM95, BDG96, MSS+97]. However, the performance of those algorithms is heavily reduced compared to the single-channel case due to the strong correlation of the loudspeaker signals

[SMH95, BMS97, BMS98b]. It was shown in [BMS98b, GB02] that no unique solution for the problem of stereo AEC exists in theory and that the solution depends not only on the RIRs in the near-end environment but also on the RIRs in the far-end environment.

One solution for stereo AEC is the decorrelation of the two transmission channels by introducing nonlinear distortions [GB00, MHB01, GB02, SMZ08] which are barely audible at least for speech signals and which have only negligible impact on the stereo image. Partial filter coefficient updates also improve convergence [KN04]. Other techniques such as time-variant filtering of one loudspeaker channel [JS98], all-pass structures [Ali98] or the use of comb filters [BMS98a] showed worse results.

For use in real-time systems, gradient algorithms working in the block-frequency-domain [BM00, BK01, BBK03, Gau03, Her05] showed promising results due to the decorrelation property of the DFT and the possibility of an efficient implementation by the FFT. A further complexity reduction can be obtained if multi-channel post-filters are used [Kal07, GKMK06b, FFK+08a, WQW11].

The AEC stereo problem is of particular importance for the multi-channel system identification which is needed for multi-loudspeaker LRC systems (cf. Section 4.4.3).

3.1 Objective Quality Measures for Echo Reduction Algorithms

Acoustic echo cancellers that will be the topic of this chapter aim at the reduction of the echo signal $\psi[k]$ that is picked up by the microphone as depicted in **Figure 3.2**. The echo cancellation filter $\boldsymbol{c}_{\mathrm{AEC}}[k]$ generates an echo estimate $\hat{\psi}[k]$ that is subtracted from the microphone signal. For the case of perfect echo cancellation, the residual echo signal $\xi[k] = \psi[k] - \hat{\psi}[k]$ vanishes.

Two different objective measures for assessment of the performance of echo reduction are common [VM06] and will be introduced briefly in the following. These measures are called system misalignment (cf. Section 3.1.1) and echo return loss enhancement (ERLE) (cf. Section 3.1.2).

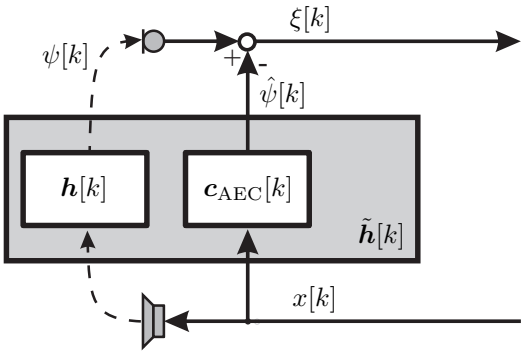

Figure 3.2: Schematic of an acoustic echo canceller. The system misalignment between RIR $\boldsymbol{h}[k]$ and AEC filter $\boldsymbol{c}_{\mathrm{AEC}}[k]$ is denoted as $\tilde{\boldsymbol{h}}[k]$.

3.1.1 AEC System Misalignment

The system misalignment vector $\tilde{\boldsymbol{h}}[k]$ is defined as the distance between the vector $\boldsymbol{h}[k]$ consisting of the RIR coefficients and the RIR estimate vector $\hat{\boldsymbol{h}}[k]$ that consists of the AEC coefficients as illustrated in Figure 3.2, i.e. $\hat{\boldsymbol{h}}[k] = \boldsymbol{c}_{\mathrm{AEC}}[k]$.

$$\tilde{\boldsymbol{h}}[k] = \boldsymbol{h}[k] - \hat{\boldsymbol{h}}[k] \tag{3.1.1}$$

$$\boldsymbol{h}[k] = [h_0[k]\,,\, h_1[k]\,,\, ...\,,\, h_{L_h-1}[k]]^T \tag{3.1.2}$$

$$\hat{\boldsymbol{h}}[k] = [c_{\mathrm{AEC},0}[k],\, c_{\mathrm{AEC},1}[k],\, ...,\, c_{\mathrm{AEC},L_{\mathrm{AEC}}-1}[k],\, 0,\, ...,\, 0]^T \tag{3.1.3}$$

Since generally the length of the RIR L_h is greater than the length of the AEC filter L_{AEC}, the vector $\hat{\boldsymbol{h}}[k]$ is filled with zeros to have the same length as $\boldsymbol{h}[k]$.
The so-called relative system misalignment $D_{\mathrm{dB}}[k]$ is commonly used for assessment of AEC algorithms. It is defined as the squared vector norm $||\tilde{\boldsymbol{h}}[k]||_2^2 = \tilde{\boldsymbol{h}}^T[k] \cdot \tilde{\boldsymbol{h}}[k]$ of the system misalignment according to (3.1.1) in dB normalized by the squared vector norm of the RIR[1].

$$D_{\mathrm{dB}}[k] = 10 \cdot \log_{10} \frac{||\tilde{\boldsymbol{h}}[k]||_2^2}{||\boldsymbol{h}[k]||_2^2} \tag{3.1.4}$$

[1] Cf. (3.2.13) for the general norm definition.

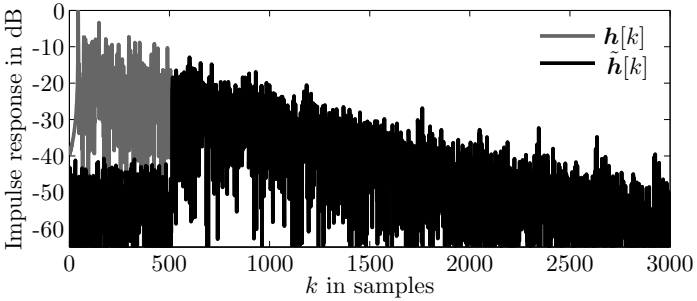

Figure 3.3: RIR vector $\boldsymbol{h}[k]$ and system misalignment vector $\tilde{\boldsymbol{h}}[k]$ for a converged AEC of order 511.

Figure 3.3 exemplarily shows an RIR (grey) and the corresponding system misalignment (black) for an AEC filter of order 511. The early samples of the RIR are compensated by the AEC filter. The tail of the RIR that cannot be modelled by the AEC filter due to its limited influence length contributes to the system misalignment. Thus, the minimum achievable system misalignment depends on the AEC order [VM06]. Since the RIR is unknown in general for real-world systems the system misalignment only can be evaluated during simulation when the RIR is known.
An objective measure to assess the AEC performance which also can be calculated if the RIR is unknown is called echo return loss enhancement (ERLE) and will be introduced in the following section.

3.1.2 Echo Return Loss Enhancement (ERLE)

The ratio between echo energy and residual echo energy is known as echo return loss enhancement (ERLE) [VM06].

$$\mathrm{ERLE}|_{\,\mathrm{dB}}\,[k] = 10 \cdot \log_{10} \frac{\mathrm{E}\left\{\psi^2[k]\right\}}{\mathrm{E}\left\{(\psi[k] - \hat{\psi}[k])^2\right\}} \tag{3.1.5}$$

$$= 10 \cdot \log_{10} \frac{\mathrm{E}\left\{\psi^2[k]\right\}}{\mathrm{E}\left\{\xi^2[k]\right\}} \tag{3.1.6}$$

In contrast to the system misalignment, the ERLE can be measured even if the RIR is unknown since all signals in (3.1.6) are available in a practical situation. The expectation operators in (3.1.6) can be replaced by their short-term expectations. For white noise as an input signal,

ERLE and the system misalignment $D_{\mathrm{dB}}[k]$ show the same results [VM06], i.e. $\mathrm{ERLE}|_{\mathrm{dB}}[k] = -D_{\mathrm{dB}}[k]$, as exemplarily shown in **Figure 3.4**. However, while for a non-white input signal a small system distance is equivalent to a high echo reduction in terms of ERLE, a high ERLE does not automatically result in a low system distance as it can be seen from Figure 3.4 for speech as an input signal. Thus, from a system-theoretical point of view the system distance is the more general objective measure. However, firstly it may not be measurable and secondly the performance perceived by the listener is more correlated to the ERLE measure.

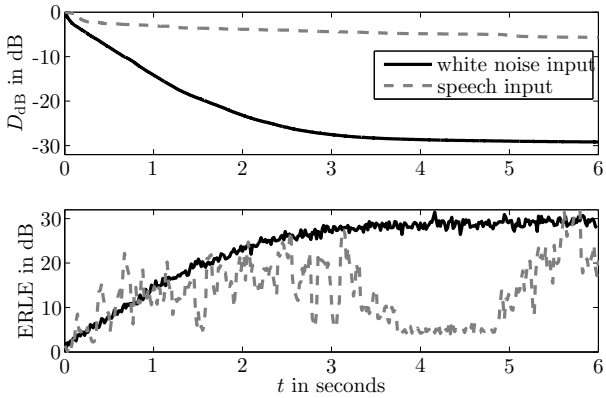

Figure 3.4: System misalignment D_{dB} and ERLE.

3.2 Gradient Algorithms for System Identification

A schematic of a single-channel system for acoustic echo cancellation is shown in **Figure 3.5**. The LRM system in Figure 3.5 is described by the unknown vector $\boldsymbol{h}[k]$. If the AEC filter $\boldsymbol{c}_{\mathrm{AEC}}[k]$ perfectly models the impulse response, its output $\hat{\psi}[k]$ equals the acoustic echo $\psi[k]$ contained in the microphone signal $y[k]$. For this case, the error signal $e_{\mathrm{AEC}}[k]$ is zero for an inactive near-end user $s_n[k]$. Please note that perfect cancellation of the error signal $e_{\mathrm{AEC}}[k]$ is only possible if the length of the AEC L_{AEC} at least equals the length of the RIR L_h. Otherwise a residual echo remains due to the unmodelled tail of the RIR.
Various algorithms have been proposed for solving the problem of acoustic echo cancellation. The most prominent one is the normalized version of the

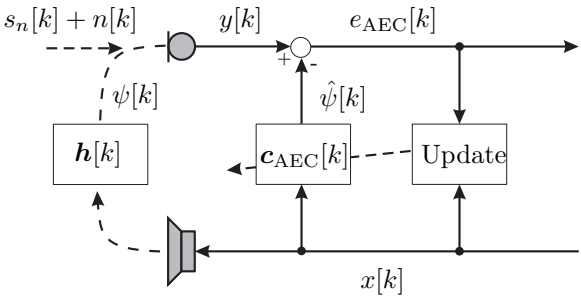

Figure 3.5: Schematic of a single-channel acoustic echo canceller (AEC).

LMS algorithm (cf. Section 3.2.1) and its extension to the affine projection algorithm (APA) [Hay02]. Proportionate update schemes which converge faster for *sparse* impulse responses as well as frequency-domain implementations which reduce the computational load will be discussed in Sections 3.2.2 and 3.3, respectively.
Most gradient algorithms are based on the *Wiener-Hopf* equation [Hay02, HS04]

$$\boldsymbol{R}_{\boldsymbol{xx}}\, \boldsymbol{c}_{\mathrm{AEC}} = \boldsymbol{r}_{\boldsymbol{x}\psi} \tag{3.2.1}$$

Here, $\boldsymbol{c}_{\mathrm{AEC}}$ is the coefficient vector of the AEC filter according to (3.2.2) of length L_{AEC} which is excited by a input signal vector $\boldsymbol{x}[k]$ containing the last L_{AEC} input samples

$$\boldsymbol{c}_{\mathrm{AEC}} = [c_{\mathrm{AEC},0},\, c_{\mathrm{AEC},1},\, \ldots,\, c_{\mathrm{AEC},L_{\mathrm{AEC}}-1}]^T, \tag{3.2.2}$$

$$\boldsymbol{x}[k] = [x[k],\, x[k-1],\, \cdots,\, x[k-L_{\mathrm{AEC}}+1]]^T. \tag{3.2.3}$$

$\boldsymbol{R}_{\boldsymbol{xx}}$ is the autocorrelation matrix of size $L_{\mathrm{AEC}} \times L_{\mathrm{AEC}}$ of the real-valued input signal

$$\boldsymbol{R}_{\boldsymbol{xx}} = \mathrm{E}\left\{\boldsymbol{x}[k] \cdot \boldsymbol{x}^T[k]\right\} \tag{3.2.4}$$

and $\boldsymbol{r}_{\boldsymbol{x}\psi}$ is the cross correlation vector

$$\boldsymbol{r}_{\boldsymbol{x}\psi} = \mathrm{E}\left\{\boldsymbol{x}[k] \cdot \psi[k]\right\} \tag{3.2.5}$$

between the filter input signal vector $\boldsymbol{x}[k]$ and the echo signal $\psi[k]$ which acts as a reference for the filter. The vector of the filter coefficients can be calculated by solving (3.2.1) for $\boldsymbol{c}_{\mathrm{AEC}}$:

$$\boldsymbol{c}_{\mathrm{AEC}} = \boldsymbol{R}_{\boldsymbol{xx}}^{-1} \boldsymbol{r}_{\boldsymbol{x}\psi} \tag{3.2.6}$$

Equation (3.2.6) is known as Wiener solution or *normal equation*. Typically, $\boldsymbol{R}_{\boldsymbol{xx}}$ and $\boldsymbol{r}_{\boldsymbol{x}\psi}$ are time-variant and, thus, a-priori unknown. Furthermore, the expectations in (3.2.4) and (3.2.5) have to be estimated by proper averaging.

3.2.1 The LMS and NLMS Algorithm

The least-mean-squares (LMS) algorithm is by far the most popular adaptive algorithm for designing adaptive filters [WS85, Hay02]. It was introduced by Widrow and Hoff in 1960 [WH60, WMB75]. Since then the LMS algorithm found applications in many areas, such as interference cancellation, equalization, and system identification [Son67]. The LMS algorithm belongs to the class of gradient algorithms which update the coefficient vector iteratively following the direction of the negative gradient

$$-\nabla_{\boldsymbol{c}_{\mathrm{AEC}}} = -\left[\frac{\partial}{\partial c_{\mathrm{AEC},0}}, \frac{\partial}{\partial c_{\mathrm{AEC},1}}, \ldots, \frac{\partial}{\partial c_{\mathrm{AEC},L_{\mathrm{AEC}}-1}}\right]^T, \tag{3.2.7}$$

which points towards the minimum of a properly chosen cost function (typically the mean squared error $\mathrm{E}\left\{|e_{\mathrm{AEC}}[k]|^2\right\}$). The so-called *deterministic gradient algorithms* avoid the inversion of $\boldsymbol{R}_{\boldsymbol{xx}}$ in (3.2.6) which would lead to a high computational effort for common AEC filter lengths of several thousand coefficients,

$$\begin{aligned} \boldsymbol{c}_{\mathrm{AEC}}[k+1] &= \boldsymbol{c}_{\mathrm{AEC}}[k] - \nabla_{\boldsymbol{c}_{\mathrm{AEC}}} e^2_{\mathrm{AEC}}[k], & (3.2.8)\\ &= \boldsymbol{c}_{\mathrm{AEC}}[k] + \mu\left(\hat{\boldsymbol{r}}_{\boldsymbol{x}\psi} - \hat{\boldsymbol{R}}_{\boldsymbol{xx}}\boldsymbol{c}_{\mathrm{AEC}}[k]\right). & (3.2.9) \end{aligned}$$

This leads to the well-known LMS update equation which results from estimating the autocorrelation matrix $\boldsymbol{R}_{\boldsymbol{xx}}$ and the cross-correlation vector $\boldsymbol{r}_{\boldsymbol{x}\psi}$ using instantaneous values $\hat{\boldsymbol{R}}_{\boldsymbol{xx}} = \boldsymbol{x}[k]\boldsymbol{x}^T[k]$ and $\hat{\boldsymbol{r}}_{\boldsymbol{x}\psi} = \boldsymbol{x}[k]\psi[k]$ [Hay02]:

$$\hat{\boldsymbol{c}}_{\mathrm{AEC}}[k+1] = \hat{\boldsymbol{c}}_{\mathrm{AEC}}[k] + \mu_{\mathrm{LMS}}\,\boldsymbol{x}[k]\,e_{\mathrm{AEC}}[k] \tag{3.2.10}$$

The simplicity of (3.2.10) contributed to the great success of the LMS filter algorithm. The computational load is minimum since only L_{AEC} multiplications are needed for each update. Unfortunately, the maximum possible step-size μ_{LMS} depends on the largest eigenvalue of $\boldsymbol{R}_{\boldsymbol{xx}}$ which is difficult to estimate in a real-world algorithm and fluctuates a lot depending on the current input samples of $x[k]$. To achieve convergence which is independent of the input signal's power, the step-size can be normalized resulting in the normalized least-mean-squares (NLMS) algorithm proposed in 1967

[NN67, AJ67]:

$$\hat{\boldsymbol{c}}_{\mathrm{AEC}}[k+1] = \hat{\boldsymbol{c}}_{\mathrm{AEC}}[k] + \frac{\mu_{\mathrm{NLMS}}[k]}{\boldsymbol{x}^T[k]\boldsymbol{x}[k] + \delta_{\mathrm{NLMS}}}\,\boldsymbol{x}[k]\,e[k] \tag{3.2.11}$$

In (3.2.11) δ_{NLMS} is a regularization parameter which prevents division by zero if the energy of the input signal $\boldsymbol{x}[k]$ is too low. The step-size $\mu_{\mathrm{NLMS}}[k]$ in (3.2.11) can be chosen independently of the input signal as $0 \leq \mu_{\mathrm{NLMS}}[k] \leq 2$. Common step-sizes for highly correlated and non-stationary signals like speech are $0.1 \leq \mu_{\mathrm{NLMS}}[k] \leq 0.5$ depending on the expected ambient noise which disturbs convergence of the algorithm. The step-size $\mu_{\mathrm{NLMS}}[k]$ has to be heavily reduced towards 0 if a near-end speaker is active to prevent cancellation of his or her signal. If echoes have to be cancelled in reverberant rooms, such as office environments, the convergence rate of the LMS algorithms is not sufficient to fulfill the ITU recommendations for echo suppression [ITU93a, ITU93b, ITU88]. The update of the coefficient vector always points into the direction of the input vector $\boldsymbol{x}$ which may be different from the direction towards the minimum of the cost function. If consecutive vectors $\boldsymbol{x}$ of the input signal are strongly correlated, as it is the case for speech input, their directions may differ only slightly from each other. This leads to slower convergence because more steps are needed to reach the global minimum of the cost function in the L_{AEC}-dimensional space of the filter coefficients. Algorithms which allow for faster convergence but raise the computational load are the recursive least-squares (RLS) algorithm and the affine projection algorithm [Hay02]. For the specific nature of equalized impulse responses (IRs) the class of so-called proportionate update schemes will be briefly reviewed and analyzed in the following.

3.2.2 Proportionate Filter Update

The previously discussed adaptive filter algorithms evenly spread their update energy over all filter coefficients. Proportionate update algorithms, which have been developed for network echo cancellation allow for faster convergence if the impulse response, which has to be identified, is sparse [Dut00, BGM$^+$01, NCB06]. An impulse response can be considered to be sparse if a small percentage of its samples have a significant magnitude while the rest are zero or small [BHCN06]. In proportionate update schemes, an individual step-size is calculated for each filter coefficient allowing for faster convergence for filter coefficients with higher energy. The idea of incorporating knowledge about the impulse response for defining an individual step-size for each filter coefficient was presented in [KMK93] for typical RIRs having an exponential decay. By choosing the step-size of the AEC filter

proportional to the expected exponential decay of the RIR, earlier filter coefficients converged faster than later filter coefficients [KMK93]. This led to a faster convergence in total. Proportionate update schemes adaptively estimate the optimum step-size for each coefficient. By this, they allow for faster convergence independently of the structure of the sparse impulse response [Dut00, KLDN08].

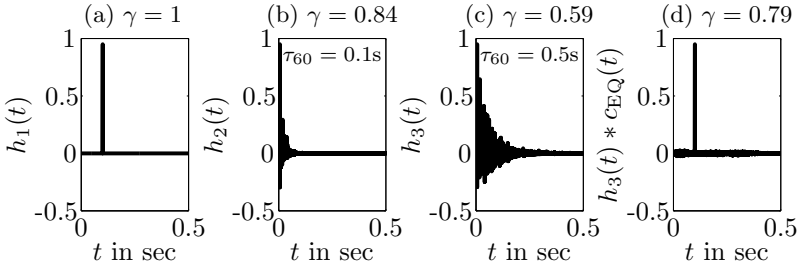

Figure 3.6: Examples of impulse responses that have to be identified by an AEC.

Four different examples of impulse responses and their corresponding sparsity measures [BHCN06]

$$\gamma(\boldsymbol{h}) = \frac{L_h}{L_h - \sqrt{L_h}} \left(1 - \frac{\|\boldsymbol{h}\|_1}{\sqrt{L_h}\|\boldsymbol{h}\|_2}\right) \tag{3.2.12}$$

are shown in **Figure 3.6**. In (3.2.12), $\|\boldsymbol{h}\|_1$ and $\|\boldsymbol{h}\|_2$ are the l_1-norm and the l_2-norm, respectively, defined as

$$\|\boldsymbol{h}\|_p = \left(\sum_{i=0}^{L_h-1} |h_i|^p\right)^{\frac{1}{p}} \tag{3.2.13}$$

for $p = \{1, 2\}$. The sparsity measure defined in (3.2.12) can take values between $0 \leq \gamma(\boldsymbol{h}) \leq 1$ and equals 1 for the (delayed) dirac function which is the most sparse possible impulse response and which is shown in Figure 3.6 (a). The sparsity measure equals 0 for a uniform impulse response $\boldsymbol{h} = [1, 1, ..., 1]^T$ and is independent of the scaling of $\boldsymbol{h}$ [BHCN06]. Common RIRs with room reverberation times of $\tau_{60} = 100$ ms and $\tau_{60} = 500$ ms are shown in Figure 3.6 (b) and (c), respectively. Figure 3.6 (d) shows the equalized impulse response[2] $h_3[k] * c_{\mathrm{EQ}}[k]$ after processing by a least-squares equalizer $c_{\mathrm{EQ}}[k]$ of order $L_{\mathrm{EQ}} = 2048$ as illustrated in **Figure 3.7**.

[2]The equalization of RIRs will be discussed in detail in Chapter 4.

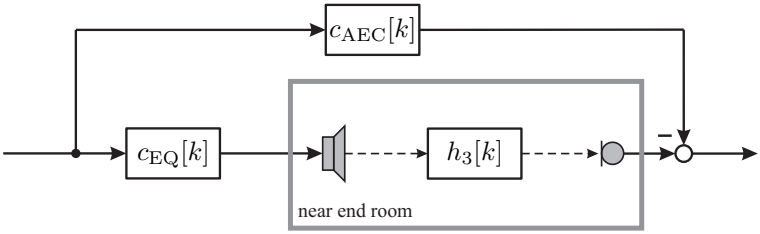

Figure 3.7: Identification of an equalized impulse response by an outer AEC.

It is obvious that the equalized impulse response in Figure 3.6 (d) is quite sparse which motivates the discussion of the class of proportionate normalized least-mean-squares (PNLMS) algorithms in the following for identification of such impulse responses[3].

The Proportionate NLMS Algorithm (PNLMS)

The sparse nature of impulse responses as they occur e.g. in network echo cancellation, feedback cancellation in hearing aids or for equalized systems as depicted in Figure 3.6 (d) causes standard adaptive algorithms including the NLMS to perform poorly since every filter coefficient is updated with the same step-size. The proportionate normalized least-mean-squares (PNLMS) algorithm calculates an individual step-size for each filter coefficient based on the estimated energy of the corresponding coefficient. This allows for faster convergence of filter coefficients with higher energy [Dut00, Gay98].

The update rule of the PNLMS algorithm is given by [BHCN06]

$$\boldsymbol{c}_{\mathrm{AEC}}[k+1] = \boldsymbol{c}_{\mathrm{AEC}}[k] + \mu_{\mathrm{PNLMS}} \frac{\boldsymbol{M}_{\mathrm{PNLMS}}[k]\,\boldsymbol{x}[k]\,e[k]}{\boldsymbol{x}^T[k]\boldsymbol{M}_{\mathrm{PNLMS}}[k]\boldsymbol{x}[k] + \delta_{\mathrm{PNLMS}}}. \quad (3.2.14)$$

In (3.2.14) the matrix $\boldsymbol{M}_{\mathrm{PNLMS}}[k]$ containing the step-sizes is defined as

$$\boldsymbol{M}_{\mathrm{PNLMS}}[k] = \mathrm{diag}\,\{\boldsymbol{\mu}_{\mathrm{PNLMS}}[k]\} \quad (3.2.15)$$

[3]Please note, that equalizers need an estimate of the RIR $\boldsymbol{h}$ which means that an additional inner AEC is be needed in Figure 3.7. System identification for equalizers will be discussed in Section 4.4.2 and combined systems with inner and outer AECs will be discussed in Chapter 5.

with

$$\boldsymbol{\mu}_{\mathrm{PNLMS}}[k] = \left[\frac{\mu'_{0,\mathrm{PNLMS}}[k]}{\bar{\mu}'_{\mathrm{PNLMS}}[k]}, \ldots, \frac{\mu'_{L_{\mathrm{AEC}}-1,\mathrm{PNLMS}}[k]}{\bar{\mu}'_{\mathrm{PNLMS}}[k]}\right]^T, \quad (3.2.16)$$

$$\mu'_{i,\mathrm{PNLMS}}[k] = \max\{\rho\, l'_{\infty}[k], |c_{\mathrm{AEC},i}[k]|\}, \quad (3.2.17)$$

$$l'_{\infty}[k] = \max\{\upsilon, l_{\infty}[k]\}, \quad (3.2.18)$$

$$l_{\infty}[k] = \|\boldsymbol{c}_{\mathrm{AEC}}[k]\|_{\infty} \quad (3.2.19)$$

$$= \max\{|c_{\mathrm{AEC},0}[k]|, \ldots, |c_{\mathrm{AEC},L_{\mathrm{AEC}}-1}[k]|\}, \quad (3.2.20)$$

$$\bar{\mu}'_{\mathrm{PNLMS}}[k] = \frac{1}{L_{\mathrm{AEC}}} \sum_{i=0}^{L_{\mathrm{AEC}}-1} \mu'_{i,\mathrm{PNLMS}}[k]. \quad (3.2.21)$$

The regularization parameter δ_{PNLMS} can be chosen as $\delta_{\mathrm{PNLMS}} = \delta_{\mathrm{NLMS}}/L_{c,\mathrm{AEC}}$ compared to the NLMS algorithm [BHCN06]. The parameters ρ and υ in equations (3.2.17) and (3.2.18) control the proportionate behaviour of the PNLMS algorithm. Here, the parameter ρ is the more important one [Dut00] since it controls the amount of *proportionateness* of the PNLMS algorithm. If ρ is chosen to have values of $\rho \geq 1$ it leads to a degeneration of the PNLMS algorithm to the NLMS algorithm because $\rho\, l'_{\infty}[k]$ is always greater than $|c_{\mathrm{AEC},i}[k]|$ in this case and all step-sizes $\mu'_{i,\mathrm{PNLMS}}[k]$ become equal. Therefore, ρ should be chosen smaller than 1 since, in general. As ρ decreases, the initial convergence speed will become faster [Dut00]. The parameter υ is of minor importance since it just prevents a dead-lock of the update if all coefficients have (initial) values of 0 and it ensures equal convergence in the very beginning. Once one coefficient of $|c_{\mathrm{AEC},i}[k]|$ is greater than υ it will become ineffective as it can be seen from (3.2.18).
In general, the larger the respective coefficient in a PNLMS update scheme is, the more adaptation speed it gets. This leads to a fast initial convergence. However, other coefficients converge more slowly and, furthermore, the gradient noise for large coefficients is greater than for the conventional NLMS algorithm. This leads to a slightly worse final convergence as it will be shown later in this section after describing the so-called improved proportionate NLMS (IPNLMS) algorithm in the following.

The Improved PNLMS Algorithm (IPNLMS)

The PNLMS algorithm discussed above leads to faster convergence for sparse impulse responses as they occur e.g. in the field of network echo cancellation or for the case of equalized impulse responses. However, the PNLMS algorithm converges slower than the conventional NLMS algorithm

for more dispersive impulse responses [NCB06] (e.g. as depicted in Figure 3.6 (c)). Several extensions of the PNLMS algorithm have been proposed in literature to solve this problem. Examples are the so-called PNLMS++ algorithm that alternates the adaptation rule between PNLMS and NLMS [Gay98] at each iteration step and the so-called μ-law PNLMS (MPNLMS) [DD05]. Both show improved convergence. An algorithm which directly offers the possibility to smoothly switch between the convergence behavior of the NLMS algorithm and the PNLMS algorithm is the so-called improved proportionate NLMS (IPNLMS) [BG02b] which will be briefly introduced in the following.

The reason for the performance loss of the PNLMS algorithm for the case of a dispersive impulse response is the strong focus on the most prominent coefficient in (3.2.17). A more relaxed consideration of the proportionate idea can be formulated by [BG02b]

$$\mu'_{i,\mathrm{IPNLMS}}[k] = (1-\alpha)\frac{\|\boldsymbol{c}_{\mathrm{AEC}}[k]\|_1}{L_{\mathrm{AEC}}} + (1+\alpha)|c_{\mathrm{AEC},i}[k]|. \tag{3.2.22}$$

In (3.2.22), the parameter $-1 \leq \alpha < 1$ allows for a trade off between NLMS update ($\alpha = -1$) and PNLMS update ($\alpha \approx 1$). Similarly to the definition of the PNLMS algorithm in (3.2.14) the IPNLMS update is given by [BG02b]:

$$\boldsymbol{c}_{\mathrm{AEC}}[k+1] = \boldsymbol{c}_{\mathrm{AEC}}[k] + \mu_{\mathrm{IPNLMS}}\frac{\boldsymbol{M}_{\mathrm{IPNLMS}}[k]\boldsymbol{x}[k]e[k]}{\boldsymbol{x}^T[k]\boldsymbol{M}_{\mathrm{IPNLMS}}[k]\boldsymbol{x}[k] + \delta_{\mathrm{IPNLMS}}} \tag{3.2.23}$$

$$\boldsymbol{M}_{\mathrm{IPNLMS}}[k] = \mathrm{diag}\{\boldsymbol{\mu}_{\mathrm{IPNLMS}}[k]\} \tag{3.2.24}$$

$$\boldsymbol{\mu}_{\mathrm{IPNLMS}}[k] = \left[\frac{\mu'_{0,\mathrm{IPNLMS}}[k]}{\|\boldsymbol{\mu}'_{\mathrm{IPNLMS}}[k]\|_1}, \ldots, \frac{\mu'_{L_{\mathrm{AEC}}-1,\mathrm{IPNLMS}}[k]}{\|\boldsymbol{\mu}'_{\mathrm{IPNLMS}}[k]\|_1}\right]^T \tag{3.2.25}$$

With (3.2.13) and (3.2.22), the l_1 norm $\|\boldsymbol{\mu}_{\mathrm{IPNLMS}}[k]\|_1$ in (3.2.25) can expressed by

$$\|\boldsymbol{\mu}'_{\mathrm{IPNLMS}}[k]\|_1 = 2\|\boldsymbol{c}_{\mathrm{AEC}}[k]\|_1 \tag{3.2.26}$$

and, by this, the elements $\mu_{i,\mathrm{IPNLMS}}[k], i = 0, 1, \ldots, L_{\mathrm{AEC}}-1$ of the step-size vector in (3.2.25) can be calculated by

$$\begin{aligned}\mu_{i,\mathrm{IPNLMS}}[k] &= \frac{\mu'_{i,\mathrm{IPNLMS}}[k]}{\|\boldsymbol{\mu}'_{\mathrm{IPNLMS}}[k]\|_1} \\ &= \frac{1-\alpha}{2L_{\mathrm{AEC}}} + (1+\alpha)\frac{|c_{\mathrm{AEC},i}[k]|}{2\|\boldsymbol{c}_{\mathrm{AEC}}[k]\|_1 + \varepsilon}.\end{aligned} \tag{3.2.27}$$

Please note, that in order to avoid a division by zero in equation (3.2.27), especially at the beginning of the adaptation when all filter taps are initialized by zero, a small positive constant ε is added to the denominator

[BG02b]. The regularization parameter δ_{IPNLMS} in (3.2.23) can be chosen to $\delta_{\text{IPNLMS}} = (1-\alpha)/(2L_{\text{AEC}})\delta_{\text{NLMS}}$ [BG02b].
Variants of the APA with proportionate update schemes also exist [GBGS00, HGS04]. However, they will not be discussed here, since simulations showed worse performance than the conventional APA for highly coloured and non-stationary input signals, such as speech.

Performance Comparison of Proportionate Gradient Algorithms

Before a comparison of NLMS, PNLMS and IPNLMS will be given by means of achievable ERLE and system distance D_{dB} a comparative illustration of the initial convergence of the three algorithms will be shown exemplarily for the RIR depicted in **Figure 3.8** which is characterized by a short room reverberation time of $\tau_{60} \approx 50$ ms and a sparsity measure of $\gamma(\boldsymbol{h}) = 0.91$. Such an impulse response can be observed e.g. in a car.

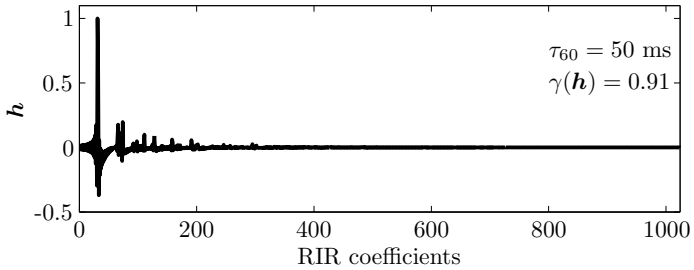

Figure 3.8: Example of a sparse RIR.

The states of convergence after several filter updates are shown in **Figure 3.9** for the NLMS algorithm in panels (a)-(e), for the PNLMS algorithm in panels (f)-(j) and for the IPNLMS algorithm in panels (k)-(o), respectively. The absolute values of the filter coefficients $|\boldsymbol{c}_{\text{AEC}}[k]|$ are indicated by a thick solid black line at update steps $k = \{30, 40, 50, 200, 250\}$. The first 100 AEC filter coefficients $|c_{\text{AEC},i}[k]|$ are shown for the update steps $k = \{30, 40, 50\}$ in panels (a)-(c), (f)-(h) and (k)-(m) while the first 250 AEC filter coefficients $|c_{\text{AEC},i}[k]|$ are shown for the update steps $k = \{200, 250\}$ in panels (d)-(e), (i)-(j) and (n)-(o), respectively.

NLMS algorithm: Panels (a)-(e) show the absolute values of the first AEC filter coefficients $|\boldsymbol{c}_{\text{AEC}}[k]|$ and the corresponding absolute values of the RIR $\boldsymbol{h}[k]$ (dotted grey line) which has to be identified. Please note, that for better illustration, the RIR $\boldsymbol{h}[k]$ is time-invariant for all panels in Figure 3.9,

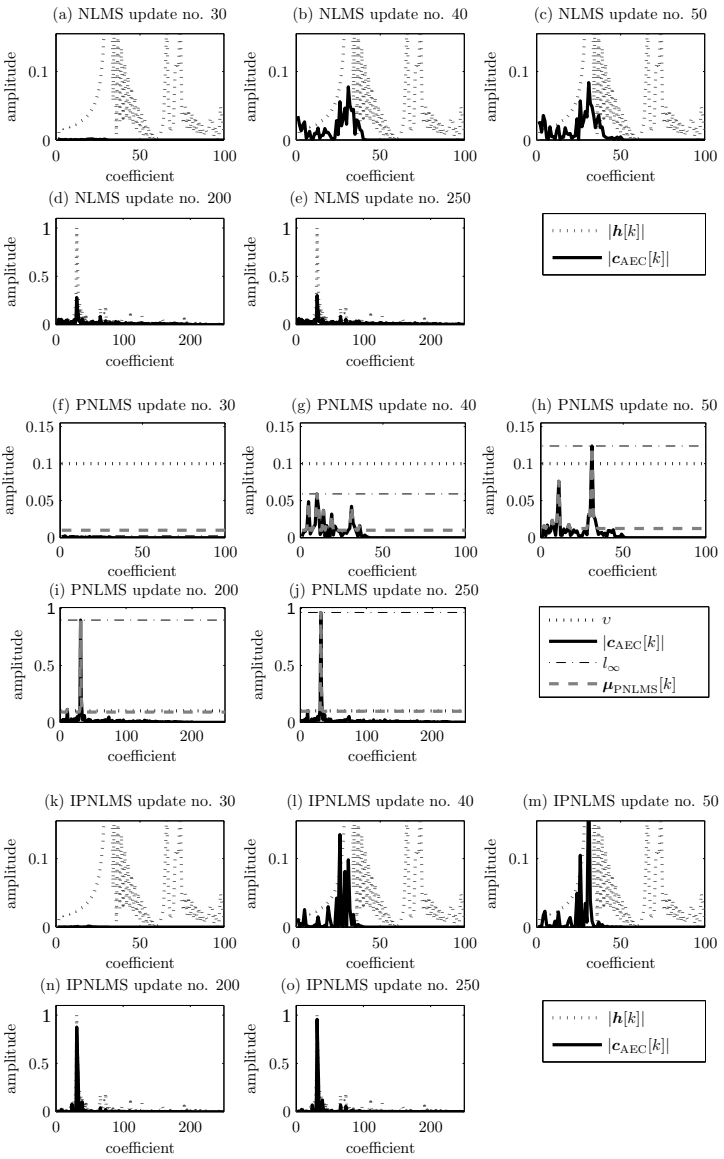

Figure 3.9: Comparison of the convergence of the NLMS algorithm (panels (a)-(e)), the PNLMS algorithm (panels (f)-(j)) and the IPNLMS algorithm (panels (k)-(o)). AEC filter length is $L_{\mathrm{AEC}} = 1024$. Room reverberation time of the RIR is $\tau_{60} \approx 50$ ms.

i.e. $\boldsymbol{h}[k] = \boldsymbol{h}$.

PNLMS algorithm: The PNLMS algorithm's convergence depicted in panels (f)-(j) shows that the update at prominent peaks (e.g. at coefficient $i = 30$) is much faster if the PNLMS algorithm is used compared to the NLMS algorithm's performance shown in panels (a)-(e). Panels (f)-(j) additionally visualize the influence of the parameter υ, given in (3.2.18). As visible in panels (f)-(h), a value of $\upsilon = 0.1$ has been chosen. The parameter υ prevents a dead-lock of the update if all coefficients have (initial) values of 0 and it ensures equal convergence in the very beginning. Once one coefficient of $|c_{\mathrm{AEC},i}[k]|$ is greater than υ, it will become ineffective which happens in panel (h)-(j) (also cf. (3.2.18)).
In the very beginning of the PNLMS algorithm's convergence, all coefficients $|c_{\mathrm{AEC},i}[k]|$ are smaller than $\upsilon \cdot \rho = 0.01$ (cf. (3.2.17) and (3.2.18)) as depicted in panel (f). Thus, all step-sizes $\mu'_{i,\mathrm{PNLMS}}[k]$ are equal. After a few update steps some coefficients become larger than $\upsilon \cdot \rho$ but are still smaller than υ as depicted in panel (g). Thus, the proportionate step-size $\boldsymbol{\mu}_{\mathrm{PNLMS}}[k]$ speeds up the convergence of those coefficients. After at least one coefficient becomes greater than υ as depicted in panels (h)-(j), the threshold is raised from $\upsilon \cdot \rho$ to $l_\infty[k] \cdot \rho$ since $l_\infty[k] \geq \upsilon$. By this, the convergence speed of the smaller coefficients is increased slightly as it can be observed from comparing $\boldsymbol{\mu}_{\mathrm{PNLMS}}[k]$ in panels (g) and (h).

IPNLMS algorithm: Panels (k)-(o) of Figure 3.9 show AEC filter coefficients $|\boldsymbol{c}_{\mathrm{AEC}}[k]|$ (solid black line) and the RIR $\boldsymbol{h}$ to be identified (dotted grey line) for the IPNLMS algorithm. The performance is comparable to that of the PNLMS algorithm and considerably better than for the NLMS algorithm.

A convergence comparison by the objective measures ERLE and system distance D_{dB} (cf. Section 3.1) is given in **Figures 3.10** to **3.13** for the four impulse responses depicted in Figure 3.6. Panels (a) of Figures 3.10 to 3.13 show the respective impulse response that is to be identified by the gradient algorithms. The ERLE and system distance D_{dB} are shown for a white input signal (left panels, (c) and (e)) and for a speech input (right panels, (d) and (f)), respectively. Panel (b) shows the speech input signal for the algorithms that was used to obtain the simulation results shown in the right panels. The following parameters were chosen for the simulations: the lengths of the respective impulse responses were $L_h = 4096$, the AEC filter lengths were $L_{\mathrm{AEC}} = 1024$, the step-sizes were $\mu[k] = 0.2$. For the PNLMS algorithm $\rho = 5/L_{\mathrm{AEC}}$ and $\upsilon = 0.01$ were chosen and for the IPNLMS $\alpha = -0.5$. The

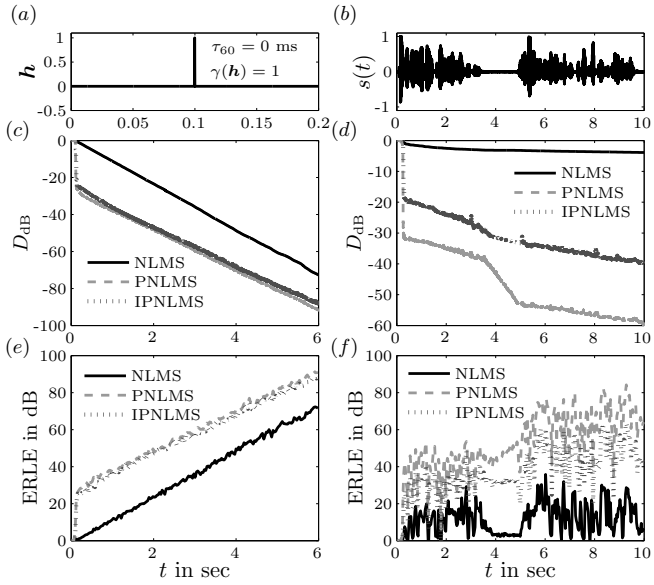

Figure 3.10: Convergence of NLMS, PNLMS and IPNLMS for the delayed delta function shown in Figure 3.6 (a); (a) first 200 ms of IR shown in Figure 3.6 (a); (b) speech excitation signal; (c) system distance in dB for white noise excitation; (d) system distance in dB for speech excitation; (e) ERLE in dB for white noise excitation; (f) ERLE in dB for speech excitation signal.

regularization parameter were $\delta_{\mathrm{NLMS}} = 0.01$, $\delta_{\mathrm{PNLMS}} = \delta_{\mathrm{NLMS}}/L_{\mathrm{AEC}}$ and $\delta_{\mathrm{IPNLMS}} = \delta_{\mathrm{NLMS}}/(2L_{\mathrm{AEC}})$.

The performance comparison for the most sparse impulse response, i.e. a delayed delta function, is shown in Figure 3.10. It can be seen that the initial convergence of PNLMS and IPNLMS is much faster than that of the NLMS. Especially for the speech input (right panels), the performance is drastically increased by the proportionate algorithms. Since the PNLMS is optimized for this maximally sparse impulse response its performance is even better than the performance of the IPNLMS which is more obvious for the speech input signal (right panels) than for the white input signal (left panels).

Figure 3.11 shows the performance comparison for the somewhat more dispersive RIR depicted in Figure 3.6 (b) whose first 200 ms are also shown in panel (a) of Figure 3.11. The reverberation time of $\tau_{60} = 100$ ms cor-

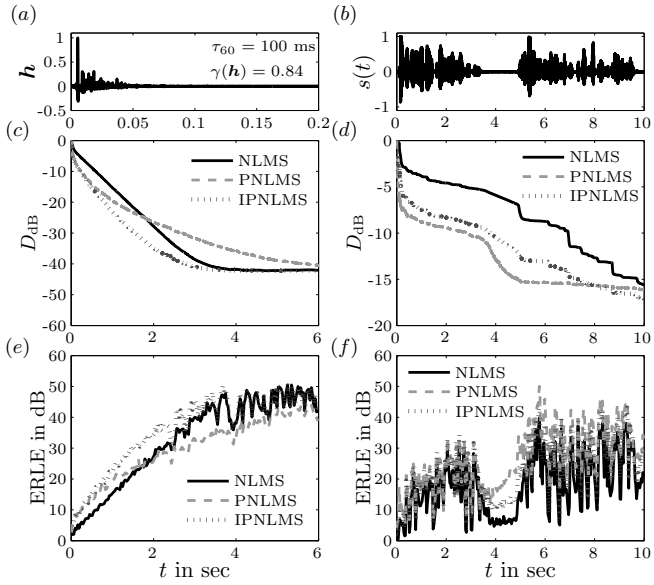

Figure 3.11: Convergence of NLMS, PNLMS and IPNLMS for the RIR shown in Figure 3.6 (b); (a) first 200 ms of RIR shown in Figure 3.6 (b) ($\tau_{60} = 100$ ms); (b) speech excitation signal; (c) system distance in dB for white noise excitation; (d) system distance in dB for speech excitation; (e) ERLE in dB for white noise excitation; (f) ERLE in dB for speech excitation signal.

responds to a small and acoustically dry room. Initial convergence of the proportionate algorithms still is faster than for the NLMS algorithm. However, at least for the white input signal (left panels) the performance of the PNLMS algorithms becomes worse than that of the NLMS algorithm after a few seconds. The IPNLMS algorithm as a trade-off between NLMS and PNLMS always shows better performance than the NLMS algorithm.
Simulation results for a RIR characterised by a room reverberation time of $\tau_{60} = 500$ ms are shown in Figure 3.12. Such an RIR can e.g. be observed in a common office environment. Here, the tendency already observed in Figure 3.11 can also be observed for the speech signal (right panels), i.e. that the PNLMS algorithm performs worse than the conventional NLMS algorithm. Still, the IPNLMS algorithm shows good performance for white noise input as well as for speech input.
Chapter 5 of this thesis discusses combinations of AECs and equalizers. A

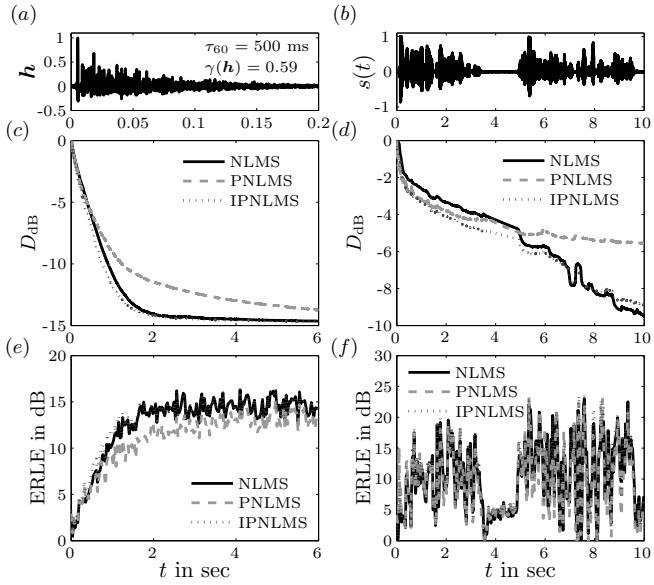

Figure 3.12: Convergence of NLMS, PNLMS and IPNLMS for the RIR shown in Figure 3.6 (c); (a) first 200 ms of RIR shown in Figure 3.6 (c) ($\tau_{60} = 500$ ms); (b) speech excitation signal; (c) system distance in dB for white noise excitation; (d) system distance in dB for speech excitation; (e) ERLE in dB for white noise excitation; (f) ERLE in dB for speech excitation signal.

schematic for an AEC that aims at identification of an equalized IR was already shown in Figure 3.7. Simulation results in Figure 3.13 compare the performance of NLMS algorithm, PNLMS algorithm and IPNLMS algorithm for such an equalized IR (cf. also Figure 3.6 (d)).
It can be seen from Figure 3.13 that the performance of PNLMS and IPNLMS is similar, but both algorithms outperform the conventional NLMS.
From the previously shown simulation results, it is not difficult to draw the conclusion that PNLMS behaves better than NLMS only if the RIR is sparse, while IPNLMS converges better than PNLMS when the RIR is dispersive. Actually, IPNLMS performs best independent of the nature of the RIR for Gaussian white noise excitation. For speech as input signal, IPNLMS with $\alpha = 0$ always leads to a good performance, however, not to the best performance in any case. An optimum α for the IPNLMS depends

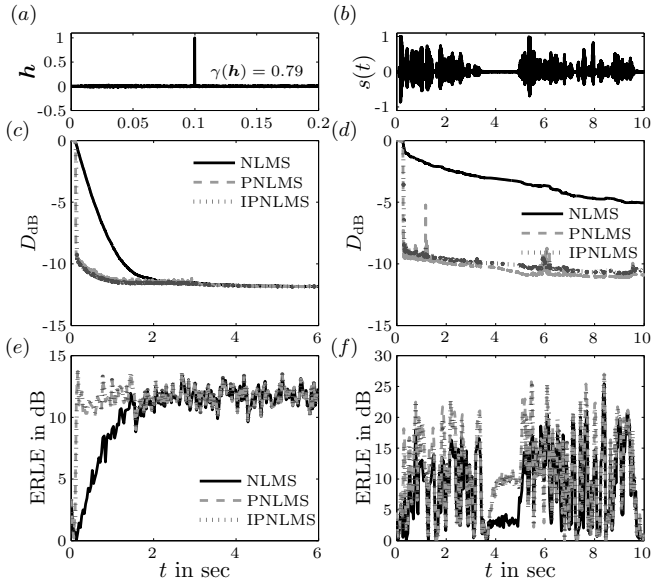

Figure 3.13: Convergence of NLMS, PNLMS IPNLMS and for the equalized IR shown in Figure 3.6 (d); (a) first 200 ms of IR shown in Figure 3.6 (d); (b) speech excitation signal; (c) system distance in dB for white noise excitation; (d) system distance in dB for speech excitation; (e) ERLE in dB for white noise excitation; (f) ERLE in dB for speech excitation signal.

on the nature of the RIR. However, for equalized IRs such as in Figure 3.13 proportionate update schemes are clearly preferable over conventional algorithms.

3.3 Post-Filters for Residual Echo Suppression

Although the acoustic echo $\psi[k]$ theoretically can be removed from the microphone signal $y[k]$ by the previously described AEC filter approaches, in general, a residual echo

$$\xi[k] = \psi[k] - \hat{\psi}[k] \tag{3.3.1}$$

remains in the AEC error signal

$$e_{\mathrm{AEC}}[k] = s_n[k] + n[k] + \xi[k], \tag{3.3.2}$$

$$= s_n[k] + n[k] + \psi[k] - \hat{\psi}[k] \tag{3.3.3}$$

after the compensation point of the AEC. This is mainly due to the facts that (i) the AEC filter length L_{AEC} is too low to model the RIR and (ii) that the convergence of the AEC filter is imperfect, in general, due to time-varying impulse responses and correlated input signals [Hay02]. The residual echo signal $\xi[k]$ in $e_{\mathrm{AEC}}[k]$ can be further reduced by so-called acoustic echo suppression (AES) filters $\boldsymbol{p}[k]$ which are also known as post-filters [MV96, TGS97a, GMV98, HS00, HS04, GKK05, GKMK06b] since they generally succeed the AEC filter as depicted in **Figure 3.14**.

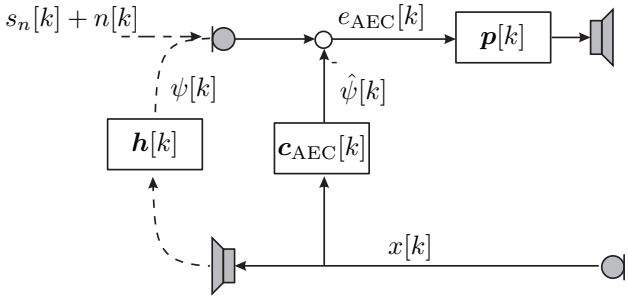

Figure 3.14: Schematic of an acoustic echo reduction system composed of AEC filter $\boldsymbol{c}_{\mathrm{AEC}}[k]$ and an AES post-filter $\boldsymbol{p}[k]$.

It was shown in [AF95, FB95b, BF96, MV96, BSFB01] that this arrangement of AEC and AES filter is mathematically optimal for echo suppression. Structures applying suppression filters in front of conventional AEC filters are usually not used, mainly because of the different adaptation speeds of conventional AEC filters and AES filters.
The filter coefficients

$$\boldsymbol{p}[k] = \big[p_0[k], p_1[k], \ldots, p_{L_p-1}[k]\big]^T \tag{3.3.4}$$

for a post-filter of length L_p can be obtained by minimizing the mean squared error [Hay02]

$$\mathrm{E}\left\{e_{\mathrm{PF}}^2[k]\right\} = \mathrm{E}\left\{|\boldsymbol{p}^T[k]\boldsymbol{e}_{\mathrm{AEC}}[k] - s_n[k]|^2\right\} \stackrel{!}{=} \min \tag{3.3.5}$$

with the AEC error vector

$$\boldsymbol{e}_{\mathrm{AEC}}[k] = [e_{\mathrm{AEC}}[k], \ldots, e_{\mathrm{AEC}}[k - L_p + 1]]^T \tag{3.3.6}$$

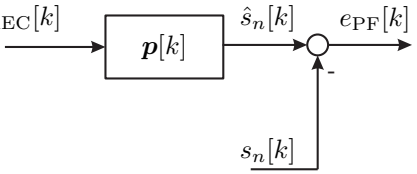

Figure 3.15: Schematic for AES post-filter design.

as illustrated in **Figure 3.15**.
The obtained MMSE solution

$$\boldsymbol{p}[k] = \mathrm{E}\left\{\boldsymbol{e}_{\mathrm{AEC}}[k]\,\boldsymbol{e}_{\mathrm{AEC}}^{T}[k]\right\}^{-1}\mathrm{E}\left\{\boldsymbol{e}_{\mathrm{AEC}}[k]\,s_n[k]\right\} \tag{3.3.7}$$

usually is applied in the short-term frequency-domain. The frequency-domain AES weighting function reads [Hay02]

$$\mathbf{p}[\ell] = \boldsymbol{\Phi}_{\mathsf{e}_{\mathrm{AEC}}\,\mathsf{s}_n}[\ell] \oslash \boldsymbol{\Phi}_{\mathsf{e}_{\mathrm{AEC}}\,\mathsf{e}_{\mathrm{AEC}}}[\ell], \tag{3.3.8}$$

with $\boldsymbol{\Phi}_{\mathsf{e}_{\mathrm{AEC}}\,\mathsf{s}_n}[\ell]$ and $\boldsymbol{\Phi}_{\mathsf{e}_{\mathrm{AEC}}\,\mathsf{e}_{\mathrm{AEC}}}[\ell]$ being the cross power spectral density (CPSD) vector of the AEC error signal and the near-end speaker's signal and the auto power spectral density (APSD) vector of the AEC error signal, respectively. The symbol $\oslash$ represents the element-by-element division of two vectors. Assuming absence of noise disturbance $n[k]$ and that the AEC error signal and the signal of the near-end speaker are uncorrelated, i.e. that

$$\boldsymbol{\Phi}_{\mathsf{e}_{\mathrm{AEC}}\,\mathsf{s}_n}[\ell] = \boldsymbol{\Phi}_{\mathsf{s}_n\,\mathsf{s}_n}[\ell], \tag{3.3.9}$$

and

$$\boldsymbol{\Phi}_{\mathsf{e}_{\mathrm{AEC}}\,\mathsf{e}_{\mathrm{AEC}}}[\ell] = \boldsymbol{\Phi}_{\mathsf{s}_n\,\mathsf{s}_n}[\ell] + \boldsymbol{\Phi}_{\boldsymbol{\xi}\,\boldsymbol{\xi}}[\ell], \tag{3.3.10}$$

Eq. (3.3.8) can be rewritten,

$$\mathbf{p}[\ell] = (\boldsymbol{\Phi}_{\mathsf{e}_{\mathrm{AEC}}\,\mathsf{e}_{\mathrm{AEC}}}[\ell] - \boldsymbol{\Phi}_{\boldsymbol{\xi}\,\boldsymbol{\xi}}[\ell]) \oslash \boldsymbol{\Phi}_{\mathsf{e}_{\mathrm{AEC}}\,\mathsf{e}_{\mathrm{AEC}}}[\ell]. \tag{3.3.11}$$

For practical application (3.3.11) is often generalized to

$$\mathbf{p}[\ell] = ((\mathrm{E}\left\{|\mathbf{e}_{\mathrm{AEC}}[\ell]|^{\alpha}\right\} - \beta\mathrm{E}\left\{|\boldsymbol{\xi}[\ell]|^{\alpha}\right\}) \oslash \mathrm{E}\left\{|\mathbf{e}_{\mathrm{AEC}}[\ell]|^{\alpha}\right\})^{\gamma} \tag{3.3.12}$$

with α, β and γ being filter design parameters to control the echo suppression performance. Eqs. (3.3.11) and (3.3.12) are commonly known as spectral subtraction which exists in various slightly different versions, cf. e.g. [Bol79, BSM79, MS97, GKMK06b]. Since expectations necessary

for calculating the post-filter weighting rules have to be estimated properly, the weighting rule is normally limited to $\mathbf{p}_{min}[\ell] \leq \mathbf{p}[\ell] \leq 1$ for all frequency bins to avoid amplification on the one hand and negative values or too high attenuation on the other hand. The maximum suppression $\mathbf{p}_{min}[\ell]$ can be chosen to a fixed value (20-40 dB is often used for AES) or in dependence of psychoacoustical findings to reduce the so-called musical-noise problem [Gus99, Fal03, GMK06a].

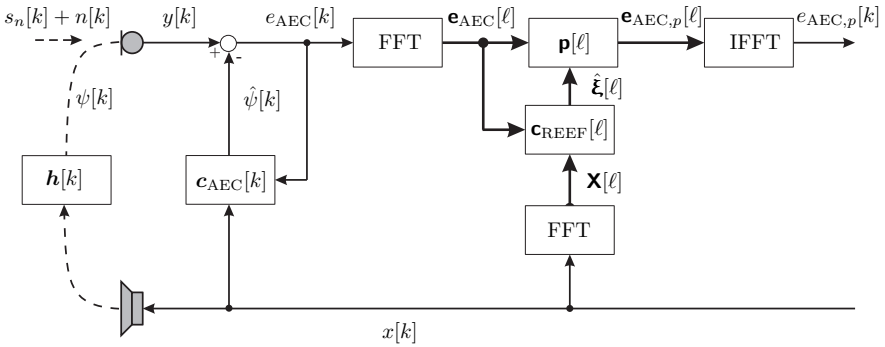

Figure 3.16: Schematic of an time-domain acoustic echo cancellation filter $\boldsymbol{c}_{\mathrm{AEC}}[k]$ with subsequent frequency-domain post-filter $\mathbf{p}[\ell]$. An estimate of the residual echo which is used by the post-filter to calculate the residual echo PSD is generated by a residual echo estimation filter (REEF) $\mathbf{c}_{\mathrm{REEF}}[\ell]$.

A schematic of the acoustic echo reduction system containing an AEC filter $\boldsymbol{c}_{\mathrm{AEC}}[k]$ and an AES post-filter $\mathbf{p}[\ell]$ is depicted in **Figure 3.16**. The AEC filter is applied and updated in time-domain in Figure 3.16 since only time-domain AEC gradient algorithms have been described so far. Of course a frequency-domain update rule for the AEC, e.g. as in [Her05], would be also possible in Figure 3.16. The loudspeaker signal $x[k]$ and the AEC output signal $e_{\mathrm{AEC}}[k]$ are transformed to the frequency-domain to result in $\mathbf{X}[\ell]$ as defined in (2.2.15) on page 22 and

$$\mathbf{e}_{\mathrm{AEC}}[\ell] = \boldsymbol{F}_{2L\times 2L}\boldsymbol{W}^{01}_{2L\times L}\boldsymbol{e}_{\mathrm{AEC}}[\ell]. \tag{3.3.13}$$

The echo suppression filter $\mathbf{p}[\ell]$ is applied to the matrix

$$\mathbf{E}_{\mathrm{AEC}}[\ell] = \mathrm{diag}\{\boldsymbol{F}_{2L\times L}\boldsymbol{e}_{\mathrm{AEC}}[\ell] + \tilde{\boldsymbol{I}}_{2L\times 2L}\,\boldsymbol{F}_{2L\times L}\boldsymbol{e}_{\mathrm{AEC}}[\ell-1]\} \tag{3.3.14}$$

containing the short-term spectra of the last two blocks of the AEC error signal $e_{\mathrm{AEC}}[k]$, after it has been transformed to the frequency-domain.

$$\mathbf{e}_{\mathrm{AEC},p}[\ell] = \boldsymbol{G}\,\mathbf{E}_{\mathrm{AEC}}[\ell]\mathbf{p}[\ell] \tag{3.3.15}$$

Since the filter $\mathbf{p}[\ell]$ is applied in the signal path, it always affects both, the residual echo and the desired signal part. Hence, post-filters always lead to a certain amount of distortion of the desired signal. On the other hand, convergence of post-filters is generally much faster than for AEC filters and, in general, the calculation of the weighting rule is not restricted to system identification [Fal03, FFK+08b] since $\mathbf{p}[\ell]$ only depends on reliably estimated PSDs of input signal and residual echo as it can be seen from (3.3.11). In the following, a method to obtain the residual echo PSD $\mathbf{\Phi}_{\hat{\xi}\,\hat{\xi}}[\ell]$ by means of a so-called REEF $\mathbf{c}_{\mathrm{REEF}}[\ell]$ which performs system identification is chosen. In general, various ways exist to obtain $\mathbf{\Phi}_{\hat{\xi}\,\hat{\xi}}[\ell]$ also without REEF, however, since the information about the impulse response will be necessary for the LRC algorithms described in Chapters 4 and 5 [GKK05, Kal07, GMK06a], the REEF will be used in the following to calculate an estimate of the residual echo $\hat{\boldsymbol{\xi}}[\ell]$ as input for the AES filter $\mathbf{p}[\ell]$.
To obtain an estimate of the residual echo, the system misalignment vector $\tilde{\boldsymbol{h}}[k]$ as defined in (3.1.1) and depicted in Figures 3.2 and 3.3 on page 34f. is the IR that has to be identified by an adaptive filter. For this purpose, a simplified structure to calculate an update rule for the REEF $\mathbf{c}_{\mathrm{REEF}}[\ell]$ is shown in **Figure 3.17**.

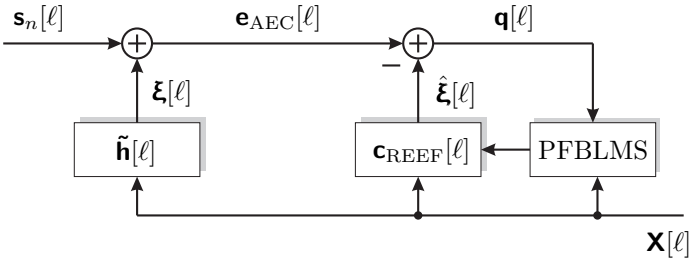

Figure 3.17: Definition of block-error signal $\boldsymbol{q}[\ell]$ for post-filter design in frequency-domain updated by the partitioned frequency block LMS (PFBLMS) algorithm.

To derive an update rule for the REEF, the time-domain block error signal

$$\begin{aligned} \boldsymbol{q}[\ell] &= \boldsymbol{e}_{\mathrm{AEC}}[\ell] - \hat{\boldsymbol{\xi}}[\ell] && (3.3.16)\\ &= \boldsymbol{e}_{\mathrm{AEC}}[\ell] - \boldsymbol{W}^{01}_{L\times 2L}\boldsymbol{F}^{-1}_{2L\times 2L}\mathbf{X}[\ell]\mathbf{c}_{\mathrm{REEF}}[\ell-1] && (3.3.17)\end{aligned}$$

is transformed to the frequency-domain by multiplication with $\boldsymbol{F}_{2L\times 2L}\boldsymbol{W}^{01}_{2L\times L}$ including zero-padding as described in (2.2.7) and using the definition of the constraining matrix $\boldsymbol{G}$ in (2.2.21).

$$\mathbf{q}[\ell] = \mathbf{e}_{\mathrm{AEC}}[\ell] - \boldsymbol{G}\mathbf{X}[\ell]\mathbf{c}_{\mathrm{REEF}}[\ell-1] \qquad (3.3.18)$$

The RLS-like frequency-domain criterion for the optimization of the REEF $\mathbf{c}_{\mathrm{REEF}}[\ell]$ can be defined as [Hay02, BM01]

$$J[\ell] = (1-\alpha)\sum_{i=0}^{\ell}\alpha^{\ell-i}\mathbf{q}^H[i]\mathbf{q}[i]. \tag{3.3.19}$$

To find the minimum of the error function in (3.3.19), the gradient [BR72]

$$\nabla_{\mathbf{c}_{\mathrm{REEF}}} J[\ell] = 2\frac{\partial J[\ell]}{\partial \mathbf{c}^*_{\mathrm{REEF}}} \tag{3.3.20}$$

$$= 2\frac{\partial}{\partial \mathbf{c}^*_{\mathrm{REEF}}}(1-\alpha)\sum_{i=0}^{\ell}\alpha^{\ell-i}\left(\mathbf{e}_{\mathrm{AEC}}[i] - \boldsymbol{G}\mathbf{X}[i]\mathbf{c}_{\mathrm{REEF}}\right)^H \cdot \left(\mathbf{e}_{\mathrm{AEC}}[i] - \boldsymbol{G}\mathbf{X}[i]\mathbf{c}_{\mathrm{REEF}}\right) \tag{3.3.21}$$

has to be calculated. Using the Wirtinger calculus [Hay02],

$$\frac{\partial \mathbf{c}^*_{\mathrm{REEF}}}{\partial \mathbf{c}^*_{\mathrm{REEF}}} = \boldsymbol{I}, \qquad \frac{\partial \mathbf{c}_{\mathrm{REEF}}}{\partial \mathbf{c}^*_{\mathrm{REEF}}} = \mathbf{0}, \tag{3.3.22}$$

and setting (3.3.21) to zero we obtain

$$0 \stackrel{!}{=} 2(1-\alpha)\sum_{i=0}^{\ell}\alpha^{\ell-i}\left(-\mathbf{X}^H[i]\boldsymbol{G}^H\right)\left(\mathbf{e}_{\mathrm{AEC}}[i] - \boldsymbol{G}\mathbf{X}[i]\mathbf{c}_{\mathrm{REEF}}\right). \tag{3.3.23}$$

With $\boldsymbol{G}^H\boldsymbol{G} = \boldsymbol{G}$ and $\boldsymbol{G}^H\mathbf{e}_{\mathrm{AEC}}[\ell] = \mathbf{e}_{\mathrm{AEC}}[\ell]$ (cf. Appendix D.1 and D.2 for proof) we obtain the frequency-domain normal equation

$$\hat{\boldsymbol{\Phi}}_{\mathsf{xx}}[\ell]\mathbf{c}_{\mathrm{REEF}}[\ell] = \hat{\boldsymbol{\phi}}_{\mathsf{xe}}[\ell] \tag{3.3.24}$$

with the CPSD vector between loudspeaker signal and AEC error signal

$$\hat{\boldsymbol{\phi}}_{\mathsf{xe}}[\ell] = (1-\alpha)\sum_{i=0}^{\ell}\alpha^{\ell-i}\mathbf{X}^H[i]\mathbf{e}_{\mathrm{AEC}}[i] \tag{3.3.25}$$

and the APSD matrix of the loudspeaker signal

$$\hat{\boldsymbol{\Phi}}_{\mathsf{xx}}[\ell] = (1-\alpha)\sum_{i=0}^{\ell}\alpha^{\ell-i}\mathbf{X}^H[i]\boldsymbol{G}\mathbf{X}[i]. \tag{3.3.26}$$

To obtain an iterative update equation, (3.3.25) can be rewritten in its recursive form that can easily be obtained by extracting $\alpha \mathbf{X}^H[i]\mathbf{e}_{\mathrm{AEC}}[i]$ from the sum in (3.3.25) and reintroducing $\hat{\boldsymbol{\phi}}_{\mathsf{xe}}[\ell-1]$ to result in (3.3.27). Similarly, (3.3.28) can be obtained from (3.3.26).

$$\hat{\boldsymbol{\phi}}_{\mathsf{xe}}[\ell] = \alpha\hat{\boldsymbol{\phi}}_{\mathsf{xe}}[\ell-1] + (1-\alpha)\mathbf{X}^H[\ell]\mathbf{e}_{\mathrm{AEC}}[\ell] \tag{3.3.27}$$

$$\hat{\boldsymbol{\Phi}}_{\mathsf{xx}}[\ell] = \alpha\hat{\boldsymbol{\Phi}}_{\mathsf{xx}}[\ell-1] + (1-\alpha)\mathbf{X}^H[\ell]\boldsymbol{G}\mathbf{X}[\ell]. \tag{3.3.28}$$

In (3.3.27) and (3.3.28), $0 \leq \alpha \leq 1$ is an exponential forgetting factor which is usually chosen close to one for speech PSD estimation. The normalization factor $(1-\alpha)$ assures an asymptotically unbiased estimate [Bri75, Her05]. Introducing (3.3.24) in terms of ℓ and $\ell-1$ for $\hat{\boldsymbol{\phi}}_{\mathsf{xe}}[\ell]$ and $\hat{\boldsymbol{\phi}}_{\mathsf{xe}}[\ell-1]$ in (3.3.27) leads to

$$\hat{\boldsymbol{\Phi}}_{\mathsf{xx}}[\ell]\mathbf{c}_{\mathrm{REEF}}[\ell] = \alpha\hat{\boldsymbol{\Phi}}_{\mathsf{xx}}[\ell-1]\mathbf{c}_{\mathrm{REEF}}[\ell-1] + (1-\alpha)\mathbf{X}^H[\ell]\mathbf{e}_{\mathrm{AEC}}[\ell] \tag{3.3.29}$$

where the dependency from $\hat{\boldsymbol{\Phi}}_{\mathsf{xx}}[\ell-1]$ can be eliminated using (3.3.28). With the definition of the frequency-domain error vector as already defined in (3.3.18), the update equation can be given as

$$\mathbf{q}[\ell] = \mathbf{e}_{\mathrm{AEC}}[\ell] - \boldsymbol{G}\mathbf{X}[\ell]\mathbf{c}_{\mathrm{REEF}}[\ell-1] \tag{3.3.30}$$

$$\mathbf{c}_{\mathrm{REEF}}[\ell] = \mathbf{c}_{\mathrm{REEF}}[\ell-1] + (1-\alpha)\boldsymbol{M}_{\mathrm{REEF}}[\ell]\hat{\boldsymbol{\Phi}}_{\mathsf{xx}}^{-1}[\ell]\mathbf{X}^H[\ell]\mathbf{q}[\ell] \tag{3.3.31}$$

The diagonal matrix $\boldsymbol{M}_{\mathrm{REEF}}[\ell]$ contains the step-size vector on its main diagonal. The theoretical optimum step-size is $\boldsymbol{M}_{\mathrm{REEF}}[\ell] = 2\boldsymbol{I}$ [Her05], which is decreased in periods of an active near-end signal to decrease adaptation speed of the REEF. For an overview of algorithms for step-size control the interested reader is referred to [MPS00]. For the following simulations a so-called shadow filter approach is applied for step-size control.
Simulation results for the combined system consisting of conventional AEC filter and AES filter are shown in **Figure 3.18**. The lengths of the AEC filter and the REEF were chosen to $L_{\mathrm{AEC}} = 1024$ and $L_{\mathrm{REEF}} = 2048$, respectively, at a sampling rate of $f_s = 8000$ Hz. Panel (a) of Figure 3.18 shows the echo part $\psi[k] = h[k] * x[k]$ contained in the microphone signal in dark grey and the the near-end speaker's signal $s_n[k]$ in light grey (cf. also Figure 3.16). The captured microphone signal is, thus, the superposition of both signals depicted in panel (a) and the aim of the echo reduction system (cf. Figure 3.14) is to remove the echo, i.e. the far-end speaker's signal, without affecting the near-end speaker's signal $s_n[k]$ (light grey). Panels (b) and (c) show the achieved ERLE for AEC filter (dark grey), AES post-filter (medium grey) and the overall system (light grey) and the corresponding

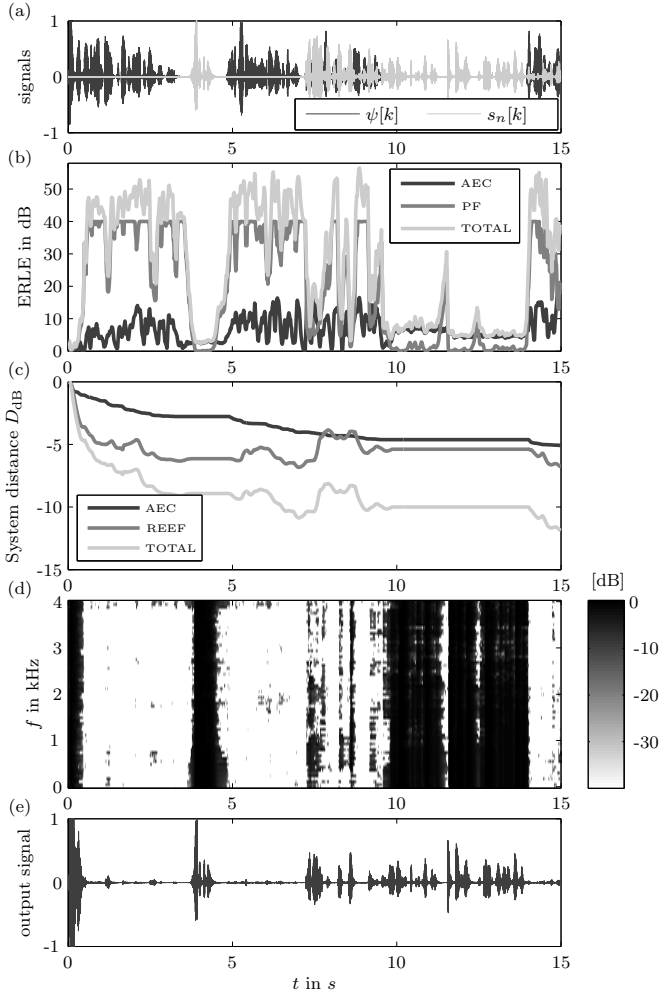

Figure 3.18: Performance of AEC filter and AES post-filter. Near-end signal $s_n[k]$ and echo signal $\psi[k]$ are shown in panel (a). Achieved performance in terms of ERLE and normalized system distance D_{dB} is shown in panels (b) and (c), respectively. Panel (d) shows the post-filter's transfer function over time and panel (e) the resulting output signal $e_{\mathrm{AEC},p}[k]$ after processing by AEC filter and AES post-filter.

system distance D_{dB} of AEC filter (dark grey), REEF (medium grey) and the combines system distance (light grey), respectively. It can be seen from panels (b) and (c) that the echo reduction of the conventional AEC filter is limited and that most of the performance (especially in terms of ERLE) is achieved by the AES filter. Since the evaluation of the ERLE measure and the system distance alone may not be sufficient to assess the performance of the echo reduction system, the AES transfer function in dependence of time and frequency is depicted in panel (d) and the output signal $e_{\mathrm{AEC},p}[k]$ is shown in panel (e). After initial convergence, the echo part is reduced by the AEC/AES system while the desired signal, i.e. the near-end speaker's signal $s_n[k]$, is transmitted without being affected too much. As visible from the system distances in panel (c), the REEF converges much faster than the conventional AEC despite its higher filter order.

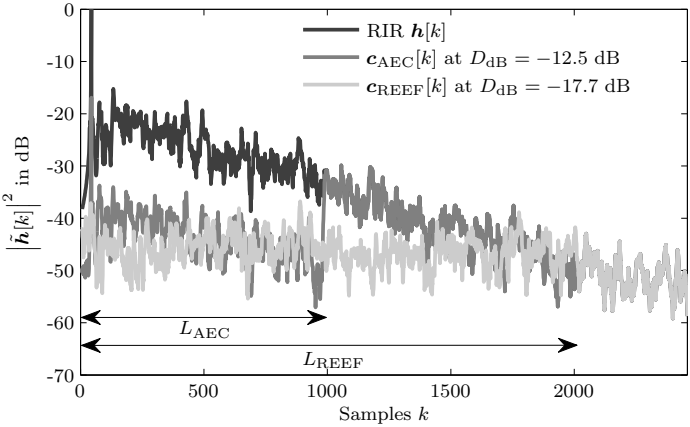

Figure 3.19: System distance of AEC and REEF.

Figure 3.19 compares the system distance vectors after a certain period of filter convergence. It is obvious that the system identification capability of the REEF is higher due to its higher filter order ($L_{\mathrm{REEF}} > L_{\mathrm{AEC}}$). In the following chapters a quickly converging system identification of sufficient quality is essential for RIR equalization. The developed structure is capable to archive the desired performance and is already designed to be integrated into the block-frequency LRC filters which will be developed in the next chapters.

3.4 Chapter Summary

This chapter introduced different algorithms for acoustic echo cancellation that will be combined in Chapter 5 with the algorithms for listening-room compensation introduced in the following chapter. Algorithms for LRC need knowledge about the RIR which can be obtained by the AEC approaches described in this chapter. However, RIR estimates obtained by AEC filters or AES post-filters always are erroneous, e.g. due to the AEC tail-effect. AEC filters suffer from slow convergence, especially for long RIRs. The discussed AES filters converge faster, however still do not lead to a perfect system identification. Effects of these practical limitations of AEC and AES filters on the LRC sub-systems will thus be analyzed in the following chapters.

For the system identification of equalized systems, the discussed proportionate update schemes (cf. Section 3.2.2) seem to be particularly suited [GXJ+11]. A thorough evaluation of these algorithms for system identification of equalized IRs will be the topic of Chapter 5, as well as the use of the described conventional AECs [GKMK08d] and post-filters (cf. Section 3.3).

Chapter 4

Dereverberation by Listening-Room Compensation

Reverberation occurs naturally in enclosed spaces such as offices or living rooms due to multi-path propagation of the sound signal from the acoustic source to the microphone (cf. Figure 2.2 (a) on page 10). Reverberant speech can be described as sounding distant characterized by colouration and echo [NG05]. If a sound signal is transmitted in a reverberant environment, reflections at walls, ceiling and floor change the perceived sound signal in amplitude as well as in phase [RK00] (cf. also Figures 2.3 and 2.4 on pages 12f.). This influence can be described by the room impulse response (RIR) which can be modelled by a linear finite impulse response (FIR) system [Kut00] (cf. Section 2.1.1). Although humans are used to a moderate amount of reverberation, higher amounts of reverberation lead to a decreased speech intelligibility in hands-free scenarios [All82, Ber80, Hän92, IEC98] as it can be typically observed e.g. from speech signals in a church or gymnasium. In music signal processing, *adding* reverberation may be advantageous but as far as speech communication is concerned whose aim usually is to transmit information unaffectedly, *removing* reverberation from the speech signal and, hence, restoring the original, non-reverberant signal normally is desired.

In general, two distinct dereverberation classes exist, viz. reverberation suppression and reverberation cancellation. Reverberation suppression approaches focus on removing the reverberant part of the speech signal by cal-

culating a spectral weighting rule for each time-frequency coefficient similar to well-known approaches for noise reduction [Hab07, PN10]. Reverberation cancellation approaches remove the influence of the acoustic channel between the sound source and the listener by equalizing the corresponding RIR. Knowledge about the RIR can be obtained either by means of blind [YHC05] or non-blind [GKMK08d, EN89, Mou94] channel/system identification. Furthermore, filters for dereverberation of a speech signal can be applied at two different positions in a hands-free scenario aiming at dereverberation of either the microphone signal $y[k]$ which is a reverberant version of the near-end speaker's signal $s_n[k]$ or the far-end speakers signal $s_f[k]$ as it should be perceived at the position of the near-end listener. Filters that can be used for dereverberation were already shown in Figure 1.2 on page 2, denoted there as *post-filter* and *equalizer*. A filter $\boldsymbol{c}_{\mathrm{EQ}}[k]$ aiming at the removal of reverberation from the microphone signal $y[k]$ to provide a dereverberated signal to the far-end listener either by reverberation suppression or by reverberation cancellation is depicted in **Figure 4.1**.

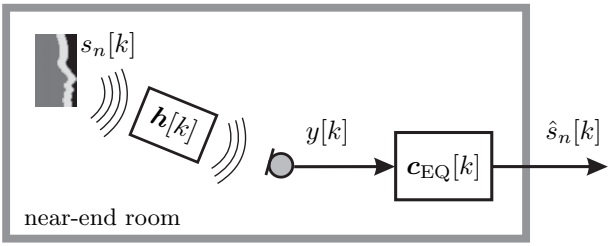

Figure 4.1: Dereverberation of the near-end speaker's signal $s_n[k]$.

Since neither the non-reverberant signal $s_n[k]$ nor the RIR between the speaker and the microphone are known, reverberation suppression and reverberation cancellation as depicted in Figure 4.1 usually leads to a blind estimation problem. Please note that the RIR between the near-end user's signal $s_n[k]$ and the microphone is time variant if the user moves and that an IR identification filter may estimate not only the RIR but also the mouth-room-impulse response if this impulse response is identified blindly. Here, the IR part corresponding to the human speech production system, i.e. the vocal tract, of course must not be equalized by the filter $\boldsymbol{c}_{\mathrm{EQ}}[k]$.

In contrast to the dereverberation of the microphone signal $y[k]$, also dereverberation of the far-end speaker's signal $s_f[k]$ can be desired aiming at a non-reverberant signal at the position of the near-end listener as depicted in **Figure 4.2**. Since for this approach the influence of the RIR $\boldsymbol{h}[k]$ between loudspeaker and near-end listener has to be cancelled by equaliza-

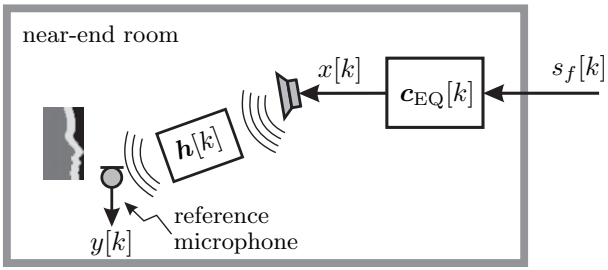

Figure 4.2: Listening-room compensation: Dereverberation of the far-end speaker's signal $s_f[k]$ at the position of the near-end listener.

tion, this reverberation cancellation approach is known as listening-room compensation (LRC). For LRC, the equalizer is applied to the signal that is emitted by the loudspeaker such that the influence of reverberation on the perceived signal is reduced at the position of a reference microphone where the near-end listener is assumed to be located. In order to compute the equalizer, knowledge of the RIR is required, which, in the context of LRC, is often obtained using non-blind system identification [GKMK08d, EN89, Mou94].

Although mathematically both structures in Figures 4.1 and 4.2 are equivalent at a first glance regarding the problem of reverberation cancellation, they behave differently in real-world systems considering imperfect channel knowledge, noise disturbances or spatial mismatch regarding assumed locations of microphones, loudspeakers and system users. The structure in Figure 4.2 naturally does not allow for reverberation suppression since the filter is located in front of the acoustic channel and reverberation suppression approaches usually only influence the magnitude of the signal spectrum.

This thesis focuses on the problem of LRC while reverberation suppression as well as blind reverberation cancellation as depicted in Figure 4.1 is out of the scope of this thesis. The interested reader is referred to the literature, e.g. [Hab07, HE08, PN10] and the references therein.

The problem of LRC is similar to pre-equalization approaches for data transmission used in mobile communications [WK03, Kam94, FM73]. Unfortunately, approaches developed for data transmission are only partly applicable since acoustic channels are generally of much higher length than those of e.g. discrete multitone (DMT) or orthogonal frequency division multiplexing (OFDM) systems [MYR96, KM05a] and sending training sequences via the acoustic channel is practically not feasible.

Algorithms for acoustic LRC try to reconstruct the non-reverberant signal $s_f[k]$ at the position of a reference microphone by designing an equalization

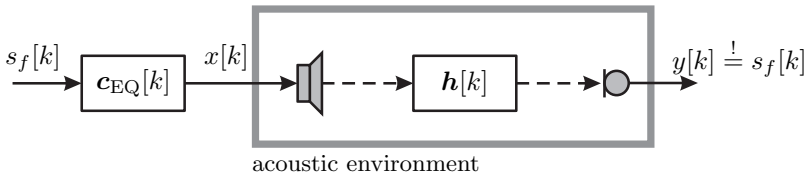

Figure 4.3: General setup for listening-room compensation.

filter $\boldsymbol{c}_{\mathrm{EQ}}[k]$ as depicted in **Figure 4.3** aiming at reconstruction of the non-reverberant signal $s_f[k]$ at the position of the reference microphone where the human listener is assumed to be located. A detailed description of LRC approaches will be given in Section 4.3. Since in real-world systems the user of the system, or more precisely his or her ears that *receive* the sound, generally will be located in a certain spatial distance from the reference microphone(s), spatial robustness issues will be discussed in Section 4.4.2.

4.1 Literature Survey on Speech Dereverberation

In the following, a brief literature survey on different techniques for dereverberation of speech signals will be given, without claim of completeness, since there is an enormous and still growing number of relevant contributions that could be considered. However, the following survey hopefully provides a basic overview about the possibilities for removing reverberation from speech signals and provides the possibility to class the following work on LRC into the broader field of dereverberation research.

4.1.1 Inverse Filtering

Since the influence of the RIR can be modelled by a convolution with a linear FIR system [Kut00], the most straightforward dereverberation approach is deconvolution by the inverse system of the RIR [NA79]. Unfortunately, room impulse responses are, in general, mixed-phase systems having thousands of zeros inside and outside the unit circle in the z-domain (cf. Section 2.1.6). Thus, only their minimum-phase part can be inverted by a stable causal filter [NA79, RK00]. By allowing an additional delay, the causality may be guaranteed also if the maximum-phase part of the RIR is taken into account and the inversion may be done by least-squares approaches or homomorphic filtering [MCH82, OS89, Mou94]. However, it depends on

the specific application, whether the delay introduced is acceptable (cf. also Section 4.4.1). Furthermore, zeros of the RIR close to the unit circle in z-domain lead to deep dips in the corresponding RTF (cf. Figure 2.3 on p. 12) which have to be compensated by high peaks in the transfer function of the equalization filter, i.e. a high amount of energy has to be spent for amplification of single frequencies.

The direct inversion of RIRs is only valid for the specific RIR for which an equalizer is calculated for. Thus, it depends on the exact position of sound source and sound pick-up. Unfortunately, it is very error-prone in relation to changes in the acoustic environment. Spatial restrictions have been investigated e.g. in [Mou85, RWK99, RWK00, GKMK08c], and it has been shown that when inverting an RIR, small deviations in relation to the positions of source and microphone lead to severe loss in quality (cf. also Section 4.4.2 for analysis of spatial robustness). To tackle this issue clustering of all possible RIRs by means of vector quantization has been proposed in [Mou94] aiming at the inversion of the spatially closest RIR available. However, the effort for measuring and grouping of all RIRs for any source-microphone combination for each region is extremely high and, thus, impracticable. In [MP91, HMK94, HMK97] the assumption is made that common acoustical poles that correspond to resonance properties of a RTF will only change little, when changing the positions of source and microphone. However, this assumption does not hold for general case as described in [RWK00].

Another problem in practice is that many authors assume perfect knowledge of the room impulse response to calculate the inverse. However in real systems the RIR has to be identified either by gradient algorithms [EN89, ZKN08, GKMK08d, GKMK08b] or by measuring [BA83, RV89b, Van94] and, hence, usually the identified RIR differs from the true one. Therefore, approaches that incorporate a regularization parameter in the filter design [HDM07, GKMK08d] (cf. also Section 4.4.2), the use of a truncated singular value decomposition approach [NTSS04] or only partly inverting the acoustic channel [KD12] have been proposed. Sections 4.4.2 and 4.5 of this thesis will focus on the influence of non-perfect RIR estimates.

4.1.2 Multi-channel Inverse Filtering

Exploiting spatial diversity by using multiple loudspeakers and multiple microphones can increase the performance of the equalizer as well as the spatial robustness. For that reason the inverse filtering approach [NA79] was extended in [MK86] to a single input multiple output (SIMO) system using one loudspeaker and several reference microphones. By this, parallel equalization for spatially separated microphone positions is achieved result-

ing in higher spatial robustness [GKMK08c, GKMK08b]. In [MK88] this approach is extended to multiple input single output (MISO) systems and the general case of multiple input multiple output (MIMO) systems (cf. also Section 4.4.3). For MIMO systems several loudspeakers and microphones are placed at different spatial positions. This approach is known as multiple input/output inverse theorem (MINT) and allows the exact inversion of RIRs if the assumption holds that their z-transforms do not have common zeros [MK88, Wan95]. However, exactly this requirement leads to problems in practical systems because common RIRs have lengths of several thousand coefficients and have thousands of zeros that are very close to each other (cf. Figure 2.7 on page 17). Thus, the probability of joint or closely spaced zeros is very high [GBN05, LGN06, KLN08].

4.1.3 Equalization

Methods for LRC that are based on minimum mean squared error (MMSE) approaches are better suited for practical applications for dereverberation [EN89, NOBH95, Ged98, KN99, KNHOb98, Fie01, GKMK08b]. They are at least partly able to prevent the problems of the previously described deconvolution approaches by means of direct RIR inversion. MMSE approaches minimize the Euclidean distance between a given desired system and the overall impulse response of the concatenated system of RIR and equalizer filter (cf. Section 4.4). The desired system usually is a delayed impulse, band-pass or high-pass [GKMK08c] (cf. also Section 4.4.1 for a proper choice of this delay). In [EN89], a concept for adaptive MMSE equalization of several selected discrete points in space is presented.

The human auditory system is capable to jointly perceive the influence of the first 50 ms of an IR. Energy arriving within 50 ms increases intelligibility of speech signals while energy that arrives later than 50 ms decreases speech intelligibility [ISO97] since it is perceived as reverberation. The *definition* measure D_{50} as defined in (A.1.1) is capable to predict speech intelligibility by calculating the ratio of the energy of an impulse response within the first 50 ms to its total energy. Therefore, methods for RIR shortening to a limit of 50 ms are investigated in [KM05c, KM05b] and are extended to the more general approach of RIR shaping in [KM06, GKMK08c, MKM09b, MMK10]. Some of these approaches have been adapted from research in the field of mobile communications [FM73, Kam94, Mer99, SK00, AD01, Mer01, Sch01, WK03, Wüb06]. However, while typical radio channels can be modelled with a few coefficients [Kam08], typically several thousand coefficients are required for RIRs [BDH+99]. Furthermore, for mobile communication systems assumptions

about the signal statistics of the transmitted data like stationarity or Gaussianity may be made [Wüb06] while speech signals are non-stationary and highly correlated, in general. Furthermore, short-term statistics of speech are unknown, in general [VM06]. A further problem of RIR shortening, e.g. by simply optimizing objective measures such as the time-domain D_{50} measure, is that this procedure not necessarily leads to a better listening experience as reported in [KM05b], due to possible spectral distortions [JMGM11]. Thus, masking effects of the human auditory system [Fie01] should be considered that ensure a perceptually acceptable result [MMK10, JMGM11]. Outside the influence-length of an equalization filter or an impulse-response shortening filter, an increase in energy that becomes annoyingly perceivable may occur [GAK+10, GAR+10b]. This is due to the fact that late parts of the resulting equalized impulse response are no longer covered by temporal masking of the human auditory system [TO88, OT89, ZF99, Fie01, BMB01] and cause additional reverberation. A shaping of the impulse response to its approximately exponentially decaying character is, thus, preferable to the simple shortening [MMK10].

If perfect equalization of a given transfer function is desired, in other words the overall transfer function has to result in a flat frequency response, deep dips in the room transfer function caused by zeros close to the unit circle in z-domain have to be compensated by large peaks in the equalizer's transfer function. Therefore, a large amount of energy has to be spent for the equalization filter at those frequency points. Thus, e.g. [KNHOb98, KN99, KRF99] proposed a regularization of the equalizer design to avoid unnecessarily high transmission power [HDM07, GKMK08d, KGD12a, KGD13b].

4.1.4 Dereverberation by Means of Spatial Filtering

Beamforming microphone arrays [MM80, GZ91, VM06] and their extensions by multi-channel post-filters [e.g., BS01, SBM01, GMK06a, RGH+08a] are common approaches to exploit spatial information by spatially sampling a given sound field. They are commonly used for reduction of ambient noise, estimation of direction of arrival of a specific acoustic signal [e.g., KC76, Dob06, GRH+08, Roh08] or for reducing spatial interferences. Only signal components impinging from the assumed direction of arrival are added in phase by beamformers, while signal components from other directions are damped. Thus, besides the capability of microphone arrays to spatially separate sound sources also a certain degree of dereverberation can be achieved by beamforming and spatial post-filtering [CMS96, AG97, GSO98, SLS01, DM01, BM03, HBCG09, HBG+09]. Beamforming is a robust method to dereverberate signals even in environments

with high ambient noise since the transfer function of a beamformer can be designed based on the desired direction of sound arrival only. Moreover, common beamforming algorithms are easy to implement.

Most adaptive post-filters, which are capable to significantly improve the noise reduction capabilities of conventional beamformers [SBM01] only marginally contribute to dereverberation of a speech signal. Such algorithms were originally evaluated in [ABB77] and [BC82] for a 2-microphone setup. The reason for the poor capability to reduce reverberation is the strong correlation of the signal parts arriving at the microphones via the direct path and the early reflections of the RIR [CMS96, CMS98]. It should be noted that the previous statement is restricted to the class of post-filters evaluated in [ABB77] and [BC82] and that post-filters in general are capable to significantly contribute to dereverberation suppression if they are designed accordingly [Hab07].

It should be mentioned here that beamforming approaches, unlike the discussion in previous sections, dereverberate the signal of the near-end speaker that is picked up by the microphones. Thus, they aim at dereverberation of the near-end speaker's signal for the far-end listener. Since beamformers rely on information about the direction of the desired source that is unknown a priori and has to be estimated [e.g., KC76, Dob06, GRH$^+$08, Roh08], they can be considered as partly blind approaches.

4.1.5 Blind Dereverberation Approaches

Various approaches exist for the problem of blind dereverberation [AG97, BM03, Hab07, Hab08, PN10], e.g. by changing the prediction error signal [YM00, GB01c, BYR02, NG03, GA03] of a predictor filter or by exploiting the harmonic structure of speech [NM03, TNK03, TNM06]. Blind dereverberation is still a topic of active current research [Hab07, HCGS08, HE08, PN10] and will not be within the scope of this thesis.

4.1.6 Combined Approaches for Dereverberation and Suppression of other Disturbances

A 2-microphone system for joint suppression of echo, noise and reverberation was proposed in [MV93, MV94]. For that purpose a filter for suppression of noise and reverberation in the signal path is combined with a conventional AEC filter (cf. Section 3.2). As already mentioned before, such post-filters that were originally proposed for noise reduction only remove the uncorrelated parts of the reverberation [SBM01, GMK06b].

Contributions [SKR03a, SKR03b, SBR04c, SBR04a, SBR04b, SRR05] propose an approach for room equalization in combination with wave-field synthesis (WFS). This approach is extended in [BSKR02] to a system consisting of acoustic echo canceller, beamformer and room equalization for WFS [BdV93, SBR04c]. A further extension of the proposed WFS system is described in [HBK04, BSK04] that provides the possibility to reduce the number of AEC filter coefficients and, by this, exploits synergies between the echo cancellation and the WFS sub-system.
Further work on combination of dereverberation and noise reduction can be found in [DM01] and the combination of AEC and dereverberation suppression is tackled e.g. in [Hab07, Hab08, HCGS08]. The combination of AEC and LRC [GKMK07, GKMK08c, GKMK08b, GKMK08a, GKMK08d, GKMK09, GXJ+11] will be main topic of this thesis and will be discussed in Chapter 5.

4.2 Subjective and Objective Assessment of Quality for LRC algorithms

Whenever signal processing strategies change a signal e.g. to enhance speech quality, speech intelligibility, listening effort, etc., the question arises how to assess the achieved enhancement. Among the given examples speech intelligibility can be assessed by standardized listening tests [Wag03, WWB07]. However, an unambiguous rating for *speech quality* is much harder to obtain since the perceived quality may depend heavily on the listener and his or her subjective definition of a good quality. Generally, either subjective listening tests or technical measures can be applied to assess an enhancement. In the following, the term *subjective* denotes all test methods that involve subjects (human listeners) while quality assessment by means of technical measures is denoted by the term *objective*. During subjective listening tests human listeners are asked for their preferences. These tests lead to reliable assessment of quality if a large number of representatively chosen subjects are interviewed and the test itself is set up properly. However, subjective listening tests may depend on the experience of the subjects, are time-consuming, and costly. Ideally, phonetically balanced speech material produced by different speakers has to be used. Especially for national and international standardization processes subjective listening tests lead to huge efforts in terms of time and money. Thus, especially during development of algorithms, technical measures are needed which at least should give a basic idea of the amount of enhancement. Technical measures lead to a reproducible rating and, thus, are called objective measures. Techni-

cal measures may also be used as target functions for adaptive algorithms [RHK06], however, this topic will not be addressed in this thesis. One major goal of this section is to identify technical measures that assess algorithms for LRC without subjective rating by humans, e.g. like the system distance that is commonly used for objective rating of acoustic echo cancellation algorithms (cf. Section 3.1). For LRC still not *the one* objective measure exists and to find technical measures that lead to the same conclusions as subjective rating, i.e. a high correlation between objective and subjective ratings, is still subject to research.
A basic schematic for the identification of a proper enhancement measure is depicted in **Figure 4.4**. Usually a *distorted signal*, which is the reverberant signal in our case, is processed by an algorithm aiming at a certain enhancement and leading to the *equalized signal* or *equalized channel*. As already stated, the *processed signal* now either can be assessed by subjects or by an objective measure. Some objective measures are not based on the output signal of the algorithm but on the equalized impulse response or transfer function (channel). Thus, objective measures are classified as signal-based or channel-based in the remainder of this work.

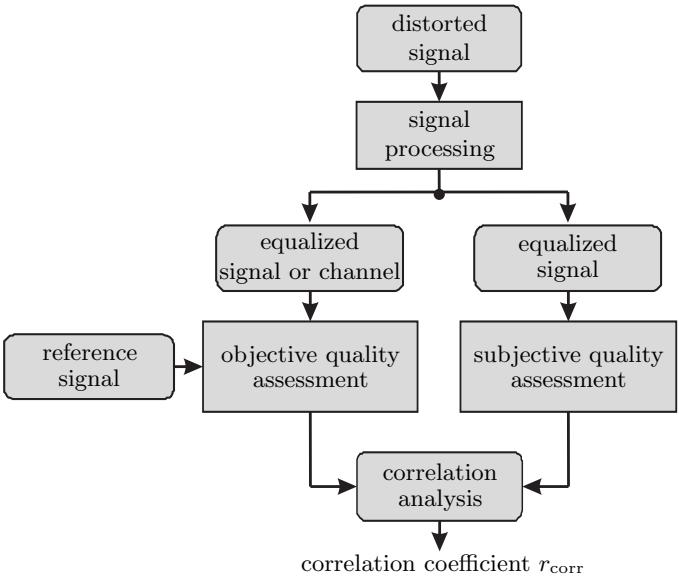

Figure 4.4: Quality assessment by means of subjective and objective testing.

If humans are asked for their opinion about the quality of a specific sound sample they are able to assess the quality based on an internal reference.

This reference is created throughout their life while listening to various sounds and allows the subject to distinguish between *good quality* and *bad quality*. However, most technical algorithms for objective quality assessment need an additional reference, which usually is the undistorted signal. Those measures are called *intrusive* measures while algorithms that perform a rating without additional reference signal are called *non-intrusive*. Since objective measures have to be determined that assess performance of dereverberation algorithms in the same manner as humans do, the correlation between the subjective and objective ratings is determined for each objective measure by the Pearson product-moment correlation coefficient (PPMCC) [RN88]

$$r_{\text{corr}} = \frac{\sum_i (a_i - \bar{a})(b_i - \bar{b})}{\sqrt{\sum_i (a_i - \bar{a})^2 \sum_i (b_i - \bar{b})^2}}, \tag{4.2.1}$$

where a_i and b_i are the subjective and objective ratings on a specific sound sample and $\bar{a}$ and $\bar{b}$ the respective mean values.

While objective quality assessment for acoustic echo cancellation is quite common and leads to easily interpretable results (cf. Section 3.1), a reliable and meaningful quality assessment e.g. for noise reduction is already harder to obtain, however possible [RHK06, Roh08]. While several commonly accepted quality measures exist to assess the performance of noise reduction algorithms or acoustic echo cancellers, the assessment of dereverberation algorithms is still an open issue and topic of current research [NG05, WGH$^+$06, WN06, Hab07, Loi07, WN07, FC08, Fal08, GAK$^+$10, ARG10, GAR$^+$10b]. Often quality measures for dereverberation are adopted from the research field of noise reduction. To determine quality measures that meaningfully assess quality of LRC algorithms, several measures that are assumed to be capable to assess quality of LRC approaches are compared to subjective ratings in the following. All measures can roughly be divided in two classes: Measures that are based on the (i) impulse response or the transfer function of a system (channel-based measures) and (ii) measures that are based on signals. Generally, for LRC algorithms as well the LRC filter's impulse response vector

$$\boldsymbol{c}_{\text{EQ}} = \left[\, c_{\text{EQ},0},\, c_{\text{EQ},1},\, \ldots,\, c_{\text{EQ},L_{\text{EQ}}-1} \,\right]^T \tag{4.2.2}$$

of length L_{EQ} as the RIR coefficient vector

$$\boldsymbol{h} = \left[h_0\,,\, h_1\,,\, \ldots,\, h_{L_h-1}\right]^T \tag{4.2.3}$$

of length L_h, and, hence, also the IR of the equalized system

$$\boldsymbol{v} = [v_0\,,\, v_1\,,\, ...\,,\, v_{L_v-1}]^T \tag{4.2.4}$$

$$= \boldsymbol{H}_{\mathrm{CM}} \cdot \boldsymbol{c}_{\mathrm{EQ}} \tag{4.2.5}$$

of length $L_v = L_h + L_{\mathrm{EQ}} - 1$ are available during simulations. In (4.2.5)

$$\boldsymbol{H}_{\mathrm{CM}} = \mathrm{convmtx}\{\boldsymbol{h}, L_{\mathrm{EQ}}\} \tag{4.2.6}$$

$$= \begin{bmatrix} h_0 & h_1 & \dots & h_{L_h-1} & & \mathbf{0} \\ & \ddots & \ddots & \ddots & \ddots & \\ \mathbf{0} & & h_0 & h_1 & \dots & h_{L_h-1} \end{bmatrix}^T \tag{4.2.7}$$

denotes the channel convolution matrix of size $L_h + L_{\mathrm{EQ}} - 1 \times L_{\mathrm{EQ}}$. It will be shown in the following that measures based on RIRs lead to high correlations with subjective rating, cf. also [GAK+10, GAR+10b].
However, if gradient algorithms (cf. [GKMK08b] and Section 4.5) are used to avoid computational complex matrix inversions, e.g. as in (4.4.6), or to track time-varying environments, or if the effect of the dereverberation algorithm cannot be characterized in terms of an linear time-invariant (LTI) impulse response, e.g. as in [GB99, YM00, Hab07], the necessary impulse responses of the room or the filter may not be accessible or it may be inappropriate to apply those measures [NGH10]. Such situations restrict the number of applicable measures to those based on signals. Whenever a proper RIR is not available, the objective rating has to rely on the signals only. In such situation most technical measures fail to assess LRC algorithms. Only such measures that apply a proper model of the human auditory system show high correlations with subjective ratings.
The remainder of this section is organized as follows. Channel-based and signal-based objective quality measures that are used in the literature to assess the quality of LRC or reverberation suppression approaches are listed in **Tables 4.1** and **4.2**, respectively.
A more detailed definition and discussion of these measures can be found in Appendix A (cf. detailed references in Tables 4.1 and 4.2). To identify measures that are highly correlated to subjective ratings of humans, subjective listening tests are conducted that are described in Section 4.2.1 and the corresponding correlation analysis is presented in Section 4.2.2.

4.2.1 Subjective Listening Tests

For the subjective listening tests, reverberant speech samples were calculated by first convolving room impulse responses generated by the image

Acronym	Objective Quality Measure	Section (page)
D_{50}, D_{80}	Definition	A.1.1 (p. 162)
C_{50}, C_{80}	Clarity Index	A.1.2 (p. 163)
CT	Center Time	A.1.3 (p. 164)
DRR	Direct to Reverberation Ratio	A.1.4 (p. 165)
VAR	Spectral Variance	A.1.5 (p. 165)
SFM	Spectral Flatness Measure	A.1.6 (p. 167)

Table 4.1: Channel-based objective quality measures.

Acronym	Objective Quality Measure	Section (page)
SSRR	Segmental Signal to Reverberation Ratio	A.2.1 (p. 168)
FWSSRR	Frequency Weighted SSRR	A.2.2 (p. 169)
WSS	Weighted Spectral Slope	A.2.3 (p. 170)
ISD	Itakura-Saito-Distance	A.2.5 (p. 171)
CD	Cepstral Distance	A.2.5 (p. 172)
LAR	Log Area Ratio	A.2.5 (p. 172)
LLR	Log Likelihood Ratio	A.2.5 (p. 171)
LSD	Log Spectral Distortion	A.2.4 (p. 170)
BSD	Bark Spektral Distortion	A.2.6 (p. 172)
OMCR	Objective Measure of Colouration in Reverberation	A.2.6 (p. 184)
RDT	Reberberation Dacay Tail Measure	A.2.6 (p. 179)
SRMR	Speech to Reverberation Modulation Energy Ratio	A.2.6 (p. 186)
PSM, PSM_t	Perceptual Similarity Measure	A.2.6 (p. 190)
PESQ	Perceptual Evaluation of Speech Quality	A.2.6 (p. 189)

Table 4.2: Signal-based objective quality measures.

method [AB79] for a room having a size of 6 m × 4 m × 2.6 m (length × width × height) with male and female utterances. Please note, that in addition to artificially generated RIRs also measured RIRs were used in parallel and no dependancy on artificial vs. measured RIR could be identified for the following results. Thus, only the results for artificially generated RIRs are shown in the following. The distance between sound source and microphone was approximately 0.8 m. Room reverberation times were approximately $\tau_{60} \approx \{500, 1000\}$ ms corresponding to normal and somewhat larger office

environments. Within this thesis several LRC approaches will be discussed in Sections 4.3 to 4.7 that have different impact on the processed signal. Four different LRC approaches were chosen to be applied to the reverberant speech samples that are named in **Table 4.3**. The LRC approaches will be discussed later in Sections 4.3 to 4.7 in more detail.

Acronym	Description of method
LS-EQ	Least-squares equalizer according to (4.4.6)
WLS-EQ	Weighted least-squares equalizer according to (4.6.9) with window function according to (4.6.1), $\alpha = 0.8$
ISwPP	Impulse response shaping (IS) according to (4.7.6) with post-processing (PP) (cf. Sec. 4.7.1) [KM06]
ISwINO	Impulse response shaping (IS) with infinity-norm optimization (INO) according to [MMK10]

Table 4.3: Different LRC approaches and the corresponding acronyms.

To generate the dereverberated speech samples that later were presented to the subjects, the reverberant speech samples were convolved with the equalization filters $\boldsymbol{c}_{\mathrm{EQ}}$ calculated by the different algorithms listed in Table 4.3. Filter lengths of these equalizers were $L_{\mathrm{EQ}} = \{1024, 2048, 4096, 8192\}$ at a sampling rate of 8000 Hz.
From all generated speech samples, 21 audio samples were chosen which represented a wide variety of acoustic conditions and possible distortions. These audio samples had a length of 8 s and were scaled to have the same root-mean-squares (RMS) value. Properties of the sound samples and the selected systems are depicted in Appendix B and an audiovisual presentation of the samples and the corresponding systems can be found in [GAR10a]. They were presented diotically to 24 normal-hearing listeners via headphones (Sennheiser HD650) in quiet and in random order. A graphical user interface (GUI) for the subjective listeining test was developed based on ITU recommendations [ITU96, ITU03] (with slight differences) asking to assess the attributes *reverberant, coloured/distorted, distant* and *overall quality* on a continuous 5-point *mean opinion score (MOS)* scale as shown in **Figure 4.5**. Before the subjective test, the human listeners had to make themselves familiar with the test material in a three-step training period. First, all audio samples hat to be listened to in random order to build internal references and anchors. After this, the listeners' attention regarding the dimensions *reverberation* and *colouration* was trained by presenting the audio samples again ordered according to this dimensions as shown in **Figure 4.6**.

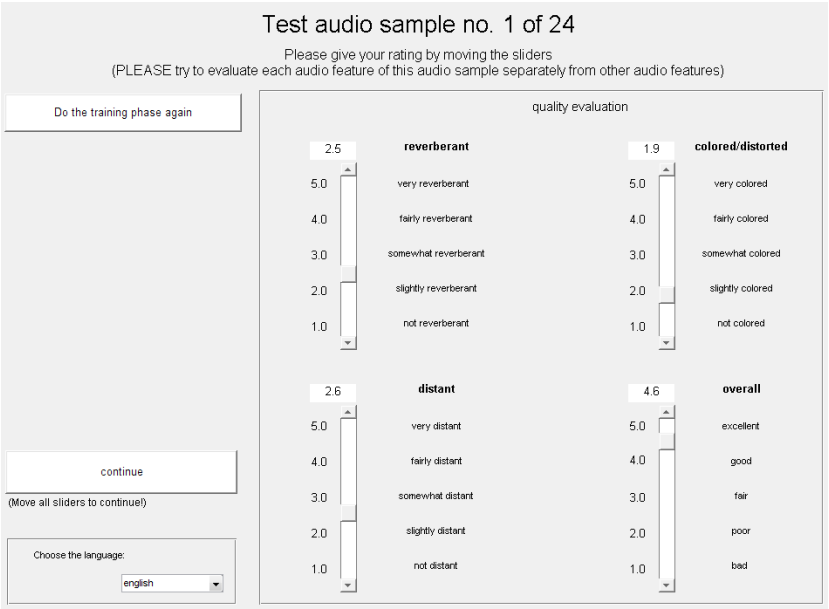

Figure 4.5: Speech quality evaluation of the first audio sample for the attributes *reverberant, coloured/distorted, distant* and *overall quality.*

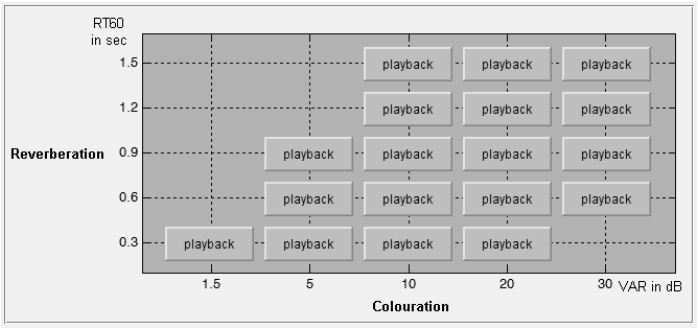

Figure 4.6: Training phase 2 aiming at distinction of two dimensions reverberation and colouration.

After this, in a third training phase three different sound samples were presented that were chosen to represent a sample of each, *very good, medium* and *very bad* quality to generate internal anchors. After this training period the 21 sound samples were presented. No further (hidden) anchors were used

during the listening tests. Training and listening could be repeated as often as desired. The subjective ratings of all sound samples will be presented in Section 4.8 with focus of comparing the different LRC algorithms described in the following Sections 4.4 to 4.7. In the following of this section, only an analysis of the subjective test itself, the chosen attributes and the correlation to objective quality measures will be presented.
For the algorithms under test, it was expected that attributes *reverberant* and *distant* would lead to similar results. Since for LRC algorithms frequency distortion is perceptually much more prominent than what usually is understood as colouration, the listeners were asked to assess colouration/distortion as one spectral attribute. This leads to the fact that common measures that were designed to assess colouration may not correlate well to the subjective data. However, these distortions dominate the spectral perception of subjective quality.

Attribute	Coloured/distorted	Distant	Overall
Reverberant	0.44	0.91	0.94
Coloured/distorted	-	0.29	0.66
Distant	-	-	0.86

Table 4.4: Inter-attribute correlations.

Table 4.4 shows the inter-attribute correlations for the given set of speech samples. As expected, the attributes *reverberant* and *distant* show high inter-attribute correlation although the attribute *distant* leads to a higher inter quartile range (IQR) (cf. results in Section 4.8). Furthermore, the correlation between the attributes *overall quality* and the attributes *distant* as well as *reverberant* is high. Thus, the perceived audio quality is strongly influenced by reverberation (including late reverberation).

4.2.2 Correlation Analysis

In the following the correlations between subjective ratings and objective measures are presented. The correlations of subjective rating for the four attributes and the channel-based objective measures are shown in **Table 4.5** while correlations with signal-based objective measures are shown in **Tables 4.6** and **4.7**.
For each objective measure, correlations with the subjective ratings are shown for the case that all LRC approaches of Section 4.3 are considered (Method: All EQs) and for the case that only one LRC approach is used. For the latter case no correlation was calculated for the impulse-response

Measure	Method	Reverberant	Col./dist.	Distant	Overall
D50	All EQs	-0.86	-0.63	-0.94	0.91
	LS-EQ	-0.71	-0.33	-0.79	0.79
	WLS-EQ	-0.94	-0.73	**-0.99**	**0.98**
	ISwPP	-0.94	-0.61	-0.94	0.93
D80	All EQs	-0.9	-0.5	-0.91	0.9
	LS-EQ	-0.73	-0.31	-0.82	0.82
	WLS-EQ	-0.94	-0.59	-0.98	0.93
	ISwPP	-0.85	-0.55	-0.84	0.84
C50	All EQs	**-0.93**	**-0.67**	**-0.94**	**0.94**
	LS-EQ	-0.78	-0.32	-0.85	0.86
	WLS-EQ	**-0.96**	-0.76	-0.98	0.97
	ISwPP	**-0.98**	-0.58	-0.96	0.93
C80	All EQs	**-0.93**	-0.61	-0.89	0.91
	LS-EQ	-0.8	-0.3	-0.86	0.88
	WLS-EQ	-0.98	-0.69	-0.99	0.96
	ISwPP	-0.92	-0.54	-0.9	0.88
CT	All EQs	0.85	0.61	0.93	-0.91
	LS-EQ	**0.91**	0.29	**0.94**	**-0.95**
	WLS-EQ	0.86	0.79	0.96	-0.97
	ISwPP	0.97	0.67	**0.98**	**-0.97**
DRR	All EQs	0.24	-0.1	0.18	-0.13
	LS-EQ	-0.77	-0.33	-0.83	0.84
	WLS-EQ	-0.4	**-0.86**	-0.6	0.7
	ISwPP	-0.25	**-0.69**	-0.27	0.36
VAR	All EQs	-0.03	0.37	0.23	-0.16
	LS-EQ	0.62	**0.42**	0.71	-0.69
	WLS-EQ	0.69	0.81	0.84	-0.88
	ISwPP	0.6	0.46	0.61	-0.65
SFM	All EQs	0.13	-0.27	-0.13	0.05
	LS-EQ	-0.69	-0.38	-0.77	0.76
	WLS-EQ	-0.71	-0.82	-0.86	0.9
	ISwPP	-0.88	-0.66	-0.88	0.91

Table 4.5: Correlations r_{corr} of mean opinion score (MOS) values of subjective ratings and channel-based objective measures (maxima are indicated in boldface).

shaping approach based on infinity-norm optimization because the number of sound samples was too low for a reliable correlation analysis. The highest correlation for each attribute and approach is highlighted in boldface in the tables.

The reason for additionally calculating correlations for each LRC approach separately is exemplarily illustrated in **Figure 4.7** for the spectral flatness measure (SFM) (cf. also Appendix A.1.6).

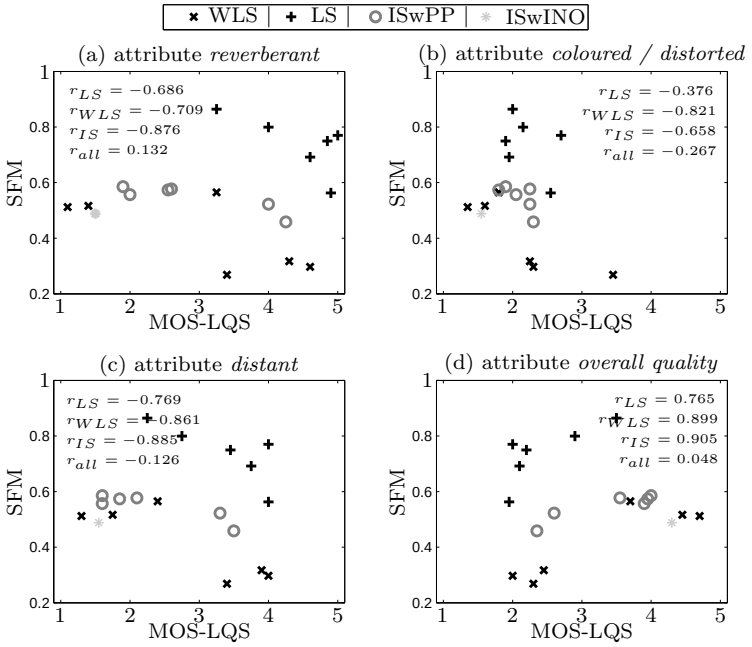

Figure 4.7: Correlations of subjective ratings (MOS for listening quality (subjective), MOS-LQS) and spectral flatness measure (SFM) for all four attributes.

The SFM shows much higher correlation to the MOS values of the subjective test data when a single rather than all LRC approaches are considered. However, the time-domain channel-based measures in Table 4.5 show consistent correlations for all LRC approaches. The interested reader is referred to Appendix C for an overview of all correlation patterns. It can be seen from Table 4.5 that the time-domain channel-based objective measures show high correlation with the subjective data for the attributes

reverberation, distance and *overall quality* (with the exception of the DRR measure). The frequency-domain channel-based measures VAR and SFM show much lower correlation. However, as stated before, they may show somewhat higher correlation for single LRC approaches such as SFM for the WLS-EQ. In general, and this is also true for the signal-based measures (cf. Table 4.6), only low correlation was obtained with the attribute *coloured/distorted* for all measures. An explanation for this finding may be that the source-receiver distance for our experiment (0.8 m) is larger than the critical distance (cf. Section 2.1.5, p. 15).

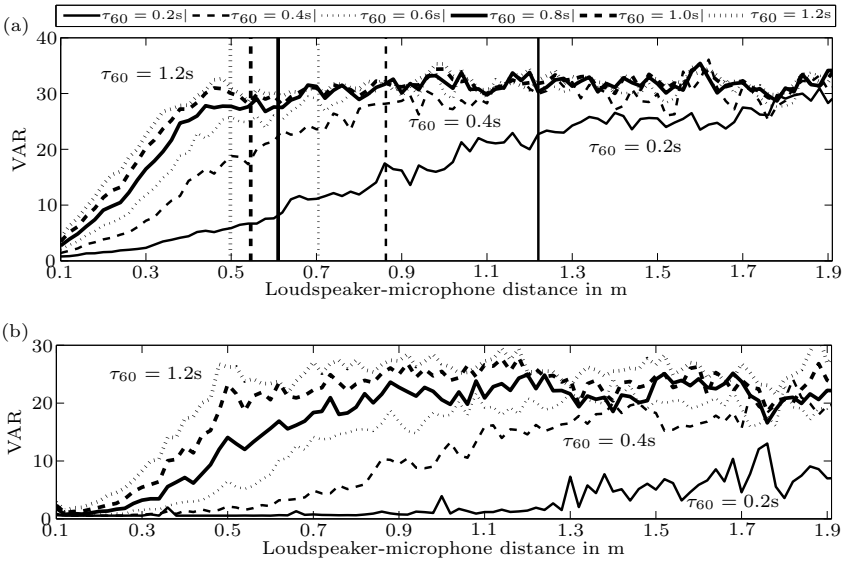

Figure 4.8: VAR measure of (a) RIR h_k and (b) equalized channel v_k over loudspeaker-microphone distance for different room reverberation times (critical distances (cf. Section 2.1.5, p. 15) are indicated as dashed vertical lines). Sub-figure (b) shows the VAR measure for an equalized system using an LS-EQ with $L_{\mathrm{EQ}} = 2048$ at $f_s = 8$ kHz.

To illustrate the dependency of the spectral measures on the distance between acoustic source and microphone, **Figure 4.8 (a)** depicts the variance measure (cf. also Appendix A.1.5) over the source-microphone distance for different room reverberation times $\tau_{60} = 200$ ms ... 1200 ms. Additionally, the respective critical distances of the RIRs are depicted as vertical lines. The variance measure for the respective equalized channels is depicted in panel (b) of Figure 4.8. It can be seen that the variance measure (cf. also Appendix A.1.5) does not increase once it reaches its maximum value at

about 31 dB in panel (a). These results are in consilience with the findings in [Jet79, Hab07] where the maximum reachable variance was calculated to be at about 31 dB for common RIRs [Jet79]. This point is approximately reached at the critical distance as it is shown in Figure 4.8 (a).

Another reason for lower correlations for the spectral measures VAR and SFM may be that they equally asses spectral peaks which are perceived as being very annoying [KM06] and spectral dips that are common for RIRs and do not decrease the perceived quality to a great extent [TO88, Buc81, Fry75]. Here, more research has to be undertaken to find appropriate technical measures to asses frequency-domain quality criteria.

Tables 4.6 and 4.7 show the correlations of subjective ratings with signal-based objective measures. Again the maxima for each attribute and each class of LRC filters is highlited by bold letters jointly for Tables 4.6 and 4.7. It can be seen that the signal-based measures show lower correlation to subjective data than the channel-based measures in general. The LPC-based measures (cf. also Appendix A.2.5) outperform purely signal-based measures like the SSRR. The analysis of SSRR in perceptually motivated frequency bands, however, already significantly increases the correlation with subjective ratings. By far, the highest correlations are obtained by the measures PSM and PSMt that rely on auditory models. PSMt, in addition to PSM, evaluates short-time behaviour of the correlations of internal signal representations and focuses on low correlations as it is done by human listeners [HK06]. The auditory-model based measures show even higher correlation than RDT, SRMR and OMCR although the latter were designed to explicitly assess reverberation. The performance of RDT and OMCR measures can be adjusted by changing internal parameters. By this, higher correlation to the specific set of samples can be obtained. However, we used standard values for these parameters given in the respective literature [WN06, WN07]. Furthermore, it has to be emphasized that the attribute *colouration/distortion* is most difficult to assess by objective measures at least for the discussed LRC algorithms, since distortions are perceptually relevant and measures like OMCR try to assess colouration effects only (the same holds for SFM and the variance measure). They succeed in doing so (cf. e.g. Figure A.21), but colouration alone is not well correlated to our subjective data due to distortions like late echoes and pre-echoes which are much more prominent than the colouration effect. All tested measures are not capable to explicitly assess those influences and further development of objective measures is required.

An example for such late echoes is shown in **Figure 4.9**. Although the spectral characteristics are clearly enhanced (cf. panel (b)) and in time-domain (cf. panel (a)) much energy of the impulse response is suppressed,

Measure	Method	Reverberant	Col./dist.	Distant	Overall
SSRR	All EQs	-0.33	-0.29	-0.43	0.4
	LS-EQ	-0.6	0.15	-0.65	0.67
	WLS-EQ	-0.8	-0.74	-0.83	0.8
	ISwPP	-0.7	-0.34	-0.65	0.64
FWSSRR	All EQs	-0.44	-0.4	-0.57	0.55
	LS-EQ	-0.79	-0.04	-0.82	0.85
	WLS-EQ	-0.94	-0.78	-0.99	**0.98**
	ISwPP	-0.81	-0.46	-0.76	0.75
WSS	All EQs	0.6	0.58	0.76	-0.71
	LS-EQ	0.79	0.44	0.87	-0.85
	WLS-EQ	0.89	0.76	0.96	-0.98
	ISwPP	0.91	0.58	0.87	-0.86
ISD	All EQs	0.64	0.35	0.69	-0.68
	LS-EQ	0.35	-0.44	0.36	-0.41
	WLS-EQ	**0.96**	0.71	**0.99**	-0.98
	ISwPP	0.7	0.37	0.67	-0.68
CD	All EQs	0.63	0.41	0.7	-0.67
	LS-EQ	0.45	-0.37	0.48	-0.52
	WLS-EQ	0.89	0.81	0.94	-0.93
	ISwPP	0.8	0.42	0.75	-0.73
LAR	All EQs	0.52	0.38	0.61	-0.59
	LS-EQ	0.33	-0.5	0.36	-0.42
	WLS-EQ	0.93	0.78	0.99	-0.98
	ISwPP	0.75	0.39	0.7	-0.69
LLR	All EQs	0.66	0.43	0.75	-0.71
	LS-EQ	0.47	-0.36	0.5	-0.54
	WLS-EQ	0.89	0.85	0.96	-0.96
	ISwPP	0.84	0.45	0.8	-0.78
LSD	All EQs	0.74	0.48	0.81	-0.78
	LS-EQ	0.75	0.07	0.81	-0.83
	WLS-EQ	0.87	0.83	0.92	-0.92
	ISwPP	0.87	0.5	0.83	-0.82

Table 4.6: Correlations r_{corr} of MOS values of subjective ratings and signal-based objective measures (maxima of signal-based measures in Tables 4.6 and 4.7 are indicated in boldface).

Measure	Method	Reverberant	Col./dist.	Distant	Overall
BSD	All EQs	0.04	0.3	0.24	-0.2
	LS-EQ	0.53	0.47	0.63	-0.6
	WLS-EQ	0.85	0.64	0.94	-0.94
	ISwPP	0.91	0.64	0.93	-0.94
OMCR	All EQs	0.05	0.13	0.03	0.05
	LS-EQ	0.52	**0.83**	0.62	0.54
	WLS-EQ	0.63	0.23	0.64	0.65
	ISwPP	0.16	0.45	0.24	0.26
RDT	All EQs	0.67	0.51	0.79	-0.75
	LS-EQ	0.69	0.43	0.78	-0.77
	WLS-EQ	0.81	0.74	0.88	-0.93
	ISwPP	0.94	0.57	0.92	-0.9
SRMR	All EQs	-0.53	-0.24	-0.59	0.51
	LS-EQ	-0.44	0.15	-0.51	0.54
	WLS-EQ	-0.75	**-0.88**	-0.73	0.8
	ISwPP	-0.78	-0.45	-0.72	0.69
PSM	All EQs	-0.8	**-0.63**	-0.9	0.87
	LS-EQ	-0.84	-0.64	-0.9	0.88
	WLS-EQ	-0.84	-0.83	-0.92	0.97
	ISwPP	**-0.98**	-0.65	-0.96	0.94
PSMt	All EQs	**-0.91**	-0.61	**-0.95**	**0.94**
	LS-EQ	**-0.89**	-0.56	**-0.96**	**0.92**
	WLS-EQ	-0.9	-0.76	-0.96	**0.98**
	ISwPP	-0.98	**-0.79**	**-0.97**	**0.96**
PESQ	All EQs	-0.6	-0.35	-0.69	0.63
	LS-EQ	-0.47	0.35	-0.5	0.55
	WLS-EQ	-0.84	-0.77	-0.9	0.87
	ISwPP	-0.89	-0.46	-0.85	0.82

Table 4.7: Correlations r_{corr} of MOS values of subjective ratings and signal-based objective measures (maxima of signal-based measures in Tables 4.6 and 4.7 are indicated in boldface).

a high amount of late reverberation occurs after sample 5000 that is small in amplitude but perpetually relevant since the temporal masking effect of the main peak is less distinct for those late taps.
Figure 4.9 (c) depicts the subjective rating for the given system in Figure 4.9 (a) and (b). It clearly shows that a high amount of reverberation is perceived

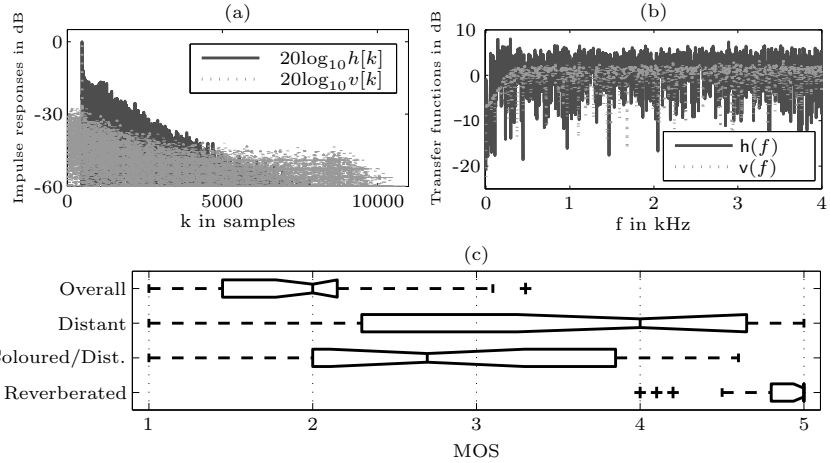

Figure 4.9: (a) RIR ($\tau_{60} \approx 1$ s) and equalized system in time-domain (LRC filter length was 8192 samples), (b) corresponding transfer functions, (c) subjective rating for equalized system in (a), (b).

by the subjects as well as relatively high spectral colouration/distortion given the fact that the transfer function is clearly enhanced compared to the unprocessed room transfer function. Furthermore, depending on the delay that is introduced by the equalizer, perceptually disturbing pre-echoes occur as observable in Figure 4.9 (a). None of the tested measures is capable to explicitly assess those influences. Thus, development of a measure capable to assess the described affects would be valuable future work.

In general, it can be stated from the previous analysis that objective quality measures based on the impulse response (like the common C50 measure) show much higher correlation between objective and subjective data than most of the tested measures that are based on the signals only. However, if impulse responses are not properly accessible, e.g. as for blind dereverberation algorithms, measures that incorporate sophisticated auditory models should be used for quality assessment. The so-called perceptual similarity measure PSM showed highest correlations to the subjective data.

4.3 Listening-Room Compensation

The general setup for single-channel listening-room compensation (LRC) was already depicted in Figure 4.3 on page 64. A more general setup for an arbitrary number of J source channels, P loudspeakers and Q microphones

is shown in **Figure 4.10**. Here, PQ RIRs between the loudspeakers and the microphones are equalized by JP filters preceding the loudspeakers.

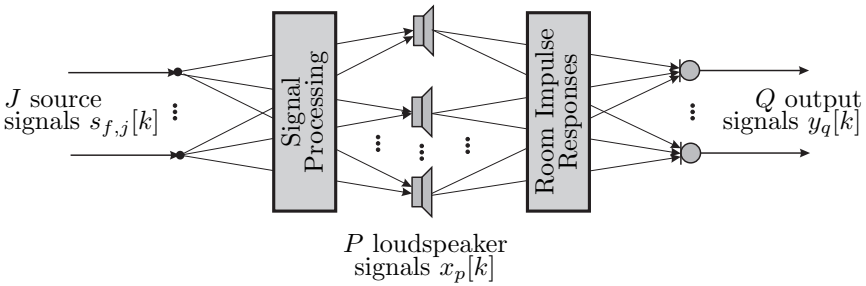

Figure 4.10: Multi-Channel setup for listening-room compensation.

A number of $J \geq 1$ source channels is e.g. used for the so-called *cross-talk-cancellation* algorithms (cf. e.g. [YHC07, KM07]) where the number of source signals J usually equals the number of the reference microphones Q to achieve reproduction of every channel of $s_f[k]$ at one microphone. This is e.g. needed for binaural sound reproduction without using headphones. For this thesis the number of source channels is restricted to $J = 1$. Thus, the aim of the equalizer is reconstruction of a single-channel non-reverberant speech signal $s_f[k]$ at the Q positions of the reference microphones using P loudspeakers.

An increased number of reference positions Q leads to an increased spatial robustness of the equalization system (cf. Section 4.4.2). If the number of loudspeakers P is greater than the number of microphones Q spatial diversity can be exploited which leads to a better equalization.

4.4 Least-Squares Equalization

Equalization concepts are widely used in the field of mobile communications. Here, a transfer system consisting of a transmitter, transmission line (channel) and receiver is considered. The influence of the channel on the transmitted data is compensated by an equalization filter which is often located at the receiving end. As stated before, for an LRC scenario, the equalizer has to be placed in front of the transmission channel (in this case the room which is described by the RIR).

4.4.1 Single Channel LS-Equalizer

An equalization scheme which aims at minimizing the Euclidean distance between the overall system of the concatenation of RIR and equalizer to a desired system $\boldsymbol{d}[k]$ is depicted in **Figure 4.11**. The equalization filter tries to reduce the influence of the RIR at the position of the reference microphone where the human user is assumed to be located. Thus, the goal of LRC is that differences between the signal $y[k]$, which a human listener at the position of the reference microphone perceives, and the original unreverberated signal $s_f[k]$ should be minimized.

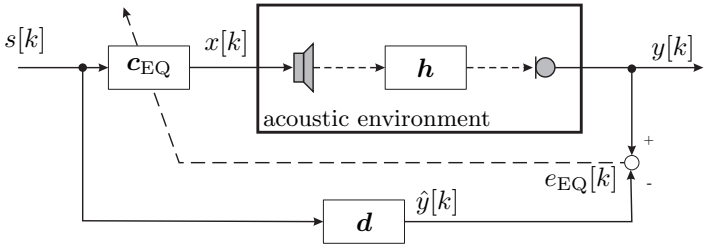

Figure 4.11: Least-squares equalizer for listening-room compensation.

From Figure 4.11 the error signal which has to be minimized in the MMSE sense can be calculated as

$$e_{\mathrm{EQ}}[k] = \boldsymbol{s}^T[k]\boldsymbol{H}_{\mathrm{CM}}\boldsymbol{c}_{\mathrm{EQ}} - \boldsymbol{s}^T[k]\boldsymbol{d} \tag{4.4.1}$$

with $\boldsymbol{c}_{\mathrm{EQ}}$ and $\boldsymbol{H}_{\mathrm{CM}}$ being the coefficient vector of the equalizer and the convolution matrix of the RIR as defined in (4.2.2) and (4.2.7), respectively. The input signal vector

$$\boldsymbol{s}[k] = [s[k], s[k-1], s[k-L_h-L_{\mathrm{EQ}}+2]]^T \tag{4.4.2}$$

and the coefficient vector of the desired system

$$\boldsymbol{d} = [\underbrace{0, ..., 0}_{\tilde{k}_0}, d_0, ..., d_{\lfloor L_d/2 \rfloor}, ..., d_{L_d-1}, \underbrace{0, ..., 0}_{L_h+L_{\mathrm{EQ}}-1-L_d-\tilde{k}_0}]^T \tag{4.4.3}$$

are of length $L_h + L_{\mathrm{EQ}} - 1$. The lengths of the RIR, the LRC filter and the desired system vector $\boldsymbol{d}$ are denoted by L_h, L_{EQ} and L_d, respectively. The desired system $\boldsymbol{d}$ usually is chosen as a delayed unit impulse, a delayed band pass or a delayed high pass as exemplarily shown in **Figure 4.12** in time- and frequency-domain. *Perfect* equalization is achieved by the

(delayed) unit impulse shown in the left panels of Figure 4.12. However, since real-world hardware like loudspeakers and microphones usually does not have perfectly flat transfer characteristics especially in very low and high frequency ranges a frequency response correction in this frequency ranges would unnecessarily boost filter coefficient and signal energy. Therefore, the delayed high pass or band pass systems in Figure 4.12 may be more suitable for real-world systems.

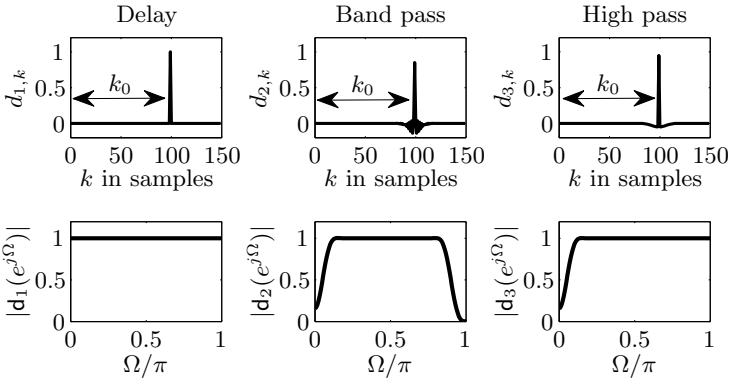

Figure 4.12: Possible desired system vectors for the EQ design. Delayed unit impulse $\boldsymbol{d}_1$, delayed band-pass $\boldsymbol{d}_2$ and high pass $\boldsymbol{d}_3$ (40^{th} order FIR filters with band limits at 200 Hz and 3700 Hz at sampling frequency of $f_s = 8$ kHz) and the respective frequency-domain representations (lower panels).

The delay introduced by the equalizer is denoted as k_0. It corresponds directly to the position of the one for the delayed impulse. For desired systems of length $L_d > 1$ the delay k_0 corresponds to the middle position of the desired system $k_0 = \tilde{k}_0 + \lfloor L_d/2 \rfloor$.

Minimization of $\mathrm{E}\left\{e_{\mathrm{EQ}}^2[k]\right\}$ by solvoing $\frac{\partial \mathrm{E}\left\{e_{\mathrm{EQ}}^2[k]\right\}}{\partial \boldsymbol{c}_{\mathrm{EQ}}^T} \stackrel{!}{=} 0$ leads to

$$\boldsymbol{c}_{\mathrm{EQ}} = \left(\boldsymbol{H}_{\mathrm{CM}}^T \boldsymbol{R}_{ss}[k] \boldsymbol{H}_{\mathrm{CM}}\right)^{-1} \boldsymbol{H}_{\mathrm{CM}}^T \boldsymbol{R}_{ss}[k] \boldsymbol{d} \tag{4.4.4}$$

with $\boldsymbol{R}_{\boldsymbol{ss}}[k] = \mathrm{E}\left\{\boldsymbol{s}[k]\boldsymbol{s}^T[k]\right\}$ being the covariance matrix of the input signal of size $L_h+L_{\mathrm{EQ}}-1 \times L_h+L_{\mathrm{EQ}}-1$. With the assumption of a white Gaussian input signal,

$$\boldsymbol{R}_{ss}[k] = \boldsymbol{I}, \tag{4.4.5}$$

the well-known least-squares equalizer is obtained.

$$\begin{aligned} \boldsymbol{c}_{\mathrm{EQ}} &= \left(\boldsymbol{H}_{\mathrm{CM}}^{T}\boldsymbol{H}_{\mathrm{CM}}\right)^{-1}\boldsymbol{H}_{\mathrm{CM}}^{T}\boldsymbol{d} \\ \boldsymbol{c}_{\mathrm{EQ}} &= \boldsymbol{H}_{\mathrm{CM}}^{+}\boldsymbol{d} \end{aligned} \tag{4.4.6}$$

In (4.4.6), $\boldsymbol{H}_{\mathrm{CM}}^{+}$ denotes the Moore-Penrose pseudoinverse of the channel convolution matrix. Equalization results of the LS equalizer according to (4.4.6) are examplarily shown in **Figure 4.13**. Panel (a) of Figure 4.13

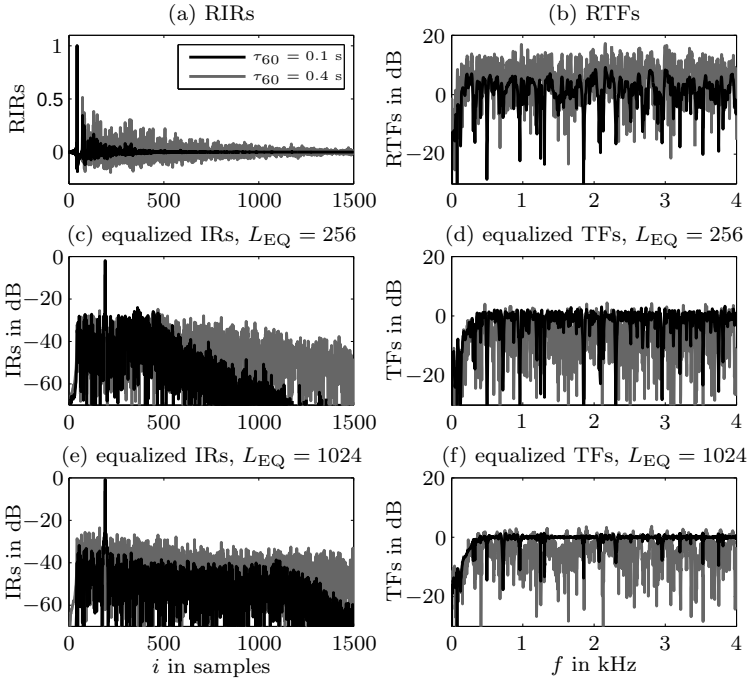

Figure 4.13: (a) RIRs with $\tau_{60} = 100$ ms and 400 ms, (b) corresponding transfer functions in dB, (c) equalized impulse responses for filter length: $L_{\mathrm{EQ}} = 256$, (d) corresponding transfer functions in dB, (e) equalized impulse responses for filter length $L_{\mathrm{EQ}} = 1024$, (f) corresponding transfer functions in dB. Sampling rate: $f_s = 8$ kHz.

shows two RIRs characterized by reverberation times of $\tau_{60} = 100$ ms and $\tau_{60} = 400$ ms and panel (b) the corresponding RTFs. Panels (c) and (d) show the equalized IRs in dB and the corresponding transfer functions (TFs)

in dB after least-squares (LS) equalization using an filter of length $L_{\mathrm{EQ}} = 256$, respectively. Panels (e) and (f) show the same results for an equalizer of length $L_{\mathrm{EQ}} = 1024$. It can be seen that the targeted high pass characteristic is archived more easily for RIRs with shorter reverberation time and that even for high LRC filter orders and relatively short reverberation times perfect equalization is not achieved by the filter. Furthermore, the natural shape of RIRs, i.e. the linear decay in logarithmic time-domain, has been changed and due to the delay k_0 introduced by the LRC filter a certain amount of energy of the equalized IR occurs in front of the main peak that may be perceived as pre-echo or pre-ringing.

Estimation of the Equalizer Delay

Many contributions in the literature suggest to use a *good guess* for the delay k_0 which has to be introduced in least-squares equalization approaches to achieve a maximum amount of dereverberation. In general, the decision how to choose this delay in an optimum way is not easy since a mathematical relation between delay k_0 and optimum LRC performance is unknown and even the definition of a proper target function is difficult since psychoacoustic properties regarding the perception of pre-echoes would have to be considered additionally to a purely mathematical optimization. Despite this problems, some experiments will be accomplished in the following to obtain a proper LRC filter delay, since designing one LRC filter for each possible delay and choosing the best one of course is not practically feasible. Therefore, the dependence of the optimum equalizer delay of different measures characterizing RIRs will be evaluated in the following, since the equalizer performance depends on the specific RIR $\boldsymbol{h}$ that has to be equalized. The performance of the LRC filter will be assessed by means of the bark spectral distortion (BSD) measure [WSG92] (cf. (A.2.30)) and the signal-to-reverberation-ratio-enhancement (SRRE) [GKMK08d, NG05] (cf. (A.2.2)) in the following. The LRC filter performance and, thus, both measures depend on (i) the specific RIR to be equalized $\boldsymbol{h}$, (ii) the LRC filter order L_{EQ}, and (iii) the chosen equalizer delay k_0.
To find a general rule for an optimum delay $k_{0,\mathrm{opt}}$ the equalizer delay k_0 in (4.4.3) which leads to a minimum achievable BSD for a given RIR

$$k_{0,\mathrm{opt,BSD}} = \underset{k_0}{\operatorname{argmin}}\{\mathrm{BSD}(\boldsymbol{h}, L_{\mathrm{EQ}}, k_0)\} \tag{4.4.7}$$

and the equalizer delay which leads to a maximum SRRE in (4.4.3)

$$k_{0,\mathrm{opt,SRRE}} = \underset{k_0}{\operatorname{argmax}}\{\mathrm{SRRE}(\boldsymbol{h}, L_{\mathrm{EQ}}, k_0)\} \tag{4.4.8}$$

are calculated for various RIRs and LRC filter orders. Please note, that a small BSD indicates a good performance while for the SRRE a high value indicates good performance.

A set of 270 different RIRs characterized by room reverberation times ranging from $\tau_{60} = 50$ ms to $\tau_{60} = 1200$ ms was generated for this evaluation by taking (i) artificially simulated RIRs generated by the so-called image method [AB79], (ii) RIRs measured using the sweep-sine method [MM01], (iii) RIRs taken from the MARDY database [WGH$^+$06], and (iv) RIRs modelled by an exponentially damped Gaussian noise according to (2.1.2). The optimum delays defined by the maximum SRRE and the minimum BSD were calculated for each RIR and for the LRC filter orders $L_{\mathrm{EQ}} = \{256, 512, 1024\}$. Please note, that both measures (BSD and SRRE) lead to similar optimum delays for all RIRs and LRC filter lengths tested ($k_{0,\mathrm{opt,BSD}} \approx k_{0,\mathrm{opt,SRRE}} \,\forall\, \boldsymbol{h}, L_{\mathrm{EQ}}$).

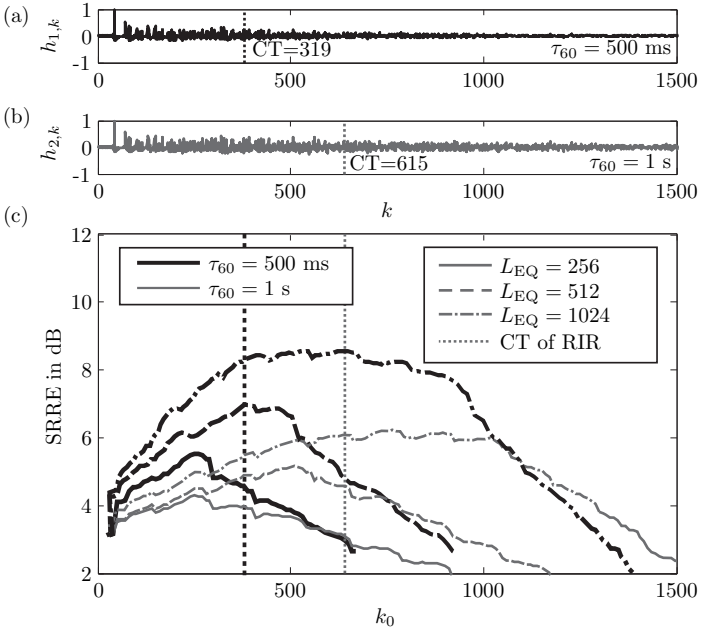

Figure 4.14: (a) RIR with reverberation times of $\tau_{60} = 500$ ms and it's central time (CT) (cf. (A.1.6)) in samples. (b) RIR with $\tau_{60} = 1$ s and it's CT. (c) equalizer performance in dependence of delay k_0 of the desired system for different equalizer filter lengths L_{EQ} and RIRs (a) (thicker black lines) and (b) (thinner grey lines).

Figure 4.14 exemplarily shows the SRRE in panel (c) in dependence of the delay k_0 and different LRC filter orders for the two RIRs $\boldsymbol{h}_1$ and $\boldsymbol{h}_2$ that are depicted in panels (a) and (b). The dotted vertical lines indicate the central times (CTs) (cf. (A.1.6)) of the two RIRs.
The LRC filter performance is shown for the different equalizer lengths $L_{\mathrm{EQ}} = \{256, 512, 1024\}$ by solid lines, dashed lines and dash-dotted lines, respectively. Thicker black lines show the LRC filter performance if the RIR $\boldsymbol{h}_1$ is equalized and thinner grey lines show performance for equalization of $\boldsymbol{h}_2$. It can be clearly seen from Figure 4.14 (c) that the equalizer performance depends on the LRC filter delay k_0 and that a certain optimum exists that depends on the RIR to be equalized.
To find a parameter that may indicate how to choose k_0, the correlation between different measures describing the RIRs and the corresponding optimum LRC filter delays $k_{0,\mathrm{opt}}$ are analyzed in the following. Six objective measures characterizing an RIR were, thus, calculated for each of the 270 RIRs, i.e. reverberation time τ_{60}, the delay of direct path of the RIR $k_{h_{max}} = \mathrm{argmax}_k\{|\boldsymbol{h}|\}$, direct-path-to-reverberation-ratio (DRR) according to (A.1.7), definition D_{50} according to (A.1.1), clarity index (CI) C_{80} according to (A.1.5), and the central time (CT) according to (A.1.6).

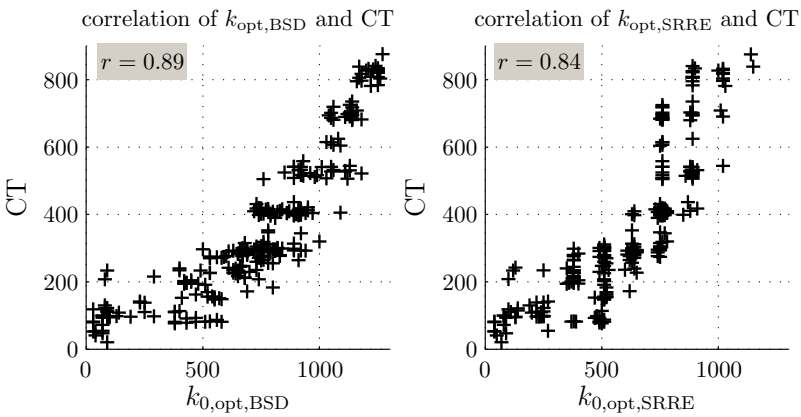

Figure 4.15: Correlation between central time (CT) and optimum equalizer delay given by the minimum of the BSD (left) and maximum of the SRRE (right) for an equalizer length of $L_{\mathrm{EQ}} = 1024$.

Figure 4.15 exemplarily shows the CT for all 270 RIRs over the optimum equalizer delays $k_{0,\mathrm{opt,BSD}}$ (left panel) and $k_{0,\mathrm{opt,SRRE}}$ (right panel). Correlations (PPMCCs) are $r_{\mathrm{corr}} = 0.89$ and $r_{\mathrm{corr}} = 0.84$ for BSD and SRRE,

respectively.
Table 4.8 summarizes the Pearson product-moment correlation coefficients (PPMCCs) according to (4.2.1) between the different measures characterizing the RIRs and $k_{0,\mathrm{opt,BSD}}$ and **Table 4.9** the respective correlations between the different measures characterizing the RIRs and $k_{0,\mathrm{opt,SRRE}}$.

	PPMCC r_{corr} between $k_{0,\mathrm{opt,BSD}}$ and ...					
L_{EQ}	τ_{60}	$k_{h_{max}}$	DRR	D_{50}	C_{80}	CT
256	0.37	**0.84**	0.84	0.37	0.56	0.82
512	0.39	0.70	0.74	0.23	0.58	**0.75**
1024	0.53	0.66	0.80	0.11	0.63	**0.89**

Table 4.8: Correlation coefficients r_{corr} between optimum equalizer delay according to minimum BSD and RIR properties for varying equalizer length.

	PPMCC r_{corr} between $k_{0,\mathrm{opt,SRRE}}$ and ...					
L_{EQ}	τ_{60}	$k_{h_{max}}$	DRR	D_{50}	C_{80}	CT
256	0.28	**0.89**	0.86	0.47	0.49	0.80
512	0.39	0.78	**0.85**	0.30	0.57	**0.85**
1024	0.36	0.74	0.83	0.27	0.50	**0.84**

Table 4.9: Correlation coefficients r_{corr} between optimum equalizer delay according to maximum SRRE and RIR properties for varying equalizer length.

The highest correlations are indicated by bold letters in Tables 4.8 and 4.9. It can be seen that the central time (CT) seems to be a good indicator for the optimum equalizer delay $k_{0,\mathrm{opt}}$ for both, BSD and SRRE, especially for higher LRC filter orders. The somewhat lower correlation for short equalizer lengths in Tables 4.8 and 4.9 can be explained by taking a closer look at Figure 4.14. If the CT is greater than the equalizer length, the equalizer may not be capable to introduce the desired delay. Hence, the LRC filter delay should be chosen as

$$\hat{k}_{0,\mathrm{opt}} = \min\{\mathrm{CT}, L_{\mathrm{EQ}}\}. \tag{4.4.9}$$

Using the criterion in (4.4.9) to determine the equalizer delay leads to 94.4% of the performance that is achieved if the optimum delay k_0 that maximizes SRRE and minimizes BSD would be perfectly known a priori for the given test corpus of the 270 RIRs (90.5% is achieved if the CT is used directly as a criterion for determining $k_{0,\mathrm{opt}}$).

In real-world systems the delay k_0 has to be chosen without a priori information about the RIR. The RIR is unknown and, due to this, also it's CT is unknown. A method to obtain CT without identification of the whole RIR was described in [GKMK09]. With the assumption that the RIR can be modelled by the stochastic RIR model described in Section 2.1.3 the PDP of the RIR model (cf. Figure 2.5 on page 14) can be calculated with the knowledge of reverberation time τ_{60} and initial delay k_{init} using (2.1.2) and (2.1.3). Both parameters can be obtained by identifying only the very first part of the RIR by means of an acoustic echo canceller. By this, the initial delay k_{init} is directly obtained by the major peak of the AEC filter and the reverberation time can be obtained from least-squares fitting of the AEC decay [GKMK09, SRHE09, SGR$^+$11]. The CT can then be calculated from the RIR model (2.1.2). Please note that estimates of the room reverberation time τ_{60} and of the initial RIR delay k_{init} have to be calculated only once for a specific room, since they do not vary too much for different spatial positions. The length of the AEC can be restricted to a few taps since only the position of the initial RIR coefficients is needed to fit the power delay profile by a least-squares approach [SRHE09]. Thus, the AEC will converge extremely fast and has a very low computational complexity. While identification of the initial delay is quite robust the least-squares fitting to obtain τ_{60} may also be replaced by different methods for blind reverberation time estimation known in the literature [e.g., RJW$^+$03, SRHE09, LV08, CLD01, SGR$^+$11] to increase robustness.

Switching of the LRC filter delay while the system is running is possible, e.g. in case that the estimate of the optimum delay is only available after a certain period needed for estimation, as it is visualized in **Figure 4.16**. The convergence of an LRC filter of length $L_{\mathrm{EQ}} = 1024$ updated by the so-called decoupled filtered-X least mean square (dFxLMS) algorithm [GKMK08b] (cf. Section 4.5.3) is shown in Figure 4.16. It compares the LRC filter convergence for the case of perfect knowledge of the best possible delay k_0 (upper curve) to the case that a *poor guess* was made for the delay (lower curve). The solid curve in the middle shows the convergence behavior if the equalizer delay is switched at about 1.5 seconds from the *poor guess* to the proposed estimate according to (4.4.9).

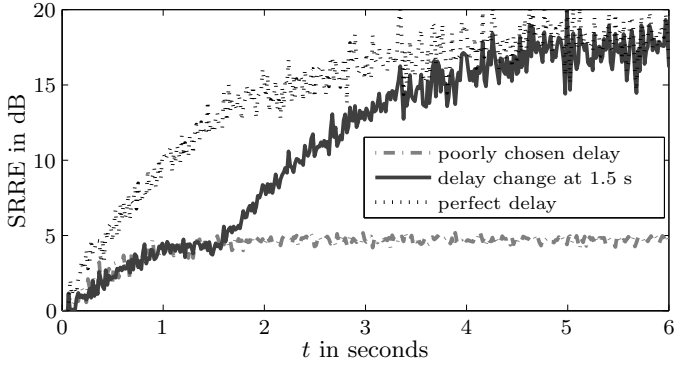

Figure 4.16: Performance comparison of equalizers using different delays k_0 in terms of SRRE.

4.4.2 Robustness Issues

As already mentioned, single-channel RIR equalization is not straightforward since

(i) the length L_h of the RIR to be equalized usually exceeds several thousand taps [BDH+99, Kut00] as depicted in Figure 2.6 on p. 16;

(ii) RIRs are mixed-phase systems [NA79] as depicted in Figure 2.7 and, therefore, direct inversion does not lead to a stable causal solution;

(iii) the average difference between maxima and minima in RTFs typically exceeds 10-20 dB [RWK00, Kut00] as depicted in Figure 2.3 and RTFs contain spectral nulls that, after equalization, give strong peaks in the LRC filter's transfer function causing narrow band noise amplification;

(iv) equalization filters designed from inaccurate RIR estimates will cause distortion in the equalized signal [RWK00].

The abovementioned problem (iii) is visualized in **Figure 4.17**. Panel (a) shows an RTF ($\tau_{60} \approx 300$ ms) and panel (b) shows a cut-out of the frequency range 2100 Hz to 2500 Hz of the same RTF (solid grey line), the corresponding TF of the LRC filter (dashed line) and the equalized TF (thick solid black line). The dips in the RTF are compensated for by large peaks in the equalizer's TF to achieve an overall flat spectrum of the equalized system (see e.g. around 2300 Hz in Figure 4.17 (b)). In such areas noise which is usually introduced at the microphone may be strongly amplified.

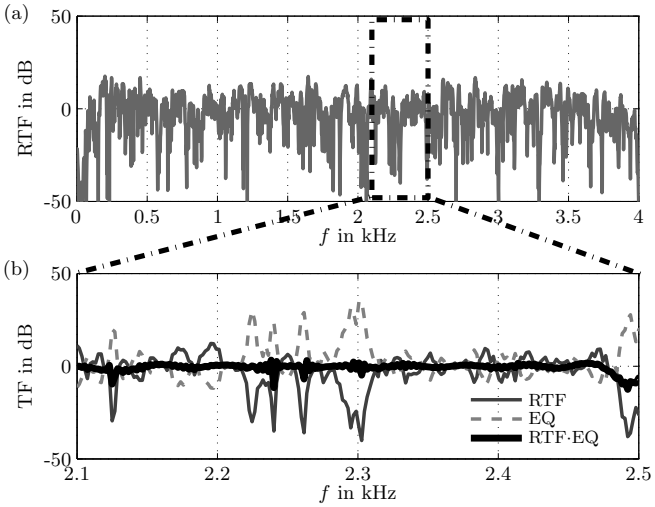

Figure 4.17: (a) Common room transfer function (RTF) (reverberation time $\tau_{60} \approx 300$ ms) (b) RTF, corresponding LRC filter transfer function and equalized system.

Problem (iv), i.e. mismatch between the *true RIR* $\boldsymbol{h}$, which is generally unknown in real-world systems, and the *RIR estimate* $\hat{\boldsymbol{h}}$ may lead to a severe degradation of the LRC performance.

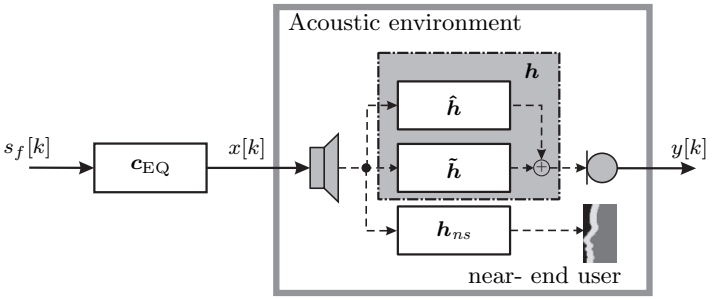

Figure 4.18: Visualization of RIR mismatch. The RIR $\boldsymbol{h}$ can be split into a part which is correctly identified $\hat{\boldsymbol{h}}$ and the system misalignment $\tilde{\boldsymbol{h}}$. A further error for the equalizer is introduced by the fact that the RIR to the reference microphone $\boldsymbol{h}$ may be different from the RIR to the near-end user $\boldsymbol{h}_{ns}$ due to spatial mismatch.

The two main reasons for mismatch are visualized in **Figure 4.18**, i.e. the error introduced to the equalizer due to the fact that the filter is designed for the RIR between loudspeaker and microphone $\boldsymbol{h}$ and not for the RIR between loudspeaker and the ear(s) of the user $\boldsymbol{h}_{ns}$,

$$\tilde{\boldsymbol{h}}' = \boldsymbol{h}_{ns} - \boldsymbol{h}, \tag{4.4.10}$$

(spatial mismatch) and the estimation error $\tilde{\boldsymbol{h}} = \boldsymbol{h} - \hat{\boldsymbol{h}}$ introduced by imperfect identification of $\boldsymbol{h}$ as defined in (3.1.1). Both errors can be combined to

$$\tilde{\boldsymbol{h}}' = \boldsymbol{h}_{ns} - \left(\tilde{\boldsymbol{h}} + \hat{\boldsymbol{h}}\right). \tag{4.4.11}$$

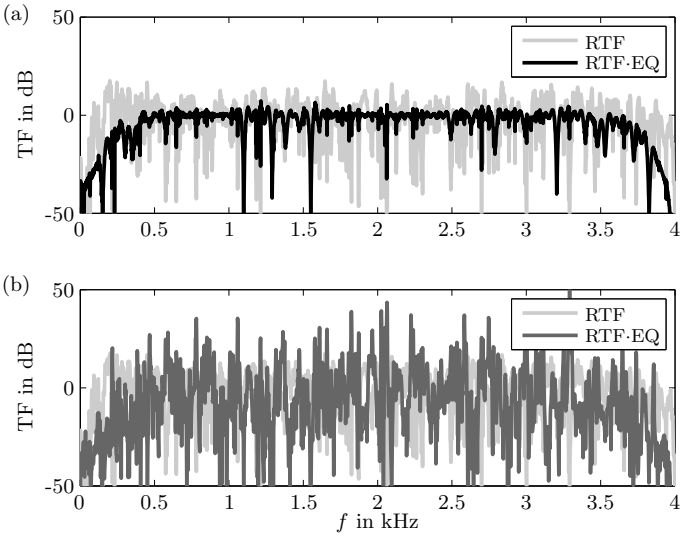

Figure 4.19: Robustness of LRC. (a) RTF and equalized TF without mismatch. (b) RTF and equalized TF with spatial mismatch of ≈ 10 cm.

The performance degradation that occurs without countermeasures is exemplarily shown in **Figure 4.19**. Panel (a) shows an RTF in light grey and the corresponding equalized TF without mismatch in black, i.e. $\tilde{\boldsymbol{h}}' = \boldsymbol{0}$, and panel (b) shows the same RTF and an equalized TF that was calculated with spatial mismatch of ≈ 10 cm, i.e. the position of the listener was at ≈ 10 cm distance from the reference microphone and, thus, $\tilde{\boldsymbol{h}}' \neq \boldsymbol{0}$. It can be clearly seen that mismatch between the correct RIR that has to

be equalized and the identified one may lead to severe distortions in the equalized impulse response and the corresponding transfer function.

Robustness regarding spatial mismatch can be increased by multi-channel approaches [Mou85, RWK99, RWK00, GKMK08c] and will be tackled in the following section. The influence of RIR estimation errors [ZGN08, GKMK08d, GKMK08b] will be topic of Chapter 5.

4.4.3 MIMO LS-Equalizer

The previously described problems can be partly tackled by the extension to a multiple input multiple output (MIMO) system [MK88] as depicted in **Figure 4.20**. Multi-channel LRC is superior to single-channel LRC due to the following reasons: (i) If spatial diversity can be exploited by the use of multiple loudspeakers perfect inversion may be possible [MK88] by exploiting the so-called multiple input/output inverse theorem (MINT) if the RIRs do not have common zeros in the z-domain. (ii) Multi-microphone systems increase spatial robustness compared to single-channel LRC systems [GKMK08d].

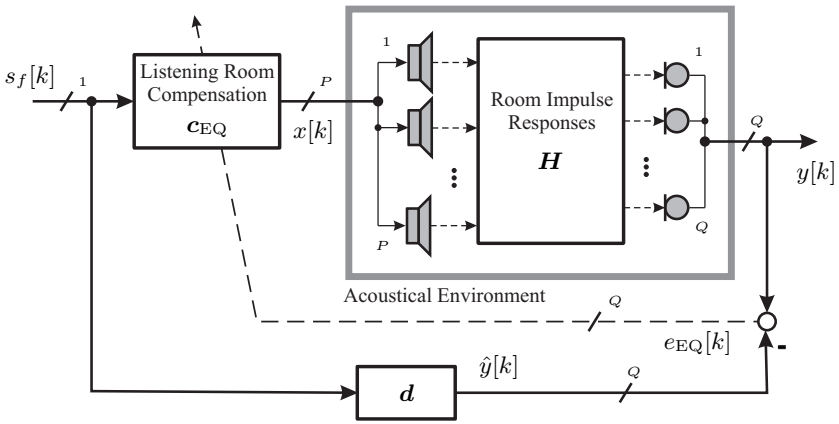

Figure 4.20: Multi-channel sound reproduction system.

Derivation of the MIMO LS-equalizer is straightforward by minimizing the Euclidean distance between the concatenated overall system of LRC filters and RIRs $\boldsymbol{H}_{\mathrm{CM}}\boldsymbol{c}_{\mathrm{EQ}}$ and the desired target systems $\boldsymbol{d}$.

$$\hat{\boldsymbol{c}}_{\mathrm{EQ}} = \underset{\boldsymbol{c}_{\mathrm{EQ}}}{\operatorname{argmin}} \, ||\boldsymbol{H}_{\mathrm{CM}}\boldsymbol{c}_{\mathrm{EQ}} - \boldsymbol{d}||^2 \tag{4.4.12}$$

Solving (4.4.12) in the same way as in Section 4.4.1 leads to

$$\boldsymbol{c}_{\mathrm{EQ}} = \boldsymbol{H}_{\mathrm{CM}}^{+}\boldsymbol{d} \tag{4.4.13}$$

with the following vector and matrix definitions:

$$\boldsymbol{c}_{\mathrm{EQ}} = \left[\, \boldsymbol{c}_{\mathrm{EQ},1}^{T},\, \boldsymbol{c}_{\mathrm{EQ},2}^{T},\, \ldots,\, \boldsymbol{c}_{\mathrm{EQ},P}^{T} \,\right]^{T} \tag{4.4.14}$$

$$\boldsymbol{c}_{\mathrm{EQ},p} = \left[\, c_{\mathrm{EQ},p,0},\, c_{\mathrm{EQ},p,1},\, \ldots,\, c_{\mathrm{EQ},p,L_{\mathrm{EQ}}-1} \,\right]^{T} \tag{4.4.15}$$

$$\boldsymbol{H}_{\mathrm{CM}} = \begin{bmatrix} \boldsymbol{H}_{\mathrm{CM},11} & \cdots & \boldsymbol{H}_{\mathrm{CM},P1} \\ \vdots & \ddots & \vdots \\ \boldsymbol{H}_{\mathrm{CM},1Q} & \cdots & \boldsymbol{H}_{\mathrm{CM},PQ} \end{bmatrix} \tag{4.4.16}$$

$$\boldsymbol{H}_{\mathrm{CM},pq} = \mathrm{convmtx}\left\{ \boldsymbol{h}_{pq}^{T}, L_{\mathrm{EQ}} \right\} \tag{4.4.17}$$

$$\boldsymbol{h}_{pq} = [h_{pq,0}, h_{pq,1}, ..., h_{pq,L_h-1}]^{T} \tag{4.4.18}$$

$$\boldsymbol{d} = \left[\, \boldsymbol{d}_{1}^{T},\, \boldsymbol{d}_{2}^{T},\, \ldots,\, \boldsymbol{d}_{Q}^{T} \,\right]^{T} \tag{4.4.19}$$

$$\boldsymbol{d}_q = [\underbrace{0,\, ...,\, 0}_{\tilde{k}_{0,q}},\, d_0,\, d_1,\, ...,\, d_{L_d-1},\, \underbrace{0,\, ...,\, 0}_{L_h+L_{\mathrm{EQ}}-1-L_d-\tilde{k}_{0,q}}\,]^{T}. \tag{4.4.20}$$

The stacked coefficient vector of the LRC filter(s) and the channel convolution matrix of size $Q(L_h + L_{\mathrm{EQ}} - 1) \times PL_{\mathrm{EQ}}$ built from the RIR coefficients are denoted by $\boldsymbol{c}_{\mathrm{EQ}}$ and $\boldsymbol{H}_{\mathrm{CM}}$, respectively. Channel matrix $\boldsymbol{H}_{\mathrm{CM}}$ is composed of single input single output (SISO) sub-matrices $\boldsymbol{H}_{\mathrm{CM},pq}$ as defined in (4.2.7). For each microphone position an individual desired system may be defined in $\boldsymbol{d}_q$, e.g. to compensate for different delays due to different sound transmission times. Differing delays for different channels q can be advantageous if the theoretical delay differences between loudspeakers and microphones are known from the geometry [EN89].

Multiple Loudspeakers (MISO LRC)

If more reproduction channels are available as depicted in **Figure 4.21**, exact equalization is possible using Bezout's theorem [MK88, YHC05]. The following MISO equalization was introduced in [MK86, MK88] as the multiple input/output inverse theorem (MINT). Given a set of P RTFs $H(z)$ in the z-domain which do not have common zeros, a set of filters $C_{\mathrm{EQ}}(z)$ can be found such that [GTN07, HDM07, MK88]

$$\sum_{p=1}^{P} H_p(z) C_{\mathrm{EQ},p}(z) = 1. \tag{4.4.21}$$

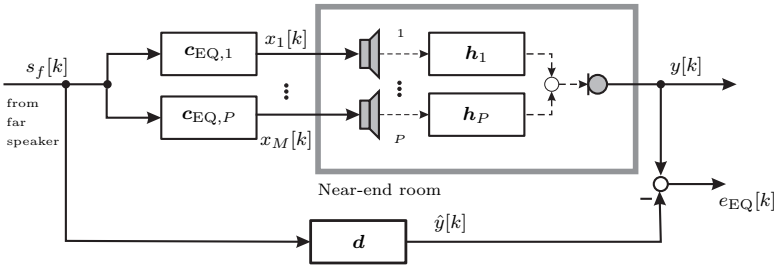

Figure 4.21: Multi-channel setup for listening-room compensation and AEC.

As shown in **Figure 4.22**, a nearly perfect equalization can be achieved using multiple loudspeakers. Panels (e) and (f) show the performance of an

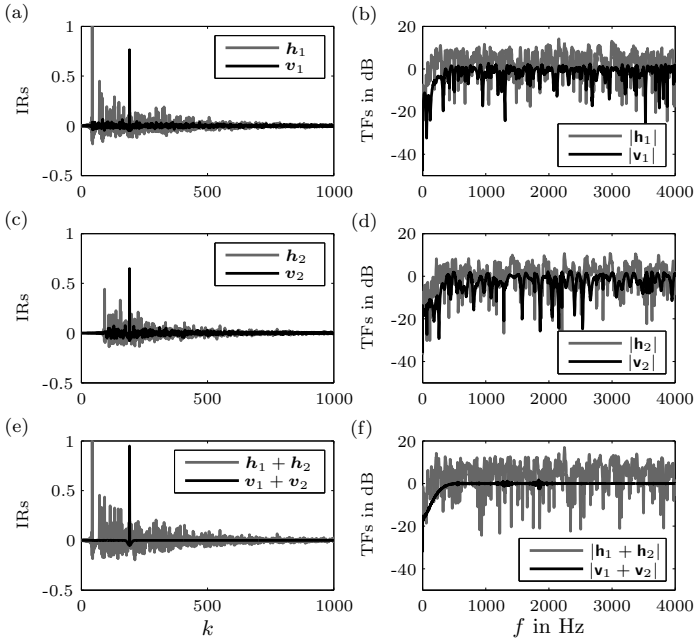

Figure 4.22: Multi-channel (MISO) equalization of room impulse responses. Room reverberation time is $\tau_{60} = 250$ ms, RIR length is $L_h = 2048$ and EQ length is $L_{\mathrm{EQ}} = 2048$ in panels (a)-(d) and $L_{\mathrm{EQ}} = 1024$ in panels (e) and (f), respectively.

LRC system in time-domain (left panels) and in frequency-domain (right panels) for $P = 2$ loudspeakers and $Q = 1$ microphone. The equalized system $\boldsymbol{v} = \boldsymbol{v}_1 + \boldsymbol{v}_2$ (cf. (4.2.5) on p. 72) is very close to the desired high-pass chosen for $\boldsymbol{d}$. If the results are compared to the performance of one single LRC filter, the equalization is much better as it is visible from panels (a) and (b) for RIR $\boldsymbol{h}_1$ as well as panels (c) and (d) for RIR $\boldsymbol{h}_2$. Please note that for the two channel LRC filter only half of the filter coefficients is used in each channel for a fair comparison

For the results in Figure 4.22 knowledge of the true RIR is assumed which in real-world systems is unknown or at least erroneously estimated. Although much better equalization can be achieved by exploiting spatial diversity using multiple loudspeakers, robustness in terms of estimation errors decreases. Furthermore, channel identification for multi-source systems may not have a unique solution, in general, which is known as the *stereo-problem of acoustic echo cancellation* [BMS98b] and which can be easily seen from the error signal of a stereo AEC filter

$$||e_{\mathrm{AEC}}[k]||^2 = ||\boldsymbol{x}_1^T[k](\boldsymbol{h}_1 - \boldsymbol{c}_{\mathrm{AEC}_1}) + \boldsymbol{x}_2^T[k](\boldsymbol{h}_2 - \boldsymbol{c}_{\mathrm{AEC}_2})||^2 \qquad (4.4.22)$$

as visualized in **Figure 4.23**.

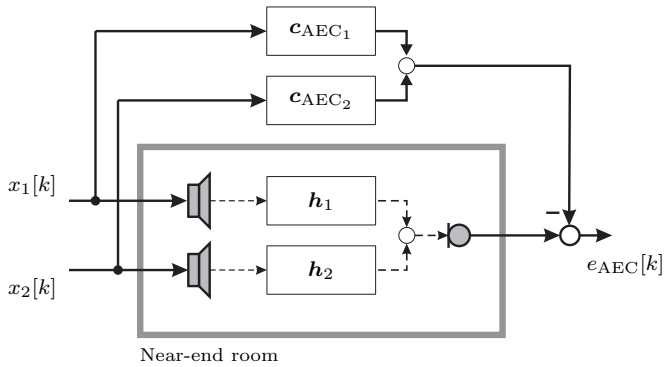

Figure 4.23: 2-channel system identification by AEC filters.

Several solutions exist to minimize (4.4.22) besides the desired solution $\boldsymbol{c}_{\mathrm{AEC},1} = \boldsymbol{h}_1$ and $\boldsymbol{c}_{\mathrm{AEC},2} = \boldsymbol{h}_2$. Approaches for better system identification in case of multiple loudspeakers have been proposed, e.g. decorrelation of the loudspeaker signals, such as adding (masked) uncorrelated noise, nonlinear processing, etc [BMS98b]. However, the system identification performance of multi-loudspeaker AEC systems is not sufficient for an equalizer relying on this information as it is illustrated in **Figure 4.24**. While panels

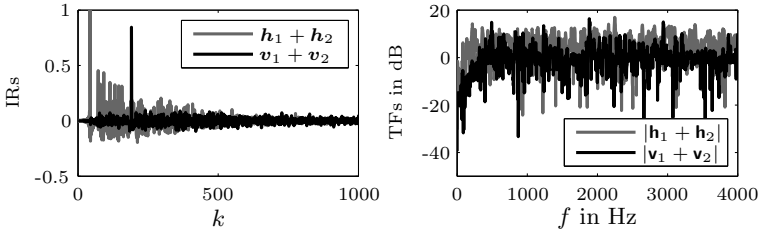

Figure 4.24: AEC Stereo-Problem for multi-channel (MISO) equalization of room impulse responses. Room reverberation time is $\tau_{60} = 250$ ms, RIR length is $L_h = 2048$ and EQ length is $L_{\mathrm{EQ}} = 1024$.

(e) and (f) of Figure 4.22 show the equalization results for correct system identification $\boldsymbol{c}_{\mathrm{AEC},1} = \boldsymbol{h}_1[k]$ and $\boldsymbol{c}_{\mathrm{AEC},2}[k] = \boldsymbol{h}_2$, Figure 4.24 shows results for $\boldsymbol{c}_{\mathrm{AEC},1} = \boldsymbol{h}_1 + \boldsymbol{n}$ and $\boldsymbol{c}_{\mathrm{AEC},2} = \boldsymbol{h}_2 - \boldsymbol{n}$. Since the same disturbance vector $\boldsymbol{n}$ was used, $||e_{\mathrm{AEC}}[k]||^2$ in (4.4.22) still equals zero. Figure 4.24 shows that inversion fails even if the disturbance $\boldsymbol{n}$ is of very low power. Since estimation errors always occur during filter convergence and while tracking of time-variant RIRs and due to the so-called *tail-effect* of stereo acoustic echo cancellation, multi-loudspeaker system inversion is often not sufficiently robust to be used in quickly changing real-world systems.

Multiple Microphones (SIMO LRC)

Spatial robustness can be increased by using multiple microphones as depicted in **Figure 4.25**.
The LRC filter in Figure 4.25 aims at equalization of the RIRs to all reference microphones and, by this, at a mean equalization for different spatial positions. For the following simulations a microphone geometry as depicted in **Figure 4.26** was chosen.
In **Figure 4.27**, the LRC filter is designed for a single loudspeaker system ($P = 1$) and for $Q \in \{1, 12, 28\}$ reference microphones lying on a rectangle in the center of the specific panel. The room dimensions are (5.6 m x 4.375 m x 3.5 m) and the loudspeaker position is at (1.7 m, 2.0 m, 1.0 m).
It can be seen that the use of multiple microphones increases spatial robustness while the maximum achievable SRR enhancement decreases slightly from 14.8 dB to 12 dB. This is due to the fact that a multi-microphone LRC system leads to a mean equalization for the given spatial positions of the reference microphones.
Taking a closer look at the equalized systems for SIMO channels that are depicted in **Figure 4.28** reveals that joint equalization of several RIRs

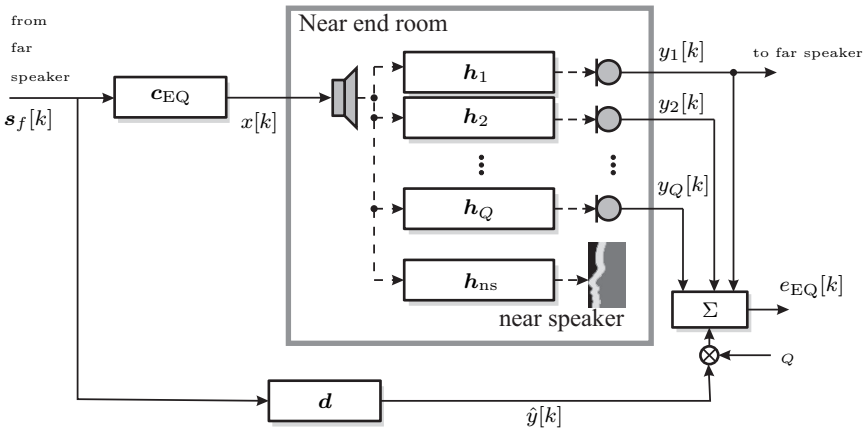

Figure 4.25: Multi-channel setup for listening-room compensation.

Figure 4.26: Example for $Q = 12$ microphones placed on a rectangle.

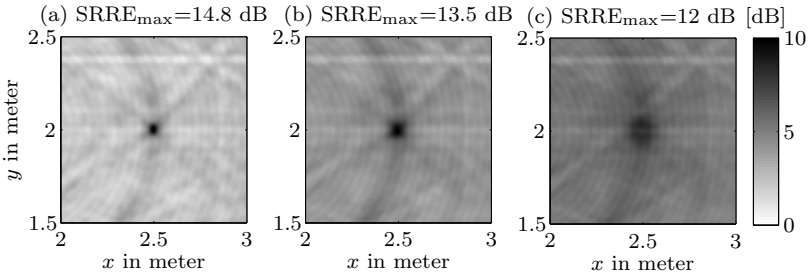

Figure 4.27: Spatial robustness of LRC filter in terms of SRRE as a function of the location for $P = 1$ loudspeaker and varying number of microphones. (a) 1 reference microphone, (b) 12 reference microphones, and (c) 28 reference microphones.

by means of one single LRC filter may not be possible especially in higher frequency regions. Panels (a)-(d) of Figure 4.28 show two RIRs $\boldsymbol{h}_{\{1,2\}}$ and the corresponding RTFs $\mathbf{h}_{\{1,2\}}$ in grey as well as the corresponding equalized systems $\boldsymbol{v}_{\{1,2\},\text{ind}} = \boldsymbol{H}_{\text{CM},\{1,2\}}\boldsymbol{c}_{\text{EQ},\{1,2\},\text{ind}}$ and $\mathbf{v}_{\{1,2\},\text{ind}}$ in black in

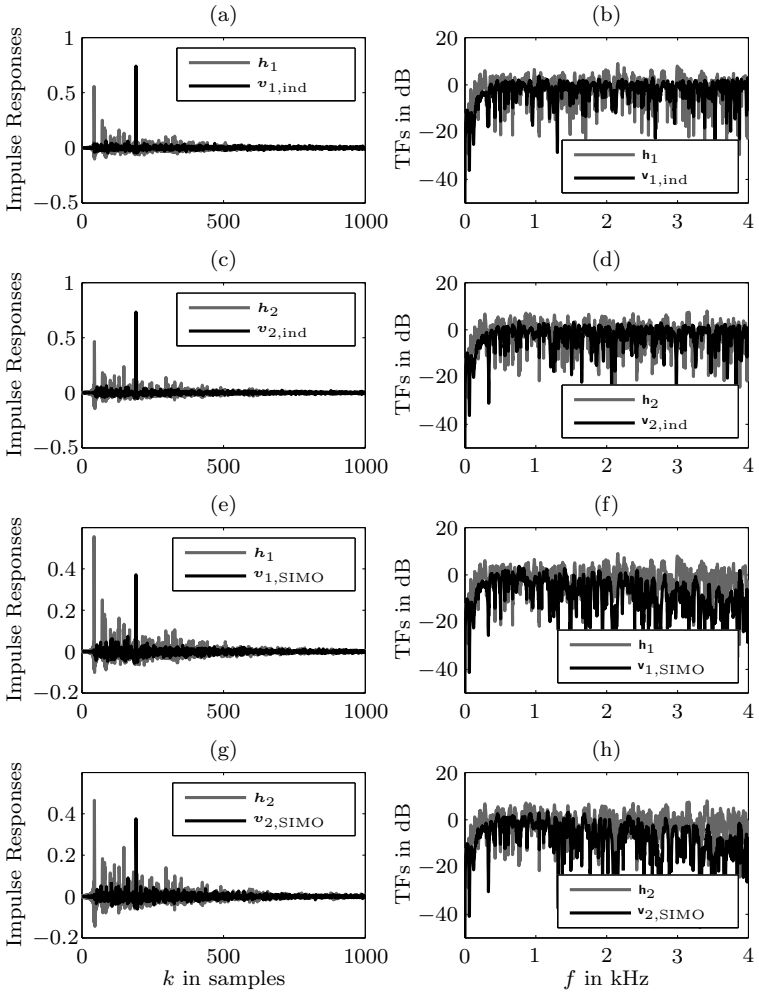

Figure 4.28: Equalization of SIMO channel: (a) and (c) show RIRs and equalized system for independently designed equalizers $\boldsymbol{c}_{\mathrm{EQ},\{1,2\},\mathrm{ind}}$ for filter length $L_{\mathrm{EQ}} = 1024$, (e) and (g) show the same RIRs and equalized systems for jointly designed equalizer for filter length $L_{\mathrm{EQ}} = 1024$, (b), (d), and (f) corresponding transfer functions in dB.

time- and frequency-domain, respectively, for the case that an individual

LRC filter $\boldsymbol{c}_{\mathrm{EQ},\{1,2\},\mathrm{ind}}$ is designed for each RIR. Panels (e)-(h) of Figure 4.28 show the same RIRs and RTFs in grey, however, this time one single LRC filter is designed to equalize both RIRs at the same time, resulting in the equalized systems $\boldsymbol{v}_{\{1,2\},\mathrm{SIMO}} = \boldsymbol{H}_{\mathrm{CM},\{1,2\}}\boldsymbol{c}_{\mathrm{EQ,SIMO}}$ and $\mathbf{v}_{\{1,2\},\mathrm{SIMO}} = \boldsymbol{F}\boldsymbol{v}_{\{1,2\},\mathrm{SIMO}}$ (black lines) in time- and frequency-domain, respectively. Robustness of LRC filters depends on frequency as described in [RWK00] which is clearly visible for the higher frequencies of the equalized TFs $\mathbf{v}_{\{1,2\},\mathrm{SIMO}}$ in panels (f) and (h).

Multiple Microphones and Multiple Loudspeakers (MIMO)

If a second loudspeaker is added to the system the overall performance is increased as it can be seen from the achievable maximum SRRE values in **Figure 4.29**. However, using multiple loudspeakers again leads to a loss of spatial robustness which becomes obvious by comparing panels (a) of Figures 4.27 and 4.29.

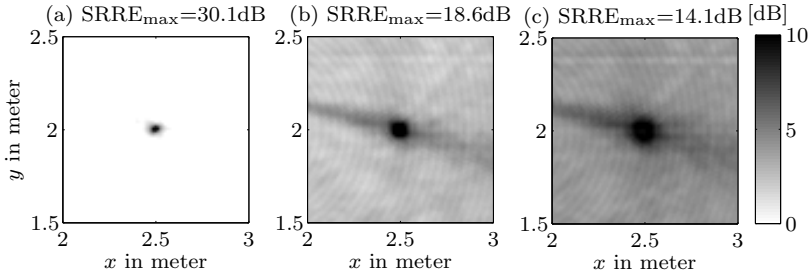

Figure 4.29: Spatial robustness of listening-room compensation in terms of SRRE as a function of the location for $P = 2$ loudspeakers and varying number of microphones. (a) 1 reference microphone, (b) 12 reference microphones, and (c) 28 reference microphones.

It can, thus, be stated that robustness of LRC filters can be increased by using multiple microphones and performance can be increased by the use of multiple loudspeakers. As shown in panel (c) of Figure 4.29 a good LRC filter performance can be achieved in a spatial area of ≈ 15 cm diameter by a system using $P = 2$ loudspeakers and $Q = 28$ microphones. Depending on the chosen microphone array geometry it is possible to decrease the number of needed microphones or to increase the spatially robust area.

An example for an equalized MIMO system using $P = 3$ loudspeakers and $Q = 2$ microphones is shown in **Figure 4.30**. It can be seen, that mathematically, a good equalization performance can be achieved by MIMO LRC

filters designed according to MINT. In frequency-domain the desired high-pass characteristic is approximated quite well and in time-domain the reflections are suppressed below ≈ 40 dB.

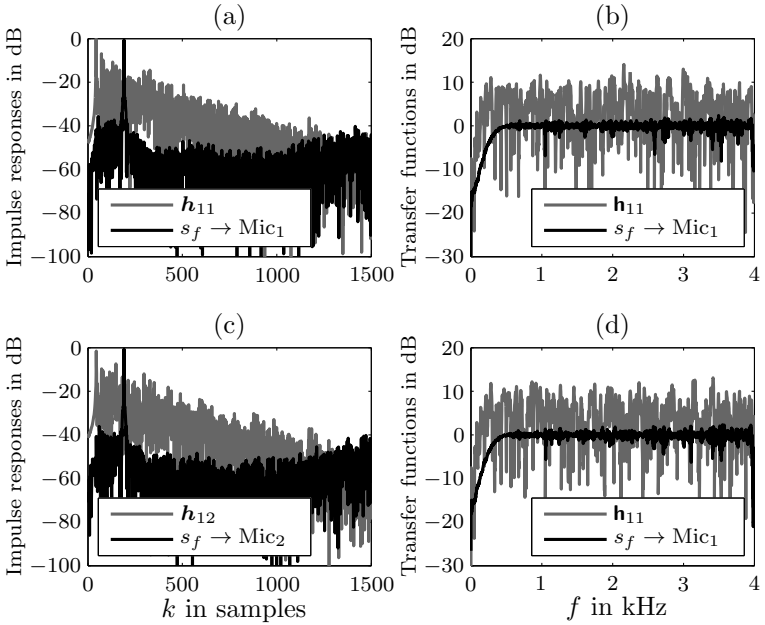

Figure 4.30: Equalization of MIMO system with P=3 loudspeakers and Q=2 microphones: (a) and (b) RIRs (grey) and equalized systems (black); (b) and (d) corresponding transfer functions in dB. $L_{\text{EQ}} = 1024$.

4.5 Gradient Algorithms for Listening-Room Compensation

The least-squares equalizer discussed in the previous sections suffers from the problem to invert the channel convolution matrix $\boldsymbol{H}_{\text{CM}}$ (cf. (4.4.16)) which may have a size of several thousand. Since the RIR is time varying, e.g. due to changes in the acoustic environment or changes caused by moving speakers, the LRC filter coefficients have to be recalculated frequently. Even minor RIR changes require a recalculation of the complete LRC filter

[RWK00]. Thus, for time-variant acoustic environments the computational complexity of the least-squares LRC approach (4.4.6) is by far too high, particularly for a real-time implementation.
Adaptive filters with appropriate learning algorithms based on the well-known least-mean-squares (LMS) algorithm are capable of tracking time-variant conditions by minimizing the time-varying error signal $e_{\mathrm{EQ}}[k]$. Examples known from the field of active noise control (ANC) are the filtered-X LMS (FxLMS) [WSS81] or the modified filtered-X LMS (mFxLMS) [Bja92, KM96] (cf. Sections 4.5.1 and 4.5.2). In this thesis, a decoupled version of the mFxLMS with a faster convergence speed will be derived which was introduced in [GKMK08a, GKMK08b] and is derived in Section 4.5.3 in detail in time-domain and block-frequency-domain. This algorithm allows for an overclocking of the filter update and, by this, for even faster convergence at the cost of additional computational load which is, however, still by far lower than the computational load of the direct least-squares approach in (4.4.6). The decoupled filtered-X least-mean-squares (dFxLMS) algorithm will be evaluated under realistic conditions including ambient noise and estimation errors of the room impulse response (RIR) in Section 4.5.3.

4.5.1 The Filtered-X LMS

A benchmark for adaptive equalization known from ANC systems is the filtered-X least-mean-squares (FxLMS) algorithm [WS85, BQ00] which is depicted in **Figure 4.31**. The FxLMS was developed independently by Widrow et al. [WSS81] and Burgess [Bur81] in the early 1980's. It differs from the conventional LMS algorithm (cf. Section 3.2.1) by prefiltering of the input signal of the update path with the RIR to be equalized (signal $r[k]$ in Figure 4.31).

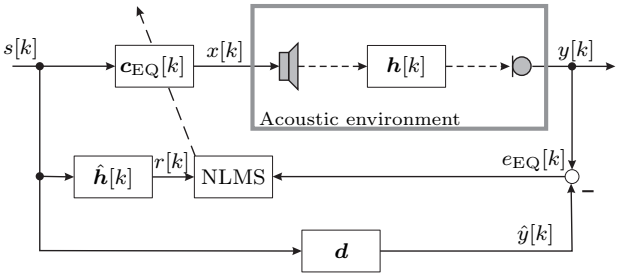

Figure 4.31: Block diagram of filtered-X LMS (FxLMS).

As depicted in Figure 4.31, the speech signal $s[k]$ is processed by the LRC filter $\boldsymbol{c}_{\mathrm{EQ}}[k]$ which precedes the acoustic channel $\boldsymbol{h}[k]$. The aim of the

equalizer is to minimize the Euclidean distance between the equalized system $\boldsymbol{v}[k] = \boldsymbol{H}_{\mathrm{CM}}[k]\boldsymbol{c}_{\mathrm{EQ}}[k]$ and the desired target system $\boldsymbol{d}$. The input signal of the LMS update path has to be filtered with the acoustic channel $\boldsymbol{h}[k]$ to ensure convergence of the algorithm [WS85]. Thus, as for the least-squares equalizer, knowledge about the RIR is needed which is not available in real-world systems. Thus, an estimate $\hat{\boldsymbol{h}}[k]$ of the RIR (known as the plant model in ANC systems) is needed. Such an estimate can be generated by an AEC filter as described in Chapter 3. The combination of LRC systems and AEC systems will be topic of Chapter 5. Since changes of the filter coefficients $\boldsymbol{c}_{\mathrm{EQ}}[k]$ using the FxLMS algorithm do not have an immediate impact on the error signal $e_{\mathrm{EQ}}[k]$ due to the delay of the RIR, a small step-size μ for the filter update is required to ensure stability. Thus, especially if a large filter length is needed, the FxLMS algorithm suffers from slow convergence.

The FxLMS algorithms can be written in matrix/vector notation as summarized in **Algorithm 1**.

Algorithm 1 Filtered-X LMS (FxLMS)

1: $\boldsymbol{r}[k] = \hat{\boldsymbol{H}}^{T}_{\mathrm{CM}}[k]\boldsymbol{s}_{\mathrm{I}}[k]$
2: $e_{\mathrm{EQ}}[k] = \boldsymbol{s}^{T}_{\mathrm{II}}[k]\boldsymbol{H}_{\mathrm{CM}}[k]\boldsymbol{c}_{\mathrm{EQ}}[k] - \boldsymbol{s}^{T}_{\mathrm{II}}[k]\boldsymbol{d}$
3: $\boldsymbol{c}_{\mathrm{EQ}}[k+1] = \boldsymbol{c}_{\mathrm{EQ}}[k] + \mu_{\mathrm{FxLMS}}\boldsymbol{r}[k]e_{\mathrm{EQ}}[k]$

In Algorithm 1,

$$\boldsymbol{c}_{\mathrm{EQ}}[k] = \left[\, c_{\mathrm{EQ},0}[k],\, c_{\mathrm{EQ},1}[k],\, \ldots,\, c_{\mathrm{EQ},L_{\mathrm{EQ}}-1}[k] \,\right]^{T} \tag{4.5.1}$$

is the time-varying coefficient vector of the LRC filter, $\boldsymbol{H}_{\mathrm{CM}}[k]$ is the channel convolution matrix as defined in (4.2.7) and

$$\hat{\boldsymbol{H}}_{\mathrm{CM}}[k] = \mathrm{convmtx}\left\{ \left[\hat{h}_0[k], \hat{h}_1[k], ..., \hat{h}_{L_{\hat{h}}-1}[k] \right]^{T}, L_{\mathrm{EQ}} \right\} \tag{4.5.2}$$

the corresponding channel convolution matrix of size $L_{\hat{h}} + L_{\mathrm{EQ}} - 1 \times L_{\mathrm{EQ}}$ generated by the coefficients of the channel estimate vector $\hat{\boldsymbol{h}}[k]$. $\boldsymbol{d}$ is the coefficient vector of the desired system as defined in (4.4.3), $\boldsymbol{r}[k]$ is the signal vector in the update path after convolution with the channel estimate and $\boldsymbol{s}_{\{\mathrm{I,II}\}}[k]$ is the input signal vector given here for two different lengths that

are needed in Algorithm 1:

$$\boldsymbol{r}[k] = [\, r[k],\, \ldots,\, r[k - L_{\mathrm{EQ}} + 1] \,]^T \tag{4.5.3}$$

$$\boldsymbol{s}_{\mathrm{I}}[k] = \left[\, s[k],\, \ldots,\, s[k - L_{\hat{h}} - L_{\mathrm{EQ}} + 2] \,\right]^T, \tag{4.5.4}$$

$$\boldsymbol{s}_{\mathrm{II}}[k] = [\, s[k],\, \ldots,\, s[k - L_h - L_{\mathrm{EQ}} + 2] \,]^T. \tag{4.5.5}$$

The lengths of the RIR, the RIR estimate, the LRC filter and the desired system are denoted by L_h, $L_{\hat{h}}$, L_{EQ}, and L_d, respectively.

4.5.2 The Modified Filtered-X LMS

To overcome the problem of a heavily reduced allowed convergence speed in comparison to the conventional LMS algorithm, several approaches have been proposed in the literature [BQ00, Bou03, ABZ07], such as a fast version of the FxLMS [Dou97], the adjoint-LMS [Wan96, BQ00], the filtered-u LMS [KM96], the filtered ϵ LMS [Bou03] or the modified filtered-X LMS (mFxLMS) algorithm [Bja92, RF94, KNKP94]. The latter is depicted in **Figure 4.32** since an enhanced version of the modified filtered-X least-mean-squares (mFxLMS) will be derived in the following Subsection 4.5.3.

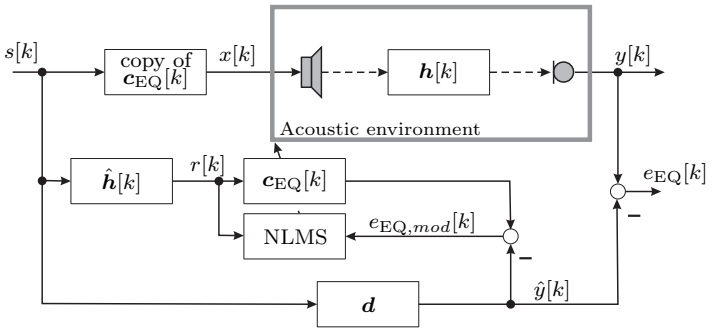

Figure 4.32: Block diagram of modified Filtered-X LMS (mFxLMS).

The reason for slower convergence of the FxLMS compared to the conventional NLMS, i.e. that the LRC filter precedes the acoustic channel and, by this, has no immediate influence on the error signal $e_{\mathrm{EQ}}[k]$, is avoided by the mFxLMS by exchanging the order of LRC filter $\boldsymbol{c}_{\mathrm{EQ}}[k]$ and RIR (estimate) $\hat{\boldsymbol{h}}[k]$ in the update path. By this, the filter update is based on a modified error signal $e_{\mathrm{EQ},mod}[k]$ which is directly influenced by the LRC filter $\boldsymbol{c}_{\mathrm{EQ}}[k]$. Please note, that the error signal $e_{\mathrm{EQ}}[k]$ in Figure 4.32 is no longer needed for the algorithm itself, but e.g. for assessment of the algorithm's performance by objective quality measures (cf. e.g. Appendix A.2.1).

The original LRC filter preceding the acoustic channel $\boldsymbol{h}[k]$ in Figure 4.32 is now just updated by copying the filter coefficients. The filter update is now based on the RIR estimate $\hat{\boldsymbol{h}}[k]$ only. Assuming a correct system identification ($\hat{\boldsymbol{h}}[k] = \boldsymbol{h}[k]$), the convergence performance of the mFxLMS depicted in Figure 4.32 is the same as for the conventional NLMS algorithm because the update of the filter coefficients has direct impact on the error signal $e_{\mathrm{EQ},mod}[k]$. In contrast to the FxLMS, the calculation of the error signal of the modified FxLMS $e_{\mathrm{EQ},mod}[k]$ is independent of the true room impulse response $\boldsymbol{h}[k]$ and, thus, independent of the microphone signal $y[k]$. **Algorithm 2** summarizes the mFxLMS.

Algorithm 2 Modified filtered-X LMS (mFxLMS)

1: $\boldsymbol{r}[k] = \hat{\boldsymbol{H}}^{T}_{\mathrm{CM}}[k]\boldsymbol{s}_{\mathrm{I}}[k]$
2: $e_{\mathrm{EQ},mod}[k] = \boldsymbol{r}^{T}[k]\boldsymbol{c}_{\mathrm{EQ}}[k] - \boldsymbol{s}^{T}_{\mathrm{I}}[k]\boldsymbol{d}$
3: $\boldsymbol{c}_{\mathrm{EQ}}[k+1] = \boldsymbol{c}_{\mathrm{EQ}}[k] + \mu_{\mathrm{mFxLMS}}\boldsymbol{r}[k]e_{\mathrm{EQ},mod}[k]$

The mFxLMS has been extended to the multi-channel case in [ESN87, Dou95, Dou97] and enhanced versions based on the RLS [BQ00] or APA have been proposed in [Dou95, Bou03, ABZ07]. An enhanced version of the mFxLMS for the purpose of LRC will be introduced in the following.

4.5.3 The Decoupled Filtered-X LMS

The mFxLMS depicted in Figure 4.32 and described by Algorithm 2 already allows for a larger step-size than the conventional FxLMS and, thus, for faster convergence. Since the filter update path is more or less independent of the system which should be equalized, i.e. the error signal in line 2 of Algorithm 2 is independent of the real RIR and only the estimate or model is used, the update path can also be excited by an independent signal $s_{dec}[k]$ as it is shown in **Figure 4.33** for switch S_1 in the depicted position. S_1 switches between modified FxLMS (mFxLMS) and decoupled FxLMS (dFxLMS).
By the possibility to arbitrarily choose the filter input signal $s_{dec}[k]$, even faster convergence can be achieved. Best convergence would be archived for so-called *perfect sequences* [AD94] as an input signal for the NLMS $r[k]$. Since generation of such a signal by crafting the excitation signal $s_{dec}[k]$ at the input of the RIR estimate is difficult, a white Gaussian excitation for $s_{dec}[k]$ can be used. An additional advantage of the proposed algorithm is the fact that with a decoupled input signal for the update path an over-clocking of the filter update is possible and, by this a trade-off between convergence speed and computational complexity. The proposed dFxLMS

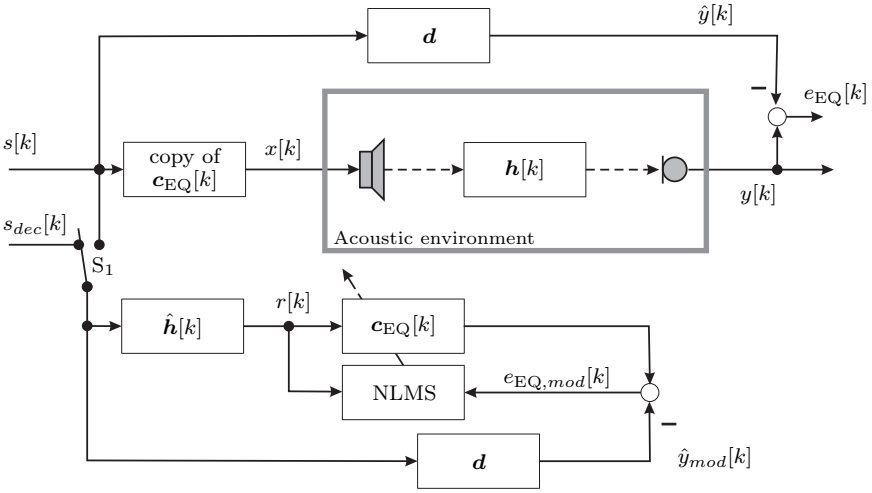

Figure 4.33: Block diagram of decoupled Filtered-X LMS (dFxLMS).

algorithm is summarized in Algorithm 3. Here, $O \geq 1, O \in \mathbb{N}$ is the overclocking factor. The term overclocking does not mean over-sampling since this would affect both, the sampling rate of the signals and the systems. The sampling rate of the systems, such as that of $\hat{\boldsymbol{h}}[k]$, remains unchanged while more than one input sample $s_{dec}[k]$ is processed before copying the filter weights $\boldsymbol{c}_{\mathrm{EQ}}[k]$ to the upper branch. **Algorithm 3** summarizes the single-channel dFXLMS algorithm in time-domain.

Algorithm 3 Decoupled version of modified filtered-X LMS (dFxLMS)

1: **for** $i = 0 : O - 1$ **do**
2: $\quad \boldsymbol{r}[k+i] = \hat{\boldsymbol{H}}_{\mathrm{CM}}^{T}[k]\boldsymbol{s}_{dec}[k+i]$
3: $\quad e_{\mathrm{EQ},mod}[k+i] = \boldsymbol{r}^{T}[k+i]\boldsymbol{c}_{\mathrm{EQ}}[k+i] - \boldsymbol{s}_{dec}^{T}[k+i]\boldsymbol{d}$
4: $\quad \boldsymbol{c}_{\mathrm{EQ}}[k+i+1] = \boldsymbol{c}_{\mathrm{EQ}}[k+i] + \mu_{\mathrm{dFxLMS}}\boldsymbol{r}[k+i]e_{\mathrm{EQ},mod}[k+i]$
5: **end for**
6: Copy updated EQ coefficients $\boldsymbol{c}_{\mathrm{EQ}}[k+i+1]$ to upper branch

Multi-Channel Frequency-Domain Implementation

To further decrease the computational complexity, the developed dFxLMS algorithm will be derived in frequency-domain in the following. Furthermore, it will be extended to the multi-channel case to allow for an

increased performance and spatial robustness (cf. Section 4.4.3). The frequency-domain description is based on the multi-delay filtering approach [Shy92, MAG95, SP90, KNHOb98, BQ00].
The multi-delay filter described in Section 2.2 can be extended to the MIMO case by [SP90, MAG95]

$$\mathbf{Y}[\ell] = \boldsymbol{G}\mathbf{X}[\ell]\mathbf{H}[\ell] \tag{4.5.6}$$

using the definitions of the DFT matrix $\boldsymbol{F}_{2L\times L}$ as defined in (2.2.5), the shifting matrix $\tilde{\boldsymbol{I}}_{2L\times 2L}$ as defined in (2.2.17), the constraining matrix $\boldsymbol{G}$ of size $2L \times 2L$ as defined in (2.2.21) and the block-frequency-domain channel matrix

$$\mathbf{H}[\ell] = \text{bdiag}\{\underbrace{\boldsymbol{F}_{2L\times L},\, \boldsymbol{F}_{2L\times L},\, ...,\, \boldsymbol{F}_{2L\times L}}_{PL'_h}\}\boldsymbol{H}[\ell], \tag{4.5.7}$$

of size $2LL'_hP \times Q$ defined by transforming all RIRs

$$\boldsymbol{H}[\ell] = \begin{bmatrix} \boldsymbol{h}_{11}[\ell] & \cdots & \boldsymbol{h}_{1Q}[\ell] \\ \vdots & \ddots & \vdots \\ \boldsymbol{h}_{P1}[\ell] & \cdots & \boldsymbol{h}_{PQ}[\ell] \end{bmatrix} \tag{4.5.8}$$

to the partitioned frequency-domain. The multi-channel signal matrix

$$\begin{aligned}\mathbf{X}[\ell] = \Big[&\breve{\mathbf{X}}_1[\ell],\, ...,\, \breve{\mathbf{X}}_1[\ell - L'_h + 1],\, ...,\\ &\breve{\mathbf{X}}_P[\ell],\, ...,\, \breve{\mathbf{X}}_P[\ell - L'_h + 1]\Big],\end{aligned} \tag{4.5.9}$$

of size $2L \times 2LL'_hP$ in (4.5.6) is generated by concatenating the single-channel/single-block input-signal matrices

$$\breve{\mathbf{X}}_p[\ell] = \text{diag}\{\boldsymbol{F}_{2L\times L}\boldsymbol{x}_p[\ell] + \tilde{\boldsymbol{I}}_{2L\times 2L}\,\boldsymbol{F}_{2L\times L}\boldsymbol{x}_p[\ell-1]\}, \tag{4.5.10}$$

$$\boldsymbol{x}_p[\ell] = [x_p[\ell L],\, x_p[\ell L + 1],\, ...,\, x_p[(\ell+1)L - 1]]^T \tag{4.5.11}$$

and, by this, is a straightforward generalization of the singe-channel definitions in (2.2.15) and (2.2.16).
The signal matrix for the update path of the block-frequency-domain adaptive filter

$$\mathbf{S}[\ell] = \Big[\breve{\mathbf{S}}[\ell],\, ...,\, \breve{\mathbf{S}}[\ell - L'_{\hat{h}} + 1]\Big], \tag{4.5.12}$$

$$\breve{\mathbf{S}}[\ell] = \text{diag}\{\boldsymbol{F}_{2L\times L}\boldsymbol{s}[\ell] + \tilde{\boldsymbol{I}}_{2L\times 2L}\,\boldsymbol{F}_{2L\times L}\boldsymbol{s}[\ell-1]\}, \tag{4.5.13}$$

is defined in dependence of switch S_1 in Figure 4.33 either based on the signal $s[k]$ as in

$$\boldsymbol{s}[\ell] = [s[\ell L],\, s[\ell L+1],\, ...,\, s[(\ell+1)L-1]]^T \tag{4.5.14}$$

or on the decoupled input signal $s_{\mathrm{dec}}[k]$ as in

$$\boldsymbol{s}[\ell] = [s_{\mathrm{dec}}[\ell L],\, s_{\mathrm{dec}}[\ell L+1],\, ...,\, s_{\mathrm{dec}}[(\ell+1)L-1]]^T\,. \tag{4.5.15}$$

The input signal $\mathbf{S}[\ell]$ has to be filtered by each block-frequency-domain channel estimate $\hat{\mathbf{h}}_{pq}[\ell] = \mathrm{bdiag}\{\underbrace{\boldsymbol{F}_{2L\times L},\, \boldsymbol{F}_{2L\times L},\, ...,\, \boldsymbol{F}_{2L\times L}}_{L'_{\hat{h}}}\}\hat{\boldsymbol{h}}_{pq}[\ell]$ to obtain

$$\mathbf{r}_{pq}[\ell] = \boldsymbol{G}\mathbf{S}[\ell]\hat{\mathbf{h}}_{pq}[\ell]. \tag{4.5.16}$$

After transforming (4.5.16) to time-domain,

$$\boldsymbol{r}_{pq}[\ell] = \boldsymbol{W}^{01}_{L\times 2L}\boldsymbol{F}^{-1}_{2L\times 2L}\mathbf{S}[\ell]\hat{\mathbf{h}}_{pq}[\ell], \tag{4.5.17}$$

the block-frequency-domain signal matrix $\mathbf{R}_{pq}[\ell]$ can be defined as

$$\mathbf{R}_{pq}[\ell] = \left[\breve{\mathbf{R}}_{pq}[\ell],\, ...,\, \breve{\mathbf{R}}_{pq}[\ell - L'_{\hat{h}} + 1]\right], \tag{4.5.18}$$

$$\breve{\mathbf{R}}_{pq}[\ell] = \mathrm{diag}\{\boldsymbol{F}_{2L\times L}\boldsymbol{r}_{pq}[\ell] + \tilde{\boldsymbol{I}}_{2L\times 2L}\,\boldsymbol{F}_{2L\times L}\boldsymbol{r}_{pq}[\ell-1]\}, \tag{4.5.19}$$

for each loudspeaker-microphone pair $\{p, q\}$ in the usual way. For the following derivations the matrices (4.5.18) can be concatenated to result in a matrix

$$\mathbf{R}_q[\ell] = [\mathbf{R}_{1q}[\ell],\, ...,\, \mathbf{R}_{Pq}[\ell]] \tag{4.5.20}$$

containing the sub matrices for all loudspeakers $p = 1..P$ and a matrix

$$\mathbf{R}_p[\ell] = [\mathbf{R}_{p1}[\ell],\, ...,\, \mathbf{R}_{pQ}[\ell]] \tag{4.5.21}$$

containing the sub matrices for all microphones $q = 1..Q$.
The block-frequency-domain error signal matrix $\mathbf{E}_{\mathrm{EQ},mod,q}[\ell]$ for each reference microphone q can then be calculated by

$$\mathbf{E}_{\mathrm{EQ},mod,q}[\ell] = \hat{\mathbf{Y}}_q[\ell] - \boldsymbol{G}\mathbf{R}_q[\ell]\mathbf{c}_{\mathrm{EQ}}[\ell]. \tag{4.5.22}$$

In (4.5.22), $\hat{\mathbf{Y}}_p[\ell]$ is the desired signal which results from filtering the input signal $\mathbf{S}[\ell]$ with the respective desired system $\mathbf{d}_q$ for each channel q.

$$\hat{\mathbf{Y}}_q[\ell] = \left[\hat{\overset{\circ}{\mathbf{Y}}}_q[\ell],\, ...,\, \hat{\overset{\circ}{\mathbf{Y}}}_q[\ell - L'_h + 1]\right], \tag{4.5.23}$$

$$\hat{\overset{\circ}{\mathbf{Y}}}_q[\ell] = \mathrm{diag}\{\boldsymbol{F}_{2L\times L}\hat{\boldsymbol{y}}_q[\ell] + \tilde{\boldsymbol{I}}_{2L\times 2L}\,\boldsymbol{F}_{2L\times L}\hat{\boldsymbol{y}}_q[\ell-1]\}, \tag{4.5.24}$$

$$\hat{\boldsymbol{y}}_q[\ell] = \boldsymbol{W}^{01}_{L\times 2L}\boldsymbol{F}^{-1}_{2L\times 2L}\mathbf{S}[\ell]\mathbf{d}_q, \tag{4.5.25}$$

$$\mathbf{S}[\ell] = \left[\check{\mathbf{S}}[\ell],\, ...,\, \check{\mathbf{S}}[\ell - L'_v + 1]\right] \tag{4.5.26}$$

$$\mathbf{d}_q = \mathrm{bdiag}\{\underbrace{\boldsymbol{F}_{2L\times L},\, \boldsymbol{F}_{2L\times L},\, ...,\, \boldsymbol{F}_{2L\times L}}_{L'_v}\}\boldsymbol{d}_q. \tag{4.5.27}$$

Please note, that the definition of $\mathbf{S}[\ell]$ in (4.5.26) slightly differs from that in (4.5.12). However, since the only difference is the number of input blocks taken into account (L'_v in (4.5.26) and $L'_{\hat{h}}$ in (4.5.12)) to match the respective length for the block-frequency-domain filtering, no explicit distinction will be made in the following to increase readability.

The block-frequency-domain LRC filter coefficients $\mathbf{c}_{\mathrm{EQ}}[\ell]$ already used in (4.5.22) are defined as a stacked vector containing all p channels.

$$\mathbf{c}_{\mathrm{EQ}}[\ell] = \mathrm{bdiag}\{\underbrace{\boldsymbol{F}_{2L\times L},\, \boldsymbol{F}_{2L\times L},\, ...,\, \boldsymbol{F}_{2L\times L}}_{PL'_{EQ}}\}\boldsymbol{c}_{\mathrm{EQ}}[\ell] \tag{4.5.28}$$

$$\boldsymbol{c}_{\mathrm{EQ}}[\ell] = \left[\, \boldsymbol{c}^T_{\mathrm{EQ},1}[\ell],\, \boldsymbol{c}^T_{\mathrm{EQ},2}[\ell],\, ...,\, \boldsymbol{c}^T_{\mathrm{EQ},P}[\ell]\,\right]^T \tag{4.5.29}$$

$$\boldsymbol{c}_{\mathrm{EQ},p}[\ell] = \left[\, c_{\mathrm{EQ},p,0}[\ell],\, c_{\mathrm{EQ},p,1}[\ell],\, ...,\, c_{\mathrm{EQ},p,L_{\mathrm{EQ}}-1}[\ell]\,\right]^T \tag{4.5.30}$$

A joint error signal

$$\mathbf{E}_{\mathrm{EQ},mod}[\ell] = \sum_{q=1}^{Q}\mathbf{E}_{\mathrm{EQ},mod,q}[\ell], \tag{4.5.31}$$

$$= \sum_{q=1}^{Q}\hat{\mathbf{Y}}_q[\ell] - \boldsymbol{G}\mathbf{R}_q[\ell]\mathbf{c}_{\mathrm{EQ}}[\ell], \tag{4.5.32}$$

$$= \hat{\mathbf{Y}}[\ell] - \boldsymbol{G}\mathbf{R}[\ell]\mathbf{c}_{\mathrm{EQ}}[\ell] \tag{4.5.33}$$

can be defined as the sum of the error signals for each channel q given in

(4.5.22) with the definitions

$$\hat{\mathbf{Y}}[\ell] = \sum_{q=1}^{Q} \hat{\mathbf{Y}}_q[\ell], \tag{4.5.34}$$

$$\mathbf{R}[\ell] = \sum_{q=1}^{Q} \mathbf{R}_q[\ell]. \tag{4.5.35}$$

A cost function $J[\ell]$ [SP90] to obtain a block-frequency-domain version of the dFxLMS can be defined as

$$J[\ell] = (1-\alpha) \sum_{i=0}^{\ell} \alpha^{\ell-i} \mathrm{tr}\{\mathbf{E}_{\mathrm{EQ},mod}^{H}[i] \mathbf{E}_{\mathrm{EQ},mod}[i]\} \tag{4.5.36}$$

with the exponential forgetting factor $0 \leq \alpha \leq 1$. The gradient $\nabla_{\mathbf{c}_{\mathrm{EQ}}} J[\ell]$ will be calculated using (4.5.33) in the following to obtain the minimum of $J[\ell]$ [BR72].

$$\begin{aligned} \nabla_{\mathbf{c}_{\mathrm{EQ}}} J[\ell] =& 2 \frac{\partial J[\ell]}{\partial \mathbf{c}_{\mathrm{EQ}}^{*}} && (4.5.37) \\ =& 2 \frac{\partial}{\partial \mathbf{c}_{\mathrm{EQ}}^{*}} (1-\alpha) \sum_{i=0}^{\ell} \alpha^{\ell-i} \left(\hat{\mathbf{Y}}[i] - \boldsymbol{G}\mathbf{R}[i]\mathbf{c}_{\mathrm{EQ}} \right)^{H} \cdot \\ & \cdot \left(\hat{\mathbf{Y}}[i] - \boldsymbol{G}\mathbf{R}[i]\mathbf{c}_{\mathrm{EQ}} \right) && (4.5.38) \end{aligned}$$

Using the Wirtinger calculus (3.3.22) [Hay02] and setting (4.5.38) to zero we obtain

$$0 \stackrel{!}{=} 2(1-\alpha) \sum_{i=0}^{\ell} \alpha^{\ell-i} \left(-\mathbf{R}^{H}[i] \boldsymbol{G}^{H} \right) \left(\hat{\mathbf{Y}}[i] - \boldsymbol{G}\mathbf{R}[i]\mathbf{c}_{\mathrm{EQ}} \right). \tag{4.5.39}$$

With $\boldsymbol{G}^{H}\boldsymbol{G} = \boldsymbol{G}$ (cf. Appendix D.1 for a proof) and $\boldsymbol{G}^{H}\hat{\mathbf{Y}}[\ell] = \hat{\mathbf{Y}}[\ell]$ (cf. Appendix D.2) we obtain the frequency-domain normal equation

$$\hat{\boldsymbol{\Phi}}_{\mathbf{R}\mathbf{R}}[\ell]\mathbf{c}_{\mathrm{EQ}}[\ell] = \hat{\boldsymbol{\Phi}}_{\mathbf{R}\hat{\mathbf{Y}}}[\ell] \tag{4.5.40}$$

with the CPSD vector between filtered-X signal $\mathbf{R}[\ell]$ and the desired signal $\hat{\mathbf{Y}}[\ell]$

$$\hat{\boldsymbol{\Phi}}_{\mathbf{R}\hat{\mathbf{Y}}}[\ell] = (1-\alpha) \sum_{i=0}^{\ell} \alpha^{\ell-i} \mathbf{R}^{H}[i] \hat{\mathbf{Y}}[i] \tag{4.5.41}$$

and the APSD matrix of the filtered-X signal $\mathbf{R}[\ell]$

$$\hat{\boldsymbol{\Phi}}_{\mathbf{RR}}[\ell] = (1-\alpha)\sum_{i=0}^{\ell}\alpha^{\ell-i}\mathbf{R}^H[\ell]\boldsymbol{G}\mathbf{R}[\ell]. \tag{4.5.42}$$

To obtain iterative update equations, (4.5.41) can be rewritten in its recursive form that can easily be obtained by extracting $\alpha\mathbf{R}^H[i]\hat{\mathbf{Y}}[i]$ from the sum in (4.5.41) and reintroducing $\hat{\boldsymbol{\Phi}}_{\mathbf{R}\hat{\mathbf{Y}}}[\ell-1]$ to result in (4.5.43). Similarly, (4.5.44) can be obtained from (4.5.42).

$$\hat{\boldsymbol{\Phi}}_{\mathbf{R}\hat{\mathbf{Y}}}[\ell] = \alpha\hat{\boldsymbol{\Phi}}_{\mathbf{R}\hat{\mathbf{Y}}}[\ell-1] + (1-\alpha)\mathbf{R}^H[\ell]\hat{\mathbf{Y}}[\ell] \tag{4.5.43}$$

$$\hat{\boldsymbol{\Phi}}_{\mathbf{RR}}[\ell] = \alpha\hat{\boldsymbol{\Phi}}_{\mathbf{RR}}[\ell-1] + (1-\alpha)\mathbf{R}^H[\ell]\boldsymbol{G}\mathbf{R}[\ell]. \tag{4.5.44}$$

The normalization factor $(1-\alpha)$ is usually chosen close to one for speech PSD estimation and assures an asymptotically unbiased estimate [Bri75, Her05].
Introducing (4.5.40) in terms of ℓ and $\ell-1$ for $\hat{\boldsymbol{\Phi}}_{\mathbf{R}\hat{\mathbf{Y}}}[\ell]$ and $\hat{\boldsymbol{\Phi}}_{\mathbf{R}\hat{\mathbf{Y}}}[\ell-1]$ in (4.5.43) leads to

$$\hat{\boldsymbol{\Phi}}_{\mathbf{RR}}[\ell]\mathbf{c}_{\mathrm{EQ}}[\ell] = \alpha\hat{\boldsymbol{\Phi}}_{\mathbf{RR}}[\ell-1]\mathbf{c}_{\mathrm{EQ}}[\ell-1] + (1-\alpha)\mathbf{R}^H[\ell]\hat{\mathbf{Y}}[\ell] \tag{4.5.45}$$

where the dependency of $\hat{\boldsymbol{\Phi}}_{\mathbf{RR}}[\ell-1]$ can be eleminated using (4.5.44).

$$\begin{aligned}\hat{\boldsymbol{\Phi}}_{\mathbf{RR}}[\ell]\mathbf{c}_{\mathrm{EQ}}[\ell] &= \left(\hat{\boldsymbol{\Phi}}_{\mathbf{RR}}[\ell] - (1-\alpha)\mathbf{R}^H[\ell]\boldsymbol{G}\mathbf{R}[\ell]\right)\mathbf{c}_{\mathrm{EQ}}[\ell-1] \\ &\quad + (1-\alpha)\mathbf{R}^H[\ell]\hat{\mathbf{Y}}[\ell] \qquad (4.5.46)\\ &= \hat{\boldsymbol{\Phi}}_{\mathbf{RR}}[\ell]\mathbf{c}_{\mathrm{EQ}}[\ell-1] \\ &\quad + (1-\alpha)\mathbf{R}^H[\ell]\left(\hat{\mathbf{Y}}[\ell] - \boldsymbol{G}\mathbf{R}[\ell]\mathbf{c}_{\mathrm{EQ}}[\ell-1]\right) \qquad (4.5.47)\end{aligned}$$

With the definition of the block-frequency-domain error vector $\mathbf{E}_{\mathrm{EQ},mod}[\ell]$ as given in (4.5.31) and after multiplying by $\hat{\boldsymbol{\Phi}}_{\mathbf{RR}}^{-1}[\ell]$ from the left-hand side the update equation can be written as

$$\mathbf{c}_{\mathrm{EQ}}[\ell] = \mathbf{c}_{\mathrm{EQ}}[\ell-1] + \mu\,\hat{\boldsymbol{\Phi}}_{\mathbf{RR}}^{-1}[\ell]\mathbf{R}^H[\ell]\mathbf{E}_{\mathrm{EQ},mod}[\ell]. \tag{4.5.48}$$

Please note, that the smoothing factor $(1-\alpha)$ has been replaced by the step-size μ in (4.5.48) to be consistent to the usual definition of a gradient update rule. The matrix $\boldsymbol{G}$, that leads to a high computational load during calculation of $\hat{\boldsymbol{\Phi}}_{\mathbf{RR}}[\ell]$ in (4.5.44) can be roughly approximated by a diagonal matrix $\boldsymbol{G} \approx \boldsymbol{I}/2$, especially for larger matrix sizes [BM01], as visualized in

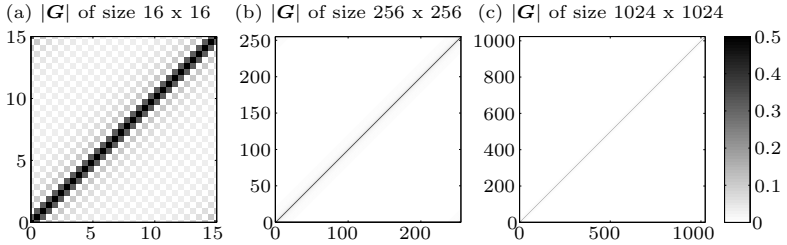

Figure 4.34: Illustration of constraining matrix $\boldsymbol{G}$ for different matrix sizes.

Figure 4.34. By this, the computational complexity of the filter update (4.5.48) is drastically reduced.
A schematic of the multi-channel dFxLMS algorithm is examplarily depicted in **Figure 4.35** for $P = 2$ loudspeaker channels and $Q = 3$ microphone channels.

4.5.4 Simulation Results

In the following, the performance of the previously described FxLMS, mFxLMS and dFxLMS algorithms will be evaluated. The simulation results are based on RIRs $\boldsymbol{h}_{pq}[\ell]$ characterized by room reverberation times of $\tau_{60} \approx 500$ ms. Although in practical environments an RIR is of infinite length, they have been truncated after $L_h = 4096$ samples due to sufficiently decay. The LRC filter length was chosen to $L_{\mathrm{EQ}} = 1024$ at a sampling rate of $f_s = 8000$ Hz. The following simulations are given for $P = 1$ loudspeakers and $Q = 3$ microphones to lead to a spatially more robust design (compare Section 4.4.3). The delay introduced by the equalizer was $k_0 = 512$ samples for all channels.
Figure 4.36 compares the dFxLMS algorithm with the FxLMS and the mFxLMS by means of the system distance

$$D_{\mathrm{dB}}[k] = 10\log_{10}\frac{||\boldsymbol{H}_{\mathrm{CM}}[k]\boldsymbol{c}_{\mathrm{EQ}}[k] - \boldsymbol{d}||^2}{||\boldsymbol{d}||^2} \tag{4.5.49}$$

between the equalized system $\boldsymbol{H}_{\mathrm{CM}}[k]\boldsymbol{c}_{\mathrm{EQ}}[k]$ and the desired system $\boldsymbol{d}$. For Figure 4.36 the RIR estimate is assumed to be perfectly known before the influence of RIR estimation errors will be analyzed in the following simulations.
It can be seen in Figure 4.36 that FxLMS and mFxLMS algorithms perform poor since their update is based on the highly correlated speech input signal

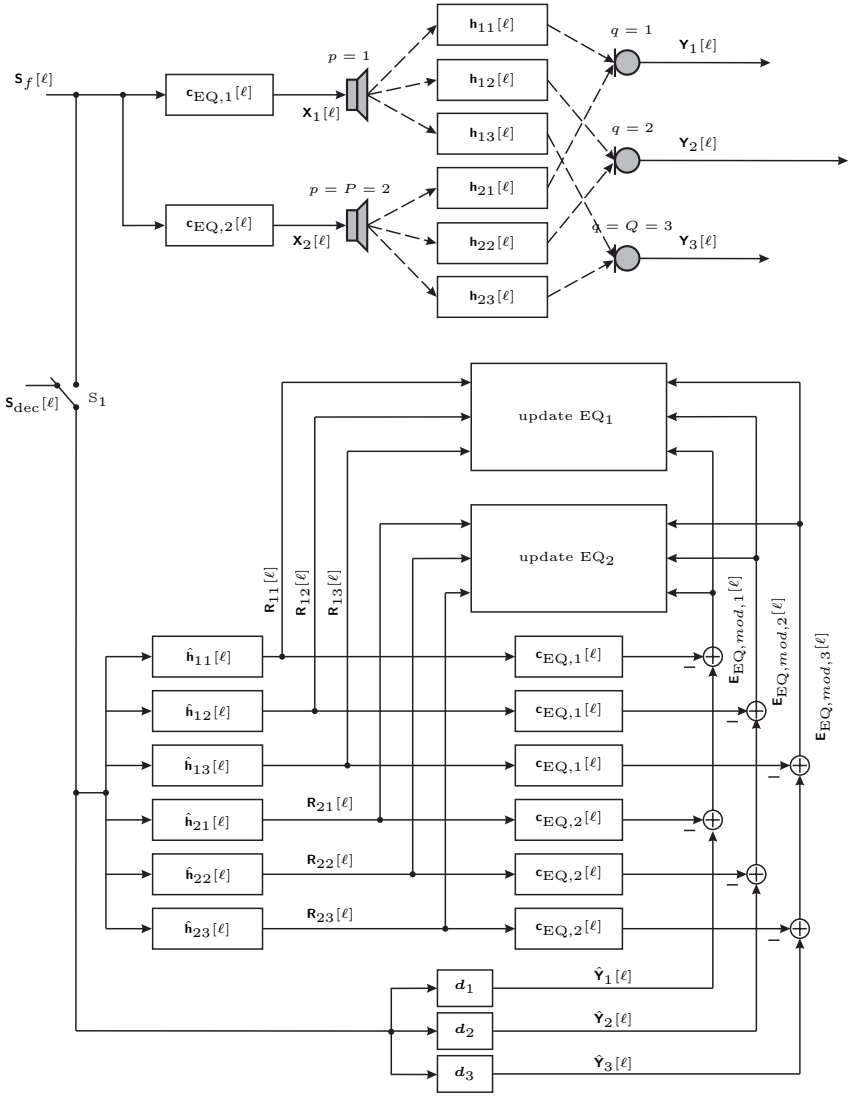

Figure 4.35: Multi-channel structure for mFxLMS and dFxLMS algorithms for $P = 3$ loudspeakers and $Q = 2$ microphones.

$\mathbf{s}[\ell]$. The mFxLMS algorithm (dashed line) performs slightly better than the conventional FxLMS algorithm. A large performance gain is achieved by the dFxLMS algorithm (dash-dotted line) even without overclocking. Please

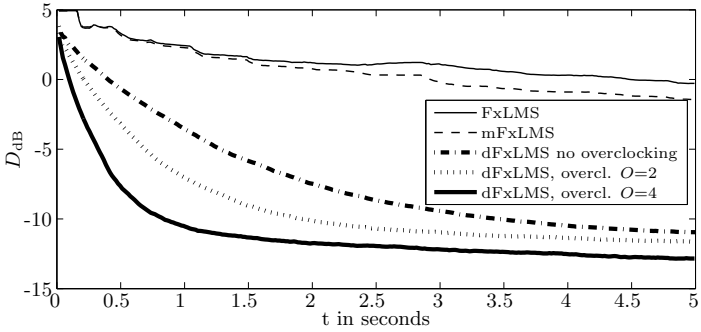

Figure 4.36: Comparison of FxLMS, mFxLMS and dFxLMS (speech input).

note, that mFxLMS and dFxLMS algorithms are the same for speech input (switch S_1 in right position in Figures 4.33 and 4.35) and no overclocking. Thus, the distance between the dashed line (mFxLMS) and the dash-dotted line (dFxLMS, $O = 1$) is due to the white excitation used for $\mathbf{S}_{\mathrm{dec}}[\ell]$. Further performance gain is achieved if an overclocking factor $O \geq 1$ is used as it can be seen from the lower two curves.

Figure 4.37 shows simulation results for the more realistic case that the RIR is only imperfectly estimated. For this purpose the RIR estimate is generated by adding white Gaussian noise to the correct RIR with different SNRs of -10 dB (left panel of Figure 4.37) and 0 dB (right panel). Here, the term SNR denotes the ratio between RIR power $||\boldsymbol{h}[k]||^2$ and error power $||\tilde{\boldsymbol{h}}[k]||^2$. Please note, that more realistic errors $\tilde{\boldsymbol{h}}[k]$ for RIR identification are described in Chapter 5 where mutual influences of the subsystems of AEC and LRC filter are analyzed (cf. e.g. Figure 5.11 on page 147). The following simulations would look very similar with a more realistic error.

As it can be seen from a comparison of Figure 4.36 and Figure 4.37, an imperfect estimate of the RIR leads to a decreased performance of the LRC filter. However, the dFxLMS algorithm still clearly outperforms FxLMS and mFxLMS algorithms. Although the dFxLMS algorithm is not more robust in terms of RIR estimation errors since it is just a quickly converging version of the mFxLMS algorithm, it becomes obvious from Figure 4.37 that the mFxLMS algorithm as well as the FxLMS algorithm are not suitable for a real-world hands-free scenario due to their slow convergence, whereas the dFxLMS algorithm can be applied in quickly changing environments. The dFxLMS converges towards the least-squares solution and, thus can be used as a real-time version for time-varying environments.

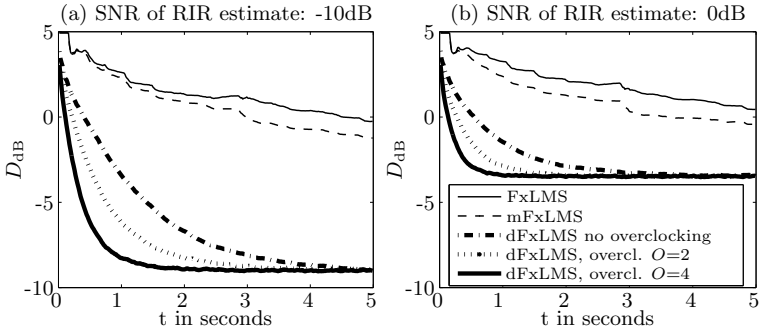

Figure 4.37: Comparison of FxLMS, mFxLMS and dFxLMS for imperfect RIR estimates.

4.6 Weighted Least-Squares Equalization

As already stated in Section 4.2, e.g. during the discussion of Figure 4.9 on page 83, least-squares approaches for LRC filters may lead to mathematically small but clearly perceivable late echoes in the equalized IR. This is due to the fact that in the typical linear decay in logarithmic time-domain of an RIR, later echoes are partly *masked* by more early echoes for the human auditory system.

The phenomenon of *masking* occurs whenever one sound is rendered inaudible by another more dominant sound which may be presented even at a different frequency and/or time instance [FZ07]. Two different kinds of masking exist, i.e. frequency masking and temporal masking. Frequency masking describes the effect that the threshold of hearing is elevated by a sinusoidal tone or narrow-band noise (the *masker*) and another tone may be below this raised threshold even if it's frequency is different from that of the masker [FZ07]. Effects of frequency masking are exploited in state-of-the-art audio coding such as MP3 or the advanced audio codec (AAC) [Bra97] as well as in noise reduction algorithms [Gus99, GMK06a].

Temporal masking, i.e. non-simultaneous masking, which will be used for the LRC filter design in the following is visualized in **Figure 4.38**.

A sound which is below the masking level is inaudible even if it is presented after the masker is switched off (forward masking a.k.a. post-masking) or even before the masker is switched on (backward masking a.k.a. pre-masking). Temporal masking is due to the temporal processing found in the organ of Corti in the inner ear of animals and humans [Moo03].

The influence of temporal masking on impulse responses was analyzed in [Fie01]. The effects of forward masking and backward masking are limited

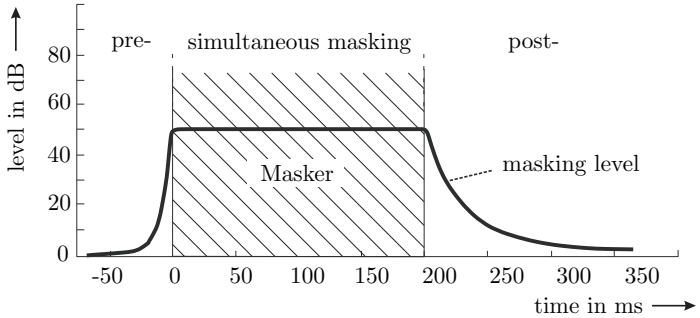

Figure 4.38: Temporal masking effects (adapted from [VM06, FZ07]).

to 100 - 200 ms [Fie01, Moo97, OT89] and 5 - 20 ms [Raa61, Fie01, Moo97, DS84], respectively. Forward masking acts like simultaneous masking for about 4 ms and after this period a fall-off in masking is roughly -35 dB per octave [Fie01]. Forward masking can be described by the window function

$$\boldsymbol{w}_{\mathrm{FW}} = [\underbrace{1,\, 1,\, ...,\, 1}_{K_1},\, \underbrace{\boldsymbol{w}_0}_{K_2}]^T \tag{4.6.1}$$

with

$$w_{0,k} = 10^{\frac{3\alpha_{w,\mathrm{WLS}}}{\log_{10}(K_0/K_1)}\log_{10}(k/K_1)+0.5} \tag{4.6.2}$$

which is adapted from [MMK10, JMGM11] with $K_0 = (t_0 + 0.2)f_s$, $K_1 = (t_0 + 0.004)f_s$ and $K_2 = L_h + L_{\mathrm{EQ}} - 1 - K_1$. The time of the direct sound is denoted by t_0 and $\alpha_{w,\mathrm{WLS}} \leq 1$ is a factor that influences the steepness of the window. For $\alpha_{w,\mathrm{WLS}} = 1$ the window corresponds to the forward masking found in human subjects.
Lags that occur in equalized IRs before the main peak are perceived as disturbing pre-echoes. Thus, a time-reversed version of $\boldsymbol{w}_0$ with steeper decay (here by using $\alpha_{w,\mathrm{WLS}} = 3$ in (4.6.1)) is used for the lags before t_0 as visualized in **Figure 4.39** for a conventional RIR $\boldsymbol{h}$ (left panel) and for an equalized IR $\boldsymbol{v} = \boldsymbol{H}_{\mathrm{CM}}\boldsymbol{c}_{\mathrm{EQ}}^{\mathrm{LS}}$, resulting from processing of the RIR by an LS equalizer of length $L_{\mathrm{EQ}} = 1024$ (right panel).
In the following, an LRC filter will be derived that allows for emphazising suppression of later parts of the equalized IR. Instead of equally reducing the energy of all lags of the impulse response in (4.4.13) the window $\boldsymbol{w}$ can be used for the LRC filter design by minimizing

$$\boldsymbol{e}_{\mathrm{EQ}}^{\mathrm{WLS}} = \boldsymbol{W}(\boldsymbol{H}_{\mathrm{CM}}\boldsymbol{c}_{\mathrm{EQ}}^{\mathrm{WLS}} - \boldsymbol{d}) \tag{4.6.3}$$

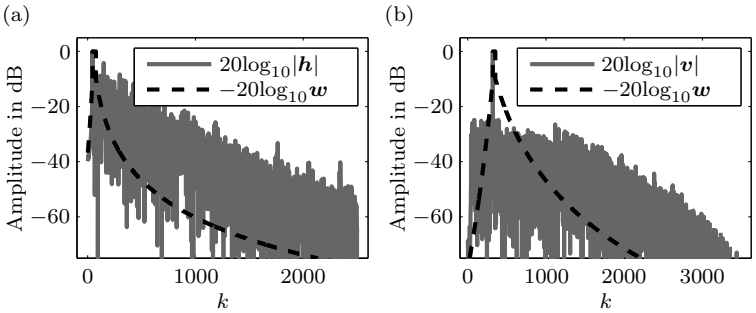

Figure 4.39: (a) RIR ($\tau_{60} \approx 300$ ms) and the temporal masking curve; (b) equalized system $\boldsymbol{v}$ obtained by LS-equalizer of length $L_{\mathrm{EQ}} = 1024$ and the temporal masking curve.

using the window function

$$\boldsymbol{W} = \mathrm{diag}\left\{\boldsymbol{w}\right\}, \tag{4.6.4}$$

$$\boldsymbol{w} = \left[w_0,\, w_1,\, ...,\, w_{L_h+L_{\mathrm{EQ}}-2}\right]^T. \tag{4.6.5}$$

The derivative of the squared error in (4.6.3) w.r.t. the LRC filter coefficients is set to $\mathbf{0}$ to obtain the generalized or weighted least-squares equalizer.

$$\begin{aligned}\frac{\partial\,||\boldsymbol{e}_{\mathrm{EQ}}||^2}{\partial\left(\boldsymbol{c}_{\mathrm{EQ}}^{\mathrm{WLS}}\right)^T} &= \boldsymbol{H}_{\mathrm{CM}}^T\boldsymbol{W}^T\boldsymbol{W}\boldsymbol{H}_{\mathrm{CM}}\boldsymbol{c}_{\mathrm{EQ}}^{\mathrm{WLS}} + \left(\left(\boldsymbol{c}_{\mathrm{EQ}}^{\mathrm{WLS}}\right)^T\boldsymbol{H}_{\mathrm{CM}}^T\boldsymbol{W}^T\boldsymbol{W}\boldsymbol{H}_{\mathrm{CM}}\right)^T \\ &\quad -\boldsymbol{H}_{\mathrm{CM}}^T\boldsymbol{W}^T\boldsymbol{W}\boldsymbol{d} - (\boldsymbol{d}^T\boldsymbol{W}^T\boldsymbol{W}\boldsymbol{H}_{\mathrm{CM}})^T \qquad (4.6.6)\\ &= 2\boldsymbol{H}_{\mathrm{CM}}^T\boldsymbol{W}^T\boldsymbol{W}\boldsymbol{H}_{\mathrm{CM}}\boldsymbol{c}_{\mathrm{EQ}}^{\mathrm{WLS}} - 2\boldsymbol{H}_{\mathrm{CM}}^T\boldsymbol{W}^T\boldsymbol{W}\boldsymbol{d} \stackrel{!}{=} 0 \qquad (4.6.7)\end{aligned}$$

$$\begin{aligned}\Rightarrow\ \boldsymbol{c}_{\mathrm{EQ}}^{\mathrm{WLS}} &= \left(\boldsymbol{H}_{\mathrm{CM}}^T\boldsymbol{W}^T\boldsymbol{W}\boldsymbol{H}_{\mathrm{CM}}\right)^{-1}\boldsymbol{H}_{\mathrm{CM}}^T\boldsymbol{W}^T\boldsymbol{W}\boldsymbol{d} \qquad (4.6.8)\\ &= (\boldsymbol{W}\boldsymbol{H}_{\mathrm{CM}})^{+}\boldsymbol{W}\boldsymbol{d} \qquad (4.6.9)\end{aligned}$$

Please note that, for $\boldsymbol{w} = [1, 1, ..., 1]^T$, the weighted least-squares equalizer in (4.6.9) reduces to the conventional least-squares equalizer as defined in (4.4.6).

Figure 4.40 shows an RIR $\boldsymbol{h}$ (room reverberation time $\tau_{60} \approx 0.5$ s) and the corresponding IR $\boldsymbol{v}$ processed by a weighted least-squares LRC filter (4.6.9) for a filter length of $L_{\mathrm{EQ}} = 4096$ at a sampling rate of $f_s = 8$ kHz in time-domain (upper panel) and frequency-domain (lower panel). The RIR

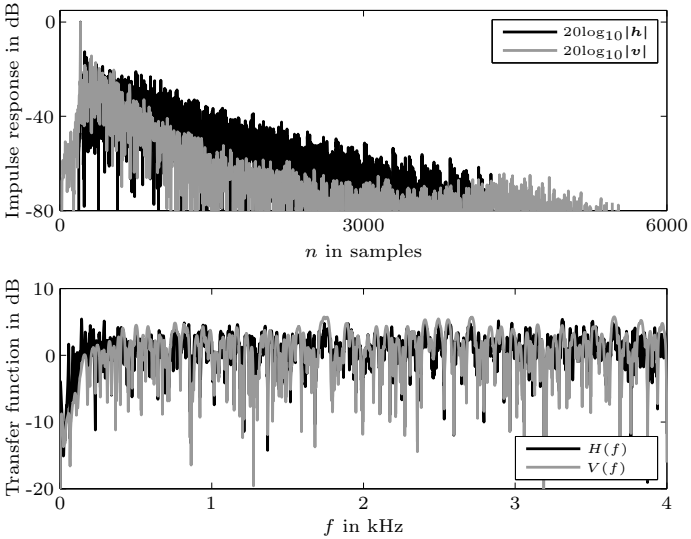

Figure 4.40: RIR $\boldsymbol{h}$ and equalized IR $\boldsymbol{v} = \boldsymbol{H}_{\mathrm{CM}}\, \boldsymbol{c}_{\mathrm{EQ}}^{\mathrm{WLS}}$ in time-domain in dB (upper panel) and the corresponding squared-magnitude spectra in dB (lower panel).

as well as the LRC filter's parameters are the same as for the LS equalizer used for Figure 4.9 on page 83 to allow for a direct comparison.
It can be seen from Figure 4.40 that disturbing late echoes as well as pre-echoes are reduced by applying the exponential window as shown in Figure 4.39 to the LRC filter (4.6.9). The weighted least-squares LRC filter *squeezes* the RIR to result in a quicker decay of the equalized IR $\boldsymbol{v}$ than the original RIR $\boldsymbol{h}$ in time-domain (upper panel). The problem of clearly perceivable late echoes above the original decay of the RIR can be reduced. On the other hand, the performance in frequency-domain is decreased as it can be seen from comparing Figures 4.40 and 4.9. Although the RIR is rather re-shaped than equalized, the term equalization is further used for all LRC approaches.

4.7 Room Impulse Response Shaping

The previously discussed weighted least-squares LRC approach already can be considered as an RIR shaping approach. Here, not perfect spectral flat-

ness of the overall system or an equalization towards one single peak or band-pass in time-domain is desired but only shortening or re-shaping of the RIR. To archive this goal, classical channel shortening concepts known from data transmission [FM73, Kam94, MYR96, MDEJ03] can be applied. For data transmission, channel-shortening concepts are usually used to shorten the effective channel at the receiving side e.g. to fit the guard interval in OFDM systems etc. In [KM05c, KM05b, KM06] these concepts were analyzed for the purpose of LRC and extended to shaping rather than shortening. By this approach the energy in the part of the RIR specified by a window $\boldsymbol{w}_d$ is maximized. As a side condition the energy in that part of the RIR specified by a window $\boldsymbol{w}_u = \mathbf{1}_{[L_{\mathrm{EQ}}+L_h-1]} - \boldsymbol{w}_d$ is kept constant. Common choices for the window function $\boldsymbol{w}_d$ are e.g. [KM05b, JMGM11, KGD12a, KGD12b]

$$\boldsymbol{w}_d = [\underbrace{0,\ ...,\ 0}_{k_0},\ 1,\ ...,\ 1,\ 0,\ ...,\ 0\,] \tag{4.7.1}$$

or a generalization of (4.7.1)

$$\boldsymbol{w}_d = \left[\, w_{d,0},\ ...,\ w_{d,L_{\mathrm{EQ}}+l_h-1} \,\right]^T, \tag{4.7.2}$$

for which the window function

$$w_{d,k} = \begin{cases} 0 & \text{for } 0 \le k \le k_0 - 1, \\ 10^{\,\alpha_{w,\mathrm{IS}}(k-k_0)} & \text{for } k_0 - 1 \le k \end{cases} \tag{4.7.3}$$

could be chosen (between others) [KM05b, JMGM11, KGD13a, KGD13b]. In (4.7.1)-(4.7.3), k_0 is the allowed delay for the LRC filter (cf. also Section 4.4.1) and the factor $\alpha_{w,\mathrm{IS}}$ has been chosen heuristically in [KM05b] to be $\alpha_{w,\mathrm{IS}} = -3 \cdot 10^{-5}$.
A desired system $\boldsymbol{d}_d$ and an undesired system $\boldsymbol{d}_u$, i.e. the desired and undesired part of the equalized IR, can then be defined as [AEK01, KM05b]

$$\boldsymbol{d}_d = \mathrm{diag}\{\boldsymbol{w}_d\}\boldsymbol{H}_{\mathrm{CM}}\boldsymbol{c}_{\mathrm{EQ}}^{\mathrm{ISwPP}} \tag{4.7.4}$$

$$\boldsymbol{d}_u = \mathrm{diag}\{\underbrace{\mathbf{1} - \boldsymbol{w}_d}_{\boldsymbol{w}_u}\}\boldsymbol{H}_{\mathrm{CM}}\boldsymbol{c}_{\mathrm{EQ}}^{\mathrm{ISwPP}}. \tag{4.7.5}$$

Maximizing the energy of $\boldsymbol{d}_d$ while keeping the energy of $\boldsymbol{d}_u$ constant leads to impulse response shortening / shaping [MYR96, MDEJ03] by solving the generalized eigenvalue problem [KM05b, MYR96]

$$\boldsymbol{B}_{\mathrm{BP}} \cdot \boldsymbol{c}_{\mathrm{EQ,opt}}^{\mathrm{ISwPP}} = \boldsymbol{A} \cdot \boldsymbol{c}_{\mathrm{EQ,opt}}^{\mathrm{ISwPP}} \cdot \lambda_{\max} \tag{4.7.6}$$

with

$$\boldsymbol{A} = \boldsymbol{H}_{\mathrm{CM}}^{H} \mathrm{diag}\left\{\boldsymbol{w}_u\right\}^{H} \mathrm{diag}\left\{\boldsymbol{w}_u\right\} \boldsymbol{H}_{\mathrm{CM}} \tag{4.7.7}$$

$$\boldsymbol{B}_{\mathrm{BP}} = \boldsymbol{H}_{\mathrm{CM,BP}}^{H} \mathrm{diag}\left\{\boldsymbol{w}_{\mathrm{BP},d}\right\}^{H} \mathrm{diag}\left\{\boldsymbol{w}_{\mathrm{BP},d}\right\} \boldsymbol{H}_{\mathrm{CM,BP}} \tag{4.7.8}$$

In (4.7.6), $\lambda_{\max}$ is the largest eigenvalue and $\boldsymbol{c}_{\mathrm{EQ,opt}}^{\mathrm{ISwPP}}$ the corresponding eigenvector, which is taken as the LRC coefficient vector [KM05b]. In (4.7.8) the modified RIR convolution matrix $\boldsymbol{H}_{\mathrm{CM,BP}}$ results from convolution of $h[k]$ with the desired system $d[k]$ (here, BP exemplarily stands for 'band-pass'). Furthermore, the length of the window $\boldsymbol{w}_{\mathrm{BP},d}$ is increased by $L_d - 1$ samples compared to the original window $\boldsymbol{w}_d$. See [KM05b] for further details.

4.7.1 Spectral Post Processing

Although the RIR shortening method in (4.7.6) is effective in time-domain, colouration may be introduced in the frequency-domain as already reported in [KM05b]. This colouration can be reduced by post processing of the equalized system $\boldsymbol{v} = \boldsymbol{H}_{\mathrm{CM}}\boldsymbol{c}_{\mathrm{EQ}}$ by a linear prediction error filter as depicted in **Figure 4.41**.

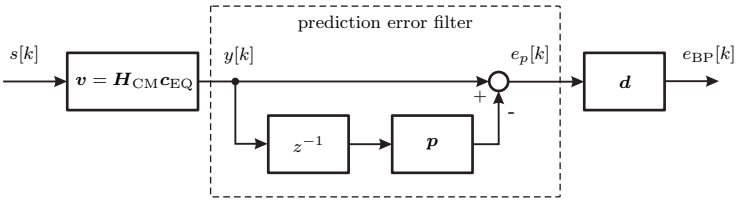

Figure 4.41: Linear prediction error filter for post processing after RIR reshaping for LRC.

The error signal $e_p[k]$ is weighted by the desired system $\boldsymbol{d}$ again to focus the predictor's performance to the spectral area of interest [KM05b]. The error signal to calculate the predictor $\boldsymbol{p}$ can be expressed as

$$e_{\mathrm{BP}}[k] = \boldsymbol{s}^{T}[k](\boldsymbol{D}\boldsymbol{v} - \boldsymbol{D}\boldsymbol{V}_{-1}\boldsymbol{p}), \tag{4.7.9}$$

with $\boldsymbol{p}$ being the coefficient vector of the predictor, $\boldsymbol{D}$ being the convolution matrix built from the coefficients of the desired system vector $\boldsymbol{d}$, $\boldsymbol{v}$ the equalized system vector and $\boldsymbol{V}_{-1}$ is a convolution matrix made of $\boldsymbol{v}$ with an additional first row of zeros to take into account the delay of one sample

as depicted in Figure 4.41 [KM05b]. The calculation of the vector $\boldsymbol{p}$ that minimizes the target function $\mathrm{E}\left\{e_{\mathrm{BP}}^2[k]\right\}$ leads to [KM05b]

$$\boldsymbol{p} = \left(\boldsymbol{V}_{-1}^H \boldsymbol{D}^H \mathrm{E}\left\{\boldsymbol{s}[k]\boldsymbol{s}^H[k]\right\} \boldsymbol{D}\boldsymbol{V}_{-1}\right)^{-1} \boldsymbol{V}_{-1}^H \boldsymbol{D}^H \mathrm{E}\left\{\boldsymbol{s}[k]\boldsymbol{s}^H[k]\right\} \boldsymbol{D}\boldsymbol{v} \tag{4.7.10}$$

Under the assumption of a white and stationary excitation signal $\boldsymbol{s}[k]$ the correlation matrices $\mathrm{E}\left\{\boldsymbol{s}[k]\boldsymbol{s}^H[k]\right\}$ vanish. The desired system $\boldsymbol{d}$ causes *don't care* region(s) outside of its cut-off frequencies. A further desired system can be applied to $e_{\mathrm{BP}}[k]$ in order to generate an accordingly weighted signal at the loudspeaker [KM05b]. The spectral post processing approach can be applied as well to the shaping approach (4.7.6) as to the weigthed least-squares approach in (4.6.9) if spectral peaks occur due to the solely time-domain minimization rules.

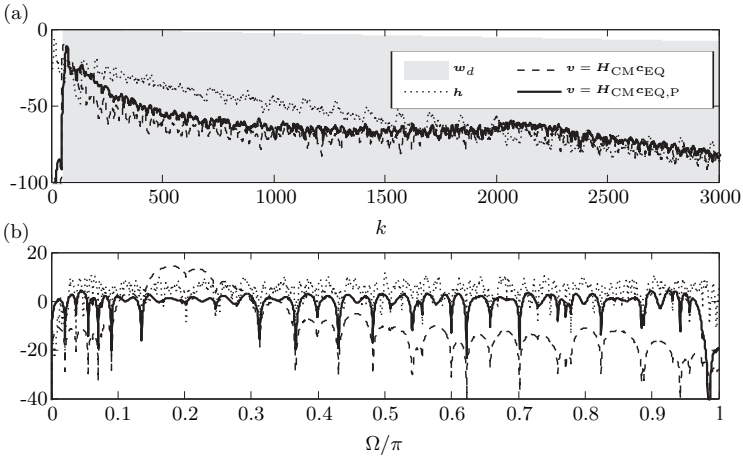

Figure 4.42: Impulse response shaping for an exponentially decreasing window $\boldsymbol{w}_d$ in dB. (a) Impulse responses in time-domain in dB and (b) transfer functions in frequency-domain in dB.

Figure 4.42 shows the described behaviour (a) in the time-domain and (b) in the frequency-domain. By shaping the original RIR (dotted line), a strong colouration is introduced (dashed line) in the frequency-domain. After post processing (solid line, $\boldsymbol{v} = \boldsymbol{H}_{\mathrm{CM}}\boldsymbol{c}_{\mathrm{EQ,P}}$) the colouration vanishes while time-domain behaviour is only slightly degraded ($\boldsymbol{v} = \boldsymbol{H}_{\mathrm{CM}}\boldsymbol{c}_{\mathrm{EQ}}$). Please note, that the IRs in Figure 4.42 (a) have been smoothed to better show the described effects.

An equalized system $\boldsymbol{v}$ after application of the impulse response shortening LRC filter with post processing (ISwPP) designed according to (4.7.6) is

shown in **Figure 4.43**, again for the same parameters and the same RIR $\boldsymbol{h}$ than for the LS equalizer (4.4.6) and the weighted LS equalizer (4.6.9). Results are similar to those depicted in Figure 4.40 for the weighted LS equalizer.

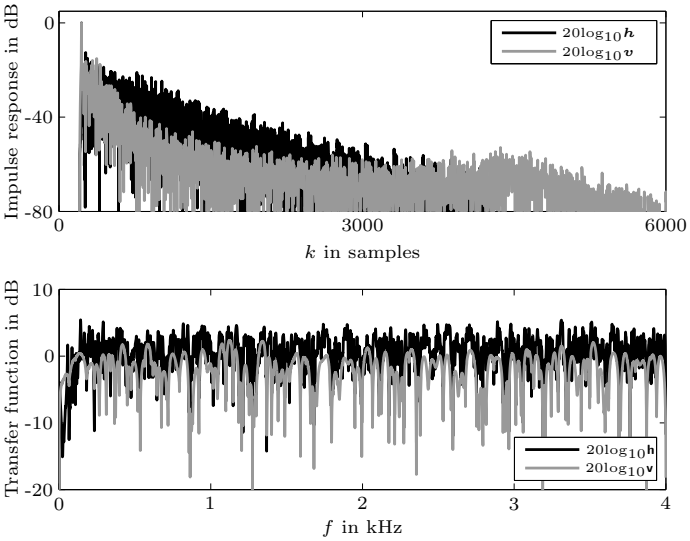

Figure 4.43: RIR $\boldsymbol{h}$ and equalized IR $\boldsymbol{v} = \boldsymbol{H}_{\mathrm{CM}}\, \boldsymbol{c}_{\mathrm{EQ}}^{\mathrm{ISwPP}}$ in time-domain in dB (upper panel) and the corresponding squared-magnitude spectra in dB (lower panel).

4.7.2 Joint Time-Frequency Processing

An approach that jointly shapes the IR of the equalized acoustic channel and minimizes spectral distortions is described in [MKM09b, MKM09a, MMK10]. By this, no explicit post processing is necessary. Additionally, the psychoacoustic property of masking is explicitly exploited in the filter design approach described in [MMK10]. Furthermore, this approach is based on a gradient update strategy which avoids computationally complex matrix operations that are needed for the other approaches, e.g. for the inverse of the matrix $\boldsymbol{H}_{\mathrm{CM}}$ in (4.4.6), the inverse of $\boldsymbol{W}\boldsymbol{H}_{\mathrm{CM}}$ in (4.6.9), both of size $(L_{\mathrm{EQ}}+L_h-1)\times L_{\mathrm{EQ}}$), or the solution of the generalized eigenvalue problem in (4.7.6).

The approach in [MMK10] is based on the p-norm optimization problem

$$\min_{\boldsymbol{c}_{\mathrm{EQ}}} f(\boldsymbol{c}_{\mathrm{EQ}}) = \min_{\boldsymbol{c}_{\mathrm{EQ}}} \log\left(\frac{f_u\left(\boldsymbol{c}_{\mathrm{EQ}}\right)}{f_d\left(\boldsymbol{c}_{\mathrm{EQ}}\right)}\right) \tag{4.7.11}$$

with

$$f_d\left(\boldsymbol{c}_{\mathrm{EQ}}\right) = \|\boldsymbol{v}_d\|_{p_d} = \left(\sum_{k=0}^{L_v-1} |v_{d,k}|^{p_d}\right)^{\frac{1}{p_d}} \tag{4.7.12}$$

and

$$f_u\left(\boldsymbol{c}_{\mathrm{EQ}}\right) = \|\boldsymbol{v}_u\|_{p_u} = \left(\sum_{k=0}^{L_v-1} |v_{u,k}|^{p_u}\right)^{\frac{1}{p_u}}. \tag{4.7.13}$$

The optimization of (4.7.11) leads to a minimization of the p-norm of the unwanted part of the equalized system $\boldsymbol{v}$ while simultaneously maximizing the p-norm of the desired part of the equalized system $\boldsymbol{v}$. By choosing $p_d = p_u = 2$, the solution of (4.7.11) reduces to the least-squares solution. The advantage of the method in (4.7.11) is that by selecting appropriately large values for p_d and p_u, the error is distributed evenly across the time coefficients in the unwanted part of the equalized IR $\boldsymbol{v}$ while favouring the production of one dominant tap in the desired part, which leads to an overall good shaping.
As visible in **Figure 4.44** the equalized system $\boldsymbol{v}$ directly follows the masking curve found in the human auditory system and a smooth decay can be observed for the whole length of the equalized system $\boldsymbol{v}$. For a more detailed discussion the interested reader is referred to the literature [MMK10, JMGM11, JMGM11, JMM12, MJM12].

4.8 Rating of the Sound Samples

In the following the four different LRC approaches (LS-EQ, WLS-EQ, ISwPP, ISwINO; cf. Table 4.3 on p. 74) will be compared by subjective listening tests. The listening tests have been described in Section 4.2.1. The sound samples were assessed regarding the four attributes *reverberant*, *coloured/distorted*, *distant*, and *overall quality* (cf. also Figure 4.5). The subjective ratings are shown in **Figure 4.45** by means of box-plots.
The sound samples are ordered according to their median value for the respective attribute. Consequently, the order is different for the different sub-figures.

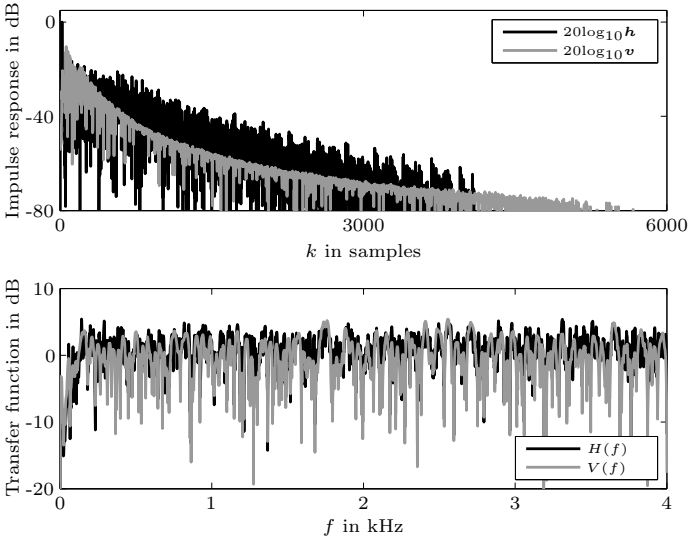

Figure 4.44: RIR $\boldsymbol{h}$ and equalized IR $\boldsymbol{v} = \boldsymbol{H}_{\mathrm{CM}}\, \boldsymbol{c}_{\mathrm{EQ}}^{\mathrm{ISwINO}}$ in time-domain in dB (upper panel) and the corresponding squared-magnitude spectra in dB (lower panel).

The subjective ratings were normally distributed which allowed for conduction of an analysis of variance (ANOVA). A two-way ANOVA revealed significant main effects of attribute type $\{F(3, 2112) = 18.8, p < 0.001\}$ and LRC approach $\{F(3, 2112) = 97.4, p < 0.001\}$. Post-hoc comparisons (Bonferroni tests with level of significance set at 5%) for the factor LRC approach showed statistical differences between all algorithms used with the highest quality for the ISwINO approach and the lowest for the LS approach. Generally, the shaping approaches (i.e. ISwPP and ISwINO) resulted in better rating scores than the least-squares approaches (i.e. LS and WLS).
Increasing the filter length of the LS approach does not necessarily improve the subjective results considerably due to the fact that despite a 'good equalization' perceptually relevant late echoes and pre-echoes are clearly perceived as disturbing by the listeners (see e.g. sound samples no. 9 ($L_{\mathrm{EQ}} = 8192$) and no. 13 ($L_{\mathrm{EQ}} = 1024$) both for an RIR with $\tau_{60} = 800$ ms).

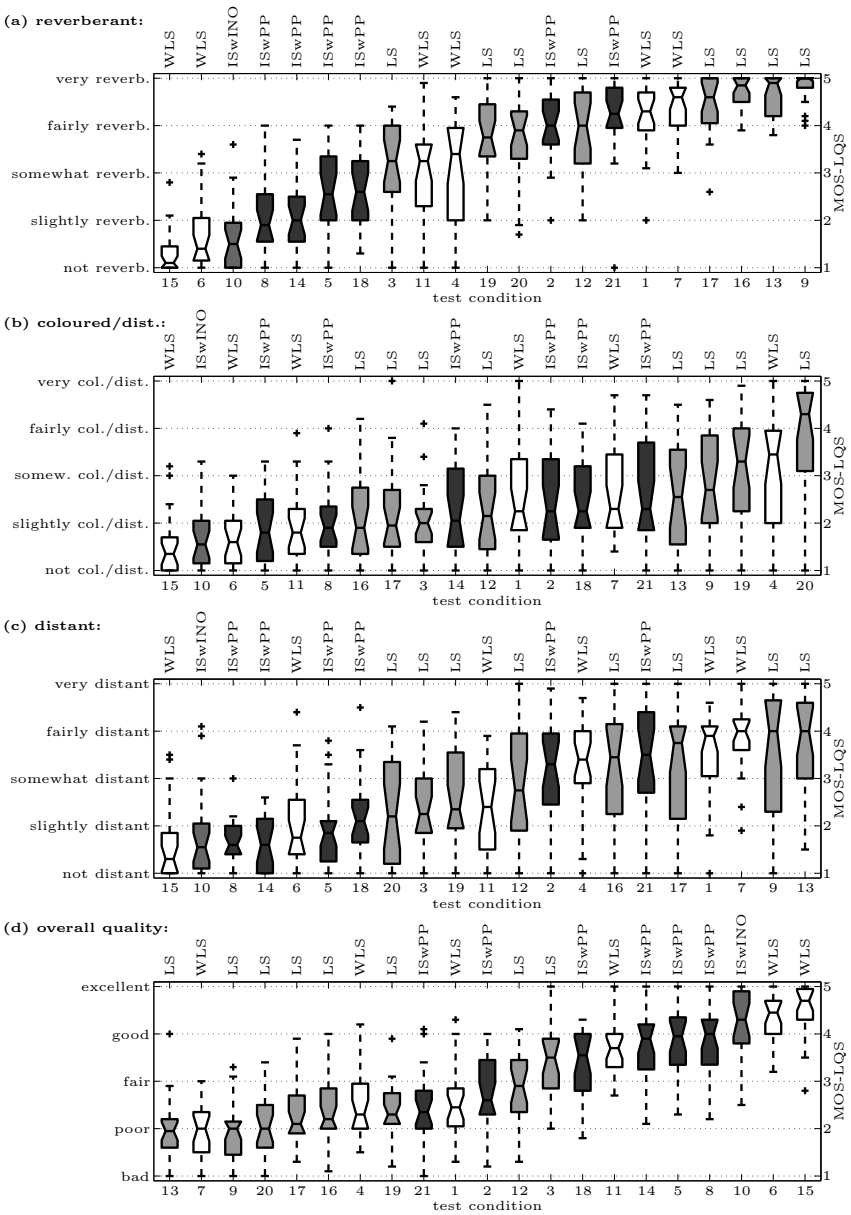

Figure 4.45: Subjective rating of sound samples for attribute (a) *reverberant*, (b) *coloured/distorted*, (c) *distant*, and (d) *overall quality*

The differences in the subjective scores between all used attributes were also statistically significant. Therefore, a separate one-way ANOVA was conducted for each attribute to test the quality of the different LRC approaches. For the attribute *reverberant*, the best ratings (indicated by the lowest rating scores) were obtained for the ISwINO algorithm with a mean value of 1.6. The ratings achieved by the ISwINO were significantly better than all remaining algorithms. The scores for the ISwPP and the WLS approach were 1.3 and 1.4 points higher than for the ISwINO approach, respectively, (meaning that signals processed by the ISwPP or WLS approach were assessed as being more reverberant than these processed by the ISwINO). No statistically significant differences in rating were found between the ISwPP and WLS approach (p=1.0). The lowest quality for the attribute *reverberant* was found for the LS approach with the mean rating score of 4.1. Exactly the same trends were observed for the attribute *overall quality*. Slightly different trends regarding the statistical dependencies of the LRC approaches were observed for the attribute *distant*. The best quality scores were again obtained for the shaping approaches, however with no significant differences between the ISwINO and ISwPP algorithm (p=0.164). Both least-squares approaches were again assessed worse than the shaping approaches and resulted in on average 0.8 points higher rating scores. A different trend between the attributes might be related to the fact that for the assessment of the attribute *distant* the differences between the four different approaches were smaller than for the attribute *reverberant* or *overall quality*. Although it seems from panels (a) and (c) of Figure 4.45 that the variance for the attribute *distant* is higher, results show similar standard errors for attributes *reverberation* and *distant*. However, for the attribute *reverberant* subjects more often decided for the maximum score of a MOS of 5 (very reverberant) which may be due to the fact that a clearer anchor for high reverberation was given in the training samples than for 'very distant'. The post-hoc comparisons for the attribute *coloured* revealed again the significantly highest quality for the ISwINO approach. No significant differences were found between the ISwPP, WLS and LS algorithm, however, from Figure 4.45 it can be seen that the LS approach usually performs worse that the other approaches which may be due to the fact that late echoes typical for the LS approaches sometimes sound like distortions.

Statistically significant differences were found regarding the room reverberation time τ_{60} ($\{F(1, 1919) = 460.659, p < 0.001\}$). In general, better LRC filter performance can be observed for shorter τ_{60}. The shorter room reverberation time $\tau_{60} \approx 500$ ms results in average in a better MOS of 1.1 than for $\tau_{60} \approx 950$ ms.

4.9 Chapter Summary

This chapter described the basic principles and problems of listening-room compensation (LRC) approaches for dereverberation of audio signals. A brief literature survey on different dereverberation techniques with a focus on methods for LRC has been given in Section 4.1.
Since for the task of quality assessment for reverberant and dereverberated signals no commonly accepted technical quality measures existed in literature, Section 4.2 focused on a evaluation of existing objective quality measures, mostly known from different fields, such as noise reduction, for the purpose of quality assessment for LRC algorithms. Hence, various objective measures have been compared to data obtained by subjective listening tests on LRC results. It was found, that quality measures exploiting information about the impulse response of the system under test perform well. If this information is not available (which is often the case for algorithms for dereverberation), objective quality measures should rely on advanced models of the human auditory system since such quality measures show highest correlation with the subjective ratings.
Different single-channel and multi-channel signal processing strategies for LRC aiming at either full equalization of the acoustic channel (cf. Sections 4.3 to 4.5) or shortening/shaping of the impulse response (cf. Section 4.6 to 4.7) have been introduced and discussed. For all LRC approaches, knowledge about the acoustic channel, i.e. the room impulse response, is necessary. Knowledge about the RIR is often assumed to be available, which, however, is usually not true in real-world systems. It is straightforward to use the system identification possibilities of acoustic echo cancellers discussed in Chapter 3 to obtain knowledge about the RIR. However, the RIR identification will always be imperfect in practical systems. Hence, problems arising due to this, i.e. the robustness against RIR perturbations caused by estimation errors as well as spatial mismatch between the RIR (from loudspeaker to listener) and the identified RIR (from loudspeaker to the reference microphone) have been analyzed. Methods to increase the robustness against estimation errors will be further topic of the following chapter.
Furthermore, a new type of gradient algorithm for LRC (cf. Section 4.5), i.e. the dFxLMS algorithm, has been proposed in Section 4.5.3 which converges quickly and is computationally efficient.
The discussed LRC algorithms were compared to each other in Section 4.8 regarding the archived signal quality and it turned out that shaping approaches in general archive a better quality than algorithms for aiming at full equalization of the acoustic channel.

Chapter 5

Combination of Systems for Acoustic Echo Cancellation and Listening-Room Compensation

The preceding Chapters 3 and 4 introduced signal processing strategies for acoustic echo cancellation and listening-room compensation, respectively. If these signal processing concepts are applied as subsystems in hands-free teleconferencing systems they mutually influence each other since both systems usually are time-varying. This chapter, thus, analyzes the mutual influences of acoustic echo cancellation and listening-room compensation.

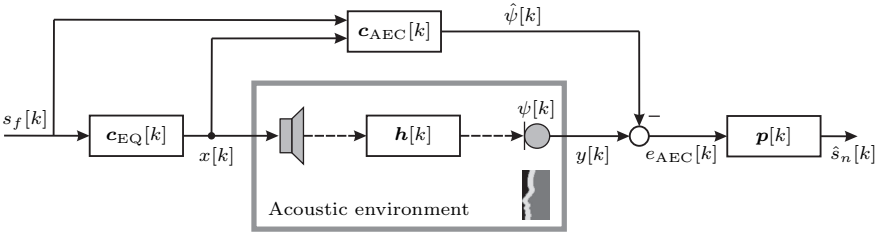

Figure 5.1: Block diagram of a hands-free system containing LRC filter $\boldsymbol{c}_{\mathrm{EQ}}[k]$, AEC filter $\boldsymbol{c}_{\mathrm{AEC}}[k]$ and post-filter $\boldsymbol{p}[k]$.

A block diagram of such a hands-free system is shown in **Figure 5.1**, containing an listening-room compensation filter $\boldsymbol{c}_{\mathrm{EQ}}[k]$ to reduce reverberation caused by the RIR $\boldsymbol{h}[k]$ and an acoustic echo cancellation filter $\boldsymbol{c}_{\mathrm{AEC}}[k]$ that can be used for identification of the RIR as well as identification of the concatenated system of LRC filter $\boldsymbol{c}_{\mathrm{EQ}}[k]$ and RIR $\boldsymbol{h}[k]$, i.e. $\boldsymbol{v}[k] = \boldsymbol{H}_{\mathrm{CM}}[k]\boldsymbol{c}_{\mathrm{EQ}}[k]$. The post-filter $\boldsymbol{p}[k]$ can be used to further reduce acoustic echoes (or even for system identification if it is properly designed).
The remainder of this chapter is organized as follows: Section 5.1 analyzes the influence of the AEC system performance on the LRC system, i.e. the influence of an imperfect estimate of the RIR on the LRC system. Furthermore, a method is proposed to increase the robustness of the LRC filter if information about the AEC system performance is available in terms of its system misalignment. Section 5.2 analyzes the influence of the LRC filter on the AEC system, i.e. the influence of the colouration of the AEC input signal introduced by the LRC filter in Section 5.2.1 and the identification of equalized impulse responses in Setion 5.2.2. A combined system consisting of two AEC filters and an LRC filter is discussed in Section 5.3, before Section 5.4 concludes the chapter.

5.1 System Identification by AEC filters

As already emphasized in Chapter 4, a reliable estimate of the acoustic channel $\boldsymbol{h}$ that has to be equalized is crucial for a satisfactory performance of the LRC filter. In the combined system depicted in **Figure 5.2**, two AEC filters $\boldsymbol{c}_{\mathrm{AEC},1}[k]$ and $\boldsymbol{c}_{\mathrm{AEC},2}[k]$ identify the time-varying acoustic channels. Please note, that the post-filter $\boldsymbol{p}[k]$ depicted in Figure 5.1 will first be neglected in the following.
An estimate of the acoustic channel $\boldsymbol{h}[k]$ can be obtained by an *inner AEC* $\boldsymbol{c}_{\mathrm{AEC},1}[k]$ lying in parallel to the RIR. The *inner AEC* $\boldsymbol{c}_{\mathrm{AEC},1}[k]$ provides an RIR estimate $\hat{\boldsymbol{h}}[k]$ of length L_{AEC} by minimizing the mean squared error signal $\mathrm{E}\left\{|e_{\mathrm{AEC},1}[k]|^2\right\}$. This estimate is inevitable for the design of the LRC filter and, thus, can not be avoided in practical systems.
Please note, that, in general, channel identification can also be obtained blindly [YHC05], i.e. without reference signal. However, if a reference signal is available, as in the given structure depicted in Figure 5.2, channel identification based on the reference channel usually leads to better results.
In addition to the *inner AEC* $\boldsymbol{c}_{\mathrm{AEC},1}[k]$, an *outer AEC* $\boldsymbol{c}_{\mathrm{AEC},2}[k]$ can further reduce the acoustic echo $\psi[k]$. To achieve this, the *outer AEC* has to identify the equalized acoustic channel. The *outer AEC* $\boldsymbol{c}_{\mathrm{AEC},2}[k]$ can either exploit the error signal of the *inner AEC* $e_{\mathrm{AEC},1}[k] = \psi[k] - \hat{\psi}_1[k]$ or work directly

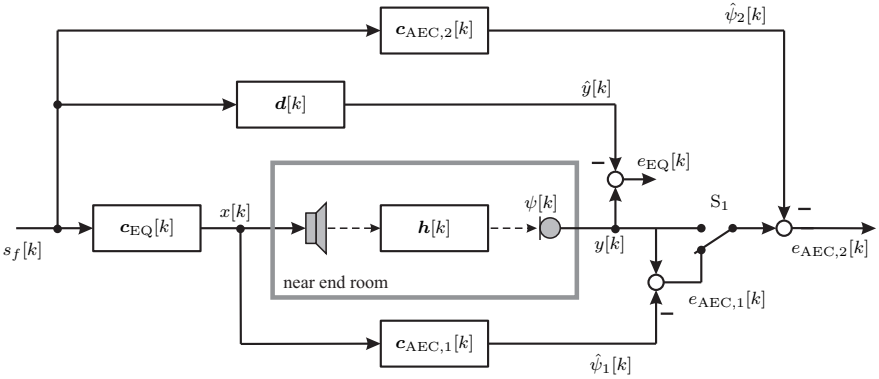

Figure 5.2: Block diagram of the combined system.

on the microphone signal $y[k]$. This can be chosen by switch S_1 in Figure 5.2. For the latter case, the *inner AEC* is solely used for system identification needed for the LRC filter, but it does not contribute to the compensation of the acoustic echo.

5.1.1 LRC Performance in Dependence of AEC System Distance of the Inner AEC

Since the LRC filter has to be placed in front of the acoustic environment, an estimate of the RIR $\boldsymbol{h}[k]$ is needed as it can be seen e.g. from (4.4.6). Even for the adaptive algorithms discussed in Section 4.5, an estimate of the RIR is needed for the update paths of the filtered-X LMS schemes. Since the LRC filter depends on a reliable estimate $\hat{\boldsymbol{h}}[k]$ of the RIR to be equalized, the current convergence state of the *inner AEC* is particularly important for the LRC filter [GKMK08d].
An AEC provides an estimate of the RIR which can be used by the equalizer. Another method to access the RIR would be, for example, ongoing measurement by means of maximum length sequences (MLS) [BA83], sweeps [MM01] or similar excitation signals. However, this would be a protracted process because averaging over time is necessary and it would result in an audible perturbation for the near-end listener. Furthermore, system identification by means of measurement as well as by using an AEC filter never leads to the true RIR $\boldsymbol{h}[k]$, but only to an erroneous version $\tilde{\boldsymbol{h}}[k]$. If an AEC is used for system identification, firstly only the first part of the RIR is identified due to the limited AEC filter length L_{AEC} while the tail of the RIR remains unidentified [BMS98b] as it is visualized in **Figure 5.3**. The

nature of the tail will of course have influence on the typical estimation error and, thus, have influence on the LRC filter error. Secondly, even the RIR estimate of the first L_{AEC} filter coefficients will be biased due to the influence of the unmodelled tail [BMS98b, Kal07, GKMK07] which further increases the estimation error $\tilde{\boldsymbol{h}}[k]$.

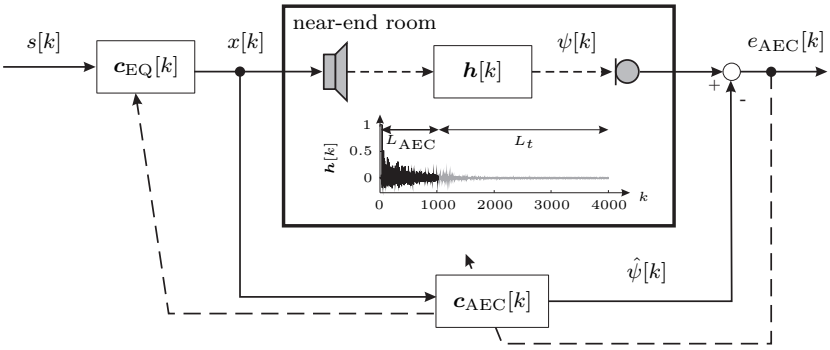

Figure 5.3: System for listening-room compensation with an acoustic echo canceller for system identification.

With the definition of the AEC system misalignment vector in (3.1.1) the RIR can be split up in two parts (cf. also Figure 4.18 in Section 4.4.2):

$$\boldsymbol{h}[k] = \boldsymbol{c}_{\mathrm{AEC}}[k] + \tilde{\boldsymbol{h}}[k] \tag{5.1.1}$$

with

$$\boldsymbol{h}[k] = [h_0[k]\,,\, h_1[k]\,,\, \ldots\,,\, h_{L_h-1}[k]]^T \tag{5.1.2}$$

$$\boldsymbol{c}_{\mathrm{AEC}}[k] = [c_{\mathrm{AEC},0}[k],\, c_{\mathrm{AEC},1}[k],\, \ldots,\, c_{\mathrm{AEC},L_{\mathrm{AEC}}-1}[k],\; 0,\, \ldots,\, 0]^T \tag{5.1.3}$$

In (5.1.1), $\boldsymbol{c}_{\mathrm{AEC}}[k]$ can be interpreted as an estimate $\hat{\boldsymbol{h}}[k]$ for the true RIR $\boldsymbol{h}[k]$, and $\tilde{\boldsymbol{h}}[k]$ is the estimation error (cf. also (3.1.1) to (3.1.3) on p. 34). **Figure 5.4** shows the system of LRC filter $\boldsymbol{c}_{\mathrm{EQ}}[k]$ and AEC $\boldsymbol{c}_{\mathrm{AEC}}[k]$ and the decomposition of the RIR $\boldsymbol{h}[k]$ into the part modelled by the AEC and the system misalignment $\tilde{\boldsymbol{h}}[k]$.
The AEC filter is updated by minimizing its error signal $\mathrm{E}\left\{e_{\mathrm{AEC}}^2[k]\right\}$ by a gradient algorithm (e.g. the PFBLMS [Shy92]). Thus, especially in periods of initial convergence or after RIR changes the system identification is insufficient and an LRC filter designed on its basis will introduce severe speech distortions.

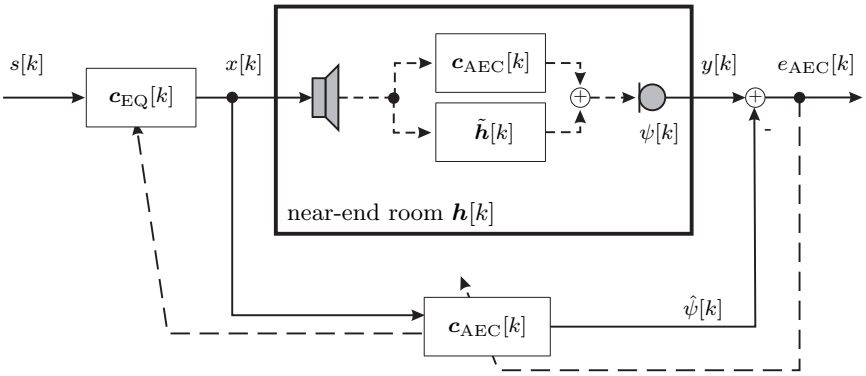

Figure 5.4: Combined system of LRC filter and acoustic echo canceller. The RIR can be split into a part modelled by the AEC $\boldsymbol{c}_{\mathrm{AEC}}[k]$ and the system misalignment $\tilde{\boldsymbol{h}}[k]$.

To visualize the performance of the LRC filter in dependence of the AEC filter's performance, the convergence curves of the dFxLMS algorithm (cf. Section 4.5.3) are shown in **Figure 5.5** in terms of signal-to-reverberation-ratio enhancement (SRRE) for different system distances $D_{\mathrm{dB}} = 10\log_{10}||\tilde{\boldsymbol{h}}||^2/||\boldsymbol{h}||^2$ of the *inner AEC*. Low values for D_{dB} are reached for a well converged AEC that delivers reliable RIR estimates. $D_{\mathrm{dB}} = 0$ dB indicates initial convergence.

The performance is shown exemplarily for the two impulse responses $\boldsymbol{h}_3$ (left panels) and $\boldsymbol{h}_4$ (right panels) which were already depicted in Figure 3.6 on page 40. It can be seen that the performance of the LRC filter increases with the convergence of the *inner AEC*. If the AEC performance is poor, the LRC performance is drastically reduced. If a certain amount of system identification is reached ($D_{\mathrm{dB}} < -2$ dB) an enhancement in terms of SSR can be obtained.

In panel (d) of Figure 5.5, two additional curves are depicted in thicker lines that show the influence of the overclocking factor O of the dFxLMS algorithm for a relative system distance of $D_{\mathrm{dB}} = -11$ dB. It can be seen that faster convergence of the LRC filter can be obtained by increasing O as already discussed in Section 4.5.4 for the dFxLMS algorithm. However, the maximum performance of the LRC filter after convergence is determined by the relative system distance of the *inner AEC* filter D_{dB}.

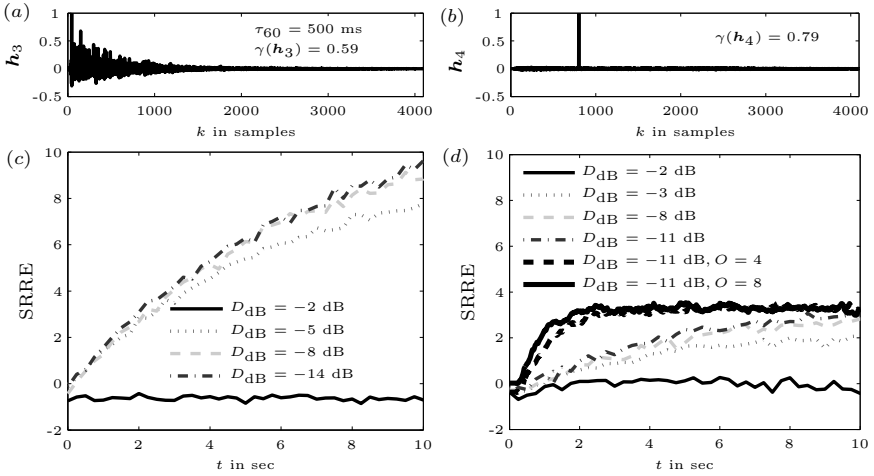

Figure 5.5: LRC performance in terms of SRRE obtained by the LRC filter for different system distances D_{dB} of the *inner AEC*. Input signal is white Gaussian noise. Panels (c) and (d) show results for the different RIRs depicted in panels (a) and (b), respectively.

5.1.2 Increasing LRC Robustness based on AEC Performance

Estimation errors $\tilde{\boldsymbol{h}}[k]$ are inevitable in practically relevant systems and, as shown before, may drastically decrease the LRC filter performance. LRC filters that have large energy may also lead to large distortions in case of estimation errors $\tilde{\boldsymbol{h}} \neq \boldsymbol{0}$. To decrease the energy of the inverse filters, a regularization term $\delta||\boldsymbol{c}_{\mathrm{EQ}}||^2$ can be incorporated in the cost function to be minimized [HDM07, KGD13b].

$$\underset{\boldsymbol{c}_{\mathrm{EQ}}}{\operatorname{argmin}}\, \boldsymbol{e}_{\mathrm{EQ},\delta} = ||\boldsymbol{H}_{\mathrm{CM}}\boldsymbol{c}_{\mathrm{EQ}} - \boldsymbol{d}||^2 + \delta||\boldsymbol{c}_{\mathrm{EQ}}||^2 \tag{5.1.4}$$

In (5.1.4), δ is a regularization parameter controlling the ratio between minimization of the error energy $||\boldsymbol{H}_{\mathrm{CM}}\boldsymbol{c}_{\mathrm{EQ}} - \boldsymbol{d}||^2$ and the energy of the LRC filter $||\boldsymbol{c}_{\mathrm{EQ}}||^2$. Setting the gradient of (5.1.4) to 0, i.e.

$$\frac{\partial \boldsymbol{e}_{\mathrm{EQ},\delta}}{\partial \boldsymbol{c}_{\mathrm{EQ}}} = 2\boldsymbol{H}_{\mathrm{CM}}^T\boldsymbol{H}_{\mathrm{CM}}\boldsymbol{c}_{\mathrm{EQ}} - 2\boldsymbol{H}_{\mathrm{CM}}^T\boldsymbol{d} + 2\delta\boldsymbol{c}_{\mathrm{EQ}} = 0, \tag{5.1.5}$$

leads to the regularized LS LRC filter [HDM07, KGD13b], also known as the MMSE solution [Kam08]

$$\boldsymbol{c}_{\mathrm{EQ}} = \left(\hat{\boldsymbol{H}}_{\mathrm{CM}}^{T}\hat{\boldsymbol{H}}_{\mathrm{CM}} + \delta\boldsymbol{I}\right)^{-1}\hat{\boldsymbol{H}}_{\mathrm{CM}}^{T}\boldsymbol{d}. \tag{5.1.6}$$

Thus, in the following, the estimation error will be incorporated in the LRC filter design to increase the filter robustness.
For a known misalignment vector $\tilde{\boldsymbol{h}}[k]$ at a fixed time instance k the LRC filter's error signal is given by

$$e_{\mathrm{EQ}}[k] = \boldsymbol{s}^{T}[k](\hat{\boldsymbol{H}}_{\mathrm{CM}} + \tilde{\boldsymbol{H}}_{\mathrm{CM}})\boldsymbol{c}_{\mathrm{EQ}} - \boldsymbol{s}^{T}[k]\boldsymbol{d}, \tag{5.1.7}$$

with the convolution matrices of the RIR estimation error

$$\tilde{\boldsymbol{H}}_{\mathrm{CM}} = \mathrm{convmtx}\{\tilde{\boldsymbol{h}}[k], L_{\mathrm{EQ}}\} \tag{5.1.8}$$

and of the RIR estimate

$$\hat{\boldsymbol{H}}_{\mathrm{CM}} = \mathrm{convmtx}\{\boldsymbol{c}_{\mathrm{AEC}}[k], L_{\mathrm{EQ}}\}. \tag{5.1.9}$$

Minimization of $\mathrm{E}\left\{e_{\mathrm{EQ}}^{2}[k]\right\}$ according to (5.1.7) leads to

$$\begin{aligned}\boldsymbol{c}_{\mathrm{EQ}} = &\left(\hat{\boldsymbol{H}}_{\mathrm{CM}}^{T}\hat{\boldsymbol{H}}_{\mathrm{CM}} + \tilde{\boldsymbol{H}}_{\mathrm{CM}}^{T}\tilde{\boldsymbol{H}}_{\mathrm{CM}} + \hat{\boldsymbol{H}}_{\mathrm{CM}}^{T}\tilde{\boldsymbol{H}}_{\mathrm{CM}} + \tilde{\boldsymbol{H}}_{\mathrm{CM}}^{T}\hat{\boldsymbol{H}}_{\mathrm{CM}}\right)^{-1}\\ &\cdot\left(\hat{\boldsymbol{H}}_{\mathrm{CM}} + \tilde{\boldsymbol{H}}_{\mathrm{CM}}\right)^{T}\boldsymbol{d}.\end{aligned} \tag{5.1.10}$$

With the simplifying assumption of $\tilde{\boldsymbol{h}}$ and $\hat{\boldsymbol{h}}$ being uncorrelated, i.e. $\mathrm{E}\{\tilde{\boldsymbol{H}}_{\mathrm{CM}}^{T}\hat{\boldsymbol{H}}_{\mathrm{CM}}\} = \boldsymbol{0}$, and a zero-mean system misalignment vector $\mathrm{E}\left\{\tilde{\boldsymbol{h}}\right\} = \boldsymbol{0}$ the LRC filter coefficients can be recalculated to a reduced solution, i.e.

$$\boldsymbol{c}_{\mathrm{EQ}} = \left(\hat{\boldsymbol{H}}_{\mathrm{CM}}^{T}\hat{\boldsymbol{H}}_{\mathrm{CM}} + \tilde{\boldsymbol{H}}_{\mathrm{CM}}^{T}\tilde{\boldsymbol{H}}_{\mathrm{CM}}\right)^{-1}\hat{\boldsymbol{H}}_{\mathrm{CM}}^{T}\boldsymbol{d}. \tag{5.1.11}$$

The system misalignment vector $\tilde{\boldsymbol{h}}[k]$ is unknown for real-world environments and difficult to estimate on its full length. However different algorithms exist for estimating the norm of the system misalignment vector $\mathrm{E}\left\{||\tilde{\boldsymbol{h}}[k]||^{2}\right\}$, often also called coupling factor [MPS00], because it describes the coupling between loudspeaker and AEC error signal $e_{\mathrm{AEC}}[k]$ if no disturbances are present. A prominent method to estimate the norm of $\tilde{\boldsymbol{h}}[k]$ is to introduce an artificial delay of $L_{\Delta} \approx 20$ to 40 samples directly after

the microphone and to extrapolate the system misalignment of the AEC filter at those coefficients to the full length of the filter. For a more detailed description see [MPS00, AGQ97]. In this contribution the estimate of the norm of the system misalignment is based on the ratio of the power of the AEC error signal $e^2_{\mathrm{AEC}}[k]$ and the power of the loudspeaker signal $x^2[k]$ which is updated in periods of an inactive near speaker ($s_n[k] = 0$).

$$\mathrm{E}\left\{||\tilde{\boldsymbol{h}}[k]||^2\right\} = \alpha_g \mathrm{E}\left\{||\tilde{\boldsymbol{h}}[k-1]||^2\right\} + (1-\alpha_g)\frac{\overline{e^2_{\mathrm{AEC}}}[k]}{\overline{x^2}[k]} \tag{5.1.12}$$

with the smoothed powers

$$\overline{e^2_{\mathrm{AEC}}}[k] = \alpha_e \overline{e^2_{\mathrm{AEC}}}[k-1] + (1-\alpha_e) e^2_{\mathrm{AEC}}[k] \tag{5.1.13}$$

$$\overline{x^2}[k] = \alpha_x \overline{x^2}[k-1] + (1-\alpha_x) x^2[k] \tag{5.1.14}$$

A voice activity detector (VAD) is implemented based on the normalized cross correlation approach by [KMK05, RGH+11]. This VAD is needed anyway by the AEC to stop the adaptation in presence of an active near speaker and thus does not lead to an increased computational load.
With the assumption of a white system misalignment, (5.1.11) can be approximated by

$$\boldsymbol{c}_{\mathrm{EQ}} = \left(\hat{\boldsymbol{H}}^T_{\mathrm{CM}}\hat{\boldsymbol{H}}_{\mathrm{CM}} + \mathrm{E}\left\{||\tilde{\boldsymbol{h}}||^2\right\}\boldsymbol{I}\right)^{-1}\hat{\boldsymbol{H}}^T_{\mathrm{CM}}\boldsymbol{d}, \tag{5.1.15}$$

only depending on accessible variables, such as the convolution matrix $\hat{\boldsymbol{H}}_{\mathrm{CM}}$ built from the AEC coefficients, the norm of the AEC misalignment vector $\mathrm{E}\left\{||\tilde{\boldsymbol{h}}||^2\right\}$ given by (5.1.12), and the desired response $\boldsymbol{d}$.
In the following section, the EQ design rules given by (5.1.10), (5.1.11), and (5.1.15) are evaluated by means of their performance to reduce reverberation introduced by the RIR.

Simulation Results

The filter length of the AEC for the following simulations is $L_{\mathrm{AEC}} = 2048$ and the LRC filter length is $L_{\mathrm{EQ}} = 1024$, respectively. For the AEC filter update, a PFBLMS algorithm [Shy92] is used. The RIR was simulated [AB79] having a length of $L_h = 4096$ for different room reverberation times of $\tau_{60} = \{200, 400, 900\}$ ms. The desired system vector $\boldsymbol{d}$ is chosen to be a 40^{th} order FIR highpass with band limit at 200 Hz at a sampling frequency of $f_s = 8000$ Hz. The delay introduced by the equalizer is $k_0 = 170$ samples. **Figure 5.6** compares the LRC filters according to Eqns. (4.4.13), (5.1.10), and (5.1.15) by means of the SRRE for the different room reverberation

times τ_{60} ranging from 200 ms (upper panels) to 900 ms (lower panels) depending on the AEC convergence state expressed by means of its normalized system misalignment D_{dB}. To avoid the non-uniqueness problem for the RIR identification [BMS98b] we restrict the number of loudspeakers to $P = 1$. Simulation results for $Q = 1$ (single-channel case) and $Q = 4$ microphones are shown in Figure 5.6 in the left and right part, respectively. For the multi-channel case (sub-plots (d)-(f)) the SRRE is averaged over all channels. The microphones were arranged in a line array with an inter-microphone distance of 5 cm.

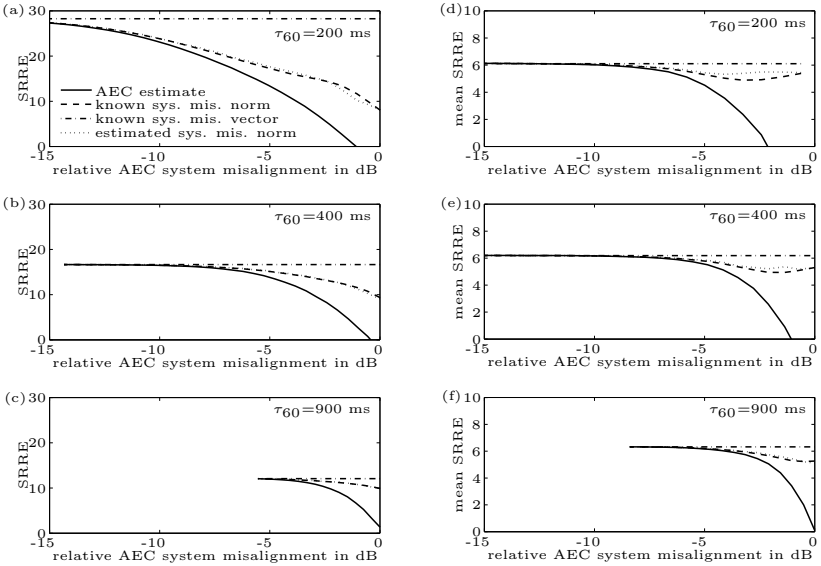

Figure 5.6: Comparison of EQ designs according to Eqns. (4.4.13), (5.1.10), and (5.1.15) by means of the SSRE in dependence of the AEC system misalignment in dB for different room reverberation times τ_{60} = 200 ms to 900 ms.

The solid lines in Figure 5.6 show the filter performance for the LRC filter design based on the RIR estimate delivered by the AEC only, which means that the least-squares LRC filter (4.4.13) is applied by taking the AEC filter coefficients as a direct estimate for the RIR $\tilde{\boldsymbol{h}}$. It can be seen, that a direct and straightforward implementation of (4.4.13) by applying an AEC for system identification may not lead to sufficient improvement or even to a deterioration of the SRR for a high system misalignment D_{dB} which will be the case most of the time for high room reverberation times.

The horizontal dash-dotted lines indicate the performance of an EQ designed according to (5.1.10) with *a-priori* knowledge of the full system misalignment vector $\tilde{\boldsymbol{h}}[k]$. Thus, they can be interpreted as upper limits for the improvement that can be achieved by the LRC filter for given RIRs and a given LRC filter order. It should be mentioned that the maximum achievable SRR improvement depends on the room reverberation time and the absolute positions of sources and microphones. Numerous positions have been simulated and Figure 5.6 shows some representative results. It can be seen that the maximum possible SRR enhancement decreases for higher room reverberation times due to the higher energy in the reverberant tail of the RIRs and also for the multi-channel case (right panels) compared to the single-channel case (left panels). The latter is due to the fact that the equalization is done by one filter for all four RIRs and, thus, a mean equalization is achieved [EN89]. This leads to a loss of SRR enhancement but to an increased spatial robustness which is very important in a hands-free scenario, since the user will not be located exactly at the positions of the microphones (cf. Section 4.4.2).

The dashed curves show the LRC filter performance if only the norm of the system misalignment is known *a-priori* and the dotted line if it is estimated by (5.1.12). It can seen that the use of $||\tilde{\boldsymbol{h}}[k]||^2$ leads to significant improvements compared to the use of the RIR estimates given by the AEC only and that it is a good approximation of the use of the misalignment vector especially for higher room reverberation times and for a multi-channel scenario. The use of (5.1.12) as an estimate for $\mathrm{E}\{||\tilde{\boldsymbol{h}}[k]||^2\}$ for the proposed EQ design (5.1.15) leads to good approximations.

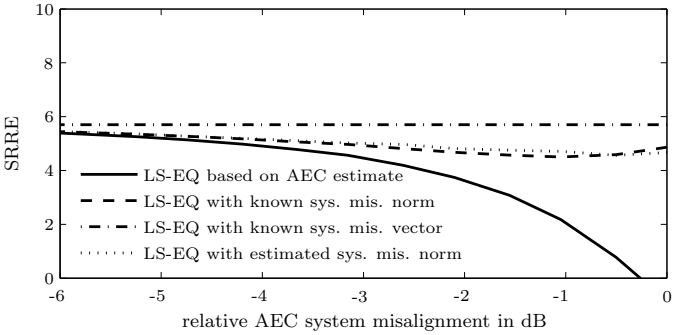

Figure 5.7: Signal-to-reverberation-ratio enhancement for spatial displacement (τ_{60} = 900 ms). Distance between user of the system and of mic array: 20 cm.

Since the user of a hands-free system will not be located directly at the position of the microphones an example for an EQ design with spatial mismatch is shown in **Figure 5.7**.
The distance between the user and the microphone array, for which the RIR identification is done, is 20 cm. As it can be seen from Figure 5.7, a spatial displacement of course degrades the performance of the system compared to Figure 5.6 (f) but an SRR gain of about 5 dB is still possible, which is quite a good result considering the findings in [RWK00].

5.1.3 Post Filter for System Identification

As described in Section 3.3, AEC filters can be supported by AES post-filters, i.e. short-term spectral suppression. **Figure 5.8** shows a system which combines an LRC filter $\mathbf{c}_{\mathrm{EQ}}[\ell]$ in block-time domain with an AEC filter $\mathbf{c}_{\mathrm{AEC}}[\ell]$ and an AES post-filter $\mathbf{p}[\ell]$. The AES filter has to rely on an estimate of the residual echo PSD $\mathbf{\Phi}_{\xi\,\xi}[\ell]$. One way to obtain the residual echo PSD is system identification by means of the residual echo estimation filter $\mathbf{c}_{\mathrm{REEF}}[\ell]$ depicted in Figure 5.8 [GKMK06b, XAG12] (cf. also Section 3.3). The adaptation speed of the identification filters in Figure 5.8 is controlled by a double-talk detection (DTD) algorithm (to which also the delay z^{-N} in Figure 5.8 belongs). However, this block is not further described in the following, the interested reader may refer e.g. to [MPS00].

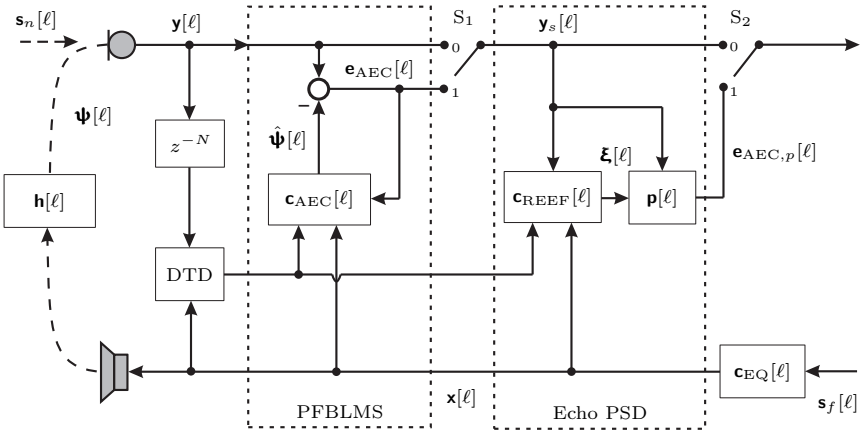

Figure 5.8: Block diagram of LRC filter $\mathbf{c}_{\mathrm{EQ}}[\ell]$, AEC filter $\mathbf{c}_{\mathrm{AEC}}[\ell]$ and AES filter $\mathbf{p}[\ell]$ including residual echo estimation filter $\mathbf{c}_{\mathrm{REEF}}[\ell]$.

In the following, the system identification performance will be analyzed in combination with the LRC filter $\mathbf{c}_{\mathrm{EQ}}[\ell]$. **Figure 5.9** shows simulation re-

sults for filter lengths of $L_{\mathrm{EQ}} = 1024$ for the LRC filter and $L_{\mathrm{AEC}} = 1024$ for AEC and residual echo estimation filter, respectively, at a sampling rate of 8 kHz. The LRC filter is updated by a dFxLMS algorithm (cf. Section 4.5.3) with overclocking factor of $O = 4$.

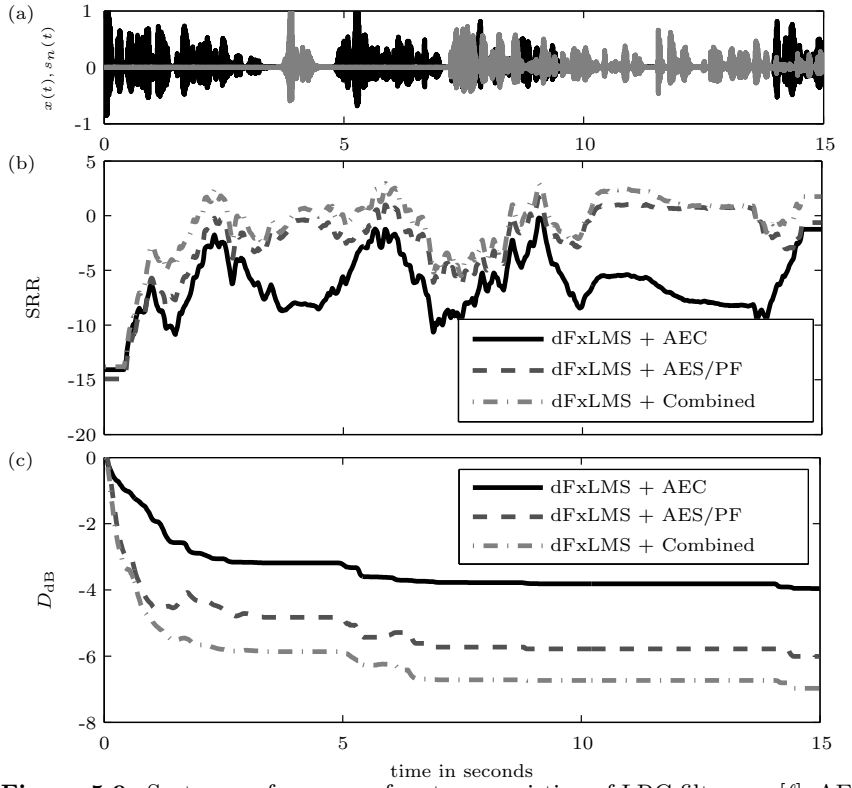

Figure 5.9: System performance of system consisting of LRC filter $\mathbf{c}_{\mathrm{EQ}}[\ell]$, AEC filter $\mathbf{c}_{\mathrm{AEC}}[\ell]$ and AES filter $\mathbf{p}[\ell]$ including residual echo estimation filter $\mathbf{c}_{\mathrm{REEF}}[\ell]$. (*a*) echo signal $x(t)$, i.e. the far-end speaker's signal (black), and near-end speaker's signal $s_n(t)$ (grey) including double-talk, (*b*) SRR performance of the LRC system, (*c*) system distance D_{dB} of the AEC/AES systems.

Figure 5.9 shows the combined system's performance for the possible system identification strategies if (i) only AEC is used for system identification (black solid lines in panels (b) and (c)), (ii) only the residual echo estimation filter is used for system identification (dashed lines), and (iii) both systems are used for system identification (dashed-dotted lines). Signal-to-

reverberation-ratio (SRR) and system distance D_{dB} are shown in panels (b) and (c), respectively, as measures for the performance of the LRC and AEC filters.
As shown in Figure 5.9 (c), the residual echo estimation filter $\mathbf{c}_{\mathrm{REEF}}[\ell]$ leads to better system identification results than $\mathbf{c}_{\mathrm{AEC}}[\ell]$ despite the fact that both filters have the same length L_{AEC}. This is mostly due to the larger step-size for $\mathbf{c}_{\mathrm{REEF}}[\ell]$ that is necessary to track changes of the inner filters. Highest performance is achieved if both filters are activated. Same tendency can also be observed for the LRC performance in panel (b). Although AEC filter $\mathbf{c}_{\mathrm{AEC}}[\ell]$ and residual echo estimation filter $\mathbf{c}_{\mathrm{REEF}}[\ell]$ lie in parallel as it can be seen in Figure 5.8, have the same filter length L_{AEC} and depend on the correlation of the same input signal $\mathbf{x}[\ell]$, their system identification, by this the influence on the LRC filter and in turn on their own input signal's correlation may be different.

5.2 AEC Performance in Dependence of LRC System

While the influence of the AEC filter on the LRC system was analyzed in the previous Section 5.1, this section will now focus on the influence of the LRC system on the AEC. Since the discussed system in Figure 5.2 uses two AEC filters, i.e. the *outer AEC* and the *inner AEC*, the following Section 5.2.1 will analyze the influence of the LRC filter on the *inner AEC* before the performance of the *outer AEC* will be analyzed in Section 5.2.2 which has to identify the system of LRC filter and RIR.

5.2.1 Performance of Inner AEC in Dependence of Equalizer

It is known that gradient algorithms for AEC perform better for uncorrelated input signals such as Gaussian white noise [Hay02]. An LRC filter which is located in front of the AEC input signal $x(t)$ will change the signal correlation, i.e. for a white input signal it will introduce correlation. **Figure 5.10** shows the performance of the *inner AEC* in terms of relative system distance $D_{\mathrm{dB}}(t)$ for different LRC filter types in panel (c) for white Gaussian input $x(t)$ and in panel (d) for the speech signal $x(t)$ depicted in panel (b), respectively. The corresponding RIR $\boldsymbol{h}$ is depicted in panel (a). The AEC filter length is $L_{\mathrm{AEC}} = 1024$, the LRC filter length $L_{\mathrm{EQ}} = 1024$, and the length of the RIR is $L_h = 4096$, respectively.
The results in panel (c) for a white Gaussian input signal $x(t)$ show that best

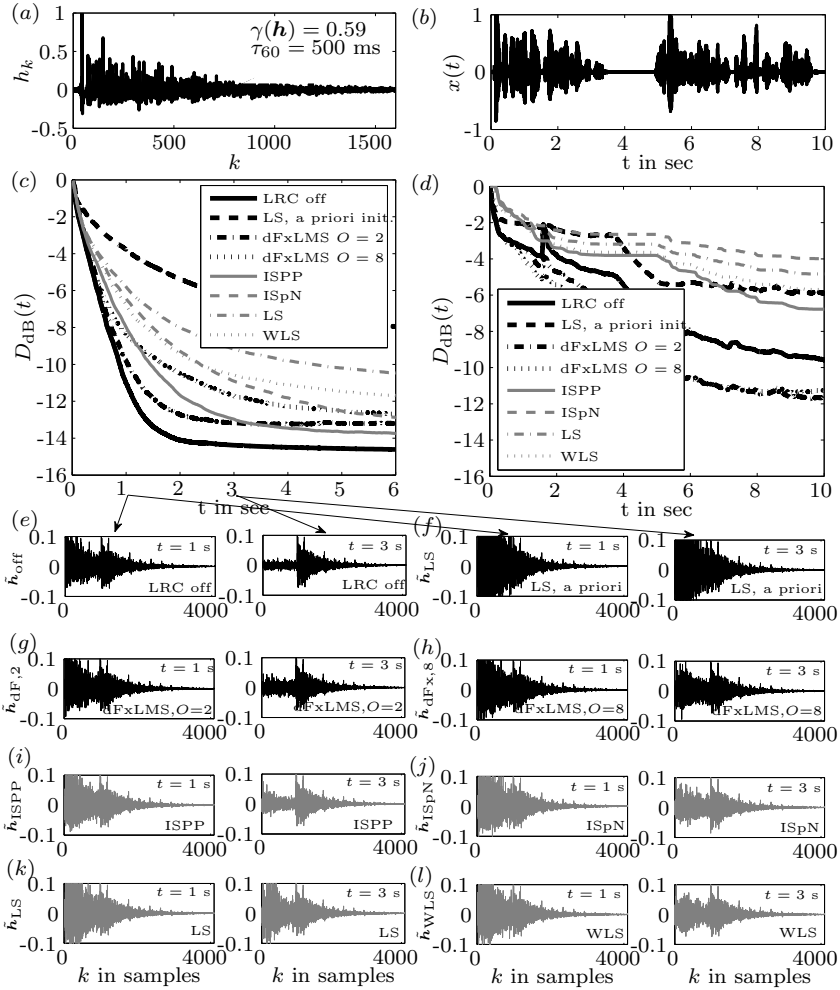

Figure 5.10: Performance of *inner AEC* in dependence of different update strategies of the preceding LRC filter. RIR to be identified is shown in panel (a), panel (b) shows speech excitation signal to generate results in panel (d). System performance in terms of the AEC's relative system distance is shown in panels (c) and (d) for white and speech excitation $x(t)$, respectively. Panels (e) to (l) show the system distance vectors $\bar{\boldsymbol{h}}$ for different LRC approaches after 1 sec and 3 sec convergence of the *inner AEC*, respectively.

performance of the AEC filter is achieved if the LRC filter is switched off while worst behaviour is archived for the least-squares (LS) and weighted least-squares (WLS) LRC approaches since these introduce most correlation. Please note that in Figure 5.10 the least-squares approach appears two times (curves labels by 'LS' and 'LS, a priori init.'). The latter is generated by providing perfect knowledge about the true RIR $\boldsymbol{h}$ immediately to the LRC filter, i.e. no convergence for the LRC filter is needed and thus the LRC filter is perfectly converged from the beginning. All other curves are generated by the more realistic assumption that both systems update their coefficients in parallel, i.e. mutually influence each other. It can be seen that perfect initialization of the LRC filter by providing the true RIR $\boldsymbol{h}$ immediately introduces correlation leading to worst performance of the *inner AEC* directly at the beginning. If both systems are updated in parallel, correlation is more slowly introduced by the LRC filter, allowing the *inner AEC* filter to initially converge faster. Furthermore, it can be seen from Figure 5.10 that the LRC filters based on the RIR shaping approaches (ISPP, ISpN, WLS) perform better than the LS approach. The AEC performance, if the dFxLMS algorithm described in Section 4.5.3 is used, is closer to the performance without LRC filter, especially for small overclocking factor O. For speech input (cf. panel (d)), convergence of the *inner AEC* while using the dFxLMS algorithm to update the LRC filter is even faster than without LRC filter while the other LRC approaches show similar performance. It seems that the dFxLMS gradient algorithm has a positive effect on the signal's statistical properties here.
Panels (e) to (l) in Figure 5.10 show the respective system misalignment vectors $\tilde{\boldsymbol{h}}$ of the *inner AEC* at two distinct time instances ($t = 1$ sec and $t = 3$ sec) for the different LRC approaches. For the case that the LRC filter is switched off (panels (e)), it can e.g. be seen that obviously after 3 sec the *inner AEC* filter shows better convergence than after 1 sec and after 3 sec, the unmodelled tail of the RIR is clearly visible for the system distance vector $\tilde{\boldsymbol{h}}_{\text{off}}$. However, for the other LRC approaches, the system distance vectors, a considerable system identification error remains also for the first L_{AEC} coefficients that should be identified by the *inner AEC* filter.

5.2.2 Performance of Proportionate Update Schemes for Outer AEC

In addition to the *inner AEC* $\boldsymbol{c}_{\text{AEC},1}[k]$, an *outer AEC* $\boldsymbol{c}_{\text{AEC},2}[k]$ can further reduce the acoustic echo $\psi[k]$ as depicted in Figure 5.2.
It could be assumed that the echo reduction task for the *outer AEC* would be *easier* if the *inner AEC* already achieved a certain echo reduction. How-

ever, the *outer AEC* has to track changes caused by adaptation of *inner AEC* and LRC filter. Therefore, a sufficiently fast adaptation is needed for the *outer AEC* especially since also the inner filters need to adapt as fast as possible, e.g. since an RIR estimate is needed quickly for the LRC filter. To achieve a higher amount of echo reduction than the *inner AEC* alone, the filter length of the *outer AEC* should be greater than that of the *inner AEC* which unfortunately leads to a decreased convergence speed [BDH+99]. Here, proportionate update schemes can be a solution.
Please note, that depending on S_1 in Figure 5.2 the system to be identified by the *outer AEC* is either the equalized system

$$\boldsymbol{v}[k] = \boldsymbol{H}_{\mathrm{CM}}[k]\boldsymbol{c}_{\mathrm{EQ}}[k] \tag{5.2.1}$$

(S_1 in Figure 5.2 in upper position) or the concatenated system of LRC filter and system distance of the *inner AEC*,

$$\boldsymbol{v}'[k] = \mathrm{convmtx}\left\{\boldsymbol{h}[k] - \left[\boldsymbol{c}_{\mathrm{AEC},1}^{T}[k], \boldsymbol{0}^{T}\right]^{T}, L_{EQ}\right\}\boldsymbol{c}_{\mathrm{EQ}}[k] \tag{5.2.2}$$

$$= \tilde{\boldsymbol{H}}[k]\boldsymbol{c}_{\mathrm{EQ}}[k] \tag{5.2.3}$$

(S_1 in Figure 5.2 in lower position). The *outer AEC* may observe a sparse IR for switch S_1 in upper position and if the LRC filter performs well, i.e. a delayed delta function could be achieved for the case of perfect least-squares equalization. In this case, proportionate update schemes [Dut00, BHCN06] as introduced and analyzed in Section 3.2.2 can be applied for the *outer AEC*. However, for the case that the system $\boldsymbol{v}'[k]$ as defined in (5.2.2) has to be identified, the equalized system may not be sparse since the equalized system $\boldsymbol{v}'[k]$ results from a convolution of the LRC filter $\boldsymbol{c}_{\mathrm{EQ}}[k]$ with the system misalignment vector $\tilde{\boldsymbol{h}}[k]$ and not with the RIR $\boldsymbol{h}[k]$ the LRC filter has been designed for. As it can be seen panels (e)-(l) in Figure 5.10, the system misalignment vector $\tilde{\boldsymbol{h}}[k]$ may look very different from a usual RIR. This is further illustrated in **Figure 5.11**.
The upper part of Figure 5.11 shows a RIR $\boldsymbol{h}$ in panel (a) and the respective equalized IR $\boldsymbol{v}$ in panel (b) after processing by an LS LRC filter of length $L_{\mathrm{EQ}} = 1024$. The lower part of Figure 5.11 shows a system distance vector $\tilde{\boldsymbol{h}}$ which results from $\boldsymbol{h}$ after identification of the first 256 coefficients by an AEC filter in panel (c) and the resulting IR $\boldsymbol{v}'$ if the LRC filter is applied to $\tilde{\boldsymbol{h}}$ in panel (d), respectively. While $\boldsymbol{v}$ can be considered as sparse, $\boldsymbol{v}'$ is more dispersive.
It was shown in Figure 3.13 on page 50 that identification of a perfectly equalized IR can be done efficiently by proportionate update schemes.
However, if an *outer AEC* is concatenated to the *inner AEC* to increase the echo reduction, i.e. switch S_1 in Figure 5.2 is in lower position, the

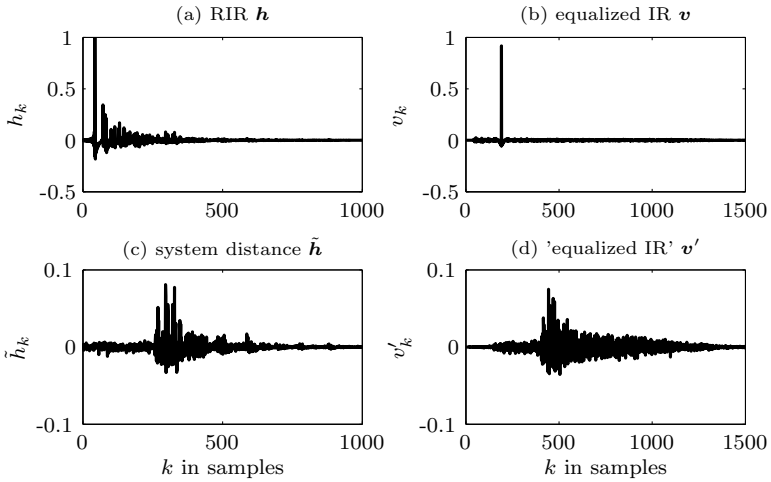

Figure 5.11: Illustration of equalized systems $\boldsymbol{v}$ and $\boldsymbol{v}'$ according to (5.2.1) and (5.2.2), respectively.

outer AEC has to identify $\boldsymbol{v}'[k]$ given in (5.2.2). The LRC filter is designed to equalize $\boldsymbol{h}[k]$, thus the resulting system $\boldsymbol{v}'[k]$ will not be as sparse as assumed, which was visualized in Figure 5.11.

Figures 5.12 and **5.13**, thus, show the performance of NLMS, PNLMS and IPNLMS for two systems $\boldsymbol{v}'[k]$ exemplarily, one obtained using white noise input after sufficient convergence of the *inner AEC* (Figure 5.12), i.e. the first 1024 samples have been identified by an *inner AEC*, and one for speech input after only partly convergence of the *inner AEC* (Figure 5.13). Although the system to be identified is not really sparse especially in Figure 5.13, the IPNLMS is still a good choice and can be used as an update scheme for the *outer AEC*.

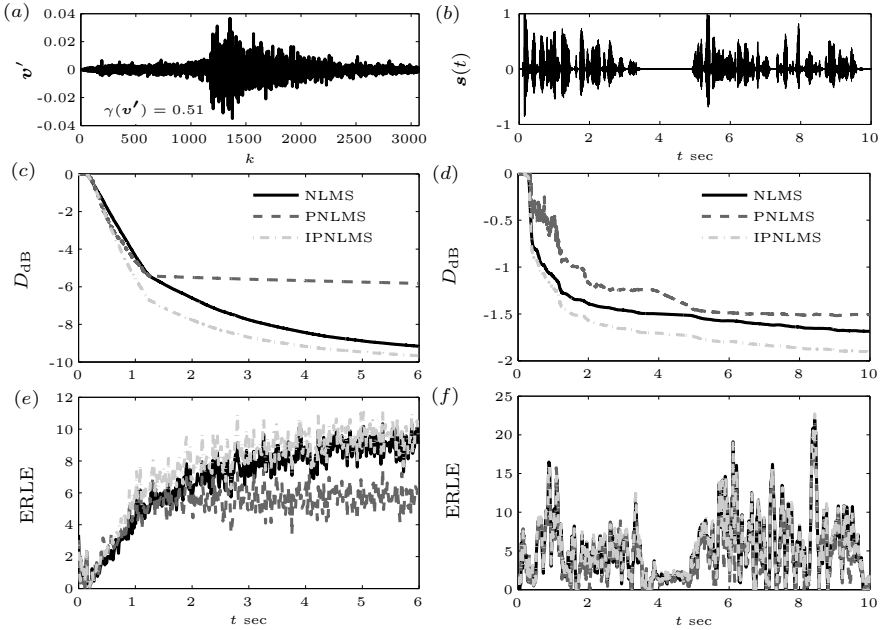

Figure 5.12: Comparison of NLMS, PNLMS and IPNLMS for impulse response $\boldsymbol{v}'[k]$ depicted in panel (a) (after sufficiently long convergence of the *inner AEC*) that may be observed by *outer AEC*. Panels (c) and (e) show the performance in terms of relative system distance D_{dB} and ERLE of the *outer AEC*, respectively, for a white excitation signal and panels (d) and (f) show the respective performance of the *outer AEC* for the speech excitation signal depicted in panel (b).

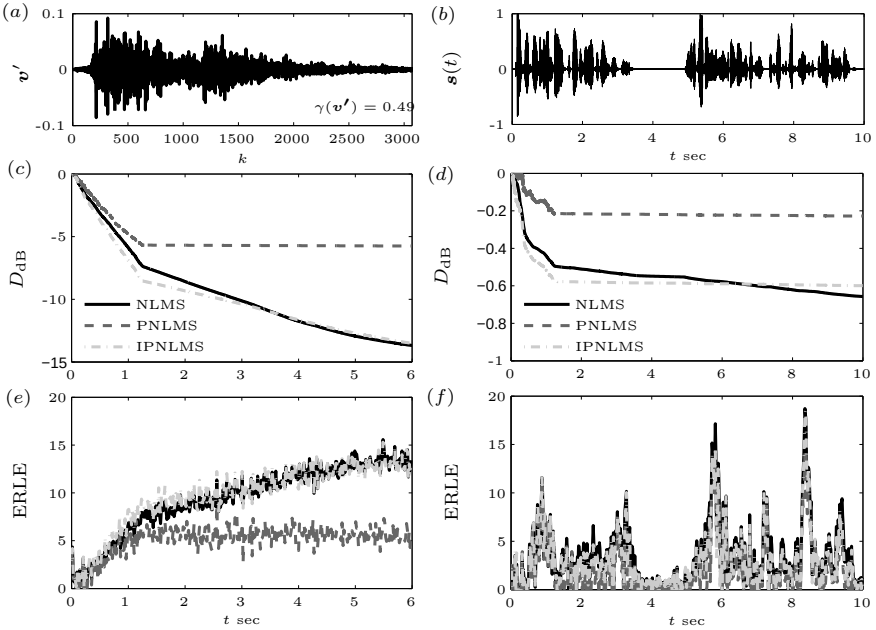

Figure 5.13: Comparison of NLMS, PNLMS and IPNLMS for impulse response $\boldsymbol{v}'[k]$ depicted in panel (a) (after insufficient convergence of the *inner AEC*) that may be observed by *outer AEC*. Panels (c) and (e) show the performance in terms of relative system distance D_{dB} and ERLE of the *outer AEC*, respectively, for a white excitation signal and panels (d) and (f) show the respective performance of the *outer AEC* for the speech excitation signal depicted in panel (b).

5.3 Combined System of LRC Filter, Inner and Outer AEC

For evaluation if an *outer AEC* should rely on the error signal of the *inner AEC* (switch S_1 in Figure 5.2 in lower position) or work independently (switch S_1 in Figure 5.2 in upper position) these two systems are compared in **Figure 5.14**. If the *outer AEC* directly depends on the error signal of the *inner AEC* the total ERLE of the combined system

$$\mathrm{ERLE}_{\mathrm{total}} = \mathrm{ERLE}_1 + \mathrm{ERLE}_2 \tag{5.3.1}$$

$$= 10 \log_{10} \frac{\mathrm{E}\{\psi^2[k]\}}{\mathrm{E}\{e^2_{\mathrm{AEC},2}[k]\}} \tag{5.3.2}$$

can be calculated from

$$\mathrm{ERLE}_1 = 10 \log_{10} \frac{\mathrm{E}\{\psi^2[k]\}}{\mathrm{E}\{e^2_{\mathrm{AEC},1}[k]\}} \tag{5.3.3}$$

achieved by the *inner AEC* and

$$\mathrm{ERLE}_2 = 10 \log_{10} \frac{\mathrm{E}\{e^2_{\mathrm{AEC},1}[k]\}}{\mathrm{E}\{e^2_{\mathrm{AEC},2}[k]\}} \tag{5.3.4}$$

achieved by the *outer AEC* as depicted in Figure 5.14.
Panels (a) and (b) of Figure 5.14 show simulation results for white excitation and speech excitation, respectively. Simulation results ERLE_1+ERLE_2 are shown for a system based on *inner AEC* updated by an NLMS algorithm and *outer AEC* updated by an IPNLMS algorithm (dashed grey line, switch S_1 in Figure 5.2 is in lower position) as well as for the case that switch S_1 in Figure 5.2 is in upper position (dash-dotted grey line). The solid black line and the dotted grey line show the contributions of the *inner AEC* and the *outer AEC* for the combined system (dashed line).
It can be seen from Figure 5.14 that, although the IR to be identified by the *outer AEC* $\boldsymbol{v}'[k]$ may not always be sparse, the system that exploits echo reduction of both filters (switch S_1 in Figure 5.2 is in lower position) leads to a higher amount of echo reduction. At all the AEC performance has been increased by about 50% by adding an *outer AEC*.
The previous simulation results (Figure 5.14) showed that a combined system of *inner AEC* and *outer AEC* relying on the error signal $e_{\mathrm{AEC},1}[k]$ which is updated by the IPNLMS algorithm shows good performance for sparse IRs and even if the system $\boldsymbol{v}'[k]$ is not always sparse, e.g. in periods of convergence of *inner AEC* or LRC filter.

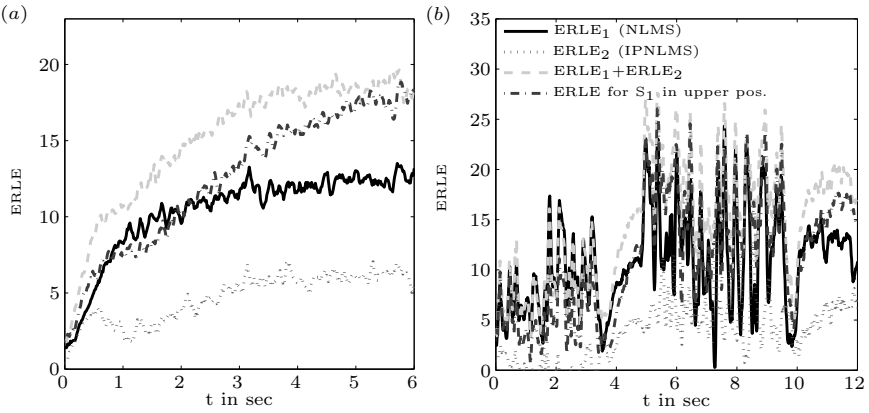

Figure 5.14: AEC performance comparison for system consisting of outer AEC and inner AEC for the two possible combination shown in Figure 5.2 (AEC filter lengths were $L_{\mathrm{AEC},1} = 1024$ and $L_{\mathrm{AEC},2} = 2048$). Left: white noise input. Right: speech input.

One further advantage of the proportionate update schemes is, that their convergence speed can be increased by a higher step-size $\mu[k]$. This will be visualized in **Figure 5.15** for a white Gaussian excitation $s_f[k]$ and in **Figure 5.16** for speech as input $s_f[k]$.
The convergence of the *inner AEC* is depicted in terms of relative system distance $D_{1,\mathrm{dB}}$ in panels (b) of Figures 5.15 and 5.16. The corresponding system distance vector $\tilde{\boldsymbol{h}}_1$ is depicted exemplary for time instances $\{1, 3, 5, 8, 10\}$ s in panels (c). Since the *outer AEC* has to identify the system $\boldsymbol{v}'[k]$, the equalizer coefficients are shown in panels (d) and the system $\boldsymbol{v}'$ is depicted in panels (e) at the respective time instances. Panels (f) show the performance of the LRC filter in terms of SRR_{out} and SRRE and panels (g) compare NLMS, PNLMS and IPNLMS for the *outer AEC* in the proposed system when all three adaptive filters are active.
If the step-size $\mu[k]$ is considered to be the same for NLMS, PNLMS and IPNLMS, $\mu = 0.05$ was found to be the highest possible step-size for the *outer AEC* to work for all algorithms. Here, the NLMS is the limiting algorithm while for PNLMS and IPNLMS higher step-sizes can be chosen. As it can be seen from Figures 5.15 and 5.16, the use of PNLMS and IPNLMS already achieves slight performance gains if the same step-size is chosen ($\mu = 0.05$). If the step-size is increased for PNLMS and IPNLMS (which is not possible for NLMS) the performance can be further increased.

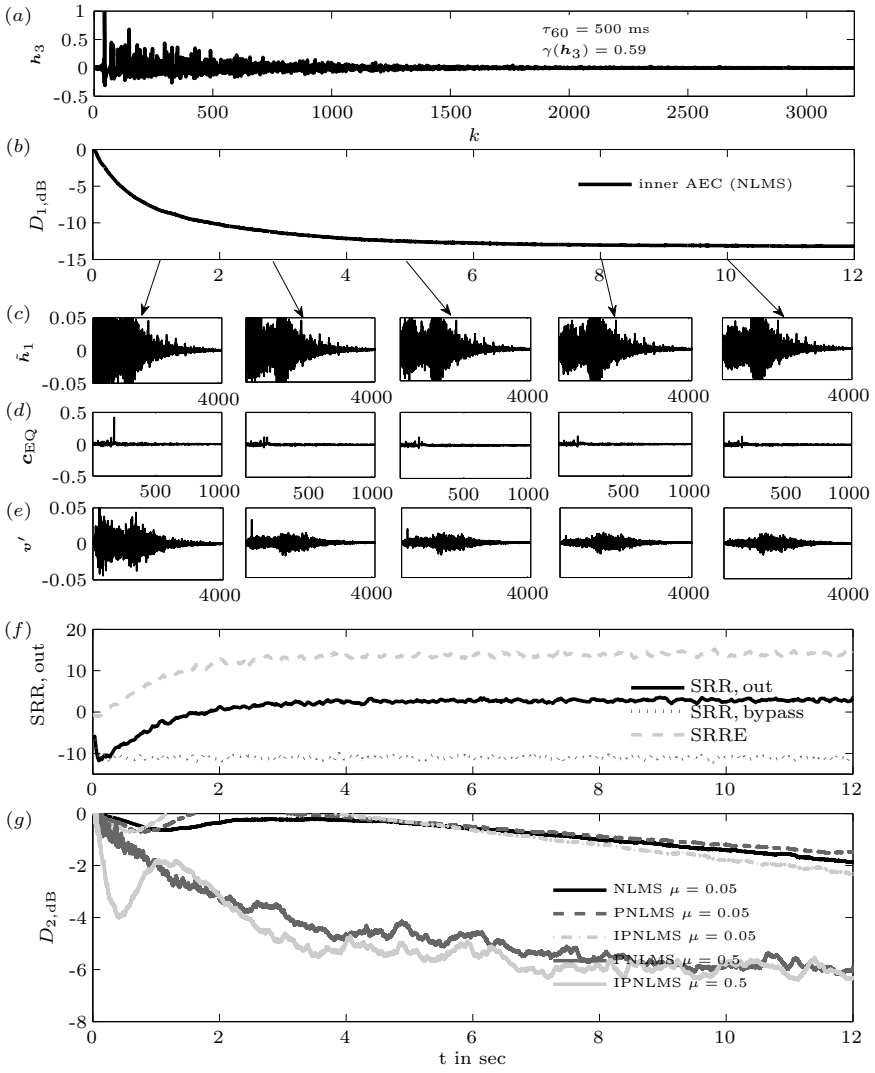

Figure 5.15: Performance of combined system as shown in Figure 5.2 (switch S_1 in lower position) for white noise as input. (a) the RIR; (b) relative system distance $D_{1,\mathrm{dB}}$ of inner AEC; (c) system distance vector $\tilde{\boldsymbol{h}} = \boldsymbol{h}[k] - \boldsymbol{c}_{\mathrm{AEC}_1}[k]$ of AEC_1 after $\{1, 3, 5, 8, 10\}$ seconds; (d) corresponding equalizer coefficients $\boldsymbol{c}_{\mathrm{EQ}}[k]$; (e) IR to be identified by *outer AEC* $\boldsymbol{v}'[k]$ for AEC_2; (f) SRR and SSRE achieved by LRC filter; (g) system distance of *outer AEC* $D_{2,\mathrm{dB}}$.

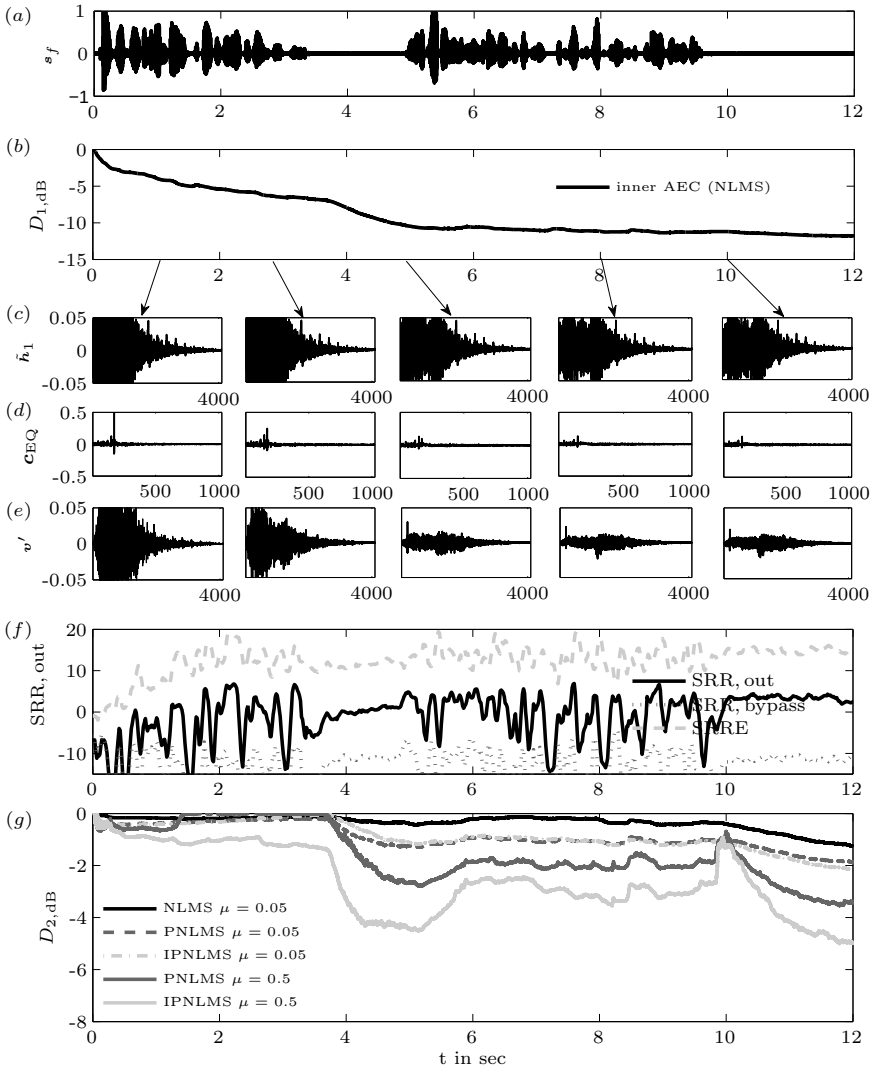

Figure 5.16: Performance of combined system as shown in Figure 5.2 (switch S_1 in lower position) for speech as input. (a) speech input signal; (b) relative system distance $D_{1,\mathrm{dB}}$ of inner AEC; (c) system distance vector $\tilde{\boldsymbol{h}} = \boldsymbol{h}[k] - \boldsymbol{c}_{\mathrm{AEC}_1}[k]$ of AEC_1 after $\{1, 3, 5, 8, 10\}$ seconds; (d) corresponding equalizer coefficients $\boldsymbol{c}_{\mathrm{EQ}}[k]$; (e) IR to be identified by *outer AEC* $\boldsymbol{v}'[k]$ for AEC_2; (f) SRR and SSRE achieved by LRC filter; (g) system distance of *outer AEC* $D_{2,\mathrm{dB}}$.

5.4 Chapter Summary

In this chapter, different possibilities for combinations of subsystems for AEC/AES and LRC and the respective mutual influences of these subsystems have been analyzed. Since all LRC approaches need knowledge about the RIR to be equalized, reliable system identification is crucial for the LRC filters. The system identification can be obtained from the *inner AEC* (cf. Section 5.1.1) or a combined system of *inner AEC* and AES post-filter if the AES filter is based on the proposed REEF (cf. Section 5.1.3).
The influence of system estimation errors on the LRC systems have been analyzed in Section 5.1. If the AEC filter's convergence state (system distance) is known or can be estimated, a method to incorporate this knowledge has been proposed in Section 5.1.2.
Not only has the convergence state of the AEC substantial influence of the performance of the LRC filter, but the LRC system also changes the input signal's correlation of the AEC filter since it is located in the AEC filter's input path and, by this, also has influence on the AEC filter's performance. This influence has been analyzed in Section 5.2.1 for the *inner AEC*.
An additional *outer AEC* was proposed to archive additional echo reduction. For the *outer AEC*, the system to be identified depends on the use of the *inner AECs* error signal. It has been shown that an additional *outer AEC* may be advantageous and that for the update of this *outer AEC* proportioned update schemes may be used, even if the assumption of a strictly sparse impulse response to be identified may not hold (cf. Section 5.2.2).
Based on these findings, a combined system based on one LRC filter and two AEC filters has been proposed and analyzed in Section 5.3.

Chapter 6

Summary and Possible Future Work

6.1 Summary

Hands-free communication systems suffer from acoustic disturbances such as ambient noise, acoustic echoes and room reverberation that decrease speech quality or even speech intelligibility and, thus, have to be removed from the transmission signals. While the combination of systems for acoustic echo cancellation and noise reduction has already been studied extensively in the literature, this thesis focused on the combinations and the mutual influences of subsystems for listening-room compensation and acoustic echo cancellation for hands-free communication systems.

Reverberation is caused in enclosed spaces by numerous reflections of a sound signal at the room boundaries (walls, floor, and ceiling) between the sound source and the receiver. A high amount of reverberation decreases speech intelligibility as it is obvious from speech in large rooms such as churches. Digital filter structures for listening-room compensation are one possibility to remove such reverberation by pre-processing a loudspeaker signal, aiming at an anechoic signal at the position of a human listener or a reference microphone. Thus, different methods for listening-room compensation have been introduced and analyzed in this thesis, especially with respect to robustness in terms of room impulse response estimation errors and spatial mismatch between the reference microphone and the human listener. It could be found that straightforward equalization of the acoustic channel, e.g. by least-squares approaches may lead to mathematically promising results in case that no errors are present. However, these ap-

proaches are very sensitive to estimation error which makes them practically infeasible. In this thesis, the influences of estimation errors and spatial mismatch has been analyzed and a method has been proposed to incorporate the knowledge about imperfect RIR estimates in the LRC filter design. Approaches for room-impulse response reshaping showed higher robustness to estimation errors and led to perceptually better results. To assess the influence on the quality of a dereverberated sound signal, extensive studies have been performed in this thesis to identify appropriate objective quality measures that show high correlation with ratings of human subjects. It has been found that most state-of-the-art objective quality measures may not be applicable for comparison of different LRC algorithms. Here, objective measures which are based on a model of the (human) auditory system showed the most promising results. A further drawback of most LRC approaches is the computational effort for the filter design. Most LRC filter designs are, thus, hardly applicable in real-time systems. Thus, a new, quickly converging gradient approach for update of the LRC filter coefficients has been developed in this thesis.

Acoustic echoes are caused by the fact that the loudspeaker signal of the system is picked up by its microphones again and transmitted back to the far-end user. By this, the far-end user perceives his or her own voice delayed by the round trip delay of the system which hampers speech communication. Acoustic echo cancellers and acoustic echo suppression filters remove this disturbance from the microphone signal and usually inherently estimate the acoustic channel at the same time. This information is crucial for the LRC system since its performance is drastically decreased for an RIR estimate of insufficient quality. However, not only the LRC approaches are influenced by the AEC filter that provides the RIR estimate needed. Since the LRC filter usually is located in front of the input path of the AEC filters, it has direct influence on the correlation of the AEC filters' input signal and, by this, on their convergence. These mutual influences have been investigated in this thesis and selection strategies for system combinations and gradient update strategies have been discussed. Especially, the application of proportionate filter update strategies has been investigated for the identification of equalized channels. Although the equalized systems to be identified not always have sparse nature, which would be optimum for proportionate update schemes, (partly) proportionate update schemes are applicable for the identification of equalized acoustic channels.

6.2 Possible Future Work

Regarding objective quality measures for listening-room compensation, no generally applicable objective measure could yet be identified in this thesis. Measures for speech intelligibility often are based on models for speech perception such as amplitude modulation. Some initial promising results could be found for speech intelligibility measures assessing speech quality which not yet have been presented in the thesis and allow for further research. Here, a combination of the so-called speech transmission index (STI), a common measure to assess speech intelligibility which is based on speech amplitude modulation, could be examined and possibly combined with the SRMR quality measure which is based on similar ideas. In general, more knowledge about the properties of the human auditory system should be incorporated in the objective quality assessment, such as spectral and temporal masking, or differences in the perceptual influence of spectral peaks and spectral dips in acoustic transfer functions. The evaluated quality measures are not only applicable to algorithms for listening-room compensation, but also to reverberation suppression approaches. Also here, further studies are necessary.

Regarding algorithms for listening-room compensation, further research is needed on increasing the robustness of the respective algorithms to estimation errors and spatial mismatch. Here, partial equalization approaches that reshape the impulse response show promising results that need further analysis.

Appendix

Appendix A

Objective Quality Measures for LRC

This appendix describes the objective quality measures that were listed in Tables 4.1 and 4.2 on page 73 without further explanation. They were used for the correlation analysis in Section 4.2.2 to identify objective quality measures for LRC that show high correlation with subjective ratings.
Section A.1 describes *channel-based* measures, i.e. algorithms that assess quality based on knowledge about the impulse response or the transfer function, and Section A.2 describes the *signal-based* measures, which can also be applied if only output signals are available but no impulse responses or transfer functions.

A.1 Channel-Based Measures

Room impulse responses (RIRs) and room transfer functions (RTFs) can be characterized by several objective measures that mostly origin from the research field of room acoustics, cf. e.g. [Kut00, ISO97, ISO06a, ISO06b, Adr06]. Most of them are based on a ratio between early and late part of the impulse response.
Since the IR of an equalized system $\boldsymbol{v}$ may look slightly different from common RIRs, **Figure A.1** exemplarily shows an equalized impulse response $\boldsymbol{v}$ of length $L_v = L_h + L_{\mathrm{EQ}} - 1$ and illustrates some definitions that will be used for the following objective measures. The position of the main peak of the impulse response is denoted by k_0 in Figure A.1. The lags corresponding to 50 ms and 80 ms later than the position of the main peak of the im-

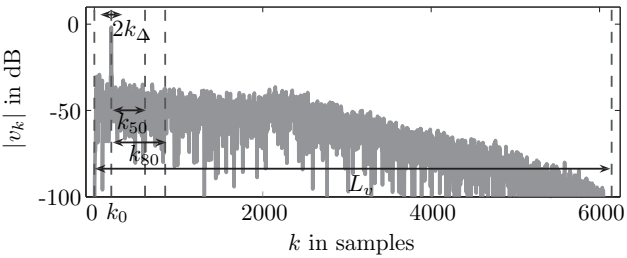

Figure A.1: Magnitude of impulse response of an equalized system $|\boldsymbol{v}| = |\boldsymbol{H}_{\mathrm{CM}}\boldsymbol{c}_{\mathrm{EQ}}|$ of length L_v in dB and the corresponding definitions of the position of the main peak k_0, the time lags following 50 ms and 80 ms after this main peak k_{50} and k_{80} and the interval of $2 \cdot k_\Delta$ around the main peak. Sampling frequency is $f_s = 8$ kHz.

pulse response k_0 are denoted by $k_{50} = \lfloor 0.05\ \mathrm{s} \cdot f_s \rfloor$ and $k_{80} = \lfloor 0.08\ \mathrm{s} \cdot f_s \rfloor$, respectively.
Channel-based measures that are widely used to characterize RIRs are defined in the following as well for common RIRs $\boldsymbol{h}$ as for equalized systems $\boldsymbol{v} = \boldsymbol{H}_{\mathrm{CM}} \cdot \boldsymbol{c}_{\mathrm{EQ}}$.

A.1.1 Definition

The ratio between the energy of the first 50 ms or the first 80 ms after the main peak of an IR to the overall energy of the RIR is called *Definition* and is denoted by D_{50} or D_{80}, respectively [Kut00].

$$\mathrm{D}_{50}(\boldsymbol{v}) = \frac{\sum_{k=k_0}^{k_0+k_{50}-1} v_k^2}{\sum_{k=0}^{L_v-1} v_k^2} \tag{A.1.1}$$

$$\mathrm{D}_{80}(\boldsymbol{v}) = \frac{\sum_{k=k_0}^{k_0+k_{80}-1} v_k^2}{\sum_{k=0}^{L_v-1} v_k^2} \tag{A.1.2}$$

Equations (A.1.1) and (A.1.2) are slightly modified compared to their usual definition in the literature [Kut00, Adr06] for the application to equalized systems $\boldsymbol{v}$ which have their maximum at the desired system delay of the equalizer k_0 (cf. Chapter 4.4.1). Please note that the coefficients v_k of the

equalized system vector $\boldsymbol{v}$ have to be replaced by the RIR coefficients h_k if the definition measure is calculated for a RIR, i.e.

$$\mathrm{D}_{50}(\boldsymbol{h}) = \frac{\sum_{k=0}^{k_{50}-1} h_k^2}{\sum_{k=0}^{L_h-1} h_k^2}. \tag{A.1.3}$$

In the following, only the definitions of channel-based objective measures for the equalized system vectors $\boldsymbol{v}$ are given, which also hold for the RIR vectors $\boldsymbol{h}$.

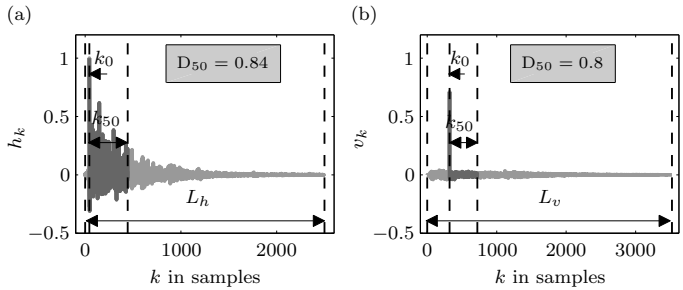

Figure A.2: Illustration of objective measure D_{50} for (a) RIR ($\tau_{60} \approx 300$ ms) and (b) equalized system $\boldsymbol{v}$ obtained by LS-equalizer of length $L_{\mathrm{EQ}} = 1024$.

Figure A.2 visualizes the calculation of D_{50}. The nominators of (A.1.1) and (A.1.3) are shown in darker grey and the demoninators span the whole IRs of length L_h and L_v, respectively.

The definition measure is motivated by the fact that the human auditory system is capable to jointly perceive the first 50 ms of an impinging speech signal. Thus, energy arriving within 50 ms increases intelligibility of speech signals while energy that arrives later than 50 ms decreases speech intelligibility [ISO97]. For music signals the D_{80} measure was found to be more suitable [Kut00, Adr06].

A.1.2 Clarity

The so-called *Clarity* [Kut00], denoted here by C_{50} or C_{80}, is the logarithmic ratio of 50 ms (80 ms) after the main peak to the rest of the impulse response.

$$\mathrm{C}_{50}(\boldsymbol{v}) = 10 \cdot \log_{10} \frac{\sum_{k=k_0}^{k_0+k_{50}-1} v_k^2}{\sum_{k=0}^{k_0-1} v_k^2 + \sum_{k=k_0+k_{50}}^{L_v-1} v_k^2} \tag{A.1.4}$$

$$\mathrm{C}_{80}(\boldsymbol{v}) = 10 \cdot \log_{10} \frac{\sum_{k=k_0}^{k_0+k_{80}-1} v_k^2}{\sum_{k=0}^{k_0-1} v_k^2 + \sum_{k=k_0+k_{80}}^{L_v-1} v_k^2} \tag{A.1.5}$$

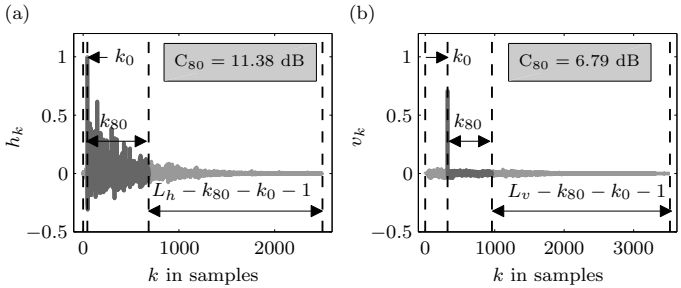

Figure A.3: Illustration of objective measure C_{80} for (a) RIR ($\tau_{60} \approx 300$ ms) and (b) equalized system $\boldsymbol{v}$ obtained by LS-equalizer of length $L_{\mathrm{EQ}} = 1024$.

Again, equations (A.1.4) and (A.1.5) are slightly modified compared to the literature [Kut00, Adr06] to account for equalized systems $\boldsymbol{v}$. The IR portions in nominators and denominators of (A.1.4) and (A.1.5) are visualized in **Figure A.3** in lighter and darker grey and the length definitions are illustrated.

A.1.3 Central Time (CT)

The so-called *Central Time* CT [Kut00] is no direct ratio but the center of gravity in terms of the energy of the RIR as visualized in **Figure A.4**. The Central Time is defined as [Kut00]

$$\mathrm{CT}(\boldsymbol{v}) = \frac{\sum_{k=0}^{L_v-1} k \cdot v_k^2}{\sum_{k=0}^{L_v-1} v_k^2}. \tag{A.1.6}$$

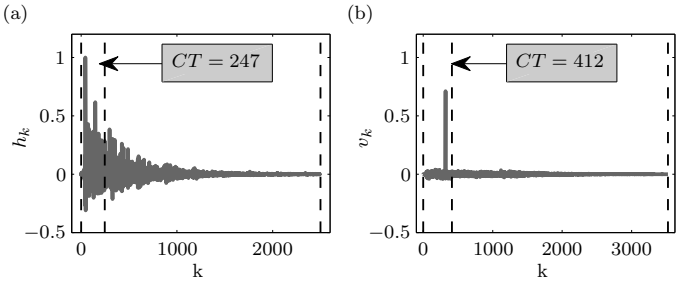

Figure A.4: Illustration of objective measure CT for (a) RIR ($\tau_{60} \approx 300$ ms) and (b) equalized system $\boldsymbol{v}$ obtained by LS-equalizer of length $L_{\mathrm{EQ}} = 1024$.

A.1.4 Direct-to-Reverberation-Ratio (DRR)

The *Direct-to-Reverberation-Ratio* DRR [TS06] is defined as the logarithmic ratio between the energy of the direct path of the impulse response and the energy of all reflections. However, since the direct path, in general, does not match the sampling grid, a small range around the main peak is considered as the direct path energy [TS06, Hab07]:

$$\mathrm{DRR}(\boldsymbol{v}) = 10 \cdot \log_{10} \frac{\sum_{k=k_0-k_\Delta}^{k_0+k_\Delta} v_k^2}{\sum_{k=0}^{k_0-k_\Delta-1} v_k^2 + \sum_{k=k_0+k_\Delta+1}^{L_v} v_k^2} \tag{A.1.7}$$

For this thesis, k_Δ was chosen as $k_\Delta = 4 \text{ ms} \cdot f_s$ (cf. Figure A.1).

A.1.5 Spectral Variance

All measures described so far assess time-domain properties of the respective IR. Spectral quality measures will be described in the following. Since equalization often aims at a flat spectrum, the *variance* (VAR) of logarithmic overall transfer function $\mathsf{v}_n = \mathsf{h}_n \mathsf{c}_{\mathrm{EQ},n}$ was proposed in [Mou94] to evaluate LRC algorithms.

$$\mathrm{VAR}(\mathbf{v}) = \frac{1}{n_{\max} - n_{\min} + 1} \sum_{n=n_{\min}}^{n_{\max}} \left(20 \log_{10}|\mathsf{v}_n| - \bar{\mathsf{v}}_{\mathrm{dB}}\right)^2 . \tag{A.1.8}$$

In (A.1.8),

$$\bar{\mathsf{v}}_{\mathrm{dB}} = \frac{1}{n_{\max} - n_{\min} + 1} \sum_{n=n_{\min}}^{n_{\max}} 20 \log_{10} |\mathsf{v}_n| \tag{A.1.9}$$

is the mean logarithmic spectrum and $n_{\min}$ and $n_{\max}$ the frequency indices that limit the considered frequency range in which the equalized transfer function is desired to be flat. Reasonable values for $n_{\min}$ and $n_{\max}$ can be the cut-off frequencies of the desired system $\boldsymbol{d}$ defined in (4.4.3), to account for the high-pass or band-pass characteristics.

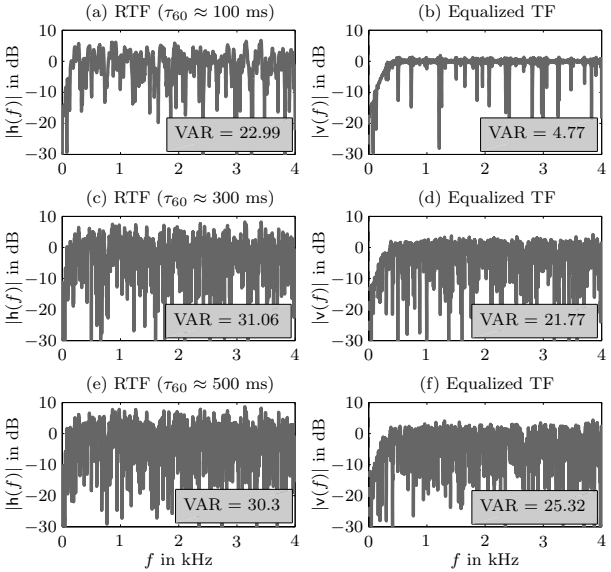

Figure A.5: Illustration of variance (VAR) as an objective measure. (a), (c), (e) RTFs $\mathbf{h}$ of different room reverberation time ($\tau_{60} \approx \{100, 300, 500\}$ ms) and their variances according to (A.1.8); (b), (d), (e) corresponding equalized systems $\mathbf{v}$ obtained by LS-equalizer of length $L_{\mathrm{EQ}} = 1024$ and their variances.

The variance measure is illustrated in **Figure A.5**, here calculated with lower and upper frequency limits at lags $n_{\min} = \lfloor L_{\mathrm{DFT}} \cdot 200 \text{ Hz}/f_s \rfloor$ and $n_{\max} = \lfloor L_{\mathrm{DFT}} \cdot 4000 \text{ Hz}/f_s \rfloor$, respectively, which were chosen to reflect the high-pass characteristic for the desired system $\boldsymbol{d}$. Left-hand panels of Figure A.5 show RTFs characterized by different room reverberation times

τ_{60} and right-hand panels show the corresponding equalized TFs generated by convolution of the respective RTF with an LS-LRC filter of length $L_{\mathrm{EQ}} = 1024$. It can be seen that higher room reverberation times lead to higher variance in the RTF and that equalization reduces the variance measure VAR.

A.1.6 Spectral Flatness Measure (SFM)

A second measure that assesses a flat overall transfer function is the so-called *spectral flatness measure* (SFM) [Joh88] that calculates the ratio of geometric mean $G(\mathbf{v})$ and the arithmetic mean $A(\mathbf{v})$ of $\mathbf{v}$.

$$\mathrm{SFM}(\mathbf{v}) = \frac{G(\mathbf{v})}{A(\mathbf{v})} = \frac{\sqrt[N]{\prod_{n=0}^{N-1} |\mathsf{v}_n|^2}}{\frac{1}{N}\sum_{n=0}^{N-1} |\mathsf{v}_n|^2} \tag{A.1.10}$$

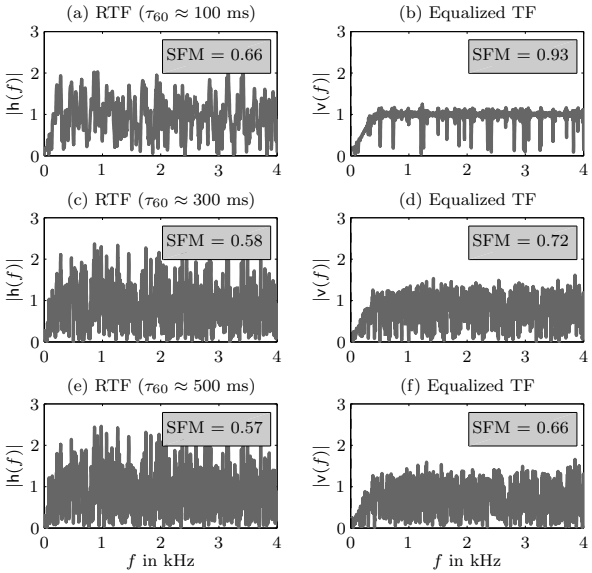

Figure A.6: Illustration of spectral flatness measure (SFM) as an objective quality measure. (a), (c), (e) RTFs $\mathbf{h}$ of different room reverberation time and the corresponding SFM; (b), (d), (e) equalized systems $\mathbf{v}$ obtained by LS-equalizer of length $L_{\mathrm{EQ}} = 1024$ and their SFM.

The SFM is illustrated in **Figure A.6** for the same RTFs (left panels) and equalized transfer functions (right panels) as in Figure A.5 (however, this time depicted with linear amplitude).

A.2 Signal-Based Quality Measures

Whenever impulse responses or transfer functions are not obtainable for objective testing, e.g. for blind dereverberation [Hab07], algorithms have to be evaluated based on the signals only. Thus, this section introduces some technical quality measures that are based on the processed signal only (mostly including a reference signal, i.e. intrusive measures).

A.2.1 Segmental Signal-to-Reverberation Ratio (SSRR)

The most simple measures are the *segmental signal-to-reverberation ratio* (SSRR) [NG05] and the *SSRR enhancement* (SSRRE) [GKMK08d] that are defined similarly to SNR-based measures known from noise reduction quality assessment. The segmental signal to reverberation ratio (SSRR) is defined as

$$\mathrm{SSRR}_{\mathrm{dB}} = \frac{1}{K/L_{\mathrm{Bl}}} \sum_{\ell=0}^{K/L_{\mathrm{Bl}}-1} 10\log_{10} \frac{\sum_{k=0}^{L_{\mathrm{Bl}}-1} \hat{y}[\ell L_{\mathrm{Bl}}+k]^2}{\sum_{k=0}^{L_{\mathrm{Bl}}-1} (\hat{y}[\ell L_{\mathrm{Bl}}+k] - y[\ell L_{\mathrm{Bl}}+k])^2} \tag{A.2.1}$$

with K being the total length of the signal, L_{Bl} the block length (typically corresponding to 16-32 ms) and ℓ the block index. The signals $y[k]$ and $\hat{y}[k]$ are the microphone signal and the reference signal, respectively (cf. e.g. Figure 4.11).
The SSRR can be normalized by

$$\mathrm{SRRE}_{\mathrm{dB}} = \mathrm{SSRR}_{\mathrm{dB}} - \mathrm{SSRR}_{\mathrm{bypass}} \tag{A.2.2}$$

with

$$\mathrm{SSRR}_{\mathrm{bypass}} = \frac{1}{K/L_{\mathrm{Bl}}} \sum_{\ell=0}^{K/L_{\mathrm{Bl}}-1} 10\log_{10} \frac{\sum\limits_{k=0}^{L_{\mathrm{Bl}}-1} \hat{y}[\ell L_{\mathrm{Bl}}+k]^2}{\sum\limits_{k=0}^{L_{\mathrm{Bl}}-1} (\hat{y}[\ell L_{\mathrm{Bl}}+k] - y_b[\ell L_{\mathrm{Bl}}+k])^2} \tag{A.2.3}$$

In (A.2.3) $y_b[k] = s_f[k] * h[k] * d[k]$ is the microphone signal processed by an LRC switched to *bypass*, i.e. $c_{\mathrm{EQ}}[k] = d[k]$. To prevent a delay between

$y[k]$ and $y_b[k]$ the loudspeaker signal is pre-filtered by the desired system $d[k]$. By this, SSRRE can be interpreted as the enhancement achieved by the LRC filter compared to the case that the filter is switched off.

Usually, signal blocks not containing speech are neglected for the SSRR measure since otherwise these blocks have strong influence on the measure itself.

A.2.2 Frequency-Weighted SSRR (FWSSRR)

The *frequency-weighted SSRR* (FWSSRR) [Loi07] represents a first step towards consideration of the human auditory system by analyzing the SSRR in different frequency bands as given in Table A.1. For each band a weighting factor w_j exist which is related to the importance of the respective band and is obtained from studies regarding the articulation index [QBC88, Loi07].

Band	f_c (in Hz)	w_j	Band	f_c (in Hz)	w_j
1	50	0.003	14	1148	0.032
2	120	0.003	15	1288	0.034
3	190	0.003	16	1442	0.035
4	260	0.007	17	1610	0.037
5	330	0.010	18	1794	0.036
6	400	0.016	19	1993	0.036
7	470	0.016	20	2221	0.033
8	540	0.017	21	2446	0.030
9	617	0.017	22	2701	0.029
10	703	0.022	23	2978	0.027
11	798	0.027	24	3276	0.026
12	904	0.028	25	3597	0.026
13	1020	0.030	-	-	-

Table A.1: Center frequencies f_c and scaling factors w_j for band pass design [QBC88, Loi07].

The FWSSRR is defined as [Loi07]

$$\mathrm{FWSSRR} = \frac{10}{N_\ell} \sum_{\ell=0}^{N_\ell-1} \frac{\sum_{j=1}^{C} \mathsf{w}_j \log_{10}\left[\left(\mathsf{x}_{\hat{y}}[\ell,j]\right)^2 / \left(\mathsf{x}_{\hat{y}}[\ell,j] - \mathsf{x}_y[\ell,j]\right)^2\right]}{\sum_{j=1}^{C} \mathsf{w}_j} \tag{A.2.4}$$

with j, ℓ, C and N_ℓ being the band index, the block index, the number of bands considered and the number of blocks considered, respectively. $\mathsf{x}_{\hat{y}}[\ell,j]$ and $\mathsf{x}_y[\ell,j]$ are the band limited versions of $y[k]$ and $\hat{y}[k]$ after filtering with the band passes according to Table A.1 for the block index ℓ and band j.

A.2.3 Weighted Spectral Slope (WSS)

The weighted spectral slope [Kla82, QBC88] assesses spectral variations between the reference signal and the signal under test. Thus, first the spectral slope for each frequency band j and block index ℓ is calculated based on the first order difference

$$\mathsf{e}_{\hat{y}}[\ell,j] = \mathsf{x}_{\hat{y}}[\ell,j+1] - \mathsf{x}_{\hat{y}}[\ell,j], \tag{A.2.5}$$

$$\mathsf{e}_y[\ell,j] = \mathsf{x}_y[\ell,j+1] - \mathsf{x}_y[\ell,j]. \tag{A.2.6}$$

With (A.2.5) and (A.2.6), the WSS is defined as

$$\mathrm{WSS} = \frac{1}{N_\ell} \sum_{\ell=0}^{N_\ell-1} \left(\sum_{j=1}^{C} \mathsf{w}[\ell,j] \left(\mathsf{e}_{\hat{y}}[\ell,j] - \mathsf{e}_y[\ell,j]\right)^2 \right). \tag{A.2.7}$$

The weighting factors $\mathsf{w}[\ell,j]$ in (A.2.7) are calculated by

$$\mathsf{w}[\ell,j] = \frac{\mathsf{k}_{max}}{\mathsf{k}_{max} + \mathsf{x}_{max}[\ell] - \mathsf{x}_{\hat{y}}[\ell,j]} \cdot \frac{\mathsf{k}_{loc,max}}{\mathsf{k}_{loc,max} + \mathsf{x}_{loc,max}[\ell] - \mathsf{x}_{\hat{y}}[\ell,j]}, \tag{A.2.8}$$

with $\mathsf{x}_{max}[\ell]$ being the largest log-spectral magnitude among all bands, $\mathsf{x}_{loc,max}[\ell]$ being the value of the peak closest to band j, and $\mathsf{k}_{max} = 20$ and $\mathsf{k}_{loc,max} = 1$ being constants which have been adjusted by linear regression analysis to maximize correlation of WSS with subjective data [Kla82, Loi07].

A.2.4 Log-Spectral Distortion (LSD)

To account for logarithmic loudness perception of the human auditory system the *log-spectral distortion* (LSD) compares logarithmically weighted signal blocks $\mathbf{y}[\ell]$ and $\hat{\mathbf{y}}[\ell]$ in frequency-domain.

$$\mathrm{LSD}[\ell] = \|\mathcal{L}\{\mathbf{y}[\ell]\} - \mathcal{L}\{\hat{\mathbf{y}}[\ell]\}\|_p \tag{A.2.9}$$

For this work, p was chosen to be 2. The operator $\mathcal{L}\{\mathcal{A}[\ell]\}$ with $\mathcal{A}[\ell] \in \{\mathbf{y}[\ell], \hat{\mathbf{y}}[\ell]\}$ is defined as

$$\mathcal{L}\{\mathcal{A}[\ell]\} = \max\{20\log_{10}(|\mathcal{A}[\ell]|), \delta\} \tag{A.2.10}$$

$$\delta = \max_{l,n}\{20\log_{10}(|\mathcal{A}[\ell]|)\} - 50. \tag{A.2.11}$$

with a limit of the short-time spectra of 50 dB.

A.2.5 LPC-based Quality Measures

Since dereverberation of speech is the aim in most scenarios, objective quality measures based on the LPC models, such as the *Log-Area Ratio* (LAR), the *Log-Likelihood Ratio* (LLR), the Itakura-Saito Distance (IS), and the *Cepstral Distance* (CD) will be introduced in the following. Since LPC based quality measures are quite common and well known e.g. in the research field of noise reduction, the respective measures are just briefly defined and the interested reader is referred to the literature, e.g. [Loi07, QBC88], for more details.

Log Likelihood Ratio (LLR)

The log-likelihood ratio for a signal block is defined as [Loi07]

$$\mathrm{LLR}[\ell] = \ln\left(\frac{\boldsymbol{a}_y^T[\ell]\ \boldsymbol{R}_{\hat{y}\hat{y}}[\ell]\ \boldsymbol{a}_y[\ell]}{\boldsymbol{a}_{\hat{y}}^T[\ell]\ \boldsymbol{R}_{\hat{y}\hat{y}}[\ell]\ \boldsymbol{a}_{\hat{y}}[\ell]}\right) \tag{A.2.12}$$

with the auto-correlation matrix $\boldsymbol{R}_{\hat{y}\hat{y}}[\ell]$ and the LPC coefficient vectors of the signal blocks $\boldsymbol{y}[\ell]$ and $\hat{\boldsymbol{y}}[\ell]$,

$$\boldsymbol{a}_{\hat{y}}[\ell] = [1, -a_{\hat{y}}[1,\ell], -a_{\hat{y}}[2,\ell], \ldots, -a_{\hat{y}}[p,\ell]]^T, \tag{A.2.13}$$

$$\boldsymbol{a}_y[\ell] = [1, -a_y[1,\ell], -a_y[2,\ell], \ldots, -a_y[p,\ell]]^T \tag{A.2.14}$$

that can e.g. obtained by Levinson-Durbin-recursion [KK09].

Itakura-Saito Distance (ISD)

The Itakura-Saito distance is defined as [QBC88]

$$\mathrm{ISD}[\ell] = \frac{G_{\hat{y}}[\ell]}{G_y[\ell]}\,\frac{\boldsymbol{a}_y^T[\ell]\ \boldsymbol{R}_{\hat{y}\hat{y}}[\ell]\ \boldsymbol{a}_y[\ell]}{\boldsymbol{a}_{\hat{y}}^T[\ell]\ \boldsymbol{R}_{\hat{y}\hat{y}}[\ell]\ \boldsymbol{a}_{\hat{y}}[\ell]} + \log\left(\frac{G_{\hat{y}}[\ell]}{G_y[\ell]}\right) - 1 \tag{A.2.15}$$

with $G_{\hat{y}}[\ell] = \boldsymbol{r}_{\hat{y}\hat{y}}^T[\ell]\boldsymbol{a}_{\hat{y}}[\ell]$ and $G_y[\ell] = \boldsymbol{r}_{yy}^T[\ell]\boldsymbol{a}_y[\ell]$ being the all-pole amplification factors depending on the LPC coefficients and the auto-correlation sequences $\boldsymbol{r}_{\hat{y}\hat{y}}[\ell]$ and $\boldsymbol{r}_{yy}[\ell]$.

Log-Area Ratio (LAR)

The log-area ratio [HP98] is defined as

$$\mathrm{LAR} = \left| \frac{1}{N_\ell} \sum_{\ell=1}^{N_\ell} \left[\log_{10}\left(\frac{1+r_{\hat{y}}[\ell]}{1-r_{\hat{y}}[\ell]}\right) - \log_{10}\left(\frac{1+r_y[\ell]}{1-r_y[\ell]}\right) \right]^2 \right|^{\frac{1}{2}}. \quad \text{(A.2.16)}$$

In (A.2.16), $r_{\hat{y}}[\ell]$ and $r_y[\ell]$ are the LP reflection coefficients [HP98, VM06].

Cepstral Distance (CD)

The cespstral distance is defined as [VHH98]

$$\mathrm{CD} = \frac{1}{K/L} \sum_{\ell=1}^{K/L} \left\{ \frac{10}{\ln 10} \cdot \sqrt{[c_{\hat{y}}[1,\ell] - c_y[1,\ell]]^2 + 2\sum_{j=2}^{m} [c_{\hat{y}}[j,\ell] - c_y[j,\ell]]^2} \right\}. \quad \text{(A.2.17)}$$

In (A.2.17), $c_{\hat{y}}[j,\ell]$ and $c_y[j,\ell]$ are the cepstral coefficients for the block ℓ that can be calculated recursively from the prediction coefficients:

$$c_y[1,\ell] = -a_y[2,\ell] \quad \text{(A.2.18)}$$

$$c_y[j,\ell] = -a_y[j,\ell] - \sum_{i=1}^{j-1} \left(1 - \frac{i}{j}\right) \cdot a_y[i,\ell] \cdot a_y[i-k,\ell] \quad \text{(A.2.19)}$$

A.2.6 Psychoacoustically Motivated Quality Measures

While LPC based measures described before are based on speech production models, the psychoacoustically motivated measures that are described in the following are based on findings in the auditory systems of humans and animals. Thus, they model, how sounds are perceived by the human and due to this may be more appropriate for assessing the perceived audio quality.

Bark Spectral Distortion (BSD)

An objective quality measure that is based on the so-called Bark bands is the *Bark spectral distortion* measure (BSD) [WSG92, Yan99] which compares perceived loudness incorporating spectral masking effects.

Inside the cochlea, the auditory hair cells are located within the organ of Corti on a thin basilar membrane. The hair cells are excited by the sound waves travelling along the organ of Corti. High-frequency components excite hair cells near the oval window while low frequency parts excite the hair cells at the end of the organ of Corti. Thus, the cochlea performs a frequency-to-place transform and, by this, a spectral analysis. The information gathered by the hair cells is fed forward to the human brain via the auditory nerves. However, this frequency analysis inside the cochlea is not a linear frequency-place transform but a logarithmic one. Due to this, the frequency resolution of the human auditory system is much better for low frequencies than for high frequencies. **Figure A.7** and **Table A.2** illustrate the so-called critical bands which are given in the pseudo-unit *Bark*[1].

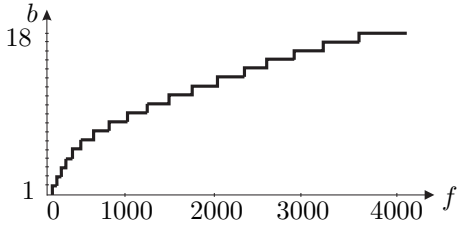

Figure A.7: Critical bands [ZF99].

Bark	f_{low} (in Hz)	f_{high} (in Hz)	Bark	f_{low} (in Hz)	f_{high} (in Hz)
0	0	100	9	1080	1265
1	100	200	10	1265	1480
2	200	300	11	1480	1715
3	300	400	12	1715	1990
4	400	510	13	1990	2310
5	510	630	14	2310	2690
6	630	770	15	2690	3125
7	770	920	16	3125	3675
8	920	1080	17	3675	4350

Table A.2: Corresponding frequency regions for critical bands from 0 to 17 according to [VHH98].

[1]The pseudo-unit bark was chosen in honour of the German physicist Heinrich Georg Barkhausen (1881 - 1956), who was born in Bremen, Germany

The critical bands approximately correspond to the frequency-to-place transform on the basilar membrane and it can be seen from Figure A.7 that the frequency resolution for low frequencies is higher than for high frequencies. The band widths given in Table A.2 have been obtained from psychoacoustical experiments, i.e. inside one Bark band two sinusoidal tones or narrow-band noises are no longer perceived independently, but only one tone is perceived.
The relation between the critical bandwidth $CB(f)$ and the frequency f can be calculated according to [VM06] as follows:

$$CB(f) = 25 + 75 \left(1 + 1.4 \left(\frac{f}{1000 \text{ Hz}} \right)^2 \right)^{0.69} \tag{A.2.20}$$

Adding the bandwidths of the adjacent critical bands leads to the scale of critical band rate b in Bark. The frequency-to-Bark transform is then obtained as follows [VM06]:

$$b = 13 \cdot \arctan \left(0.76 \frac{f}{1000 \text{ Hz}} \right) + \arctan \left(\frac{f}{7500 \text{ Hz}} \right) \tag{A.2.21}$$

If two signals (sinusoidal tones or narrowband noises) having different amplitudes occur at the same time but at different frequencies the weaker signal could be inaudible for the human auditory system. This phenomenon is called frequency masking and is illustrated in **Figure A.8**. From Figure A.8 it can be seen, that the so-called threshold of hearing around the masker is raised.

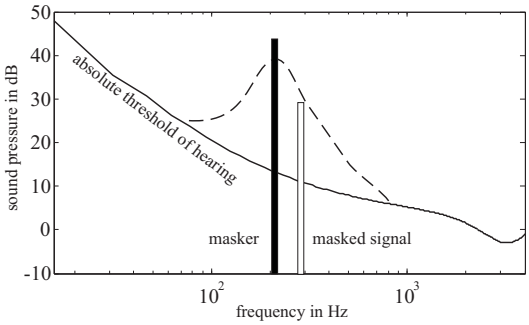

Figure A.8: Spectral masking.

If two signals (e.g. sinusoids or narrowband noise signals) are present simultaneously, one signal raises the absolute threshold of hearing. Everything

below the raised threshold (the so-called masking threshold) is inaudible for the human auditory system. The masking threshold is indicated by the dashed line in Figure A.8.
The shape of the masking threshold caused by the masker is almost triangular in the bark domain, and it declines more quickly towards lower frequencies (about 25 dB/Bark) than towards higher frequencies (10 dB/Bark). In **Figure A.9** three approximations for the so-called spreading function are depicted.

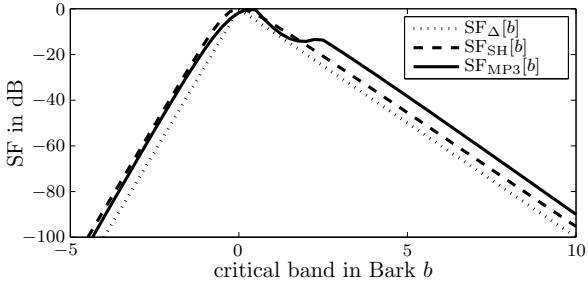

Figure A.9: Three different spreading functions.

The most simple approximation of the spreading function is the triangle function (dotted line in Figure A.9).

$$10\log_{10}\mathrm{SF}_{\Delta}[b] = \begin{cases} 25\,b, & \text{for } b < 0 \\ -10\,b, & \text{for } b \geq 0 \end{cases} \tag{A.2.22}$$

The approximation by Sekey and Hanson [SH84] (dashed line in Figure A.9) is more complex, but fits experimental results better.

$$10\log_{10}\mathrm{SF}_{\mathrm{SH}}[b] = 7 - 7.5\cdot(b-0.215) - 17.5\cdot\sqrt{0.196 + (b-0.215)^2} \tag{A.2.23}$$

The approximation used in this thesis (solid line in Figure A.9) is that defined in the MP3-Standard [Int92] which can be expressed by

$$10\log_{10}\mathrm{SF}_{\mathrm{MP3}}[b] = \begin{cases} 15.81 + 7.5x_1 - 17.5\sqrt{1.0 + x_1^2} + x_2, \\ \qquad\qquad \text{for } 0.5 \leq x_1 \leq 2.5 \\ 15.81 + 7.5x_1 - 17.5\sqrt{1.0 + x_1^2}, \\ \qquad\qquad \text{else,} \end{cases} \tag{A.2.24}$$

with

$$\tag{A.2.25}$$

$$x_1 = 1.05b \tag{A.2.26}$$

$$x_2 = 8.0[(x_1 - 0.5)^2 - 2.0(x_1 - 0.5)]. \tag{A.2.27}$$

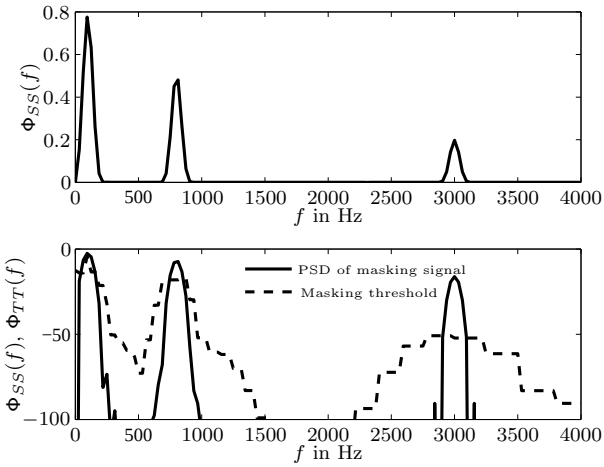

Figure A.10: PSD $\Phi_{SS}(f)$ of three sinuses and corresponding masking threshold $\Phi_{TT}(f)$.

Figure A.10 illustrates the calculation of the masking threshold $\Phi_{TT}(f)$ (dashed line in lower panel) given an excitation of three sinusoidal signal components (upper panel) by convolution of the excitation signal with the spreading function $SF_{\mathrm{MP3}}[b]$.

$$\begin{aligned} \mathsf{s}_{\hat{y}}[\ell, b] &= \sum_{j=1}^{C_b} SF_{\mathrm{MP3}}[b-j] \cdot |\hat{\mathsf{y}}[\ell, j]|^2 \\ \mathsf{s}_{y}[\ell, b] &= \sum_{j=1}^{C_b} SF_{\mathrm{MP3}}[b-j] \cdot |\mathsf{y}[\ell, j]|^2 \end{aligned} \tag{A.2.28}$$

In (A.2.28), C_b is the number of frequency groups used. As a next step the loudness level p is calculated based on the equal-loudness-curves according to [ISO03] depicted in **Figure A.11**. The loudness level is given in the pseudo-unit *phon*. The curve at 0 phon corresponds to the absolute threshold of hearing. Sounds having pressure levels on a specific equal-loudness curve are perceived by the human auditory system as if they would be equally loud [ZF99].

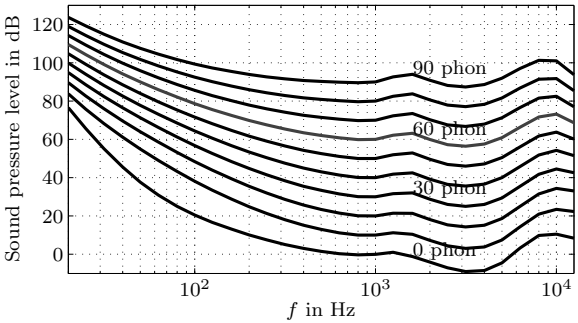

Figure A.11: Equal loudness curves according to [ISO03].

From the loudness level p in phon, the psychoacoustically motivated subjective loudness level m in sone can be calculated by [WSG92]

$$\mathsf{m} = \begin{cases} 2^{(\mathsf{p}-40)/10} & \text{for } \mathsf{p} \geq 40 \\ (\mathsf{p}/40)^{2.642} & \text{for } \mathsf{p} < 40. \end{cases} \tag{A.2.29}$$

Figure A.12 visualizes the conversion from phon to sone.

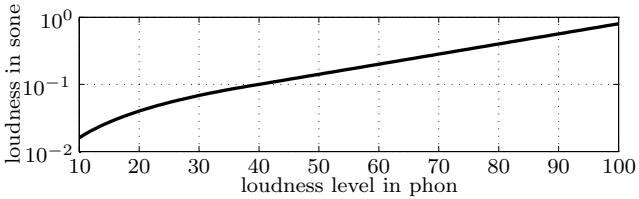

Figure A.12: Relation between sone and phon according to (A.2.29).

Eq. (A.2.29) can be used to calculate the perceived loudness $\mathsf{m}_{\hat{y}}[\ell, b]$ of $\hat{y}[k]$ in sone and the perceived loudness $\mathsf{m}_y[\ell, b]$ of $y[k]$ in sone. Now the BSD can be calculated by

$$\text{BSD} = \frac{\sum_{\ell=1}^{N_\ell} \sum_{i=1}^{C_b} \Big[\mathsf{m}_{\hat{y}}[\ell, i] - \mathsf{m}_y[\ell, i]\Big]^2}{\sum_{\ell=1}^{N_\ell} \sum_{i=1}^{C_b} \Big[\mathsf{m}_{\hat{y}}[\ell, i]\Big]^2}. \tag{A.2.30}$$

The whole signal processing chain to calculate the BSD measure is visually summarized in **Figure A.13**

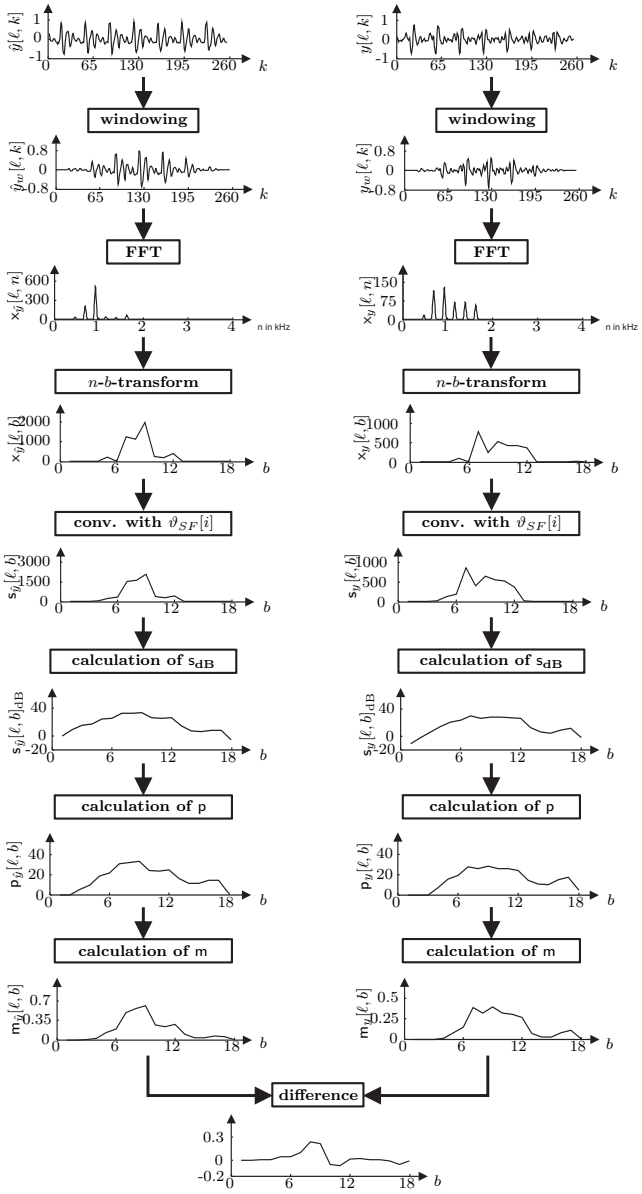

Figure A.13: Illustration of BSD calculation.

Reverberation Decay Tail (RDT)

All previous measures have not been developed specifically to assess quality of reverberant or dereverberated signals. In the following, three measures will be briefly described that have been developed to assess quality of such signals, i.e. the *reverberation decay tail (RDT)* measure that tries to estimate the amount of reverberation directly from the signal, the *objective measure for coloration in reverberation (OMCR)* that focuses on the dimension colouration (spectral changes) and the *speech-to-reverberation modulation energy ratio (SRMR)* which was designed to non-intrusively assess the quality of reverberant signals.

As depicted in Figure 2.3 on page 12, the amount of reverberation can be characterized by the room reverberation time τ_{60} (cf. also Section 2.1.4 and Figure 2.6). The RDT measure [WN06] tries to determine the decay parameter of the underlying RIR in reverberant speech from the signal. For that purpose, the algorithm searches for end-points in the signal in different frequency bands. After this end-points the decay of the signal is taken as an indicator for the influence of the RIR.

The previously described BSD measure does not distinguish between the effects of reverberation and colouration. Therefore, in [WN06] the RDT measure has been proposed which determines the reverberation effect alone based on the Bark spectra. For this, the decay of the underlying RIR is estimated from the input signal as described in the following. The exponentially decaying RIR model based on (2.1.2) and (2.1.3)

$$d[\ell, b] = A_b e^{-\lambda_b \ell}, \qquad \ell = 1, 2, ..., I + J \tag{A.2.31}$$

is assumed and estimated in the Bark domain for each Bark bin b from the difference of the Bark spectra of consecutive blocks

$$\Delta\chi[\ell, \ell', b] = \mathsf{m}_{\hat{y}}[\ell, b] - \mathsf{m}_{\hat{y}}[\ell + \ell', b]. \tag{A.2.32}$$

The decay model in (A.2.31) can be reformulated using Taylor expansion to [WN06]

$$A_b e^{-\lambda_b \ell} = A_b - A_b \lambda_b \ell. \tag{A.2.33}$$

An search algorithm is developed in [WN06] to search for so-called *end-points*, at which the signal energy abruptly decays, and for so-called *flat regions* following immediately after detected end-points. In (A.2.31), I and J are parameters of the end-point detection algorithm which will be explained in the following. $I + J$ is the length of the decay curve in blocks. The parameters of an decay model in average for all Bark bins

$$A_{avg} e^{-\lambda_b \ell} = A_{avg} - A_{avg} \lambda_{avg} \ell. \tag{A.2.34}$$

can be determined based on the average absolute decay tail energy [WN06]

$$A_{avg} = \frac{\sum_{b=1}^{C_b} A_b}{C_b}, \tag{A.2.35}$$

and the average decay tail rate [WN06]

$$\lambda_{avg} = \frac{\sum_{b=1}^{C_b} A_b \lambda_b}{C_b A_{avg}}. \tag{A.2.36}$$

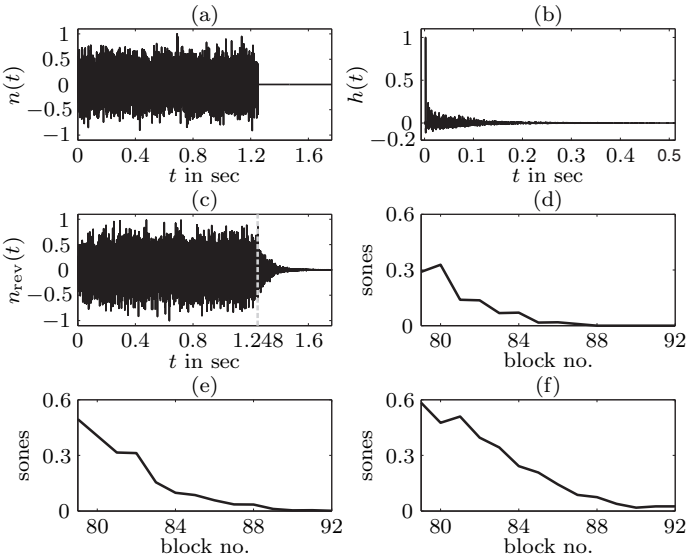

Figure A.14: (a) White noise, (b) RIR ($\tau_{60} = 400$ms), (c) Reverberated noise, (d) Bark spectral difference decay curve at Bark bin no. 7, (e) Bark spectral difference decay curve at Bark bin no. 14, (f) Bark spectral difference decay curve at Bark bin no. 16; $f_s = 8000$ Hz, $L_{\mathrm{Bl}} = \lfloor f_s \cdot 32\,\mathrm{ms} \rfloor$, overlap 50 %.

Figure A.14 illustrates the exponential decay in time-domain as well as in Bark domain. Panel (a) of Figure A.14 shows a white noise signal $n(t)$ which is reverberated by convolving with the RIR $h(t)$ depicted in panel (b)

to result in the reverberant noise $n_{\text{rev}}(t)$ depicted in panel (c). The panels (d) to (f) show the decay of the Bark spectral difference at different Bark bins. The dashed grey line in panel (c) shows the begin of the exponential decay at which the calculation for the Bark spectral difference decay curves according to (A.2.31) starts.

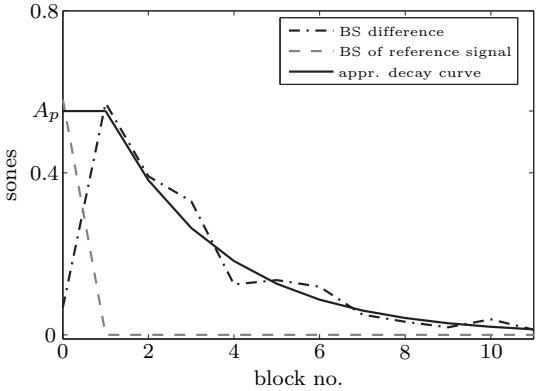

Figure A.15: Curve approximation my means of LS fitting (Bark bin no. 9).

As sown in **Figure A.15**, an exponentially decaying curve (solid black line) can be fitted to the determined curve of the Bark spectral difference (dash-dotted line).
Furthermore, the average direct path energy

$$D_{avg} = \frac{\sum_{b=1}^{C_b} D_b}{C_b} \tag{A.2.37}$$

is calculated at the determined end-points [WN06] which is estimated from the Bark spectrum of the clean reference signal.
The RDT measure is then defined as the the ratio of the amplitude and decay rate of the exponential decays normalized to the amplitude of the direct component calculated using (A.2.35), (A.2.36) and (A.2.37).

$$R_{DT} = \frac{A_{avg}}{\lambda_{avg} D_{avg}} \tag{A.2.38}$$

For realistic speech signals it may happen, that due to a local increase of the speech energy shortly after an end-point has been detected, the decay curve may not have enough time to fully decay (e.g. in very short speech

pauses). Therefore, flat regions are searched for after end-points and decay regions. In the following, the search algorithm for finding the decaying and following flat regions is briefly described.

RDT search algorithm End points which are considered to be the beginning of an decay period in the Bark spectrum difference signals can be found by

$$\Delta\chi[\ell_0, 1, b] > \delta_{\max,1}. \tag{A.2.39}$$

In (A.2.39), the parameter ℓ_0 is the block index of a possible end-point and $\delta_{\max,1} = 0.2$ is a percentage of the maximum of the Bark spectrum which is has been experimentally determined in [WN06]. **Figure A.16** illustrates decays in the Bark spectrum difference according to (A.2.39). Detected end-points are indicated by a black circle in Figure A.16.

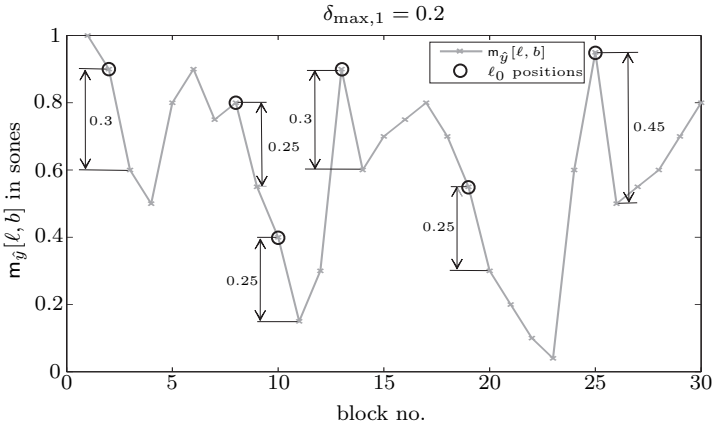

Figure A.16: End-point search, $\delta_{\max,1} = 0.2$.

Next, the the number of blocks of decay I is determined by increasing i as long as

$$\Delta\chi(\ell_0 + i, 1, b) > \delta_{\max,2} \qquad i = 0, 1, ..., I \tag{A.2.40}$$

holds which is the case for $i = I = 3$ in the example depicted in **Figure A.17**, since

$$\Delta\chi[\ell_0 + 3, 1, b] = \mathsf{m}_{\hat{y}}[\ell_0 + 3, b] - \mathsf{m}_{\hat{y}}[\ell_0 + 3 + 1, b] = 0.05 \leq 0.1 \tag{A.2.41}$$

for $\delta_{\max,2} = 0.1$ being a percentage of the maximum of the Bark spectrum difference in each Bark band b, empirically determined in [WN06].
After determining the decaying part in the Bark spectrum, the algorithm searches for flat regions for which ripples in the Bark spectrum difference

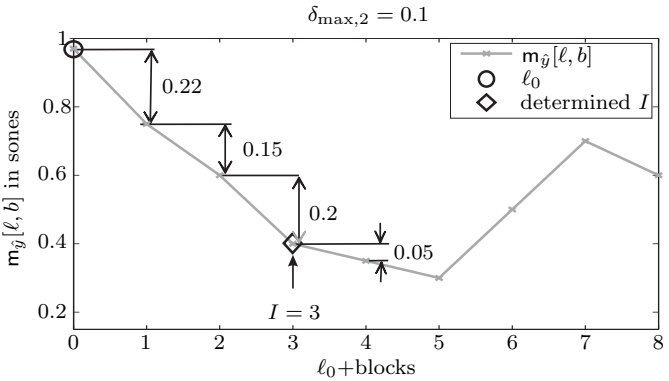

Figure A.17: Determination of parameter I, $\delta_{\max,2} = 0.1$.

are below a threshold δ_{min} and the Bark spectrum is below a threshold δ_t for J frames.

$$\left.\begin{aligned} \Delta\chi[\ell_0 + i + 1, j, b] &< \delta_{min} \\ \mathsf{m}_{\hat{y}}[\ell_0 + i + j, b] &< \delta_t \end{aligned}\right\} \qquad j = 1, 2, ..., J \tag{A.2.42}$$

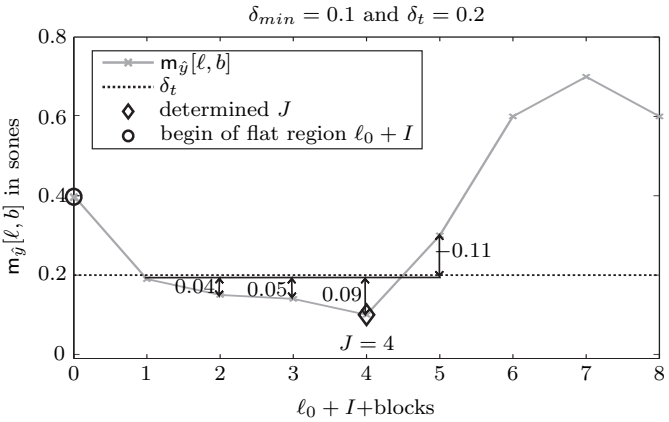

Figure A.18: Determination of the factor J in the RDT search algorithm; $\delta_{min} = 0.1$; $\delta_t = 0.2$.

This is illustrated in **Figure A.18**. Both conditions (A.2.42) are fulfilled till $J = 4$ in Figure A.18. In the search algorithm proposed in [WN06] identified flat regions are restricted to have a length of $J > 4$ blocks, thus the detected flat region in Figure A.18 would bot be considered since it would be too short.

Figure A.19 shows detected flat regions (left panels) and decay regions (right panels) in solid black lines which can be used for fitting the decay model (cf. Figure A.15) and based on that for calculation of the RTD measure (A.2.38).

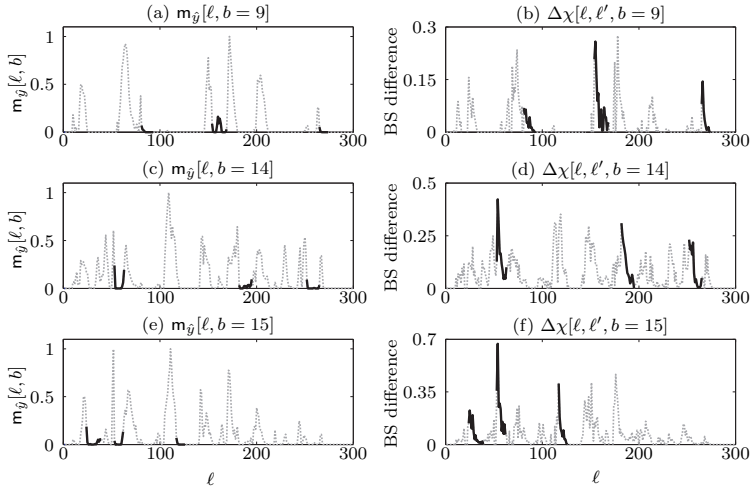

Figure A.19: (a) Bark spectrum (BS) at Bark bin $b = 9$, (b) Bark spectral difference at Bark bin $b = 9$, (c) BS at $b = 14$, (d) BS difference at $b = 14$, (e) BS at $b = 15$, (f) BS difference at $b = 15$; RIR: $\tau_{60} = 400$ ms, LS-EQ: $L_{EQ} = 256$); $f_s = 8000$ Hz.

Objective Measure for Coloration in Reverberation (OMCR)

As already discussed in Section 4.2, the perceived quality of reverberant and dereverberated speech depends on the dimensions reverberation and colouration. While the previously described RDT measure tries to estimate the amount of reverberation from a given signal, the *objective measure for coloration in reverberation (OMCR)* [WN07] focuses on the colouration, i.e. the change in the spectral characteristics of the signal. To archive this, the OMCR algorithm searches for speech-onsets, since it is assumed that colouration can be measured by analysis of the spectra at these onset points. Although the OMCR algorithm is based on *conventional* time- or frequency-domain signals, i.e. does not incorporate psychoacoustic findings, it is listed at this position due to its close relation to the RDT measure described before.

Figure A.20 visualizes the onset-detection in a specific frequency band j. The onsets found by the detection algorithm in the spectrum are marked by vertical dotted lines.

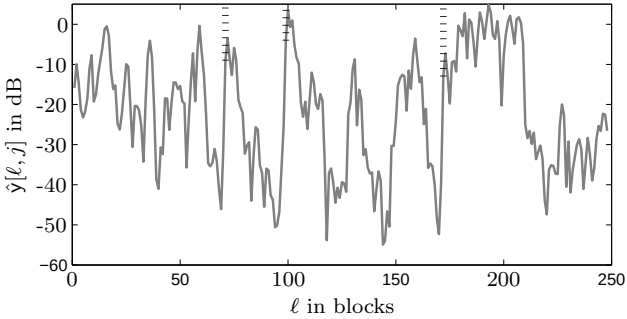

Figure A.20: Example of a speech onset detection (detected speech-onsets are marked by vertical dotted lines.

An onset is detected if the following two criteria are fullfilled [WN07].

$$\hat{\mathsf{y}}[\ell, n] \quad > \quad \max\Big\{\sum_{j=1}^{N_\ell} \hat{\mathsf{y}}[j, n]\Big\} - \gamma_1 \tag{A.2.43}$$

$$\hat{\mathsf{y}}[\ell, n] - \hat{\mathsf{y}}[\ell - i, n] \quad > \quad \gamma_2 - \gamma_3 \cdot (i - 1) \qquad \text{for } i = 1, 2, \ldots, N_\gamma \tag{A.2.44}$$

Here, $\gamma_1 = 18$ dB, $\gamma_2 = 18$ dB and $\gamma_3 = 0.05$ are design parameters of the algorithm empirically determined in [WN07]. For a more detailed description of the parameters (A.2.43) and (A.2.44) the interested reader is referred to [WN07]. From the spectral points at the detected onsets the OMCR measure is calculated by

$$\mathrm{OMCR}[n] = \sum_{\ell=1}^{N_\ell} \mathsf{w}[\ell, n](\mathsf{y}[\ell, n] - \hat{\mathsf{y}}[\ell, n])/M_n, \tag{A.2.45}$$

$$\mathrm{OMCR} = \frac{1}{N_n} \sum_{n=1}^{N_n} \mathrm{OMCR}[n]. \tag{A.2.46}$$

The factor $\mathsf{w}[\ell, n]$ in (A.2.45) equals 1 for the onset frames, and 0 otherwise. M_n is the number of onsets detected for the particular frequency bin n. **Figure A.21** shows calculated OMCR values for reverberant speech

generated by RIRs of different room reverberation time τ_{60} (left panel) and for different equalized impulse responses (LS LRC filter of different lengths L_{EQ}; right panel). It can be seen that the OMCR is in principle capable to assess the colouration effect in both cases.

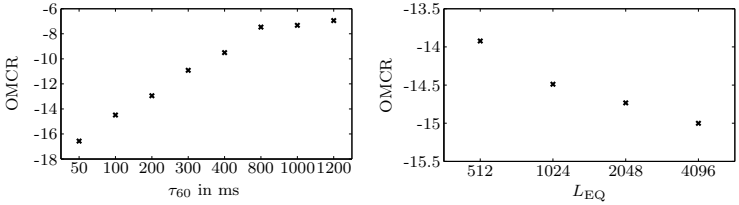

Figure A.21: OMCR values without and with eqalization. (a) OMCR values of a reverberant speech signal over τ_{60}, (b) OMCR values for equalized speech signal over LRC filter lengths L_{EQ} (RIR: $\tau_{60} = 400$ ms, LS-EQ, $d[k]$ high pass); $f_s = 8000$ Hz, $L_{Bl} = f_s \cdot 32$ ms.

Speech-to-Reverberation Modulation Energy Ratio (SRMR)

The SRMR measure [FC08, Fal08, FZC10] is based on the auditory model according to Püschel and Kollmeier and aims at non-intrusively assessing reverberation.

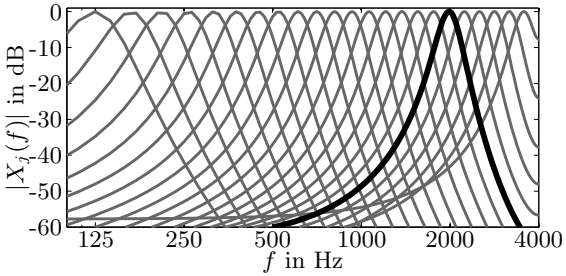

Figure A.22: Gammatone filterbank. $f_s = 8000$ Hz, 23 filters.

After analysis of the signal by a Gammatone filter bank [PAG95, Sla93] (cf. **Figure A.22**) which is designed to match the auditory filters found in the human auditory system, the envelope of the signal in the respective band j,

$$e_j[k] = \sqrt{s_j[k]^2 + \mathcal{H}\left\{s_j[k]\right\}^2}, \tag{A.2.47}$$

is calculated as visualized in **Figure A.23**. The operator $\mathcal{H}\{\cdot\}$ is the Hilbert transform [KK09].

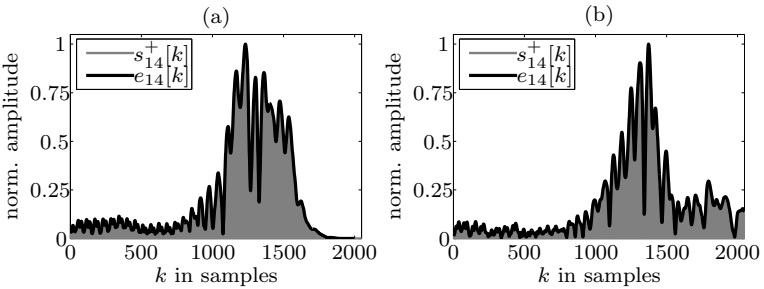

Figure A.23: Envelope $e_j[k]$ of a signal $s_j^+[k]$ in *Gammatone*-band $j = 14$ (center frequency 1.2 kHz); (a) anechoic signal, (b) reverberant signal ($\tau_{60} = 300$ ms).

In the left panel of Figure A.23 the envelope of anechoic signal is depicted while the right panel depicts the envelope of a reverberant signal is shown. It can be seen that the reverberant signal contains more modulation of the envelope. While the modulation of the envelope of the anechoic signal are in the range of $2-20$ Hz, with a maximum at about 4 Hz representing the *syllabic rate of spoken speech* [Fal08], higher modulation energy is introduced by reverberation.
After Gammatone filtering, a second filtering stage by the modulation filter bank [DPK96] which is defined in **Table A.3** and depicted in **Figure A.24** is done.

Modul. band	f_c (in Hz)	B (in Hz)	Modul. band	f_c (in Hz)	B (in Hz)
1	4.0	2.4	5	28.9	18.2
2	6.5	3.9	6	47.5	29.1
3	10.7	6.5	7	78.1	47.6
4	17.6	11.0	8	128.0	78.8

Table A.3: Center frequencies f_c and band widths B of modulation filters [Fal08].

Figure A.25 shows the resulting energy patterns of (a) the anechoic input signal, (b) after reverberation by an room impulse response of reverberation time $\tau_{60} = 300$ ms and (c) after reverberation by an room impulse response of reverberation time $\tau_{60} = 800$ ms. For analysis, the *Gammatone* filter

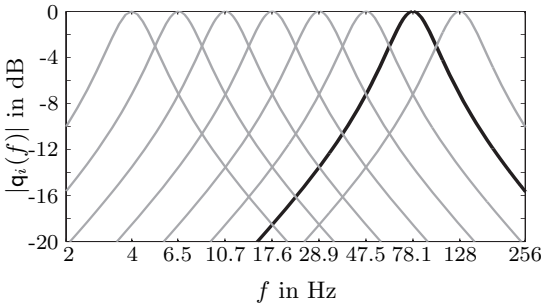

Figure A.24: Frequency responses of modulation filter bank [Fal08].

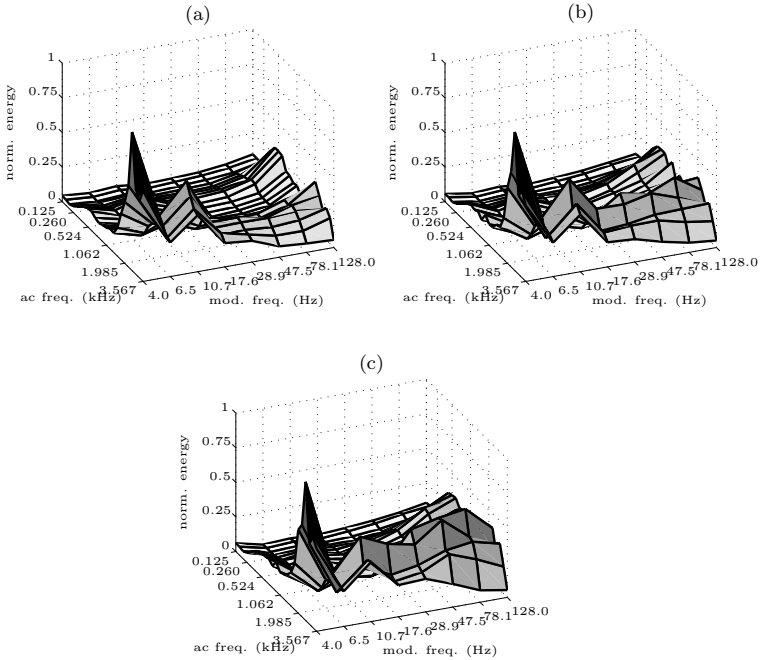

Figure A.25: Signal energy depending on analyzed acoustical frequency and modulation frequency $\bar{\mathsf{e}}_j[m]$ for (a) anechoic signal, (b) reverberant signal (τ_{60} = 300ms), and (c) reverberant signal (τ_{60} = 800ms); $f_s = 8000$ Hz.

bank as shown in Figure A.22 and the modulation filter bank as shown in Figure A.24 have been used. It can be seen from Figure A.25 that for higher reverberation time the energy in higher modulation bands raises. Therefore, the SRMR is defined as the ratio of the modulation energy in the 4 lower modulation frequency bands to the modulation energy in higher modulation frequency bands,

$$\mathrm{SRMR} = \frac{\sum_{m=1}^{4} \bar{\mathsf{e}}[m]}{\sum_{m=5}^{M} \bar{\mathsf{e}}[m]}, \tag{A.2.48}$$

with the energy

$$\bar{\mathsf{e}}_j[m] = \frac{1}{N_{act}} \sum_{i=1}^{N_{act}} \mathsf{e}_j[i,m] \tag{A.2.49}$$

for all N_{act} blocks in which speech activity has been detected and the mean energy for all 23 bands

$$\bar{\mathsf{e}}[m] = \frac{1}{23} \sum_{j=1}^{23} \bar{\mathsf{e}}_j[m]. \tag{A.2.50}$$

Perceptual Evaluation of Speech Quality (PESQ)

From quality assessment in the field of audio coding or noise reduction it is known that measures that are based on more exact models of the human auditory system show high correlation with subjective data. The *perceptual evaluation of speech quality (PESQ)* measure is a standardized measure [ITU01] that is visualized in **Figure A.26**. The detailed description of PESQ is beyond the scope of this thesis and the interested reader is referred to the literature [ITU01, Loi07].

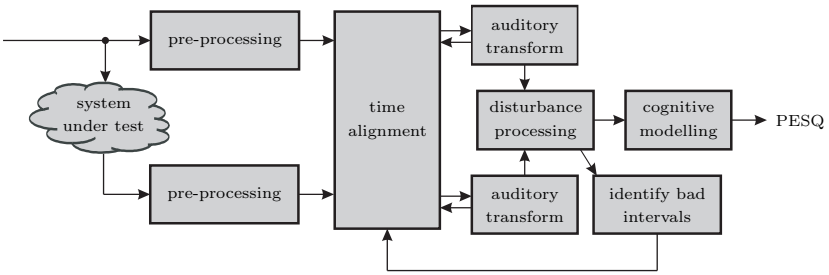

Figure A.26: Schematic of PESQ quality measure (adapted from [Loi07]).

It should be notes that the implementation of [Loi07] and the *original* implementation according to [ITU01] differ in their absolute values. However, at least for the tests performed for this work they result in the same trends.

Perceptual Similarity Measure (PSM)

A further objective quality measure originally developed to assess the quality of audio codecs is the *perceptual similarity measure* (PSM) from PEMO-Q [HK06] that uses the auditory model according to [DPK96] depicted in **Figure A.27**. The basilar membrane filtering stage and the modulation filtering stage in Figure A.27 use the same filters as used for the SRMR measure (cf. Figure A.24 and Figure A.25).

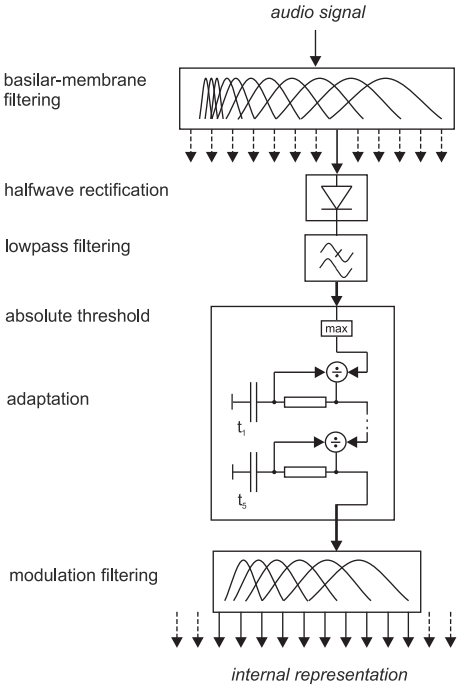

Figure A.27: Schematic of auditory model according to Dau and Püschel [DPK96].

The calculation of the PSM measure is visualized in **Figure A.28**. In a preprocesing step so-called internal representations of the reference audio signal as well as the processed audio signal are calculated by the hearing model as depicted in Figure A.27. After analysis by Gammatone filter

banks (cf. also Figure A.24) which simulate the basilar membrane filtering in the auditory system, half-wave rectifiers and low-pass filters simulate the conversion from mechanical oscillations in nerve signals in the inner ear. Then, the modulation filter bank calculates modulation frequencies (cf. also Figure A.25) to obtain the internal representations. In a post-processing step depicted in Figure A.28 below the hearing models the cross correlation is calculated as a distance measure which is used as the block-by-block quality indicator and which is combined with the current loudness of the signals. In a further processing step the internal representations are analysed by the 5% percentile which focuses on short term variations of the signals since these are perceptually relevant in the human auditory system. For more details on the PSM measure, the reader is referred to [Hub03, HK06].

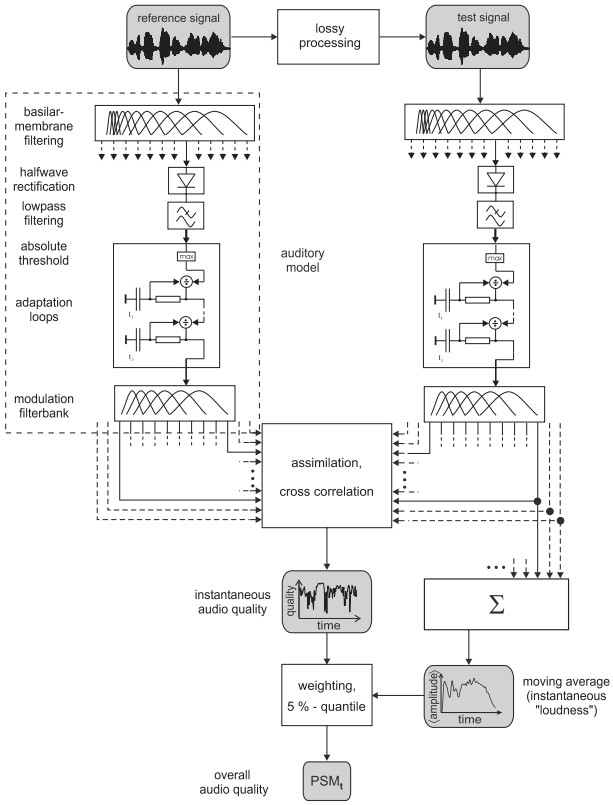

Figure A.28: Perceptual similarity measure (PSM); (Source: [Hub03]).

Appendix B

Details of Subjective Listening Tests

This appendix presents details of the subjective listening test described in Section 4.2.1. **Table B.1** summarizes some properties of the selected 21 sound samples. It shows the room reverberation time of the RIR to be equalized, the LRC filter type (cf. Sections 4.3 to 4.7 for details) and the LRC filter length L_{EQ} as well as the gender of the respective speaker.
Figures B.1 to **B.21** show systems' impulse responses and transfer functions in panels (a) and (b), respectively. Panels (c) show the subjective rating in terms of the mean opinion score (MOS) for the four attributes *reverberated*, *coloured / distorted*, *distant* and *overall quality*.

sample no.	τ_{60} of RIR	LRC filter type	L_{EQ}	gender of speaker
1	950 ms	WLS-EQ	2048	male
2	950 ms	ISwPP	4096	female
3	500 ms	LS-EQ	2048	male
4	950 ms	WLS-EQ	8192	male
5	500 ms	ISwPP	1024	male
6	500 ms	WLS-EQ	4096	male
7	950 ms	WLS-EQ	4096	female
8	500 ms	ISwPP	8192	female
9	950 ms	LS-EQ	8192	female
10	500 ms	ISwINO	4000	male
11	500 ms	WLS-EQ	1024	male
12	500 ms	LS-EQ	1024	female
13	950 ms	LS-EQ	1024	female
14	500 ms	ISwPP	4096	male
15	500 ms	WLS-EQ	8192	male
16	950 ms	LS-EQ	4096	male
17	950 ms	LS-EQ	2048	male
18	500 ms	ISwPP	2048	female
19	500 ms	LS-EQ	4096	male
20	500 ms	LS-EQ	8192	male
21	950 ms	ISwPP	1024	male

Table B.1: Properties of sound samples used for the subjective listening test in Section 4.2.1.

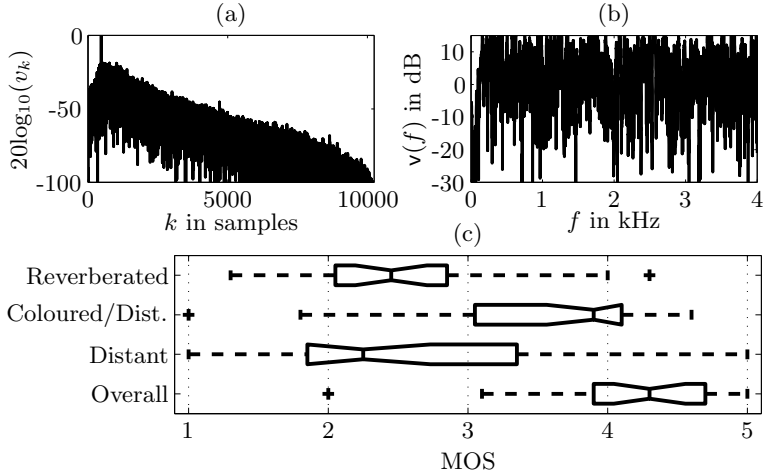

Figure B.1: Subjective assessment of audio sample no. 1 (WLS-EQ). (a) and (b) system used for generation of audio sample no. 1 in time- and frequency-domain. (c) results of subjective quality assessment.

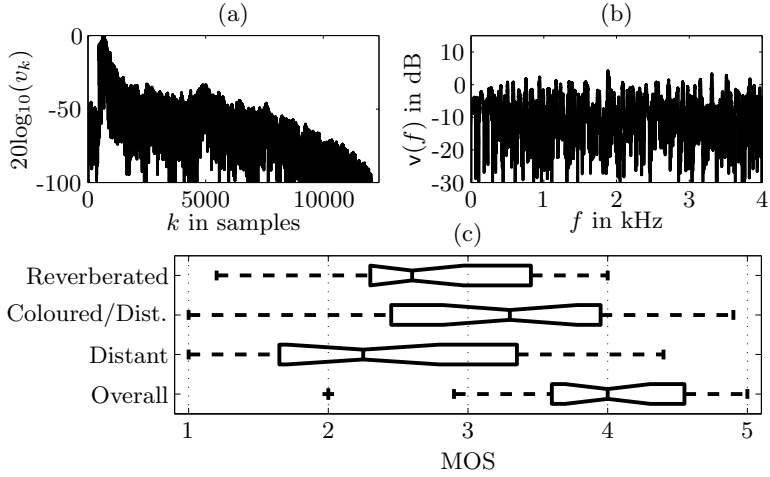

Figure B.2: Subjective assessment of audio sample no. 2 (ISwPP). (a) and (b) system used for generation of audio sample no. 2 in time- and frequency-domain. (c) results of subjective quality assessment.

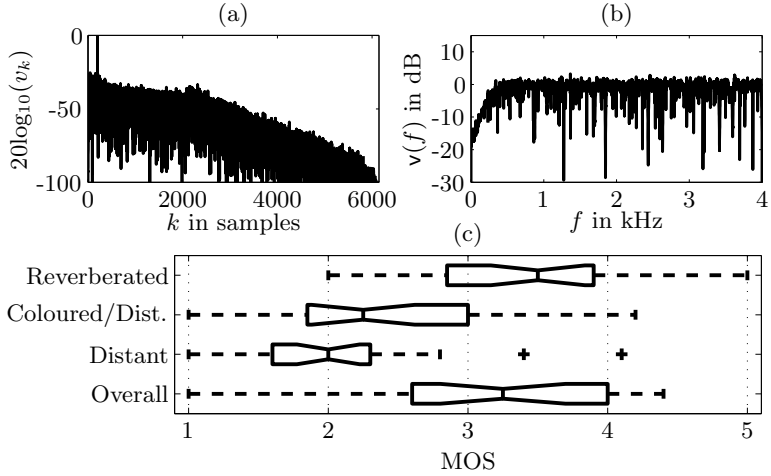

Figure B.3: Subjective assessment of audio sample no. 3 (LS-EQ). (a) and (b) system used for generation of audio sample no. 3 in time- and frequency-domain. (c) results of subjective quality assessment.

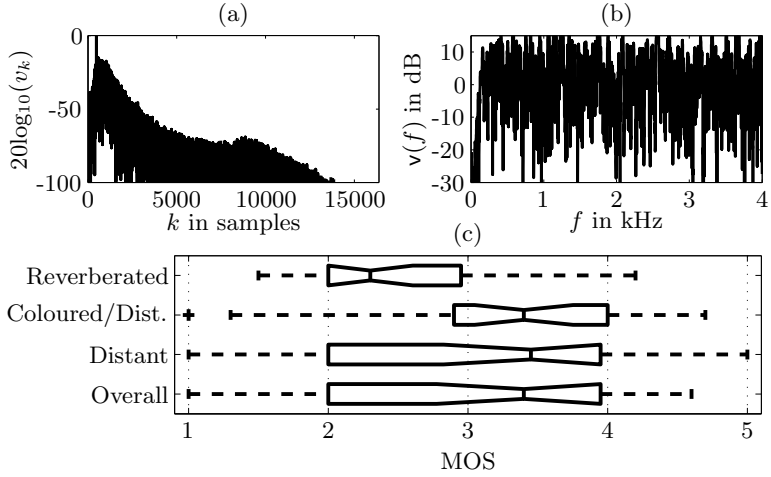

Figure B.4: Subjective assessment of audio sample no. 4 (WLS-EQ). (a) and (b) system used for generation of audio sample no. 4 in time- and frequency-domain. (c) results of subjective quality assessment.

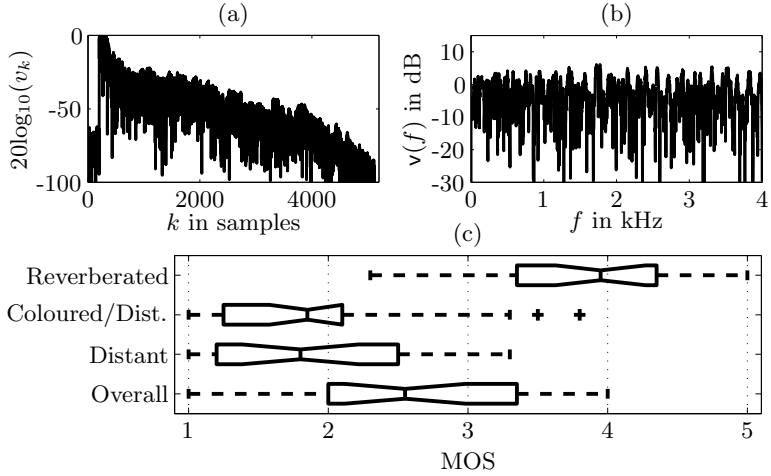

Figure B.5: Subjective assessment of audio sample no. 5 (ISwPP). (a) and (b) system used for generation of audio sample no. 5 in time- and frequency-domain. (c) results of subjective quality assessment.

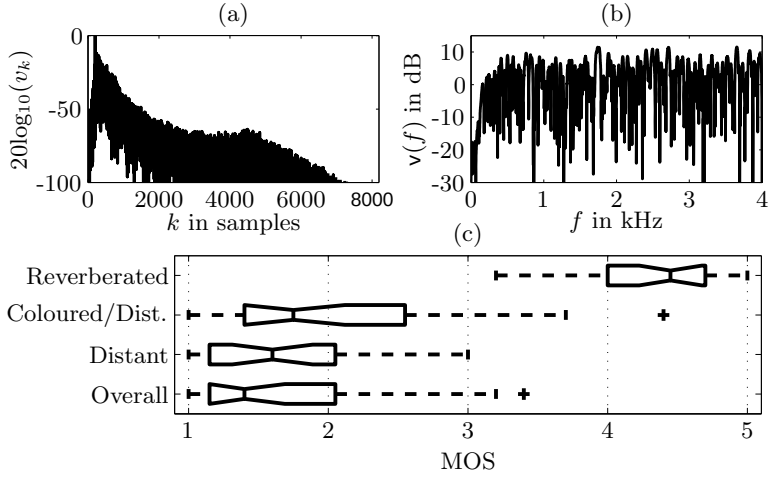

Figure B.6: Subjective assessment of audio sample no. 6 (WLS-EQ). (a) and (b) system used for generation of audio sample no. 6 in time- and frequency-domain. (c) results of subjective quality assessment.

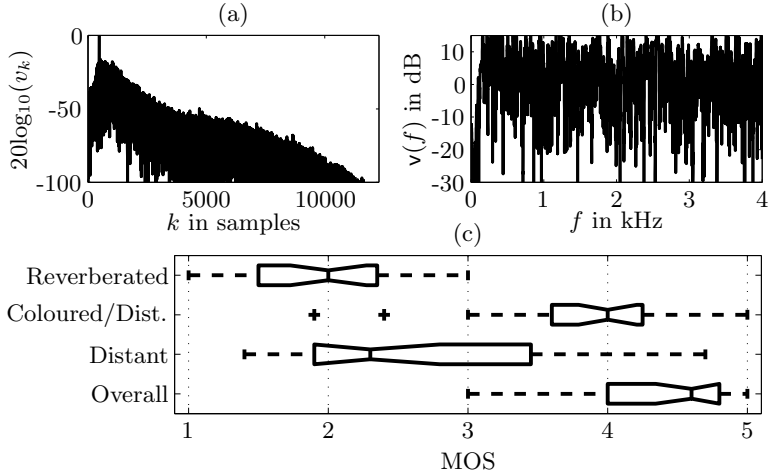

Figure B.7: Subjective assessment of audio sample no. 7 (WLS-EQ). (a) and (b) system used for generation of audio sample no. 7 in time- and frequency-domain. (c) results of subjective quality assessment.

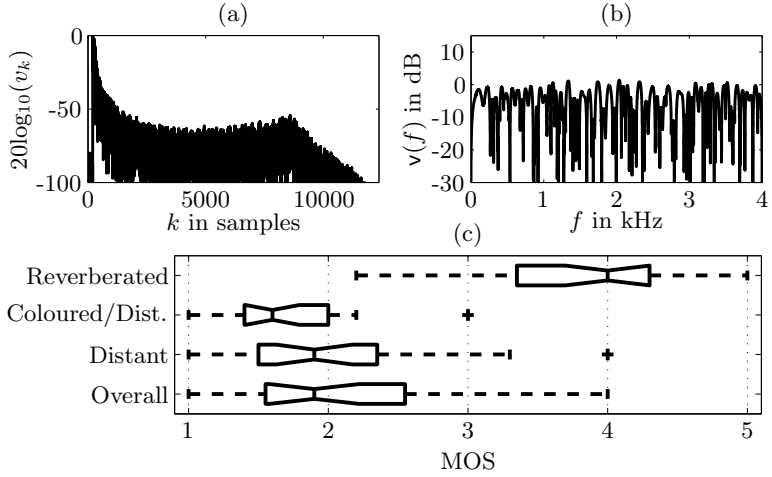

Figure B.8: Subjective assessment of audio sample no. 8 (ISwPP). (a) and (b) system used for generation of audio sample no. 8 in time- and frequency-domain. (c) results of subjective quality assessment.

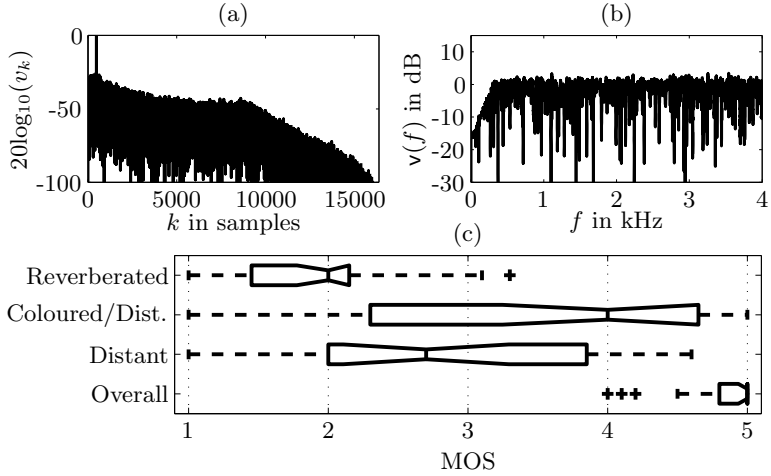

Figure B.9: Subjective assessment of audio sample no. 9 (LS-EQ). (a) and (b) system used for generation of audio sample no. 9 in time- and frequency-domain. (c) results of subjective quality assessment.

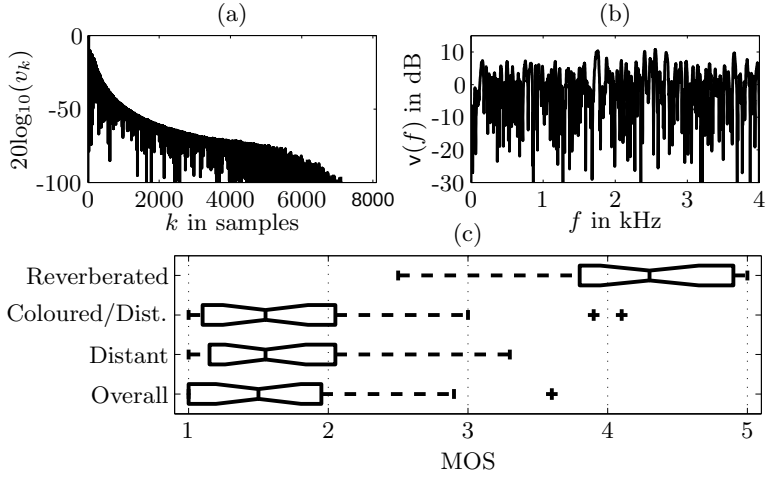

Figure B.10: Subjective assessment of audio sample no. 10 (ISwINO). (a) and (b) system used for generation of audio sample no. 10 in time- and frequency-domain. (c) results of subjective quality assessment.

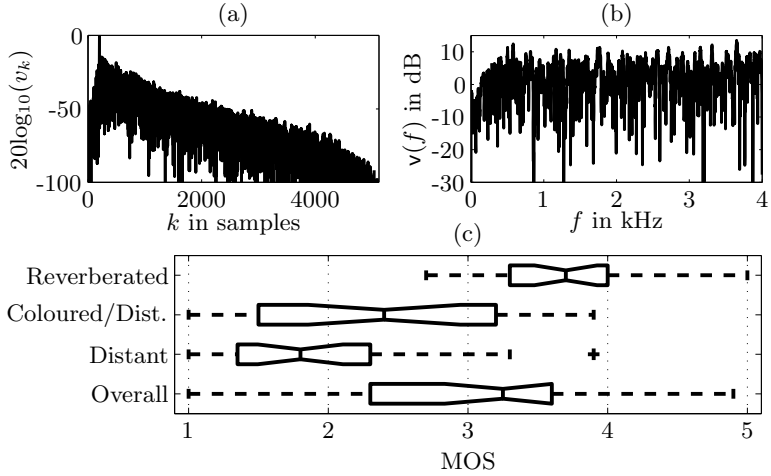

Figure B.11: Subjective assessment of audio sample no. 11 (WLS-EQ). (a) and (b) system used for generation of audio sample no. 11 in time- and frequency-domain. (c) results of subjective quality assessment.

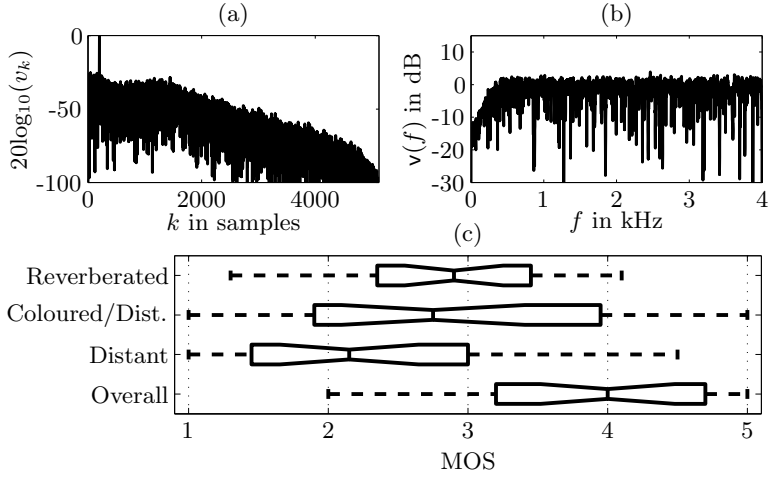

Figure B.12: Subjective assessment of audio sample no. 12 (LS-EQ). (a) and (b) system used for generation of audio sample no. 12 in time- and frequency-domain. (c) results of subjective quality assessment.

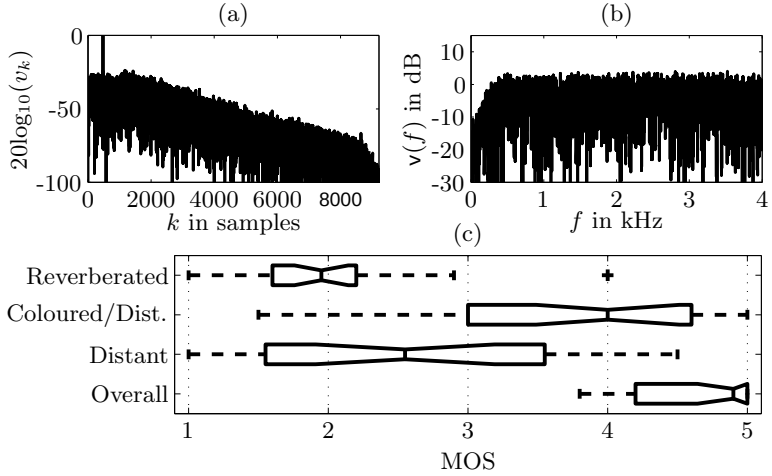

Figure B.13: Subjective assessment of audio sample no. 13 (LS-EQ). (a) and (b) system used for generation of audio sample no. 13 in time- and frequency-domain. (c) results of subjective quality assessment.

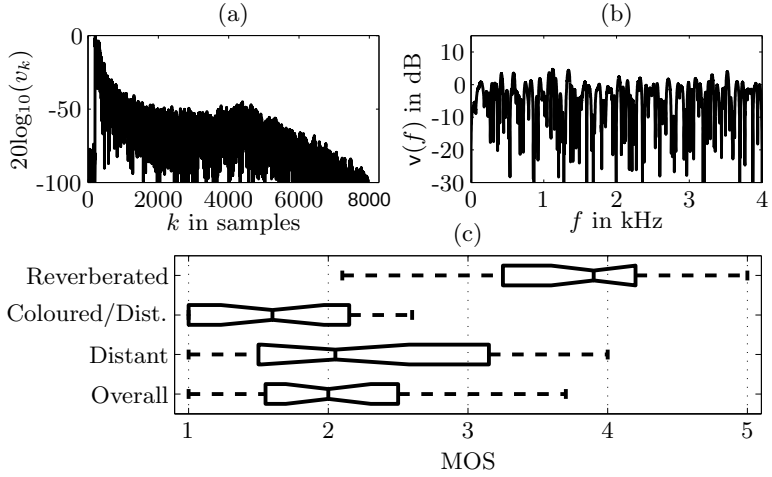

Figure B.14: Subjective assessment of audio sample no. 14 (ISwPP). (a) and (b) system used for generation of audio sample no. 14 in time- and frequency-domain. (c) results of subjective quality assessment.

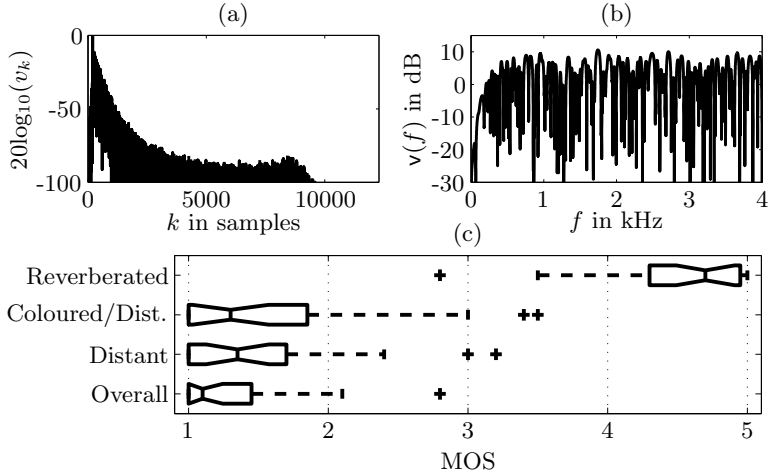

Figure B.15: Subjective assessment of audio sample no. 15 (WLS-EQ). (a) and (b) system used for generation of audio sample no. 15 in time- and frequency-domain. (c) results of subjective quality assessment.

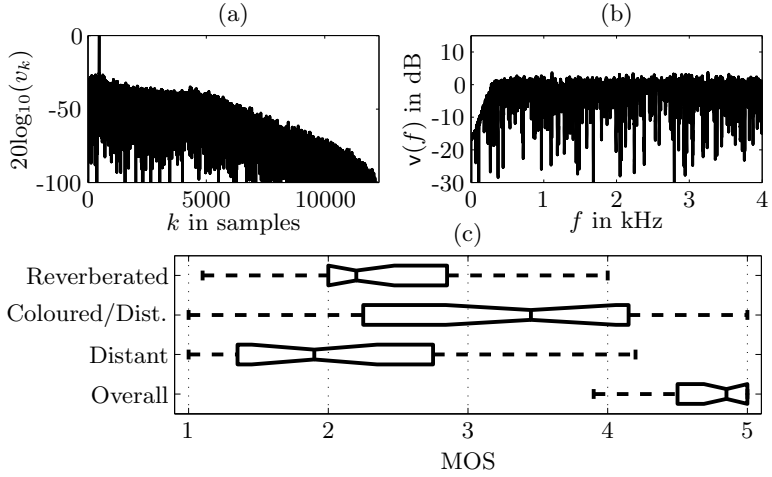

Figure B.16: Subjective assessment of audio sample no. 16 (LS-EQ). (a) and (b) system used for generation of audio sample no. 16 in time- and frequency-domain. (c) results of subjective quality assessment.

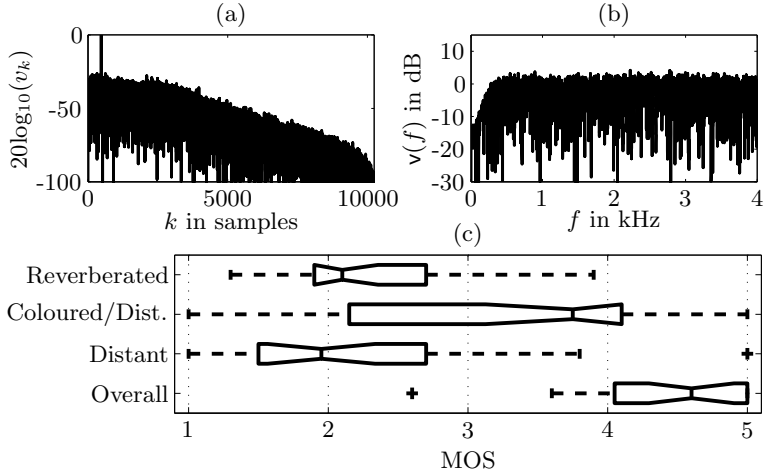

Figure B.17: Subjective assessment of audio sample no. 17 (LS-EQ). (a) and (b) system used for generation of audio sample no. 17 in time- and frequency-domain. (c) results of subjective quality assessment.

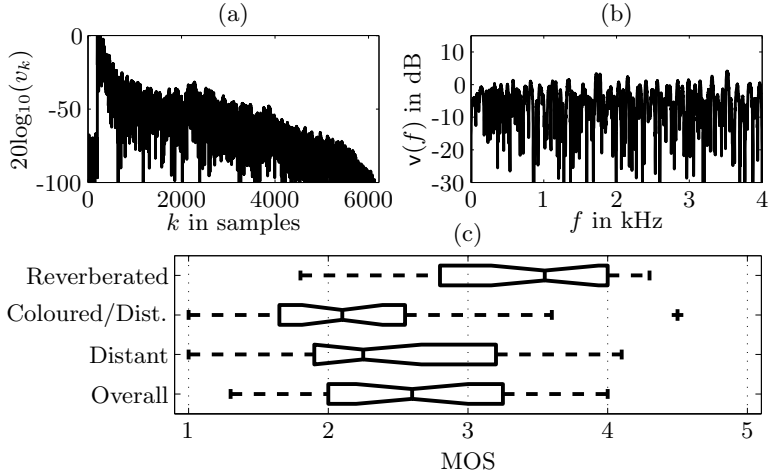

Figure B.18: Subjective assessment of audio sample no. 18 (ISwPP). (a) and (b) system used for generation of audio sample no. 18 in time- and frequency-domain. (c) results of subjective quality assessment.

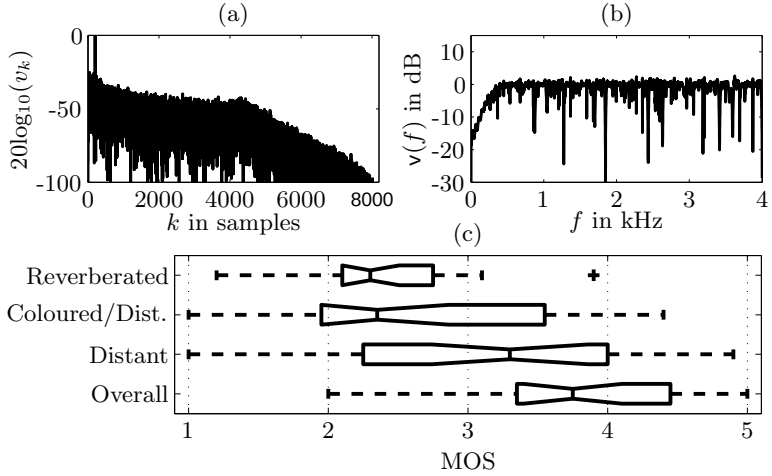

Figure B.19: Subjective assessment of audio sample no. 19 (LS-EQ). (a) and (b) system used for generation of audio sample no. 19 in time- and frequency-domain. (c) results of subjective quality assessment.

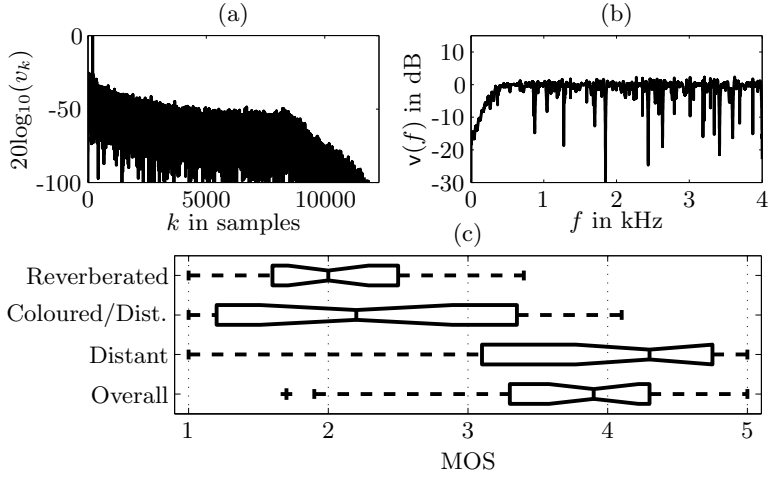

Figure B.20: Subjective assessment of audio sample no. 20 (LS-EQ). (a) and (b) system used for generation of audio sample no. 20 in time- and frequency-domain. (c) results of subjective quality assessment.

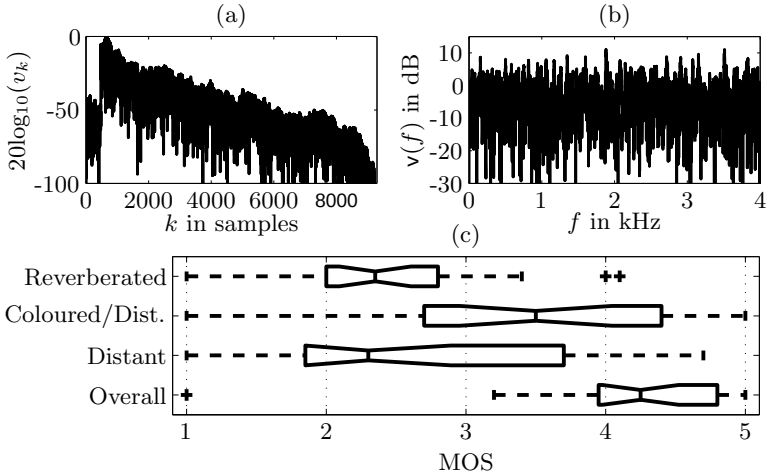

Figure B.21: Subjective assessment of audio sample no. 21 (ISwPP). (a) and (b) system used for generation of audio sample no. 21 in time- and frequency-domain. (c) results of subjective quality assessment.

Appendix C

Correlations between Objective and Subjective Quality Assessment

This appendix visualizes correlations between objective and subjective quality assessment that are discussed in Section 4.2.2.

Figures C.1 to C.22 show the subjective ratings of the human listeners in terms of MOS and the respective objective quality measure. The correlation coefficient r (cf. Tables 4.5 to 4.6 on page 77 ff.) is given for the overall correlation (r_{all}) as well as for the single LRC approaches (r_{LS}, r_{WLS}, r_{IS}) in each figure for each of the four attributes evaluated (reverberant, coloured/distorted, distant and overall quality). For more details about the quality measures please cf. Tables 4.1 and 4.2 on page 73 as well as Appendix A for a detailed description of the objective quality measures. For details on the subjective listening tests please refer to Section 4.2.1 and for more information about the correlation analysis, please refer to Section 4.2.2 and Section 4.2 in general.

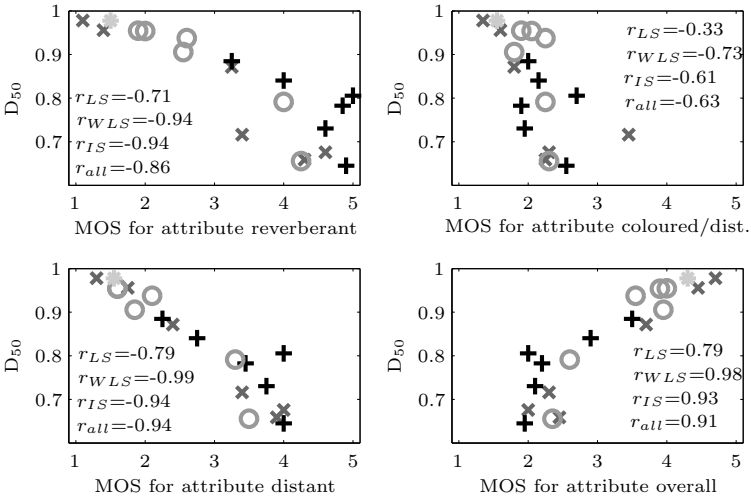

Figure C.1: Correlation analysis between D_{50} measure and subjective assessment. + LS-EQ, × WLS-EQ, ∘ ISwPP, ∗ ISwIN.

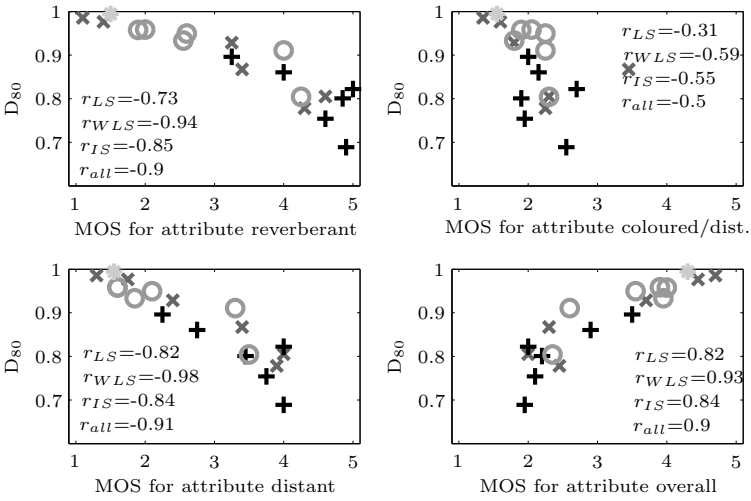

Figure C.2: Correlation analysis between D_{80} measure and subjective assessment. + LS-EQ, × WLS-EQ, ∘ ISwPP, ∗ ISwIN.

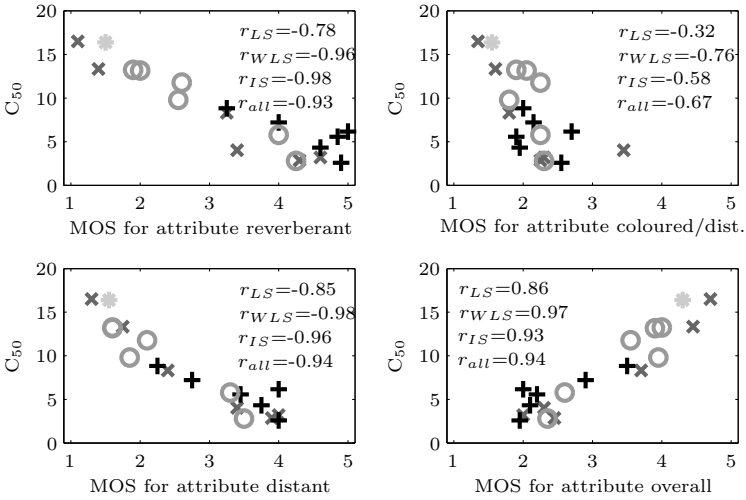

Figure C.3: Correlation analysis between C_{50} measure and subjective assessment. + LS-EQ, × WLS-EQ, ∘ ISwPP, ∗ ISwIN.

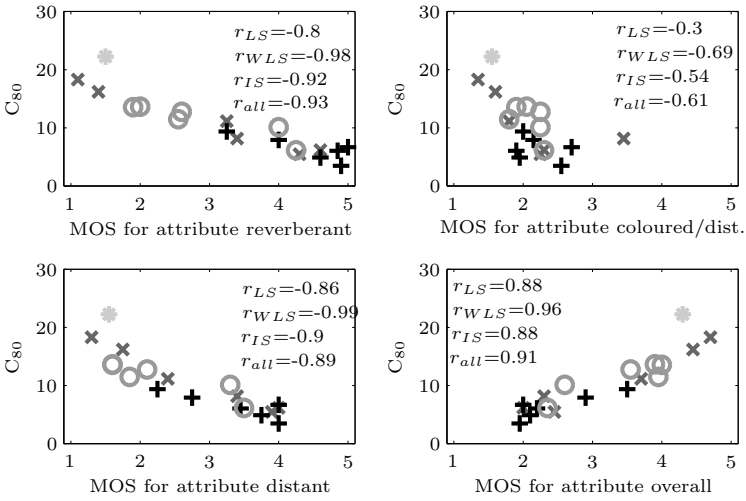

Figure C.4: Correlation analysis between C_{80} measure and subjective assessment. + LS-EQ, × WLS-EQ, ∘ ISwPP, ∗ ISwIN.

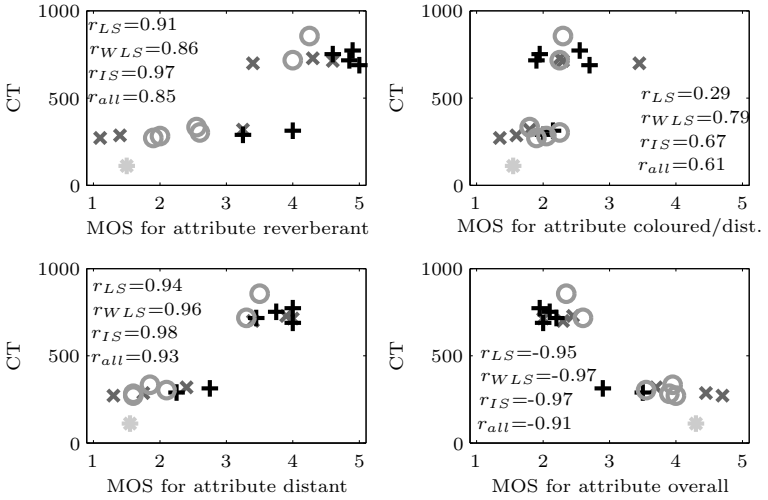

Figure C.5: Correlation analysis between CT measure and subjective assessment. + LS-EQ, × WLS-EQ, ∘ ISwPP, ∗ ISwIN.

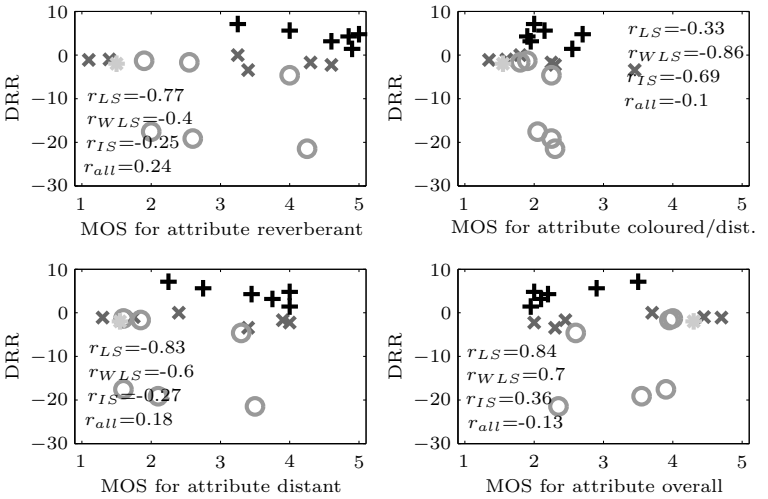

Figure C.6: Correlation analysis between DRR measure and subjective assessment. + LS-EQ, × WLS-EQ, ∘ ISwPP, ∗ ISwIN.

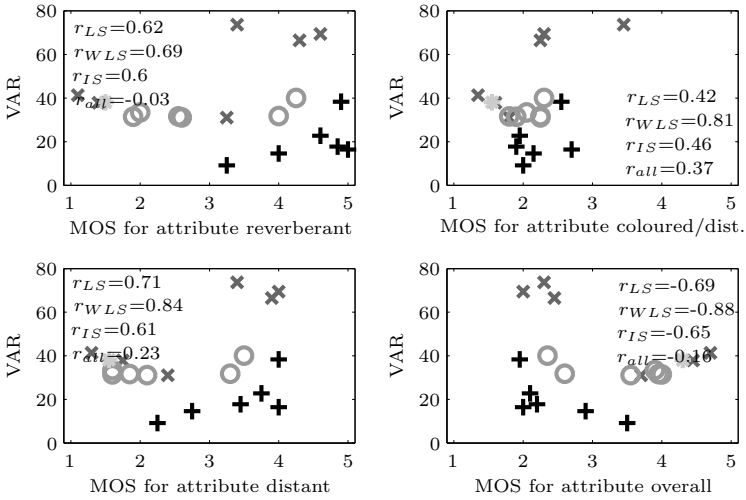

Figure C.7: Correlation analysis between VAR measure and subjective assessment. + LS-EQ, × WLS-EQ, ∘ ISwPP, ∗ ISwIN.

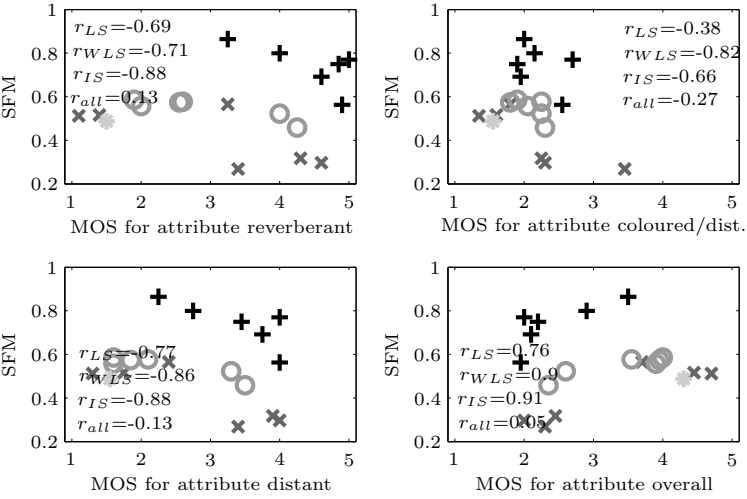

Figure C.8: Correlation analysis between SFM measure and subjective assessment. + LS-EQ, × WLS-EQ, ∘ ISwPP, ∗ ISwIN.

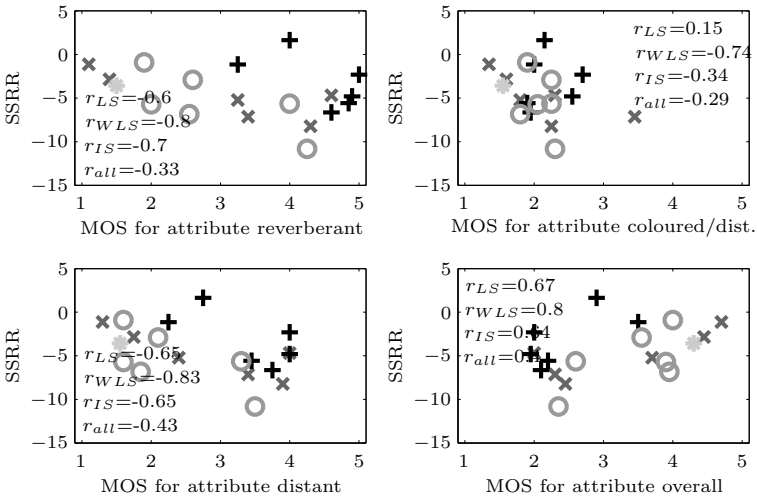

Figure C.9: Correlation analysis between SSRR measure and subjective assessment. + LS-EQ, × WLS-EQ, ○ ISwPP, ∗ ISwIN.

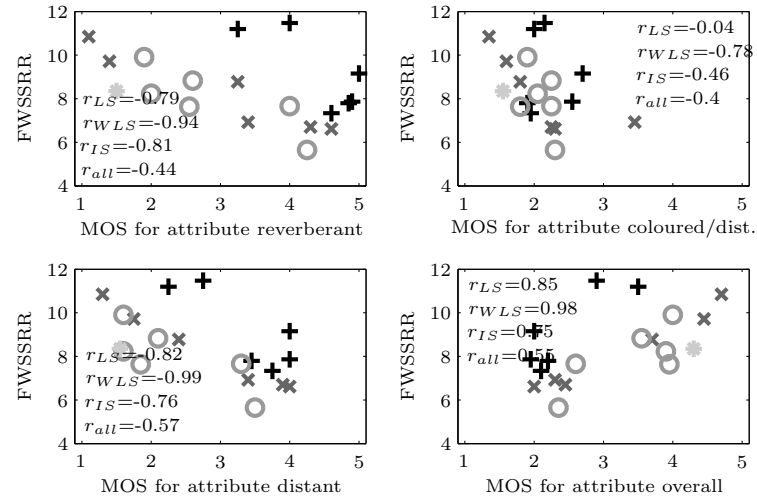

Figure C.10: Correlation analysis between FWSSRR measure and subjective assessment. + LS-EQ, × WLS-EQ, ○ ISwPP, ∗ ISwIN.

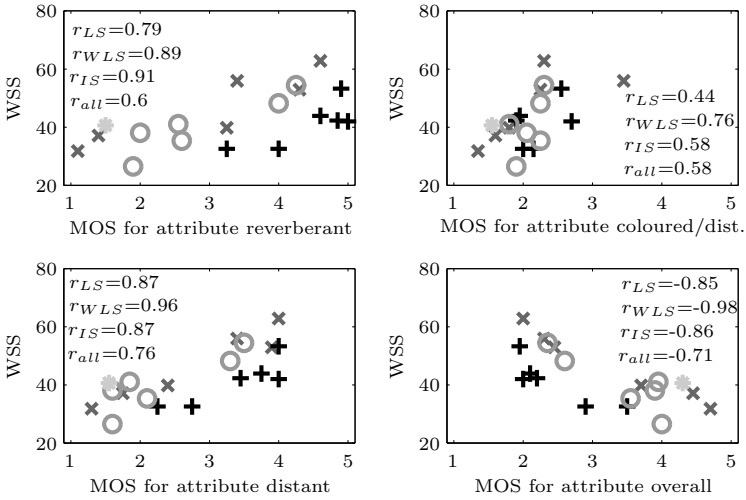

Figure C.11: Correlation analysis between WSS measure and subjective assessment. + LS-EQ, × WLS-EQ, ∘ ISwPP, ∗ ISwIN.

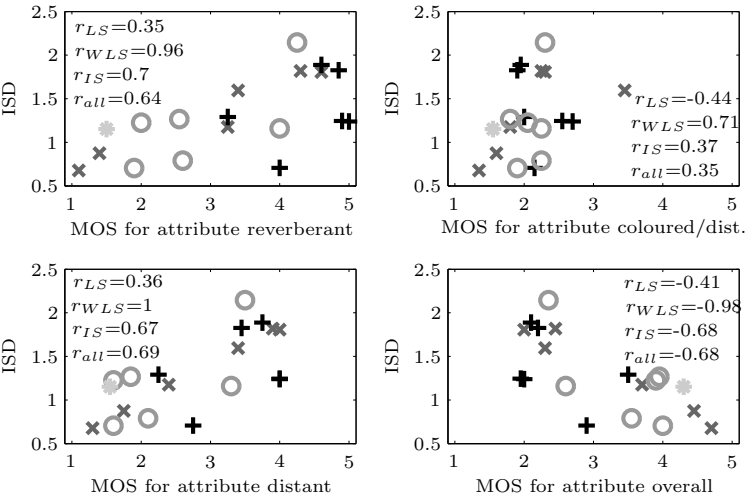

Figure C.12: Correlation analysis between ISD measure and subjective assessment. + LS-EQ, × WLS-EQ, ∘ ISwPP, ∗ ISwIN.

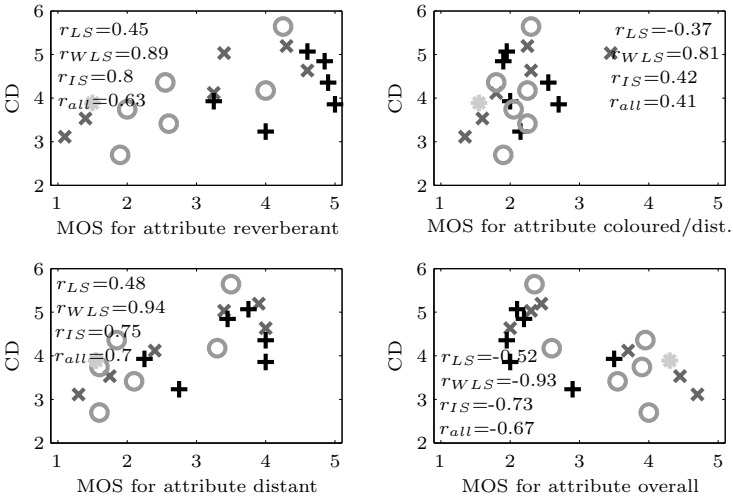

Figure C.13: Correlation analysis between CD measure and subjective assessment. + LS-EQ, × WLS-EQ, ○ ISwPP, ∗ ISwIN.

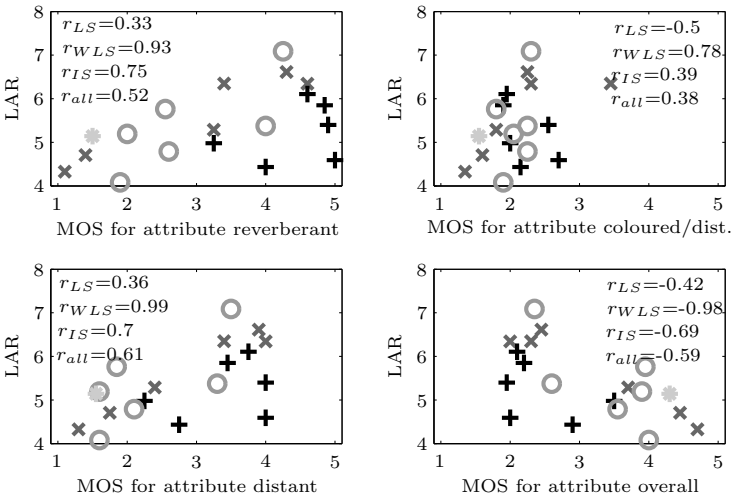

Figure C.14: Correlation analysis between LAR measure and subjective assessment. + LS-EQ, × WLS-EQ, ○ ISwPP, ∗ ISwIN.

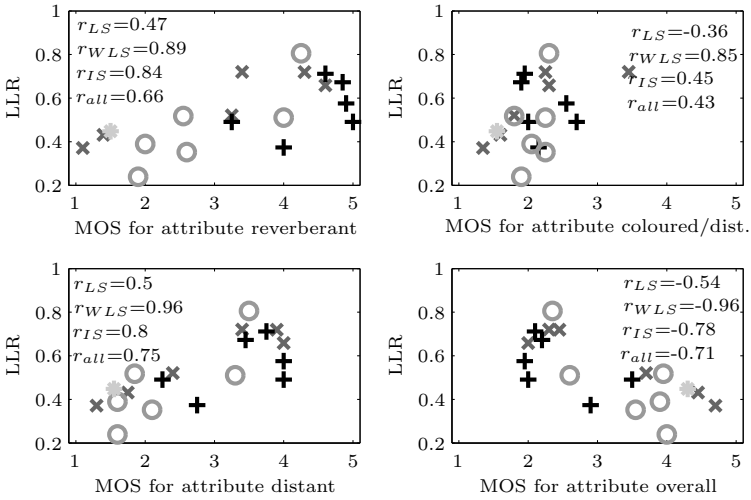

Figure C.15: Correlation analysis between LLR measure and subjective assessment. + LS-EQ, × WLS-EQ, ∘ ISwPP, ∗ ISwIN.

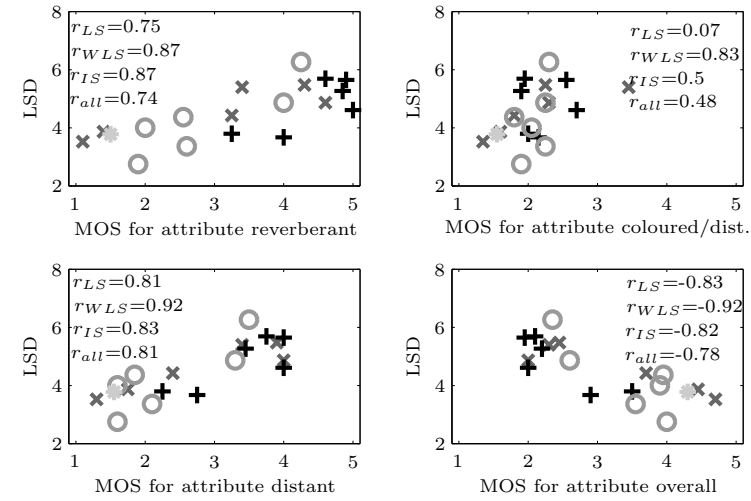

Figure C.16: Correlation analysis between LSD measure and subjective assessment. + LS-EQ, × WLS-EQ, ∘ ISwPP, ∗ ISwIN.

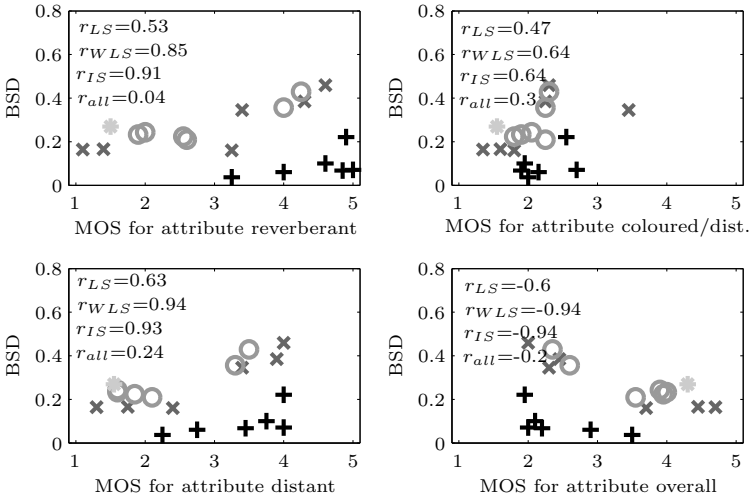

Figure C.17: Correlation analysis between BSD measure and subjective assessment. + LS-EQ, × WLS-EQ, ○ ISwPP, ∗ ISwIN.

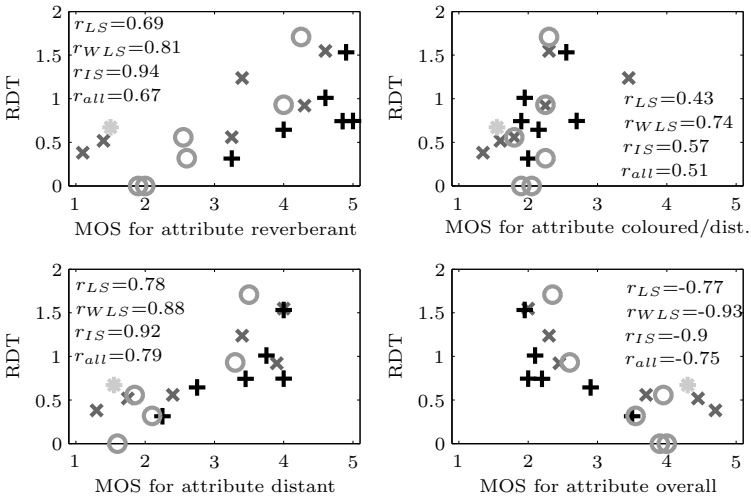

Figure C.18: Correlation analysis between RDT measure and subjective assessment. + LS-EQ, × WLS-EQ, ○ ISwPP, ∗ ISwIN.

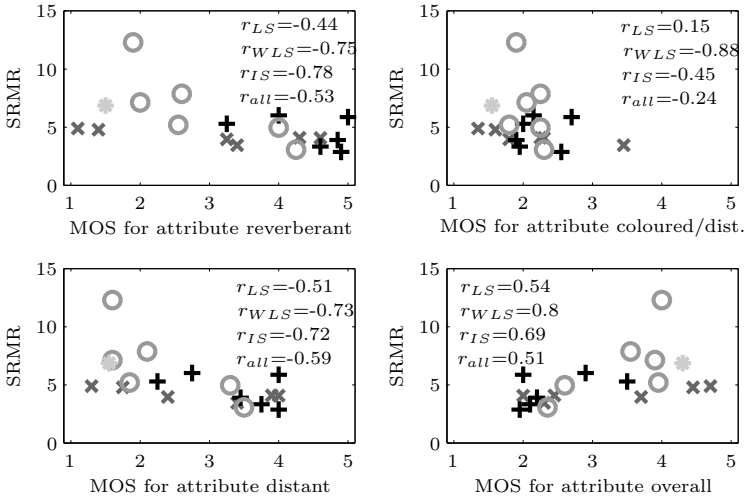

Figure C.19: Correlation analysis between SRMR measure and subjective assessment. + LS-EQ, × WLS-EQ, ∘ ISwPP, ∗ ISwIN.

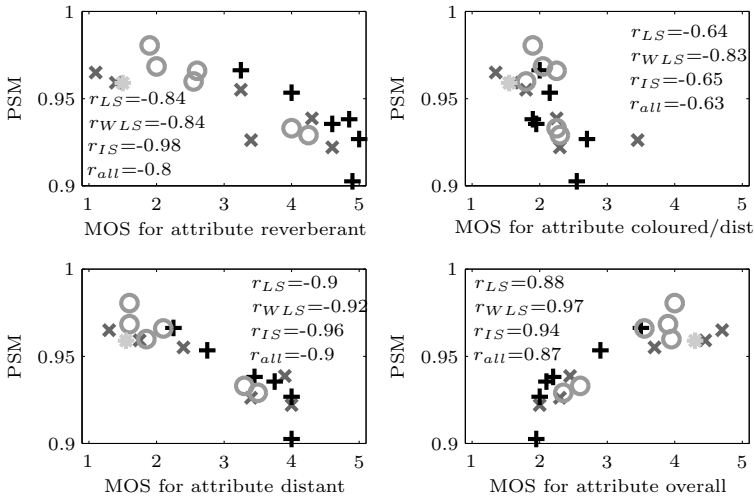

Figure C.20: Correlation analysis between PSM measure and subjective assessment. + LS-EQ, × WLS-EQ, ∘ ISwPP, ∗ ISwIN.

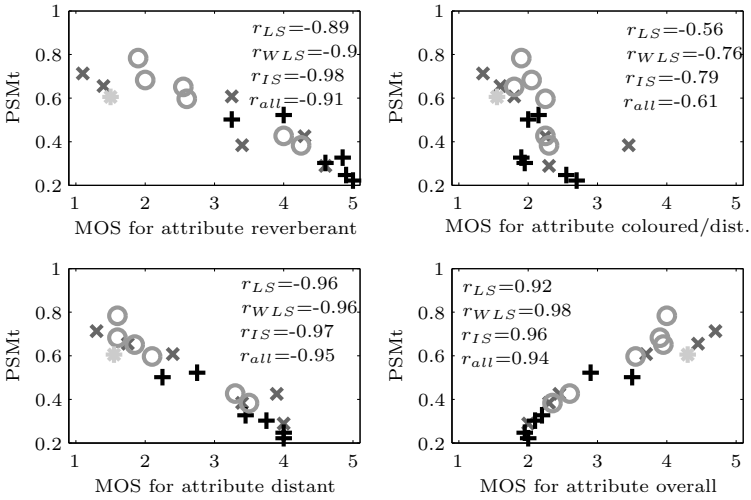

Figure C.21: Correlation analysis between PSMt measure and subjective assessment. + LS-EQ, × WLS-EQ, ∘ ISwPP, ∗ ISwIN.

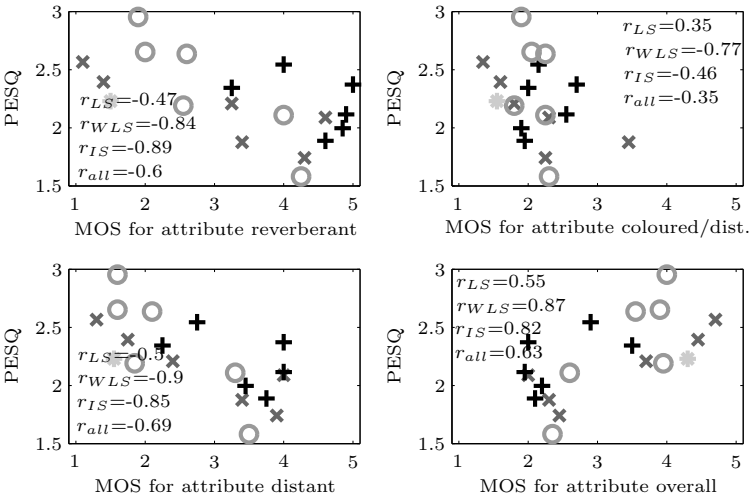

Figure C.22: Correlation analysis between PESQ measure and subjective assessment. + LS-EQ, × WLS-EQ, ∘ ISwPP, ∗ ISwIN.

Appendix D

Mathematical Proofs and Details

D.1 Proof of $\boldsymbol{G}^H\boldsymbol{G} = \boldsymbol{G}$

This section contains the proof of $\boldsymbol{G}^H\boldsymbol{G} = \boldsymbol{G}$ which is needed for the derivations in (3.3.23) and (4.5.39) on pages 55 and (4.5.39). Using the definition of $\boldsymbol{G}$ in (2.2.21) on page 23 $\boldsymbol{G}^H\boldsymbol{G}$ can be written as

$$\begin{aligned}\boldsymbol{G}^H\boldsymbol{G} = &\left(\boldsymbol{F}_{2L\times 2L}\boldsymbol{W}^{01}_{2L\times L}\boldsymbol{W}^{01}_{L\times 2L}\boldsymbol{F}^{-1}_{2L\times 2L}\right)^H\cdot\\ &\cdot\boldsymbol{F}_{2L\times 2L}\boldsymbol{W}^{01}_{2L\times L}\boldsymbol{W}^{01}_{L\times 2L}\boldsymbol{F}^{-1}_{2L\times 2L}.\end{aligned}$$

With the definition of the inverse DFT matrix $\boldsymbol{F}^{-1}_{2L\times 2L} = \boldsymbol{F}^*_{2L\times 2L}/(2L)$ and $\boldsymbol{F}^T_{2L\times 2L} = \boldsymbol{F}_{2L\times 2L}$ we obtain

$$\begin{aligned}\boldsymbol{G}^H\boldsymbol{G} =&\frac{(\boldsymbol{F}^*_{2L\times 2L})^H}{2L}\boldsymbol{W}^{01}_{2L\times L}\boldsymbol{W}^{01}_{L\times 2L}\underbrace{\boldsymbol{F}^H_{2L\times 2L}\boldsymbol{F}_{2L\times 2L}}_{2L\boldsymbol{I}_{2L\times 2L}}\boldsymbol{W}^{01}_{2L\times L}\cdot\\ &\cdot\boldsymbol{W}^{01}_{L\times 2L}\boldsymbol{F}^{-1}_{2L\times 2L}\\ =&\frac{2L}{2L}\boldsymbol{F}^T_{2L\times 2L}\boldsymbol{W}^{01}_{2L\times L}\underbrace{\boldsymbol{W}^{01}_{L\times 2L}\boldsymbol{W}^{01}_{2L\times L}}_{\boldsymbol{I}_{L\times L}}\boldsymbol{W}^{01}_{L\times 2L}\boldsymbol{F}^{-1}_{2L\times 2L}\\ =&\boldsymbol{F}_{2L\times 2L}\boldsymbol{W}^{01}_{2L\times L}\boldsymbol{W}^{01}_{L\times 2L}\boldsymbol{F}^{-1}_{2L\times 2L}\\ =&\boldsymbol{G}, \qquad \text{q.e.d.}\end{aligned}$$

D.2 Proof of $\boldsymbol{G}^H \mathbf{e}_{\mathrm{AEC}}[\ell] = \mathbf{e}_{\mathrm{AEC}}[\ell]$ and $\boldsymbol{G}^H \hat{\mathbf{Y}}[\ell] = \hat{\mathbf{Y}}[\ell]$

This section contains the proof of $\boldsymbol{G}^H \mathbf{e}_{\mathrm{AEC}}[\ell] = \mathbf{e}_{\mathrm{AEC}}[\ell]$ which is needed for the derivation in (3.3.23) on page 55.
Please note, that by replacing $\boldsymbol{e}_{\mathrm{AEC}}[\ell]$ by $\hat{\boldsymbol{y}}[\ell]$ the derivation holds to prove that $\boldsymbol{G}^H \hat{\mathbf{Y}}[\ell] = \hat{\mathbf{Y}}[\ell]$ in (4.5.39) on page 113.
Using the definition of $\boldsymbol{G}$ in (2.2.21) on page 23 we obtain

$$\boldsymbol{G}^H \mathbf{e}_{\mathrm{AEC}}[\ell] = \left(\boldsymbol{F}_{2L\times 2L} \boldsymbol{W}^{01}_{2L\times L} \boldsymbol{W}^{01}_{L\times 2L} \boldsymbol{F}^{-1}_{2L\times 2L}\right)^H \boldsymbol{F}_{2L\times 2L} \boldsymbol{W}^{01}_{2L\times L} \boldsymbol{e}_{\mathrm{AEC}}[\ell].$$

With $(\boldsymbol{W}^{01}_{L\times 2L})^T = \boldsymbol{W}^{01}_{2L\times L}$ and the definition of the inverse DFT matrix $\boldsymbol{F}^{-1}_{2L\times 2L} = \boldsymbol{F}^{*}_{2L\times 2L}/(2L)$ and $\boldsymbol{F}^{T}_{2L\times 2L} = \boldsymbol{F}_{2L\times 2L}$ we obtain

$$\begin{aligned}
\boldsymbol{G}^H \mathbf{e}_{\mathrm{AEC}}[\ell] &= \frac{(\boldsymbol{F}^{*}_{2L\times 2L})^H}{2L} \boldsymbol{W}^{01}_{2L\times L} \boldsymbol{W}^{01}_{L\times 2L} \underbrace{\boldsymbol{F}^{H}_{2L\times 2L} \boldsymbol{F}_{2L\times 2L}}_{2L\boldsymbol{I}_{2L\times 2L}} \boldsymbol{W}^{01}_{2L\times L} \boldsymbol{e}_{\mathrm{AEC}}[\ell] \\
&= \boldsymbol{F}^{T}_{2L\times 2L} \boldsymbol{W}^{01}_{2L\times L} \underbrace{[\mathbf{0}_{L\times L}\, \boldsymbol{I}_{L\times L}] \begin{bmatrix} \mathbf{0}_{L\times L} \\ \boldsymbol{I}_{L\times L} \end{bmatrix}}_{\boldsymbol{I}_{L\times L}} \boldsymbol{e}_{\mathrm{AEC}}[\ell] \\
&= \boldsymbol{F}^{T}_{2L\times 2L} \boldsymbol{W}^{01}_{2L\times L} \boldsymbol{e}_{\mathrm{AEC}}[\ell] \\
&= \mathbf{e}_{\mathrm{AEC}}[\ell], \text{q.e.d.}
\end{aligned}$$

Abbreviations and Symbols

List of abbreviations and acronyms

AAC	advanced audio codec
AD	analogue/digital
AEC	acoustic echo canceller
AES	acoustic echo suppression
AGC	automatic gain control
aka.	also known as
ANC	active noise control
ANOVA	analysis of variance
AP	affine projection
APA	affine projection algorithm
APSD	auto power spectral density
BF	beamformer
BSD	bark spectral distortion
CD	cepstral distance
cf.	confer
CI	clarity index
CPSD	cross power spectral density
CT	central time
DA	digital/analogue
dB	decibel
DFT	discrete Fourier transform

dFxLMS decoupled filtered-X least-mean-squares
DMT discrete multitone
DOA direction of arrival
DRR direct-path-to-reverberation-ratio
DTD double-talk detection
DTFT discrete time Fourier transform
EC echo cancellation
EDC energy decay curve
EIC echo- and interference canceller
e.g. exempli gratia
ERLE echo return loss enhancement
EQ equalizer
FAP fast affine projection
FDAF frequency-domain adaptive filter
FFT fast Fourier transform
FHT fast Hadamard transform
FIR finite impulse response
FT Fourier transform
FWSSRR frequency-weighted SSRR
FWSEG frequency-weighted SSRR
FxLMS filtered-X least-mean-squares
GUI graphical user interface
Hz hertz
i.e. id est / that is
IFT inverse Fourier transform
IFFT inverse fast Fourier transform
IIR infinite impulse response
IPNLMS improved proportionate NLMS
IQR inter quartile range
IR impulse response
IS Itakura-Saito (distance)
ISD Itakura-Saito distance
ISwINO impulse-response shaping with ∞-norm optimization
ISwPP impulse-response shaping with post processing

ITU International Telecommunication Union
kHz kilohertz
LAR log-area ratio
LLR log-likelihood ratio
LMS least-mean-squares
LPC linear predictive coding
LRC listening-room compensation
LRM loudspeaker room microphone
LS least-squares
LS-EQ least-squares equalizer
LSD log-spectral distortion
LTI linear time-invariant
MC multi-channel
MDF multi-delay filter
mFxLMS modified filtered-X least-mean-squares
MIMO multiple input multiple output
MINT multiple input/output inverse theorem
MISO multiple input single output
ML maximum length
MLS maximum length sequences
MMSE minimum mean squared error
MOS mean opinion score
MPNLMS μ-law PNLMS
MVDR minimum variance distortionless response
ms millisecond(s)
MSC magnitude squared coherence
MTI modulation transfer index
NR noise reduction
NLMS normalized least-mean-squares
OFDM orthogonal frequency division multiplexing
OLA overlap-add
OLS overlap-save
PDS power delay spectrum
PDP power delay profile

PESQ Perceptual Evaluation of Speech Quality
PEMO-Q perception model for quality
PFBLMS partitioned frequency block LMS
PNLMS proportionate normalized least-mean-squares
PPMCC Pearson product-moment correlation coefficient
PSD power spectral density
PSM perceptual similarity measure
q.e.d. quod erat demonstrandum
RDT reverberation decay tail
REEF residual echo estimation filter
RIR room impulse response
RLS recursive least-squares
RMS root-mean-squares
RT60 room reverberation time
RTF room transfer function
SAD speech activity detection
SAEC stereo acoustic echo canceller
SC single-channel
SFM spectral flatness measure
SIMO single input multiple output
SIR signal-to-interference ratio
SISO single input single output
SRMR speech-to-reverberation modulation energy ratio
SNR signal-to-noise ratio
SNRE signal-to-noise ratio enhancement
SRR signal-to-reverberation ratio
SRRE signal-to-reverberation ratio enhancement
SSRR segmental signal to reverberation ratio
SSRRE segmental signal to reverberation ratio enhancement
STFT short time Fourier transform
STI speech transmission index
STSA short time spectral attenuation
TF transfer function
VAD voice activity detection

viz.	videlicet
WFS	wave-field synthesis
WLS	weighted least-squares
WLS-EQ	weighted least-squares equalizer
WOLA	weighted overlap-add
w.r.t.	with respect to
WSS	weighted spectral slope

Mathematical Symbols

$(\cdot)^{-1}$	inverse of $(\cdot)$
$(\cdot)^T$	transpose of $(\cdot)$
$(\cdot)^H$	Hermitian transpose of $(\cdot)$
$(\cdot)^*$	conjugate complex of $(\cdot)$
$(\cdot)^+$	Moore-Penrose pseudoinverse of $(\cdot)$
$(\cdot) * (\cdot)$	convolution
$(\cdot) \times (\cdot)$	dimension of matrix
bdiag$\{\cdot\}$	block diagonal matrix, cf. (2.2.13)
DFT$\{\cdot\}$	discrete Fourier transform as defined in (2.2.3) or (2.2.5)
diag $\{\cdot\}$	if $(\cdot)$ is a matrix diag $\{\cdot\}$ gives the main diagonal if $(\cdot)$ is a vector diag $\{\cdot\}$ builds up a matrix with the vector's elements on the main diagonal and zeros else
E $\{\cdot\}$	expectation operator
$\mathcal{H}\{\cdot\}$	Hilbert transform
$\ln(\cdot)$	natural logarithm
$\log_{10}(\cdot)$	base 10 logarithm
$\mathcal{L}\{\cdot\}$	limiting operator, cf. (A.2.10)
tr$\{\cdot\}$	trace of a matrix
$\forall$	for all
$\|\cdot\|_p$	l_p-norm according to definition (3.2.13)
$\|\cdot\|_\infty$	l_∞-norm according to definition (3.2.19)
$\|\cdot\|$	absolute value (of each entry if applied to a vector or matrix)
$\oslash$	element-by-element division of two vectors
$\mathbb{N}^+$	positive natural number
$\nabla_{\{\cdot\}}$	gradient
$\partial\{\cdot\}$	partial derivation

Latin Symbols

$\mathbf{0}_{m\times n}$	vector or matrix of size $m \times n$ containing zeros

$\mathbf{1}_{m\times n}$	vector or matrix of size $m \times n$ containing ones
$\boldsymbol{a}_y[\ell], \boldsymbol{a}_{\hat{y}}[\ell]$	LPC coefficient vectors of the signal blocks $\boldsymbol{y}[\ell]$ and $\hat{\boldsymbol{y}}[\ell]$, cf.(A.2.13) and (A.2.14)
$\boldsymbol{A}$	auxiliary matrix used in (4.7.7)
$A(\mathbf{v})$	arithmetic mean of vector $\mathbf{v}$, cf. (A.1.10)
b	critical band rate, cf. (A.2.21)
$b[k]$	white Gaussian process used for RIR model in (2.1.2)
$\boldsymbol{B}_{\mathrm{BP}}$	auxiliary matrix used in (4.7.8)
BSD	objective quality measure Bark spectral distortion, cf. (A.2.30)
c	speed of sound ($c \approx 340$ m/s), cf. p. 15
$c_y[j,\ell], c_{\hat{y}}[j,\ell]$	cepstral coefficients of signals $y[k]$ and $\hat{y}[k]$ for block ℓ, cf. (A.2.18), (A.2.19)
$\boldsymbol{c}_{\mathrm{AEC}}, \boldsymbol{c}_{\mathrm{AEC}}[k]$	fixed and time-varying AEC filter coefficient vectors, cf. (3.2.2)
$\boldsymbol{c}_{\mathrm{EQ}}$, $\boldsymbol{c}_{\mathrm{EQ}}[k]$	fixed and time-vaying LRC filter coefficient vectors, cf. (4.2.2) and (4.5.1)
$\boldsymbol{c}_{\mathrm{EQ,opt}}^{\mathrm{ISwPP}}$	LRC by means of impulse response shortening with post processing based on (4.7.8) and after processing by (4.7.10)
$\boldsymbol{c}_{\mathrm{EQ}}^{\mathrm{ISwINO}}$	impulse response shaping based on p-norm / ∞-norm optimization, cf. Section 4.7.2
$\boldsymbol{c}_{\mathrm{EQ}}^{\mathrm{WLS}}$	weighted least-squares LRC filter coefficients, cf. (4.6.9)
$\boldsymbol{c}_{\mathrm{EQ}}[\ell], \boldsymbol{c}_{\mathrm{EQ},p}[\ell]$	LRC filter coefficient vector in block time-domain, cf. (4.5.30) and (4.5.30) for one loudspeaker channel p
$\mathbf{c}_{\mathrm{EQ}}[\ell]$	LRC filter coefficient vector in block-frequency-domain, cf. (4.5.28)
$\mathbf{c}_{\mathrm{REEF}}[\ell]$	REEF filter coefficient vector as used in Figure 3.17 and (3.3.31)
C	constant used in (2.1.5)
$\mathrm{C}_{50}(\boldsymbol{v})$	objective quality measure clarity C_{50}, cf. (A.1.4)
$\mathrm{C}_{80}(\boldsymbol{v})$	objective quality measure clarity C_{80}, cf. (A.1.5)
$CB(f)$	critical bandwidth, cf. (A.2.20)
CD	objective quality measure cespstral distance, cf. (A.2.17)
$\mathrm{CT}(\boldsymbol{v})$	objective quality measure central time, cf. (A.1.6)
$\boldsymbol{d}$	desired system for EQ, cf. (4.4.3) and (4.4.19)
$\boldsymbol{d}_q$	desired system for EQ for reference microphone q, cf. (4.4.20)
$\mathbf{d}_q$	desired system for EQ for reference microphone q in block-frequency-domain, cf. (4.5.27)
$\boldsymbol{d}_d, \boldsymbol{d}_u$	desired systems based on window functions for weighted least-squares equalization and RIR reshaping filters, cf. (4.7.4) and (4.7.5)
$\mathrm{D}_{50}(\boldsymbol{v})$	objective quality measure definition D_{50}, cf. (A.1.1)
$\mathrm{D}_{80}(\boldsymbol{v})$	objective quality measure definition D_{80}, cf. (A.1.2)

D_c	critical distance as defined in (2.1.7)
$D_{\mathrm{dB}}[k]$	relative system misalignment according to (3.1.4)
DRR	objective quality measure direct-to-reverberation-ratio, cf. (A.1.7)
$\boldsymbol{D}$	convolution matrix build from the coefficients of the desired system vector $\boldsymbol{d}$ used in Section 4.7.1
$e_{\mathrm{EQ}}[k]$	error signal for LRC filter
$e_{\mathrm{AEC}}[k]$	error signal for AEC filter (equals $\xi[k]$)
$e_{\mathrm{AEC},p}[k]$	error signal for AES filter, cf. Figure 3.16
$e_{\mathrm{BP}}[k]$	error signal to calculate the predictor for spectral post processing as in Figure 4.41
$e_p[k]$	error signal for spectral post processing
$\overline{e^2_{\mathrm{AEC}}}[k]$	smoothed power of the error signal as defined in (5.1.13)
$\boldsymbol{e}[k]$	error signal vector
$\boldsymbol{e}_{\mathrm{AEC}}[k]$	AEC error signal vector as defined in (3.3.6)
$\boldsymbol{e}_{\mathrm{EQ}}^{\mathrm{WLS}}$	error signal vector for weighted least-squares LRC filter, cf. (4.6.3)
$\boldsymbol{e}_{\mathrm{PF}}[k]$	PF error signal vector as defined in (3.3.5)
$\mathbf{e}_{\mathrm{AEC}}[\ell]$	block-frequency-domain AEC error signal, cf. (3.3.13)
$\mathbf{e}_{\mathrm{AEC},p}[\ell]$	block-frequency-domain AES error signal, cf. (3.3.15)
$\mathrm{EDC}(t)$	energy decay curve, cf. (2.1.5)
$\mathrm{ERLE}\vert_{\mathrm{dB}}[k]$	echo return loss enhancement in dB, cf. (3.1.6)
$\mathbf{E}_{\mathrm{AEC}}[\ell]$	matrix containing AEC error signal of two blocks on main diagonal, cf. (3.3.14)
$\mathbf{E}_{\mathrm{EQ},mod}[\ell]$	matrix in block-frequency-domain containing modified error signal of mFxLMS or dFxLMS equalizer, cf. (4.5.31)
$\mathbf{E}_{\mathrm{EQ},mod,q}[\ell]$	matrix in block-frequency-domain containing modified error signal of mFxLMS or dFxLMS equalizer for reference microphone channel q, cf. (4.5.22)
f	frequency
f_s	sampling frequency
FWSSRR	objective quality measure frequency-weighted SSRR, cf. (A.2.4)
$\boldsymbol{F}_{2L\times 2L}$	DFT matrix of size $2L \times 2L$, see (2.2.5)
$\boldsymbol{F}_{2L\times L}$	DFT matrix of size $2L \times L$, see (2.2.9)
$\boldsymbol{F}^{-1}_{2L\times 2L}$	inverse DFT matrix of size $2L \times 2L$, cf. p. 21
$g_i[k]$	step-size coefficient i for proportionate update schemes
$G(\mathbf{v})$	geometric mean of vector $\mathbf{v}$, cf. (A.1.10)
$\boldsymbol{G}$	constraining matrix as defined in (2.2.21)
$h(t), h[k]$	(room) impulse response
$h_{\mathrm{AEC}}[k]$	room impulse response for AEC, cf. e.g. Figure 1.2
$h_{\mathrm{EQ}}[k]$	room impulse response for EQ, cf. e.g. Figure 1.2
$h_f[k]$	room impulse response in far-end room, cf. e.g. Figure 1.2
$h_M[k]$	(room) impulse response model as defined in (2.1.2)
$\mathsf{h}(f)$	room transfer function (RTF)

$\boldsymbol{h}$	vector containing RIR coefficients as defined in (4.2.3) or (2.2.2a)
$\boldsymbol{h}_i$	vector containing one partition of the impulse response as defined in (2.2.2b)
$\boldsymbol{h}[k]$	vector containing time-variant RIR coefficients as defined in (3.1.2)
$\hat{\boldsymbol{h}}[k]$	vector containing coefficients of RIR estimates as defined in (3.1.3) - equals $\boldsymbol{c}_{\mathrm{AEC}}[k]$
$\tilde{\boldsymbol{h}}[k]$	system misalignment vector of AEC as defined in (3.1.1)
$\mathbf{h}$	vector of zero-padded block transfer function as defined in (2.2.10)
$\mathbf{h}_i$	vector containing RTF coefficients of one partition, cf. (2.2.11)
$\boldsymbol{H}_{\mathrm{CM}}$	convolution matrix build up by RIR coefficients, cf. (4.4.16) and (4.2.7)
$\boldsymbol{H}_{\mathrm{CM},pq}$	SISO convolution matrix build up by RIR coefficients, cf. (4.4.17) and (4.2.7)
$\tilde{\boldsymbol{H}}_{\mathrm{CM}}$	convolution matrix of the RIR estimation error, cf. (5.1.8)
$\hat{\boldsymbol{H}}_{\mathrm{CM}}$	convolution matrix of the estimated RIR, cf. (5.1.9)
$\mathbf{H}[\ell]$	frequency-domain MIMO channel matrix, cf. (4.5.8)
I	no of channels of the far-end signal
I	total number of room surfaces, see (2.1.4)
$\mathrm{ISD}[\ell]$	objective quality measure Itakura-Saito distance, cf. (A.2.15)
$\boldsymbol{I}_{L\times L}$	identity matrix of size $L \times L$
$\tilde{\boldsymbol{I}}_{2L\times 2L}$	shifting matrix matrix of size $2L \times 2L$, see (2.2.17)
J	no. of source signals
$J[\ell]$	error criterion function as defined e.g. in (3.3.19)
k	discrete-time index
k_0	delay introduced by the equalizer
$\tilde{k}_0$	main peak of desired system vector $\boldsymbol{d}$
$k_{0,\mathrm{opt}}$	optimum delay of the equalizer, cf. (4.4.9)
$k_{0,\mathrm{opt,BSD}}$	optimum delay of the equalizer defined by optimum BSD, cf. (4.4.7)
$k_{0,\mathrm{opt,SRRE}}$	optimum delay of the equalizer defined by optimum SRRE, cf. (4.4.8)
k_{50}	position of lag in IR corresponding to 50 ms, cf. Figure A.1
k_{80}	position of lag in IR corresponding to 80 ms, cf. Figure A.1
k_{init}	initial delay of an impulse response
k_Δ	range around the main peak of IR as used in (A.1.7)
ℓ	block-time index

$l_\infty[k]$	parameter of PNLMS algorithm in (3.2.19)
$l'_\infty[k]$	parameter of PNLMS algorithm in (3.2.18)
L	block length
L_{DFT}	DFT length
L_{AEC}	length of AEC filter $\boldsymbol{c}_{\mathrm{AEC}}$ defined in (3.2.2)
L'_{AEC}	number of blocks needed for partitioned AEC
L_{EQ}	length of EQ filter $\boldsymbol{c}_{\mathrm{EQ}}$ defined in (4.2.2)
L'_{EQ}	number of blocks needed for partitioned EQ
L_h	length of RIR vector $\boldsymbol{h}$ defined in (4.2.3)
L'_h	number of blocks of partitioned RIR
L_p	length of post-filter vector $\boldsymbol{p}$ defined in (3.3.4)
L_{REEF}	length of REEF filter $\boldsymbol{c}_{\mathrm{REEF}}$
LAR	objective quality measure log-area ratio, cf. (A.2.16)
LLR$[\ell]$	objective quality measure log-likelihood ratio, cf. (A.2.12)
LSD	objective quality measure log-spectral distortion, cf. (A.2.9)
m	loudness level in sone, cf. (A.2.29)
$\boldsymbol{M}_{\mathrm{REEF}}[\ell]$	diagonal coefficient matrix containing step-sizes for residual echo estimation filter in (3.3.31)
$\boldsymbol{M}_{\mathrm{PNLMS}}[k]$	diagonal coefficient matrix containing step-sizes for PNLMS as defined in (3.2.15)
$\boldsymbol{M}_{\mathrm{IPNLMS}}[k]$	diagonal coefficient matrix containing step-sizes for IPNLMS as defined in (3.2.24)
n	discrete frequency index
$n[k]$	disturbance / noise signal
O	overclocking factor in Algorithm 3 (dFxLMS)
OMCR, OMCR$[n]$	objective measure for coloration in reverberation, cf. (A.2.46)
$\boldsymbol{p}$	the coefficient vector of the RIR reshaping predictor, cf. (4.7.9)
$\boldsymbol{p}, p_x, p_y, p_z$	spatial position with 3-dimentional coordinates (cf. Section 2.1.1)
$\boldsymbol{p}[k]$	AES post-filter coefficients, cf. (3.3.4)
$\mathsf{p}[\ell]$	AES post-filter coefficients in block-frequency domain, cf. (3.3.8)
P	no of loudspeakers
$\boldsymbol{q}[\ell]$	the REEF time-domain block error signal as defined in (3.3.16)
Q	no of microphones
Q	directivity of sound source used in (2.1.6)
r_{corr}	correlation coefficient as defined in (4.2.1)
$\boldsymbol{r}_{\boldsymbol{x}\psi}$	crosscorrelation vector of $\boldsymbol{x}$ and ψ in (3.2.5)
$\hat{\boldsymbol{r}}_{\boldsymbol{x}\psi}$	estimated crosscorrelation vector of $\boldsymbol{x}$ and ψ, cf. (3.2.9)
$\boldsymbol{r}[k]$	signal vector in the update path of FxLMS and derivates as defined in (4.5.3)

$\boldsymbol{r}_{pq}[\ell]$	signal vector in the update path for loudspeaker channel p and microphone channel q, cf. (4.5.17)
$\mathbf{r}_{pq}[\ell]$	signal vector in the update path for loudspeaker channel p and microphone channel q in block-frequency-domain, cf. (4.5.16)
R	room constant, cf. e.g. (2.1.6)
$\boldsymbol{R}_{\boldsymbol{xx}}$	covariance matrix of $\boldsymbol{x}$ in (3.2.1)
$\hat{\boldsymbol{R}}_{\boldsymbol{xx}}$	estimated covariance matrix of $\boldsymbol{x}$ in (3.2.9)
$\mathbf{R}[\ell]$	block-frequency-domain signal used in update path of FxLMS algorithms, cf. (4.5.35)
$\mathbf{R}_p[\ell]$	block-frequency-domain signal used in update path of FxLMS algorithms for loudspeaker channel p, cf. (4.5.21)
$\mathbf{R}_q[\ell]$	block-frequency-domain signal used in update path of FxLMS algorithms for microphone channel q, cf. (4.5.20)
$\mathbf{R}_{pq}[\ell]$	block-frequency-domain signal used in update path of FxLMS algorithms for loudspeaker channel p and microphone channel q, cf. (4.5.18)
$\breve{\mathbf{R}}_{pq}[\ell]$	block-frequency-domain signal used in update path of FxLMS algorithms for loudspeaker channel p and microphone channel q containing two blocks of input, cf. (4.5.19)
$\boldsymbol{R}_{\boldsymbol{ss}}[k]$	covariance matrix of the input signal $s[k]$
$\boldsymbol{s}, s_x, s_y, s_z$	spatial position with 3-dimentional coordinates (cf. Section 2.1.1)
$s(t), s[k]$	speech signal
$s_{dec}[k]$	independent input signal defined in dFxLMS algorithm
$s_f[k]$	signal of far-end speaker
$s_n[k]$	signal of near-end speaker
$\boldsymbol{s}[k], \boldsymbol{s}_{\mathrm{I}}[k], \boldsymbol{s}_{\mathrm{II}}[k]$	vectors containing samples of signal as defined in (4.4.2), (4.5.4) and (4.5.5)
$\boldsymbol{s}[\ell]$	time-domain vector containing one block of input (speech) signal, cf. (4.5.14)
$\mathrm{S}_1, \mathrm{S}_2$	switches
S_i	wall surface area in Sabine formula, cf. (2.1.4)
$\mathrm{SF}_{\Delta}[b]$	spreading function with triangular approximation, cf. (A.2.22)
$\mathrm{SF}_{\mathrm{SH}}[b]$	spreading function according to Sekey and Hanson, cf. (A.2.23)
$\mathrm{SF}_{\mathrm{MP3}}[b]$	spreading function as used in MP3 standard, cf. (A.2.24)
$\mathrm{SFM}(\mathbf{v})$	objective quality measure spectral flatness measure (SFM), cf. (A.1.10)
$\mathrm{SRRE}_{\mathrm{dB}}$	objective quality measure segmental signal-to-reverberation-ratio enhancement in dB, cf. (A.2.2)
$\mathrm{SSRR}_{\mathrm{dB}}$	objective quality measure segmental signal-to-reverberation-ratio in dB, cf. (A.2.1)

$\mathbf{S}[\ell]$	block-frequency-domain matrix containing (speech) input data, cf. (4.5.12) or (4.5.26)
$\breve{\mathbf{S}}[\ell]$	block-frequency-domain matrix containing input data from two successive blocks, cf. (4.5.13)
t, t'	time (constant)
$u[k]$	Heaviside step function
$x(t), x[k]$	loudspeaker signal
$x_p[k]$	loudspeaker signal for channel p
$\overline{x^2}[k]$	smoothed power of the input signal, cf. (5.1.14)
$\boldsymbol{x}[k]$	vector of input samples (loudspeaker signal)
$\boldsymbol{x}[\ell]$	block time-domain input data vector, cf. (2.2.1)
$\boldsymbol{x}_p[\ell]$	block time-domain signal of loudspeaker channel p, cf. (2.2.1)
$\mathbf{X}[\ell]$	frequency-domain matrix containing input data as defined in (2.2.15), (4.5.10)
$\breve{\mathbf{X}}[\ell]$	frequency-domain matrix containing input data from two successive blocks, cf. (2.2.15)
$\breve{\mathbf{X}}_p[\ell]$	frequency-domain matrix containing input data from two successive blocks of loudspeaker channel p, cf. (4.5.10)
$y[k]$	microphone signal
$y_q[k]$	microphone signal of channel q
$\boldsymbol{y}[\ell]$	time-domain vector containing microphone signal, cf. (2.2.24) and (2.2.27)
$\tilde{\boldsymbol{y}}[\ell]$	time-domain vector of microphone signal containing cyclic convolution products, cf. (2.2.19) and (2.2.26)
$\hat{\boldsymbol{y}}_q[\ell]$	desired signal which results from filtering the input signal with the desired system, cf. (4.5.25)
$\mathbf{y}[\ell]$	block-frequency-domain microphone signal, cf. (2.2.20)
$\tilde{\mathbf{y}}[\ell]$	block-frequency-domain microphone signal containing cyclic convolution products, cf. (2.2.14)
$\mathbf{Y}[\ell]$	block-frequency-domain microphone signal, cf. (4.5.6) and (4.5.34)
$\hat{\mathbf{Y}}_q[\ell]$	desired signal in block-frequency-domain which results from filtering the input signal with the respective desired system for each channel q, cf. (4.5.23)
$\hat{\breve{\mathbf{Y}}}_q[\ell]$	block-frequency-domain matrix containing two successive blocks of signal after desired system $\mathbf{d}_q$, cf. (4.5.24)
$\boldsymbol{v}$, $\boldsymbol{v}[k]$	equalized system vector, cf. (4.2.5) and (5.2.1)
$\mathbf{v}$	transfer function vector of equalized system, cf. e.g. Section A.1.5
$\bar{\mathrm{v}}_{\mathrm{dB}}$	mean logarithmic spectrum of equalized system, cf. (A.1.9)
$\boldsymbol{v}'[k]$	equalized system vector, cf. (5.2.2)
V	volume (of a room) in Sabine formula (2.1.4)

$\boldsymbol{V}_{-1}$	convolution matrix made of $\boldsymbol{v}$ with an additional first row of zeros to take into account the delay of one sample as depicted in Figure 4.41
VAR(v)	objective quality measure VAR, cf. (A.1.8)
$\boldsymbol{w}$	window function, cf. (4.6.5)
$\boldsymbol{w}_{\mathrm{FW}}$	window function corresponding to forward time masking, cf. (4.6.1)
$w_{0,k}$	coefficients of window function $\boldsymbol{w}_{\mathrm{FW}}$, cf. (4.6.2)
w_d, w_u	window functions for weighted least-squares equalizer
$\boldsymbol{W}$	matrix containing window function vector on main diagonal, cf. (4.6.4)
$\boldsymbol{W}^{01}_{L\times 2L}$	windowing matrix of size L times $2L$, cf.(2.2.23)
$\boldsymbol{W}^{01}_{2L\times L}$	windowing matrix of size $2L$ times L, cf.(2.2.22)
$\boldsymbol{W}^{10}_{2L\times L}$	windowing matrix of size $2L$ times L, cf.(2.2.8)
WSS	objective quality measure weighted spectral slope, cf. (A.2.7)

Greek Symbols

α	smoothing constant or parameter in (3.3.12)
$\alpha_{w,\mathrm{IS}}$	design parameter for window function in (4.7.3)
$\alpha_{w,\mathrm{WLS}}$	design parameter for window function in (4.6.2)
β	absorption coefficient, cf. (2.1.4), or parameter in (3.3.12)
γ	parameter in (3.3.12)
$\gamma_1, \gamma_2, \gamma_3$	design parameters for OMCR quality measure, cf. (A.2.46)
$\gamma(\boldsymbol{h})$	sparsity measure, cf. (3.2.12)
δ	general regularization parameter used in different algorithms such as e.g. NLMS (3.2.11)
δ_{IPNLMS}	regularization parameter for PNLMS in (3.2.23)
δ_{NLMS}	regularization parameter for NLMS in (3.2.11)
δ_{PNLMS}	regularization parameter for PNLMS in (3.2.14)
ϵ	small value
η	damping constant as defined in (2.1.3)
μ_{dFxLMS}	step-size of decoupled FxLMS algorithm
μ_{FxLMS}	step-size of FxLMS algorithm
μ_{IPNLMS}	fixed step-size of IPNLMS algorithm in (3.2.23)
μ_{LMS}	step-size of LMS algorithm in (3.2.10)
μ_{mFxLMS}	step-size of modified FxLMS algorithm
$\mu_{\mathrm{NLMS}}[k]$	step-size of LMS algorithm
μ_{PNLMS}	fixed step-size of PNLMS algorithm in (3.2.14)
$\mu_{i,\mathrm{IPNLMS}}[k]$	step-size of IPNLMS for coefficient i, cf. (3.2.27)
$\mu'_{i,\mathrm{PNLMS}}[k]$	step-size parameter of PNLMS algorithm in (3.2.17)
$\boldsymbol{\mu}_{\mathrm{PNLMS}}[k]$	coefficient vector containing step-sizes of the PNLMS algorithm as defined in (3.2.16)

$\boldsymbol{\mu}_{\mathrm{IPNLMS}}[k]$	coefficient vector containing step-sizes of the IPNLMS algorithm as defined in (3.2.25)
$\boldsymbol{\mu}_{\mathrm{IPNLMS}}[k]$	IPNLMS step-size vector, cf. (3.2.25)
π	PI value, e.g. used in (2.1.7)
ρ	parameter of PNLMS algorithm in (3.2.17)
τ_{60}	room reverberation time (cf. Section 2.1.4)
υ	control parameter of PNLMS algorithm in (3.2.18)
$\psi[k]$	acoustic echo signal
$\hat{\psi}[k]$	estimate of acoustic echo signal (AEC filter output)
$\boldsymbol{\Phi}_{\mathsf{e}_{\mathrm{AEC}}\,\mathsf{e}_{\mathrm{AEC}}}[\ell]$	auto power spectral density (APSD) vector of the AEC error signal
$\boldsymbol{\Phi}_{\mathsf{e}_{\mathrm{AEC}}\,\mathsf{s}_n}[\ell]$	cross power spectral density (CPSD) vector of the AEC error signal and the near-end speaker's signal
$\boldsymbol{\Phi}_{\mathsf{s}_n\,\mathsf{s}_n}[\ell]$	auto power spectral density (APSD) vector of the near-end speaker's signal
$\hat{\boldsymbol{\Phi}}_{\mathbf{RR}}[\ell]$	auto power spectral density (APSD) vector of the update signal $\mathbf{R}$ in dFxLMS algorithm
$\hat{\boldsymbol{\Phi}}_{\mathbf{R}\hat{\mathbf{Y}}}[\ell]$	cross power spectral density (CPSD) vector of the update signal $\mathbf{R}$ and the desired signal $\hat{\mathbf{Y}}$
$\hat{\boldsymbol{\Phi}}_{\mathsf{xe}}[\ell]$	CPDS estimate of loudspeaker signal $\mathbf{X}^H[\ell]$ and AEC error signal $\mathbf{e}_{\mathrm{AEC}}[\ell]$, cf. (3.3.25)
$\hat{\boldsymbol{\Phi}}_{\mathsf{xx}}[\ell]$	APSD estimate of loudspeaker signal $\mathbf{X}^H[\ell]$, cf. (3.3.26)
$\xi[k]$	residual acoustic echo signal (equals $e_{\mathrm{AEC}}[k]$)
$\hat{\boldsymbol{\xi}}[\ell]$	output of REEF / residual echo estimate, cf. (3.3.16)

Bibliography

[AB79] J. B. Allen and D. A. Berkley. Image Method for Efficiently Simulating Small–Room Acoustics. *J. Acoust. Soc. Amer.*, 65:943–950, 1979. (Cited on pages 17, 73, 89, and 138)

[ABB77] J.B. Allen, D.A. Berkley, and J. Blauert. Multimicrophone Singalprocessing Technique to Remove Room Reverberation from Speech Signals. *Journal of the Acoustical Society of America (JASA)*, 62(4):912–915, October 1977. (Cited on page 68)

[ABZ07] F. Albu, M. Bouchard, and Y. Zakharov. Pseudo-Affine Projection Algorithms for Multichannel Active Noise Control. *IEEE Trans. on Audio, Speech and Language Processing*, 15(3):1044–1052, March 2007. (Cited on pages 107 and 108)

[AD94] C. Antweiler and M Dörbecker. Perfect Sequence Excitation of the NLMS Algorithm and its Application to Acoustic Echo Control. *Annals of Telecommunications*, 49:386–397, 1994. 10.1007/BF02999427. (Cited on page 108)

[AD01] N. Al-Dhahir. FIR Channel-Shortening Equalizers for MIMO ISI Channels. *IEEE Trans. Commun.*, 49(2):213–218, 2001. (Cited on page 66)

[Adr06] F. Adriaensen. Acoustical Impulse Response Measurement with ALIKI. In *4th International Linux Audio Conference (LAC2006)*, pages 9–14, Karlsruhe, Germany, April 2006. (Cited on pages 161, 162, 163, and 164)

[AEK01] G. Arslan, B. L. Evans, , and S. Kiaei. Equalization for discrete multitone transceivers to maximize bit rate. *IEEE Trans. on Signal Processing*, 49(12):3123–3135, Dec. 2001. (Cited on page 122)

[AF95] B. Ayad and G. Faucon. Acoustic Echo and Noise Cancelling for Hands-Free Communication Systems. In *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, pages 91–94, Roros, Norway, June 1995. (Cited on page 51)

[AG97] S. Affes and Y. Grenier. A Signal Subspace Tracking Algorithm for Microphone Array Processing of Speech. *IEEE Trans. on Speech and Audio Processing*, 5(5):425–437, September 1997. (Cited on pages 31, 67, and 68)

[AGQ97] C. Antweiler, J. Grunwald, and H. Quack. Approximation of Optimal Step Size Control for Acoustic Echo Cancellation. In *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Proc. (ICASSP)*, volume 1, pages 295–298, Munich, Germany, April 1997. (Cited on pages 31 and 138)

[AJ67] A.E. Albert and L.S. Gardner Jr. *Stochastic Approximation of Nonlinear Regression*. MIT Press Research Monograph No. 42, 1967. (Cited on page 39)

[Ali98] M. Ali. Stereophonic Acoustic Echo Cancellation System Using Time-Varying All-Pass Filtering for Signal Decorrelation. In *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 3689–3672, Seattle, USA, May 1998. (Cited on page 33)

[All82] J. B. Allen. Effects of Small Room Reverberation on Subjective Preference. *Journal of the Acoustical Society of America (JASA)*, 71(S1):S5, 1982. (Cited on page 61)

[ARG10] E. Albertin, J. Rennies, and S. Goetze. Objective Quality Measures for Dereverberation Methods based on Room Impulse Response Equalization. In *Proc. German Annual Conference on Acoustics (DAGA)*, Berlin, Germany, March 2010. (Cited on pages 5 and 71)

[AS03] L. Atlas and S. A. Shamma. Joint Acoustic and Modulation Frequency. *EURASIP Journal on Applied Signal Processing*, 2003:668–675, January 2003. (Cited on page 31)

[BA83] J. Borish and J. B. Angell. An Efficient Algorithm for Measuring the Impulse Response Using Pseudorandom Noise. *Journal of the Audio Engineering Society*, 31:478–488, Jul./Aug. 1983. (Cited on pages 65 and 133)

[BAGG95] J. Benesty, F. Amand, A. Gilloire, and Y. Grenier. Adaptive Filtering Algorithms for Stereophonic Acoustic Echo Cancellation. In *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 3099–3102, 1995. (Cited on page 32)

[BBK03] H. Buchner, J. Benesty, and W. Kellermann. Multichannel Frequency-Domain Adaptive Filtering with Application to Multichannel Acoustic Echo Cancellation. In J. Benesty and Y. Huang, editors, *Adaptive Signal Processing – Applications to*

Real-World Problems, chapter 4, pages 95–128. Springer, 2003. (Cited on pages 29 and 33)

[BC82] P.J. Bloom and G.D. Cain. Evaluation of Two-Input Speech Dereverberation Techniques. In *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 164 – 167, May 1982. (Cited on page 68)

[BDG96] J. Benesty, P. Duhamel, and Y. Granier. A Multichannel Affine Projection Algorithm with Applications to Multichannel Acoustic Echo Cancellation. *IEEE Signal Processing Letters*, 3(2):35–37, February 1996. (Cited on page 32)

[BDH+99] C. Breining, P. Dreiseitel, E. Hänsler, A. Mader, B. Nitsch, H. Puder, T. Schertler, G. Schmidt, and J. Tilp. Acoustic Echo Control – An Application of Very-High-Order Adaptive Filters. *IEEE Signal Processing Magazine*, pages 42–69, July 1999. (Cited on pages 18, 28, 29, 66, 93, and 146)

[BdV93] A.J. Berkhout, D. de Vries, and P. Vogel. Acoustic Control by Wave Field Synthesis. *Journal of the Acoustical Society of America (JASA)*, 93(5):2764–2778, May 1993. (Cited on page 69)

[Ber80] D.A. Berkley. Normal listeners in typical rooms - reverberation perception, simulation, and reduction. In Gerald A. Studebaker and Irving Hochberg, editors, *Acoustical Factors Affecting Hearing Aid Performance*, pages 3–24. University Park Press, Baltimore, 1980. (Cited on page 61)

[BF95] R. Le Bouquin–Jeannes and G. Faucon. Study of a Voice Activity Detector and its Influence on a Noise Reduction System. *EURASIP Speech Communication*, 16:245–254, 1995. (Cited on page 30)

[BF96] R. Le Bouquin–Jeannes and G. Faucon. Optimization of a Cascaded Structure for Noise and Echo Cancellation Using a Noise Filtering Preprocessing. *Ann. Telecommunication*, 51(11–12):579–584, Nov 1996. (Cited on page 51)

[BG02a] J. Benesty and T. Gänsler. A Multi-Channel Acoustic Echo Canceler Double-Talk Detector Based on a Normalized Cross-Correlation Matrix. *Eur. Trans. Telecomm.*, vol. 13, March 2002. (Cited on page 31)

[BG02b] J. Benesty and S.L. Gay. An Improved PNLMS Algorithm. In *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 1881–1884, Orlando, FL, USA, May 2002. (Cited on pages 43 and 44)

[BG03] J. Benesty and T. Gänsler. A Multidelay Double-Talk Detector combined with the MDF Adaptive Filter. *EURASIP Journal on Applied Signal Processing*, 2003(11):1056–1063, 2003. doi:10.1155/S1110865703305037. (Cited on pages 29 and 31)

[BGM+01] J. Benesty, T. Gänsler, D. R. Morgan, M. M. Sondhi, and S. L. Gay. *Advances in Network and Acoustic Echo Cancellation*. Springer, Berin, 2001. (Cited on pages 28 and 39)

[BHCN06] J. Benesty, Y. Huang, J. Chen, and P.A. Naylor. Adaptive Algorithms for the Identification of Sparse Impulse Responses. In E. Hänsler and G. Schmidt, editors, *Topics in Acoustic Echo and Noise Control*, chapter 5, pages 125–153. Springer, Berlin, 2006. (Cited on pages 39, 40, 41, 42, and 146)

[Bja92] E. Bjarnason. Active Noise Cancellation using a Modified Form of the Filtered-X LMS Algorithm. In *Proc. EURASIP European Signal Processing Conference (EUSIPCO)*, pages 1053–1056, Brussels, Belgium, 1992. (Cited on pages 105 and 107)

[BK01] H. Buchner and W. Kellermann. Acoustic Echo Cancellation for Two and More Reproduction Channels. In *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, pages 99–102, Darmstadt, Germany, Sep 2001. (Cited on pages 32 and 33)

[BM00] J. Benesty and D. R. Morgan. Frequency-Domain Adaptive Filtering Revisited, Generalization to the Multi-Channel Case, and Application to Acoustic Echo Cancellation. In *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 789–792, Istanbul, Turkey, June 2000. (Cited on pages 18, 29, and 33)

[BM01] J. Benesty and D.R. Morgan. Multi-Channel Frequency-Domain Adaptive Filtering. In S.L. Gay and J. Benesty, editors, *Acoustic Signal Processing for Telecommunication*, chapter 7, pages 121–133. Kluwer Academic Publishers, Norwell, USA, 2001. (Cited on pages 55 and 114)

[BM03] T.S. Bakir and R.M. Mersereau. Blind Adaptive Dereverberation of Speech Signals Using a Microphone Array. In *Adaptive Sensor Array Processing Workshop (ASAP)*, 11 Mar 2003. (Cited on pages 67 and 68)

[BMB01] J. M. Buchholz, J. Mourjopoulos, and J. Blauert. Room Masking: Understanding and Modelling the Masking of Room Reflections. In *Proc. AES Convention (Audio Engineering Society)*, volume 110, Amsterdam, The Netherlands, May 2001. (Cited on page 67)

[BMC00] J. Benesty, D.R. Morgan, and J.H. Cho. A New Class of Doubletalk Detectors Based on Cross-Correlation. *IEEE Trans. on Speech and Audio Processing*, 8(2):168–172, March 2000. (Cited on page 31)

[BMS97] J. Benesty, D. R. Morgan, and M. M. Sondhi. A Better Understanding and an Improved Solution to the Problems of Stereophonic Acoustic Echo Cancellation. In *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 303–306, Munich, Germany, 1997. (Cited on page 33)

[BMS98a] J. Benesty, D. R. Morgan, and J.I. Hall M. M. Sondhi. Stereophonic Acoustic Echo Cancellation Using Nonlinear Transformations and Comb Filtering. In *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 3673–3676, Seattle, USA, May 1998. (Cited on page 33)

[BMS98b] J. Benesty, D. R. Morgan, and M. M. Sondhi. A Better Understanding and an Improved Solution to the Specific Problems of Stereophonic Acoustic Echo Cancellation. *IEEE Trans. on Speech and Audio Processing*, 6(2):156–165, Mar 1998. (Cited on pages 28, 33, 99, 133, 134, and 139)

[Bol79] S. Boll. Suppression of Acoustic Noise in Speech using Spectral Subtraction. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 27(2):113–120, Apr 1979. (Cited on page 52)

[Bor89] H. Borucki. *Einführung in die Akustik (in German).* Mannheim; Wien; Zürich: BI- Wiss.-Verl., 1989. (Cited on pages 9, 14, and 15)

[Bou03] M. Bouchard. Multichannel Affine and Fast Affine Projection Algorithms for Active Noise Control and Acoustic Equalization Systems. *IEEE Trans. on Speech and Audio Processing*, 11(1):54–60, Jan 2003. (Cited on pages 107 and 108)

[BQ00] M. Bouchard and S. Quednau. Multichannel Recursive-Least-Squares Algorithms and Fast-Transversal-Filter Algorithms for Active Noise Control and Sound Reproduction Systems. *IEEE Trans. on Speech and Audio Processing*, 8(5):606–618, September 2000. (Cited on pages 105, 107, 108, and 110)

[BR72] L. W. Brooks and I. S. Reed. Equivalence of the likelihood ratio processor, the maximum signal-to-noise ratio filter, and the Wiener filter. *IEEE Trans. Aerosp. Electron. Syst.*, AES-8(5):690–692, September 1972. (Cited on pages 55 and 113)

[Bra97] M. Brandenburg, K.; Bosi. Overview of mpeg audio: Current and future standards for low bit-rate audio coding. *J. Audio Eng. Soc*, 45(1/2):4–21, 1997. (Cited on page 118)

[Bri75] D.R. Brillinger. *Time Series-Data Analysis and Theory*. Holt, Rinehart, and Winston, Inc., New York, 1975. (Cited on pages 56 and 114)

[BRX+12] J. Brümmerstedt, J. Rennies, F. Xiong, S. Goetze, and J. Bitzer. Objective Methods to Asses Speech Signals Processed by Short-Term Spectral Attenuation. In *38th Annual Convention for Acoustics (DAGA)*, Darmstadt, Germany, Mar. 2012. (Cited on page 5)

[BS01] J. Bitzer and K. U. Simmer. Superdirective microphone arrays. In M. S. Brandstein and D. Ward, editors, *Microphone Arrays: Signal Processing Techniques and Applications*, chapter 2, pages 19–38. Springer, 2001. (Cited on page 67)

[BSFB01] R. Le Bouquin–Jeannes, P. Scalart, G. Fauçon, and C. Beaugeant. Combined Noise and Echo Reduction in Hands-Free Systems: A Survey. *IEEE Transactions on Speech and Audio Processing*, 9(8):808–820, November 2001. (Cited on pages 30 and 51)

[BSK04] H. Buchner, S. Spors, and W. Kellermann. Full-Duplex Systems for Sound Field Recording and Auralization Based on Wave Field Synthesis. In *Proc. AES Convention (Audio Engineering Society)*, volume 116, pages 1–9, Berlin, Germany, May 2004. (Cited on page 69)

[BSKR02] H. Buchner, S. Spors, W. Kellermann, and R. Rabenstein. Full-Dupley Communication Systems Using Loudspeaker Arrays and Microphone Arrays. In *IEEE Int. Conference on Multimedia and Expo (ICME), Lausanne, Switzerland*, pages 509 – 512, August 2002. (Cited on page 69)

[BSM79] M. Berouti, R. Schwartz, and J. Makhoul. Enhancement of Speech Corrupted by Acoustic Noise. In *Proc. IEEE Int. Conference Acoustic, Speech and Signal Processing, ICASSP–79*, pages 208–211, Washington DC, April 1979. (Cited on page 52)

[BTSG98] C. Beaugeant, V. Turbin, P. Scalart, and A. Gilloire. New Optimal Filtering Approaches for Hands-Free Telecommunication Terminals. *EURASIP Signal Processing*, 64(1998):33–47, 1998. (Cited on page 31)

[Buc81] R. Bucklein. The Audibility of Freqzuency Response Irregularities. *(1962), Reprinted in J. Audio Eng. Soc.*, 36:126–131, March 1981. (Cited on page 80)

[Bur81] J.C. Burgess. Active Adaptive Sound Control in a Duct: A Computer Simulation. *Journal of the Acoustical Society of America (JASA)*, 70:715–726, September 1981. (Cited on page 105)

[BYR02] S. R. Mahadeva Prasanna B. Yegnanarayana and K. Sreenivasa Rao. Speech Enhancement Using Excitation Source Information. In *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 541–544, May 2002. (Cited on page 68)

[Cap94] O. Cappe. Elimination of the Musical Noise Phenomenon with the Ephraim and Malah Noise Suppressor. *IEEE Trans. on Speech and Audio Processing*, 2(2):345–349, 1994. (Cited on page 30)

[CGD12] B. Cauchi, S. Goetze, and S. Doclo. Reduction of Non-stationary Noise for a Robotic Living Assistant using Sparse Non-negative Matrix Factorization. In *Proc. Speech and Multimodal Interaction in Assistive Environments (SMIAE 2012)*, Jeju Island, Republic of Korea, Jul. 2012. (Cited on page 5)

[CKM06] J. Chang, N. S. Kim, and S. K. Mitra. Voice Activity Detection based on Multiple Statistical Models. *IEEE Trans. on Signal Processing*, 54(6):1965, 2006. (Cited on page 31)

[CLD01] T.J. Cox, F. Li, and P. Dalington. Extracting Room Reverberation Time from Speech Using Artificial Neural Networks. *Journal of the Acoustical Society of America (JASA)*, 94(4):219–230, 2001. (Cited on page 92)

[CMS96] Y. Mahieux C. Marro and K. U. Simmer. Preformance of Adaptive Dereverberation Techniques using Directivity Controlled Arrays. In *Proc. EURASIP European Signal Processing Conference (EUSIPCO)*, pages 1127–1130, Trieste, Italy, September 1996. (Cited on pages 67 and 68)

[CMS98] Y. Mahieux C. Marro and K. U. Simmer. Analysis of Noise Reduction and Dereverberation Techniques Based on Microphone Arrays with Postfiltering. *IEEE Trans. on Speech and Audio Processing*, 6(3):240–259, May 1998. (Cited on page 68)

[DD05] H. Deng and M. Doroslovački. Improving Convergence of the PNLMS Algorithm for Sparse Impulse Response Identification. *IEEE Signal Processing Letters*, 12(3):181–184, March 2005. (Cited on page 43)

[DM01] S. Doclo and M. Moonen. Combined Frequency Domain Dereverberation and Noise Reduction Technique for Multi-Microphone Speech Enhancement. In *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, pages 31–34, Darmstadt, Germany, 10.-13. September 2001. (Cited on pages 67 and 69)

[DMDC00] S. Doclo, M. Moonen, and E. De Clippel. Combined Acoustic Echo and Noise Reduction using GSVD-based Optimal Filtering.

In *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 1051–1054, Istanbul, Turkey, June 2000. (Cited on page 30)

[DMW78] M. Dentino, J. McCool, and B. Widrow. Adaptive Filtering in the Frequency Domain. *Proc. IEEE*, 66(12):1658–1659, December 1978. (Cited on pages 18 and 29)

[Dob06] G. Doblinger. Localization and Tracking of Acoustical Sources. In *Topics in Acoustic Echo and Noise Control*, chapter 6, pages 91 – 122. Springer, Berlin - Heidelberg, 2006. (Cited on pages 67 and 68)

[Dou95] S. Douglas. The Fast Affine Projection Algorithms for Active Noise Control. In *Proc. Asilomar Conf. on Signals, Systems, and Computers*, pages 1245–1249, Pacific Grove, USA, October 1995. (Cited on page 108)

[Dou96] S. Douglas. Efficient Approximate Implementation of the Fast Affine Projection Algorithm Using Orthogonal Transforms. In *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Proc. (ICASSP)*, pages 1656–1659, May 1996. (Cited on page 29)

[Dou97] S. Douglas. Fast, Exact Filtered-X LMS and LMS Algorithms for Multichannel Active Noise Control. In *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 399–402, Munich, Germany, April 1997. (Cited on pages 107 and 108)

[DPK96] T. Dau, D. Püschel, and A Kohlrausch. A Quantitative Model of the Effective Signal Processing in the Auditory System: I. Model Structure. *Journal of the Acoustical Society of America (JASA)*, 99(6):3615–3622, June 1996. (Cited on pages 187 and 190)

[DS84] T.G. Dolan and A.M. Small. Frequency Effects in Backward Masking. *Journal of the Acoustical Society of America (JASA)*, 75:932–936, March 1984. (Cited on page 119)

[Dut00] D.L. Duttweiler. Proportionate Normalized Least-Mean-Squares Adaptation in Echo Cancelers. *ITSAP*, 8(5):508–517, 2000. (Cited on pages 39, 40, 41, 42, and 146)

[EM85] Y. Ephraim and D. Malah. Speech enhancement using a minimum mean–square error log-spectral amplitude estimator. *IEEE Trans. on Acoustics, Speech and Signal Processing*, 33(2):443–445, 1985. (Cited on page 30)

[EM06] G. Evangelopoulos and P. Maragos. Multiband Modulation Energy Tracking for Noisy Speech Detection. *IEEE Transactions on Audio, Speech, and Language Processing*, 14(6):2024–2038, 2006. (Cited on page 31)

[EMV02] G. Enzner, R. Martin, and P. Vary. Unbiased Residual Echo Power Estimation for Hands–Free Telephony. In *Proc. IEEE Int. Conference on Acoustics, Speech and Signal Processing (ICASSP-2002)*, Orlando, Florida, USA, May 2002. (Cited on page 30)

[EN89] S. J. Elliott and P. A. Nelson. Multiple-Point Equalization in a Room Using Adaptive Digital Filters. *Journal of the Audio Engineering Society*, 37(11):899–907, November 1989. (Cited on pages 62, 63, 65, 66, 97, and 140)

[Enz08] G. Enzner. Kalman Filtering in Acoustic Echo Control: A Smooth Ride on a Rocky Road. In R. Martin, U. Heute, and C. Antweiler, editors, *Advances in Digital Speech Transmission*, chapter 4, pages 79–106. John Wiley & Sons, West Sussex, England, 2008. (Cited on pages 28 and 30)

[ESN87] S.J. Elliott, I.M. Stothers, and P.A. Nelson. A Multiple Error LMS Algorithm and its Application to the Active Control of Sound and Vibration. *IEEE Trans. on Acoustics, Speech and Signal Processing*, ASSP-35(10):1423–1434, October 1987. (Cited on page 108)

[EV03] G. Enzner and P. Vary. Robust and Elegant, Purely Statistical Adaptation of Acoustic Echo Canceler and Postfilter. In *International Workshop on Acoustic Echo and Noise Control (IWAENC)*, pages 43–46, Kyoto, Japan, Sep 2003. (Cited on page 30)

[Fal03] C. Faller. Perceptually Motivated Low Complexity Acoustic Echo Control. In *Preprint 114th Conv. Aud. Eng. Soc.*, Amsterdam, The Netherlands, March 2003. (Cited on pages 30, 53, and 54)

[Fal08] T.H. Falk. *Blind Estimation of Perceptual Quality for Modern Speech Communications*. PhD thesis, Queen's University, Kingston, Ont., USA, January 2008. (Cited on pages 71, 186, 187, and 188)

[FB95a] G. Faucon and R. Le Bouquin-Jeannes. Joint System for Echo Cancellation and Noise Reduction. In *European Conf. on Speech Communication and Technology (EUROSPEECH)*, pages 1525–1528, Madrid, Spain, Sep 1995. (Cited on page 30)

[FB95b] G. Faucon and R. Le Bouquin-Jeannes. Joint System for Echo Cancellation and Noise Reduction. In *Proc. ESCA European Conference on Speech Communication and Technology (EUROSPEECH)*, pages 1525–1528, Madrid, Spain, September 1995. (Cited on page 51)

[FC05] C. Faller and J. Chen. Suppressing Acoustic Echo in a Spectral Envelope Space. *IEEE Trans. on Speech and Audio Processing*, 13(5):1048–1062, September 2005. (Cited on page 30)

[FC08] T.H. Falk and W.-Y. Chan. A Non-Intrusive Quality Measure of Dereverberated Speech. In *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, Seattle, USA, September 2008. (Cited on pages 71 and 186)

[Fer80] E.R. Ferrara. Fast Implementations of LMS Adaptive Filter. *IEEE Trans. on Acoustics, Speech and Signal Processing*, 28(4):474–475, August 1980. (Cited on page 18)

[FFK+08a] A. Favrot, C. Faller, M. Kallinger, F. Kuech, and M. Schmidt. Acoustic echo control based on temporal fluctuations of short-time spectra. In *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, Seattle, USA, September 2008. (Cited on pages 30 and 33)

[FFK+08b] A. Favrot, C. Faller, M. Kallinger, F. Kuech, and M. Schmidt. Acoustic Echo Control based on Temporal Fluctuations of Short-Time Spectra. In *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, Seattle, WA, USA, September 2008. (Cited on page 54)

[Fie01] L. D. Fielder. Practical Limits for Room Equalization. In *Proc. AES Convention (Audio Engineering Society)*, volume 111, pages 1 – 20, New York, NY, USA, September 2001. (Cited on pages 66, 67, 118, and 119)

[FM73] D. D. Falconer and F. R. Magee. Adaptive Channel Memory Truncation for Maximum Likelihood Sequence Estimation. *The Bell System Technical Journal*, 52(9):1541–1562, November 1973. (Cited on pages 63, 66, and 122)

[Fry75] P.A. Fryer. Intermodulation Distortion Listening Tests. In *50th Convention, Audio Engineering Society*, London, UK, February 1975. (Cited on page 80)

[FT05] C. Faller and C. Tournery. Estimating the Delay and Coloration Effect of the Acoustic Echo. In *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, pages 53–56, Eindhoven, The Netherlands, September 2005. (Cited on page 30)

[FZ07] H. Fastl and E. Zwicker. *Psychoacoustics: Facts and Models*. Springer, Berlin, 3. edition, 2007. (Cited on pages 118 and 119)

[FZC10] T.H. Falk, C. Zheng, and W.-Y. Chan. A Non-Intrusive Quality and Intelligibility Measure of Reverberant and Dereverberated Speech. *IEEE Transactions on Audio, Speech, and Language Processing*, 18(7):1766–1774, 2010. (Cited on page 186)

[GA03] B.W. Gillespie and L.E. Atlas. Strategier for Improving Audible Quality and Speech Recognition Accuracy of Reverberant Speech. In *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 676–679, April 2003. (Cited on page 68)

[GAK+10] S. Goetze, E. Albertin, M. Kallinger, A. Mertins, and K.-D. Kammeyer. Quality Assessment for Listening- Room Compensation Algorithms. In *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, Dallas, Texas, USA, March 2010. (Cited on pages 5, 67, 71, and 72)

[GAR10a] Sound samples, correlation patterns, and MATLAB code for quality assessment available online at http://www.ant.uni-bremen.de/∼goetze/aes2010/, 2010. (Cited on page 74)

[GAR+10b] S. Goetze, E. Albertin, J. Rennies, E.A.P. Habets, and K.-D. Kammeyer. Speech Quality Assessment for Listening-Room Compensation. In *38th AES Conference*, pages 11–20, Pitea, Sweden, July 2010. (Cited on pages 5, 67, 71, and 72)

[GAR+14] S. Goetze, E. Albertin, J. Rennies, E.A.P. Habets, and K.-D. Kammeyer. Speech Quality Assessment for Listening-Room Compensation. *J. Audio Eng. Soc.*, 62(6):386–399, June 2014. (Cited on page 5)

[Gau03] M. Gauger. An Improved Method for Stereo Acoustic Echo Cancelling. In *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 596–599, April 2003. (Cited on page 33)

[Gay98] S.L. Gay. An Efficient, Fast Converging Adaptive Filter for Network Echo Cancellation. In *Proc. Asilomar Conf. on Signals, Systems, and Computers*, pages 394–398, Pacific Grove, CA, USA, November 1998. (Cited on pages 41 and 43)

[GB99] S. M. Griebel and M. S. Brandstein. Wavelet Transform Extrema Clustering for Multi-Channel Speech Dereverberation. In *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, Pocono Manor, PA, USA, September 1999. (Cited on page 72)

[GB00] T. Gänsler and J. Benesty. Stereophonic Echo Cancellation and Two-Channel Adaptive Filtering: An Overview. *Int. Journal of Adaptive Control and Signal Processing*, 14:565–586, September 2000. (Cited on page 33)

[GB01a] T. Gänsler and J. Benesty. A Frequency-Domain Double-Talk Detector Based on a Normalized Cross-Correlation Vector. *EURASIP Signal Processing*, 81(8):1783–1787, 2001. (Cited on page 31)

[GB01b] S.L. Gay and J. Benesty, editors. *Acoustic Signal Processing for Telecommunication*. Kluwer Academic Publishers, Norwell, USA, 2. edition, 2001. (Cited on pages 28 and 31)

[GB01c] S.M. Griebel and M.S. Brandstein. Microphone Array Speech Dereverberation using Coarse Channel Modeling. In *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 201–204, May 2001. (Cited on page 68)

[GB02] T. Gänsler and J. Benesty. New Insights into the Stereophonic Echo Cancellation Problem and an Adaptive Nonlinearity Solution. *IEEE Trans. on Speech and Audio Processing*, 10(5):257–267, jul 2002. (Cited on page 33)

[GBGS00] T. Gänsler, J. Benesty, S.L. Gay, and M.M. Sondhi. A Robust Proportionate Affine Projection Algorithm for Network Echo Cancellation. In *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, Istanbul, Turkey, June 2000. (Cited on page 44)

[GBN05] N. D. Gaubitch, J. Benesty, and P. A. Naylor. Adaptive Common Root Estimation and the Common Zeros Problem in Blind Channel Identification. In *Proc. EURASIP European Signal Processing Conference (EUSIPCO)*, Antalya, Turkey, September 2005. (Cited on page 66)

[Ged98] E. R. Geddes. Small Room Acoustics in the Statistical Region. In *Proc. of the AES International Conference*, volume 15, pages 51–59, Copenhagen, Denmark, October 1998. (Cited on page 66)

[GGBD11] S. Gerlach, S. Goetze, J. Bitzer, and S. Doclo. Robustness Results on Multi-Speaker Position Estimation using a Modified PoPi-Algorithm. In *Proc. 37th Annual Convention for Acoustics (DAGA)*, pages 633–634, Düsseldorf, Germany, March 2011. (Cited on page 5)

[GGD12] S. Gerlach, S. Goetze, and S. Doclo. 2D Audio-Visual Localization in Home Environments using Particle Filter. In *10. ITG Fachtagung Sprachkommunikation*, Braunschweig, Germany, Sep. 2012. (Cited on page 5)

[Gie88] H.W. Gierlich. Verfahren zur Sprachdetektion unter Störschalleinfluß. In *Digitale Sprachverarbeitung –Prinzipien und Anwendungen, ITG–Fachbericht 105*, pages 57–62, Bad Nauheim, Germany, Oct 1988. (Cited on page 30)

[GKK05] S. Goetze, M. Kallinger, and K.-D. Kammeyer. Residual Echo Power Spectral Density Estimation Based on an Optimal Smoothed Misalignment For Acoustic Echo Cancelation. In *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC-2005),*

Eindhoven, The Netherlands, pages 209–212, September 2005. (Cited on pages 5, 30, 51, and 54)

[GKKM07] S. Goetze, M. Kallinger, K.-D. Kammeyer, and A. Mertins. Spatial Sensitivity for Listening Room Compensation. In *Demonstration given at IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New Paltz, USA, October 2007. (Cited on page 5)

[GKMK06a] S. Goetze, M. Kallinger, A. Mertins, and K.-D. Kammeyer. A Study on Combining Acoustic Echo Cancelers with Impulse Response Shortening. *Abstract in Proc. 4th Joint Meeting of ASA and ASJ, Honolulu, Nov. 2006, HI, USA, in J. Acoust. Soc. Am.*, 120(5):3258, Nov. 2006. (Cited on page 5)

[GKMK06b] S. Goetze, M. Kallinger, A. Mertins, and K.-D. Kammeyer. Enhanced Partitioned Stereo Residual Echo Estimation. In *Proc. Asilomar Conf. on Signals, Systems, and Computers*, pages 1326–1330, Pacific Grove, CA, USA, October 2006. (Cited on pages 5, 28, 29, 30, 33, 51, 52, and 141)

[GKMK07] S. Goetze, M. Kallinger, A. Mertins, and K.-D. Kammeyer. Least Squares Equalizer Design under Consideration of Tail Effects. In *Proc. German Annual Conference on Acoustics (DAGA)*, pages 599–600, Stuttgart, Germany, March 2007. (Cited on pages 5, 28, 69, and 134)

[GKMK08a] S. Goetze, M. Kallinger, A. Mertins, and K.-D. Kammeyer. A Decoupled Filtered-X LMS Algorithm for Listening-Room Compensation. In *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, Seattle, USA, September 2008. (Cited on pages 5, 69, and 105)

[GKMK08b] S. Goetze, M. Kallinger, A. Mertins, and K.-D. Kammeyer. Multi-Channel Listening-Room Compensation using a Decoupled Filtered-X LMS Algorithm. In *Proc. Asilomar Conf. on Signals, Systems, and Computers*, pages 811–815, Pacific Grove, USA, October 2008. (Cited on pages 5, 6, 29, 65, 66, 69, 72, 92, 96, and 105)

[GKMK08c] S. Goetze, M. Kallinger, A. Mertins, and K.-D. Kammeyer. Room Impulse Response Shaping based on Estimates of Room Impulse Responses. In *Proc. German Annual Conference on Acoustics (DAGA)*, pages 829–830, Dresden, Germany, March 2008. (Cited on pages 6, 65, 66, 69, and 96)

[GKMK08d] S. Goetze, M. Kallinger, A. Mertins, and K.-D. Kammeyer. System Identification for Multi-Channel Listening-Room Com-

pensation using an Acoustic Echo Canceller. In *Workshop on Hands-free Speech Communication and Microphone Arrays (HSCMA)*, pages 224–227, Trento, Italy, May 2008. (Cited on pages 5, 6, 10, 59, 62, 63, 65, 67, 69, 88, 96, 133, and 168)

[GKMK09] S. Goetze, M. Kallinger, A. Mertins, and K.-D. Kammeyer. Estimation of the Optimum System Delay for Speech Dereverberation by Inverse Filtering. In *Int. Conf. on Acoustics (NAG/DAGA 2009)*, pages 976–979, Rotterdam, The Netherlands, March 2009. (Cited on pages 5, 13, 69, and 92)

[GMA+10] S. Goetze, N. Moritz, J.-E. Appell, M. Meis, C. Bartsch, and J. Bitzer. Acoustic User Interfaces for Ambient Assisted Living Technologies. *Informatics for Health and Social Care, SI Ageing & Technology*, 35(4):161–179, December 2010. (Cited on page 5)

[GMK06a] S. Goetze, V. Mildner, and K.-D. Kammeyer. A Psychoacoustic Noise Reduction Approach for Stereo Hands-Free Systems. In *Audio Engineering Society (AES), 120th Convention*, Paris, France, 20.-23. May 2006. (Cited on pages 5, 30, 53, 54, 67, and 118)

[GMK06b] S. Goetze, V. Mildner, and K.-D. Kammeyer. Comparison of Speech Enhancement Systems for Noise Fields in a Car Environment. In *German 32. Deutsche Jahrestagung für Akustik (DAGA'06)*, pages 45–46, Braunschweig, Germany, March 2006. (Cited on pages 5, 30, and 68)

[GMV98] S. Gustfsson, R. Martin, and P. Vary. Combined Acoustic Echo Control and Noise Reduction for Hands–Free Telephony. *Signal Processing*, 64:21–32, 1998. (Cited on pages 30 and 51)

[GRA10] S. Goetze, J. Rennies, and J.-E. Appell. Intelligente Konferenzsysteme für natürliche Freisprechkommunikation. In A. Schick, M. Meis, and C. Nocke, editors, *Beiträge zur psychologischen Akustik, Akustik in Büro und Objekt*. Isensee Verlag, Oldenburg, 1. edition, 2010. (Cited on page 5)

[GRH+08] S. Goetze, T. Rohdenburg, V. Hohmann, B. Kollmeier, and K.-D. Kammeyer. Direction of Arrival Estimation based on the Dual Delay Line Approach for Binaural Hearing Aid Microphone Arrays. In *Int. Symposium on Intelligent Signal Processing and Communication Systems (ISPACS)*, pages 185–188, Xiamen, China, November 2008. (Cited on pages 5, 67, and 68)

[GS99] S. Gustafsson and F. Schwarz. A Postfilter for Improved Stereo Acoustic Echo Cancellation. In *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, pages 32–35, Pocono Manor, Pennsylvania, Sep 1999. (Cited on page 30)

[GSG+12] S. Goetze, J. Schröder, S. Gerlach, D. Hollosi, J.-E. Appell, and F. Wallhoff. Acoustic Monitoring and Localization for Social Care. *Journal of Computing Science and Engineering (JCSE), SI on uHealthcare*, 6(1):40–50, March 2012. (Cited on page 5)

[GSO98] J. Gonzales-Rodriguez, J.L Sanchez-Bote, and J. Ortega-Garcia. Speech Dereverberation and Noise Reduction with a Combined Microphone Array Approach. In *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 3613–3616, 1998. (Cited on page 67)

[GT95] S.L. Gay and S. Tavathia. The Fast Affine Projection Algorithm. In *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Proc. (ICASSP)*, pages 3023–3026, May 1995. (Cited on page 29)

[GT98] A. Gilloire and V. Turbin. Using Auditory Properties to Improve the Behaviour of Stereophonic Acoustic Echo Cancellers. In *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, 1998. (Cited on page 31)

[GTN07] N.D. Gaubitch, M.R.P. Thomas, and P.A. Naylor. Subband Method for Multichannel Least Squares Equalization of Room Transfer Functions. In *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pages 14–17, New Paltz, USA, October 21-24 2007. (Cited on page 97)

[Gus99] S. Gustafsson. *Enhancement of Audio Signals by Combined Acoustic Echo Cancellation and Noise Reduction*. PhD thesis, Aachen University of Technology, Wissenschaftsverlag Mainz, Aachen, June 1999. Aachener Beiträge zu digitalen Nachrichtensystemen, Band 11. (Cited on pages 30, 31, 53, and 118)

[GXJ+11] S. Goetze, F. Xiong, J.O. Jungmann, M. Kallinger, K.-D. Kammeyer, and A. Mertins. System Identification of Equalised Room Impulse Responses by an Acoustic Echo Canceller using Proportionate LMS Algorithms. In *Proc. 130th AES Convention*, London, UK, May 2011. (Cited on pages 5, 6, 59, and 69)

[GXR+10] S. Goetze, F. Xiong, J. Rennies, T. Rohdenburg, and J.-E. Appell. Hands-Free Telecommunication for Elderly Persons Suffering from Hearing Deficiencies. In *12th IEEE International Conference on E-Health Networking, Application and Services (Healthcom'10)*, Lyon, France, July 2010. (Cited on page 5)

[GZ91] J. E. Greenberg and P. M. Zurek. Adaptive Beamformer Performance in Reverberation. In *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, Mohonk, USA, 1991. (Cited on page 67)

[GZ92] J. E. Greenberg and P. M. Zurek. Evaluation of an adaptive beamforming method for hearing aids. *Journal of the Acoustical Society of America (JASA)*, 91(3):1662–1676, March 1992. (Cited on page 31)

[Hab07] E.A.P. Habets. *Single and Multi-Microphone Speech Dereverberation using Spectral Enhancement.* PhD thesis, University of Eindhoven, Eindhoven, The Netherlands, June 2007. (Cited on pages 13, 15, 16, 30, 62, 63, 68, 69, 71, 72, 80, 165, and 168)

[Hab08] E.A.P. Habets. Multi-Microphone Speech Dereverberation using Spectral Enhancement and Statistical Reverberation Models. October 2008. (Cited on pages 68 and 69)

[Hal01] T. Halmrast. Sound Coloration From (Very) Early Reflections. In *Annual Meeting of Acoustical Society of America (ASA)*, Chicago, IL, USA, June 2001. (Cited on page 11)

[Hay02] S. Haykin. *Adaptive Filter Theory.* Prentice Hall, 2002. (Cited on pages 18, 29, 37, 38, 39, 51, 52, 55, 113, and 143)

[HBC06] Y.A. Huang, J. Benesty, and J. Chen. *Acoustic MIMO Signal Processing.* Springer, 2006. (Cited on page 10)

[HBCG09] E.A.P. Habets, J. Benesty, I. Cohen, and S. Gannot. On a Trade-off Between Dereverberation and Noise Reduction using the MVDR Beamformer. In *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 3741–3744, Taipai, Taiwan, April 2009. (Cited on page 67)

[HBG+09] E.A.P. Habets, J. Benesty, S. Gannot, P.A. Naylor, and I. Cohen. On the Application of the LCMV Beamformer to Speech Enhancement. In *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New York, USA, October 2009. (Cited on page 67)

[HBK04] S. Spors H. Buchner and W. Kellermann. Wave-Domain Adaptive Filtering: Acoustic Echo Cancellation for Full-Duplex Systems Based on Wave-Field Synthesis. In *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, volume 4, pages iv–117 – iv–120, May 2004. (Cited on page 69)

[HBNK05a] W. Herbordt, H. Buchner, S. Nakamura, and W. Kellermann. Application of a double-talk resilient dft-domain adaptive filter for binwise stepsize controls to adaptive beamforming. In *Int. Workshop on Nonlinear Signal and Image Processing (NSIP)*, Sapporo, Japan, May 2005. (Cited on page 32)

[HBNK05b] W. Herbordt, H. Buchner, S. Nakamura, and W. Kellermann. Outlier-robust dft-domain adaptive filtering for bin-wise step-size controls, and its application to a generalized sidelobe canceller. In *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, Eindhoven, The Netherlands, September 2005. (Cited on page 32)

[HBSH98] O. Hoshuyama, B. Begasse, A. Sugiyama, and A. Hirano. A Realtime Robust Adaptive Mikrophone Array Controlled by an SNR Estimate. In *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 3605 – 3608, Washington, USA, May 1998. (Cited on page 31)

[HCGS08] E.A.P. Habets, I. Cohen, S. Gannot, and P.C.W. Sommen. Joint Dereverberation and Residual Echo Suppression of Speech Signals in Noisy Environments. *IEEE Trans. on Audio, Speech and Language Processing*, 16(8):1433–1451, November 2008. (Cited on pages 68 and 69)

[HDM07] T. Hikichi, M. Delcroix, and M. Miyoshi. Inverse Filtering for Speech Dereverberation Less Sensitive to Noise and Room Transfer Function Fluctuations. *EURASIP J. on Advances in Signal Processing*, Volume 2007, Article ID 34013, 2007. doi:10.1155/2007/34013. (Cited on pages 65, 67, 97, 136, and 137)

[HE08] E. Hänsler and G. Schmidt (Eds.). *Speech and Audio Processing in Adverse Environments*. Springer, 2008. (Cited on pages 63 and 68)

[Her05] W. Herbordt. *Sound Capture for Human / Machine Interfaces – Practical Aspects of Microphone Array Signal Processing*. Springer, Berlin, Heidelberg, New York, 2005. (Cited on pages 30, 31, 32, 33, 53, 56, and 114)

[HGS04] O. Hoshuyama, R.A. Goubran, and A. Sugiyama. A Generalized Proportionate Variable Step-Size Algorithm for Fast Changing Acoustic Environments. In *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, pages IV–161–IV–164, Quebec, Canada, May 2004. (Cited on page 44)

[Hie95] P. Hietkämper. Optimization of an Acoustic Echo Canceller Combined with Adaptive Gain Control. In *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 3047–3050, May 1995. (Cited on page 31)

[HK06] R. Huber and B. Kollmeier. PEMO-Q - A New Method for Objective Audio Quality Assessment using a Model of Auditory Perception. *IEEE Trans. on Audio, Speech and Language Processing*,

14(6), 2006. Special Issue on Objectiv Quality Assessment of Speech and Audio. (Cited on pages 80, 190, and 191)

[HKN04] W. Herbordt, W. Kellermann, and S. Nakamura. Joint Optimization of LCMV Beamforming and Acoustic Echo Cancellation. In *Proc. EURASIP European Signal Processing Conference (EUSIPCO)*, pages 2003–2006, Vienna, Austria, September 2004. (Cited on page 30)

[HMK94] Y. Haneda, S. Makino, and Y. Kaneda. Common Acoustical Pole and Zero Modeling of Room Transfer Functions. *IEEE Trans. on Speech and Audio Processing*, 2(2):320–328, April 1994. (Cited on page 65)

[HMK97] Y. Haneda, S. Makino, and Y. Kaneda. Mutliple-Point Equalization of Room Transfer Functions by Common Acoustical Poles. *IEEE Trans. on Speech and Audio Processing*, 5(4):325–333, July 1997. (Cited on page 65)

[Hän92] E. Hänsler. The Hands–Free Telephone Problem: An Annotated Bibliography. *Signal Processing*, 27(3):259–271, 1992. (Cited on pages 28 and 61)

[Hän94] E. Hänsler. The Hands–Free Telephone Problem: An Annotated Bibliography Update. *Annals of Telecommunications*, 49(7–8):360–367, 1994. (Cited on page 28)

[Hän95] E. Hänsler. The Hands–Free Telephone Problem: A Second Annotated Bibliography Update. *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, pages 115–134, June 1995. (Cited on page 28)

[Hän97] E. Hänsler. From Algorithms to Systems – It's a Rocky Road. In *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, pages K1–K8, Sep 1997. (Cited on page 28)

[HNK04] W. Herbordt, S. Nakamura, and W. Kellermann. Multichannel estimation of the power spectral density of noise for mixtures of nonstationary signals. In *IPSJ SIG Technical Reports*, volume 131, pages 211 – 216, Kyoto, Japan, December 2004. (Cited on page 32)

[HP98] J.H.L. Hansen and B. Pellom. An Effective Quality Evaluation Protocol for Speech Enhancement Algorithms. In *Proc. Int. Conf. on Spoken Language Processing (ICSLP)*, volume 7, pages 2819–2822, Sydney, Australia, December 1998. (Cited on page 172)

[HS99] O. Hoshuyama and A. Sugiyama. An Adaptive Microphone Array with Good Sound Quality using Auxiliary Fixed Beamformers and

its DSP Implementation. In *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 949 – 952, Phoenix, USA, March 1999. (Cited on page 31)

[HS00] E. Hänsler and G. U. Schmidt. Hands-Free Telephones – Joint Control of Echo Cancellation and Postfiltering. *EURASIP Signal Processing*, 80(11):2295–2305, November 2000. (Cited on pages 30 and 51)

[HS04] E. Hänsler and G. Schmidt. *Acoustic Echo and Noise Control: a Practical Approach*. Wiley, Hoboken, 2004. (Cited on pages 28, 29, 30, 37, and 51)

[HSGA10] D. Hollosi, J. Schröder, S. Goetze, and J.-E. Appell. Voice Activity Detection Driven Acoustic Event Classification for Monitoring in Smart Homes. In *3rd International Symposium on Applied Sciences in Biomedical and Communication Technologies (Best Paper Award)*, Rome, Italy, November 2010. (Cited on pages 5 and 31)

[HTK03] W. Herbordt, T. Trini, and W. Kellermann. Robust spatial estimation of the signal-to-interference ratio for non-stationary mixtures. In *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, pages 247–250, Kyoto, Japan, September 2003. (Cited on page 32)

[Hub03] R. Huber. *Objective Assessment of Audio Quality Using an Auditory Processing Model*. PhD thesis, University of Oldenburg, 2003. (Cited on page 191)

[IEC98] IEC. Sound System Equipment - Part 16: Objective Rating of Speech Intelligibility by Speech Transmission Index, 1998. (Cited on page 61)

[IG07] M.A. Iqbal and S.L. Grant. Novel and Effcient Download Test for Two Path Echo Canceller. In *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New Paltz, NY, USA, October 2007. (Cited on page 31)

[IK97] M. Ihle and K. Kroschel. Integration of Noise Reduction and Echo Attenuation for Handset–Free Communication. In *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, pages 69–72, London, Great Britain, Sep 1997. (Cited on page 31)

[Int92] International Organization for Standardization. *Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to about 1.5 MBit/s, Audio Part (11172-3)*, November 1992. (Cited on pages 30 and 175)

[ISO97] ISO Norm 3382: Acoustics – Measurement of the Reverberation Time of Rooms with Reference to other Acoustical Parameters. *International Organization for Standardization (ISO), Geneva, Switzerland*, 1997. (Cited on pages 66, 161, and 163)

[ISO03] ISO Norm 226:2003: Acoustics – Normal Equal-Loudness-Level Contours. *International Organization for Standardization (ISO), Geneva, Switzerland*, 2003. (Cited on pages 176 and 177)

[ISO06a] ISO Norm 3382-1: Acoustics – Measurement of Room Acoustic Parameters – Part 1: Performance Rooms (iso/dis 3382-1:2006). *International Organization for Standardization (ISO), Geneva, Switzerland*, 2006. (Cited on page 161)

[ISO06b] ISO Norm 3382-2: Acoustics – Measurement of Room Acoustic Parameters – Part 2: Reverberation Time in Ordinary Rooms (iso/dis 3382-2:2006). *International Organization for Standardization (ISO), Geneva, Switzerland*, 2006. (Cited on page 161)

[ITU88] ITU-T P.30 Transmission Performance of Group Audio Terminals (GATs), ITU-T Recommendation P.30, 1988. (Cited on pages 29 and 39)

[ITU93a] ITU-T G.167 General Characteristic of International Telephone Connections and International Telephone Circuits – Acoustic Echo Controllers, ITU-T Recommendation G.167, 1993. (Cited on pages 29 and 39)

[ITU93b] ITU-T P.34 Transmission Characteristics of Hands-Free Telephones, ITU-T Recommendation P.34, 1993. (Cited on pages 29 and 39)

[ITU96] ITU-T P.800. Methods for Subjective Determination of Transmission Quality, ITU-T Recommendation P.8800, November 1996. (Cited on page 74)

[ITU01] ITU-T P.862. Perceptual Evaluation of Speech Quality (PESQ), an Objective Method for End-to-End Speech Quality Assessment of Narrowband Telephone Networks and Speech Codecs, ITU-T Recommendation P.862, February 2001. (Cited on pages 189 and 190)

[ITU03] ITU-T P.835. Subjective Test Methology for Evaluating Speech Communication Systems that Include Noise Suppression Algorithm, ITU-T Recommendation P.835, November 2003. (Cited on page 74)

[Jet79] J.J. Jetzt. Critical Distance Measurement of Rooms from the Sound Energy Spectral Response. *Journal of the Acoustical Society of America (JASA)*, 65(5):1204–1211, May 1979. (Cited on page 80)

[JGM11] J.O. Jungmann, S. Goetze, and A. Mertins. Room Impulse Response Reshaping by p-Norm Optimization based on Estimates of Room Impulse Responses. In *Proc. 37th Annual Convention for Acoustics (DAGA)*, pages 611–612, Düsseldorf, Germany, March 2011. (Cited on page 5)

[JMGM11] J.O. Jungmann, T. Mei, S. Goetze, and A. Mertins. Room Impulse Response Reshaping by Joint Optimization of Multiple P-Norm Based Criteria. In *Proc. 19th European Signal Processing Conference (EUSIPCO)*, pages 1658–1662, Barcelona, Spain, Aug. 2011. (Cited on pages 5, 67, 119, 122, and 126)

[JMM12] J. O. Jungmann, R. Mazur, and A. Mertins. Robust listening room compensation by optimizing multiple p-norm based criteria. In *Proc. German Annual Conference on Acoustics (DAGA)*, Darmstadt, Germany, Mar. 2012. (Cited on page 126)

[Joh88] J. D. Johnston. Transform Coding of Audio Signals using Perceptual Noise Criteria. *IEEE Journal on Selected Areas in Communication*, 6(2):314–232, February 1988. (Cited on page 167)

[Joy75] W.B. Joyce. Sabine's Reverberation Time and Ergodic Auditoriums. *Journal of the Acoustical Society of America (JASA)*, 58(3):643–655, 1975. (Cited on page 15)

[JS98] Y. Joncour and A. Sugiyama. A Stereo Echo canceler with Pre-Processing for Correct Echo-Path Identification. In *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 3677–3682, Seattle, USA, May 1998. (Cited on page 33)

[Kal07] M. Kallinger. *Neue Ansätze zur Unterdrückung akustischer Echos unter Einbeziehung einer Stereo-Sprachübertragung.* PhD thesis, Arbeitsbereich Nachrichtentechnik, Universität Bremen (FB-1), Shaker Verlag, Aachen, Deutschland, 2007. In German language. (Cited on pages 28, 30, 33, 54, and 134)

[Kam94] K. D. Kammeyer. Time Truncation of Channel Impulse Responses by Linear Filtering: A Method to Reduce the Complexity of Viterbi Equalization. *Archiv für Elektronik und Übertragungstechnik (AEÜ) – Int. Journal of Electronics and Communications*, 48(5):237–243, May 1994. (Cited on pages 63, 66, and 122)

[Kam08] K.-D. Kammeyer. *Nachrichtenübertragung.* Vieweg+Teubner, Wiesbaden, Germany, 4. edition, 2008. In German language. (Cited on pages 18, 66, and 137)

[KBK03a] M. Kallinger, J. Bitzer, and K. D. Kammeyer. Multi-Microphone Residual Echo Estimation. In *Proc. IEEE Int. Conf. on Acoustics,*

Speech, and Signal Processing (ICASSP), Hong Kong, China, Apr 2003. (Cited on pages 28 and 30)

[KBK03b] M. Kallinger, J. Bitzer, and K. D. Kammeyer. Post-Filtering for Stereo Acoustic Echo Cancellation. In *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, Kyoto, Japan, Sep 2003. (Cited on page 30)

[KC76] C. H. Knapp and G. C. Carter. The Generalized Correlation Method for Estimation of Time Delay. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 24(4):320–327, Aug. 1976. (Cited on pages 67 and 68)

[KD12] I. Kodrasi and S. Doclo. Robust Partial Multichannel Equalization Techniques for Speech Dereverberation. In *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, Kyoto, Japan, March 2012. Submitted to. (Cited on page 65)

[Kel88] W. Kellermann. Analysis and Design of Multirate Systems for Cancelling of Acoustic Echoes. In *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Proc. (ICASSP)*, pages 2570–2573, April 1988. (Cited on page 28)

[Kel97] W. Kellermann. Strategies for Combining Acoustic Echo Cancellation and Adaptive Beamforming Microphone Arrays. In *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Proc. (ICASSP)*, volume 1, pages 219–222, Munich, Germany, April 1997. (Cited on pages 30 and 32)

[Kel01] W. L. Kellermann. Acoustic Echo Cancellation for Beamforming Microphone Arrays. In M. S. Brandstein and D. Ward, editors, *Microphone Arrays: Signal Processing Techniques and Applications*, chapter 13, pages 281–306. Springer-Verlag, 2001. (Cited on page 30)

[KGD12a] I. Kodrasi, S. Goetze, and S. Doclo. Increasing the Robustness of Acoustic Multichannel Equalization by Means of Regularization. In *International Workshop on Acoustic Signal Enhancement (IWAENC 2012)*, Aachen, Germany, Sep. 2012. (Cited on pages 5, 67, and 122)

[KGD12b] I. Kodrasi, S. Goetze, and S. Doclo. Non-intrusive Regularization for Least-Squares Multichannel Equalization for Speech Dereverberation. In *2012 IEEE 27-th Convention of Electrical and Electronics Engineers in Israel*, Eilat, Israel, Nov. 2012. (Cited on pages 5 and 122)

[KGD13a] I. Kodrasi, S. Goetze, and S. Doclo. A Perceptually Constrained Channel Shortening Technique for Speech Dereverberation. In

Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP), pages 151–155, Vancouver, Canada, May 2013. (Cited on pages 5 and 122)

[KGD13b] I. Kodrasi, S. Goetze, and S. Doclo. Regularization for Partial Multichannel Equalization for Speech Dereverberation. *Accepted for publication in Trans. on Audio, Speech and Language Processing*, 2013. (Cited on pages 5, 67, 122, 136, and 137)

[KK05a] F. Küch and W. Kellermann. Partitioned Block Frequency-Domain Adaptive Second-Order Volterra Filter. *IEEE Trans. on Signal Processing*, 53(2):564–575, 2005. (Cited on page 32)

[KK05b] Y. Kida and T. Kawahara. Voice Activity Detection based on Optimally Weighted Combination of Multiple Features. In *Proc. Annual Conference of the International Speech Communication Association (INTERSPEECH)*, pages 2621–2624, Lisbon, Portugal, September 2005. (Cited on page 31)

[KK06] F. Küch and W. Kellermann. Nonlinear Acoustic Echo Cancellation. In E. Hänsler and G. Schmidt, editors, *Topics in Acoustic Echo and Noise Control*, pages 205–257. Springer, Heidelberg, Germany, 2006. (Cited on page 32)

[KK09] K.-D. Kammeyer and K. Kroschel. *Digitale Signalverarbeitung (In German langualge)*. Vieweg+Teubner, Wiesbaden, 7. edition, April 2009. (Cited on pages 18, 22, 171, and 187)

[KKS+08] F. Kuech, M. Kallinger, M. Schmidt, C. Faller, and A. Favrot. Acoustic Echo Suppression based on Separation of Stationary and Non-Stationary Echo Components. In *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, Seattle, USA, September 2008. (Cited on page 30)

[Kla82] D. H. Klatt. Prediction of Perceived Phonetic Distance from Critical-Band Spectra: A First Step. In *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 1278–1281, Paris, France, May 1982. (Cited on page 170)

[KLDN08] A.W.H. Khong, X. Lin, M. Doroslovacki, and P.A. Naylor. Frequency Domain Selective Tap Adaptive Algorithms for Sparse System Identification. In *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 229–232, Las Vegas, NV, USA, March 2008. (Cited on page 40)

[KLN08] A.W.H. Khong, X. Lin, and P.A. Naylor. Algorithms for Identifying Clusters of Near-Common Zeros in Multichannel Blind System Identification and Equalization. In *Proc. IEEE Int. Conf. on Acoustics,*

Speech, and Signal Processing (ICASSP), pages 389–392, Las Vegas, NV, USA, March 2008. (Cited on page 66)

[KM96] S.M. Kuo and D.R. Morgan. *Active Noise Control Systems - Algorithms and DSP Implementations*. Wiley Interscience, New York, 1996. (Cited on pages 105 and 107)

[KM05a] M. Kallinger and A. Mertins. Room Impulse Response Shortening by Channel Shortening Concepts. In *Signals, Systems and Computers, 2005. Conference Record of the Thirty-Ninth Asilomar Conference on*, pages 898 – 902, October 28 - November 1 2005. (Cited on page 63)

[KM05b] M. Kallinger and A. Mertins. Room Impulse Response Shortening by Channel Shortening Concepts. In *Proc. Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, CA , USA*, pages 209–212, October 2005. (Cited on pages 66, 67, 122, 123, and 124)

[KM05c] M. Kallinger and A. Mertins. Room Impulse Response Shortening for Listening Room Compensation. In *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC-2005) , Eindhoven, The Netherlands*, pages 197–200, September 2005. (Cited on pages 66 and 122)

[KM06] M. Kallinger and A. Mertins. Room Impulse Response Shaping – A Study. In *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, pages V101–V104, Toulouse, France, 2006. (Cited on pages 66, 74, 80, and 122)

[KM07] M. Kallinger and A. Mertins. A Spatially Robust Least Squares Crosstalk Canceller. In *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, volume 1, pages I–177 – I–180, 15-20 April 2007. (Cited on page 84)

[KMK93] Y. Kaneda, S. Makino, and N. Koizumi. Exponentially Weighted Step–Size NLMS Adaptive Filter Based on the Statistics of a Room Impulse Response. *IEEE Trans. on Speech and Audio Processing*, 1(1):101–108, Jan 1993. (Cited on pages 39 and 40)

[KMK05] M. Kallinger, A. Mertins, and K.-D. Kammeyer. Enhanced Double-Talk Detection Based on Pseudo-Coherence in Stereo. In *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC-2005) , Eindhoven, The Netherlands*, pages 177–180, 12.-15. Sept. 2005. (Cited on pages 30, 31, and 138)

[KN99] O. Kirkeby and P. A. Nelson. Digital Filter Design for Inversion Problems in Sound Reproduction. *Journal of the Audio Engineering Society*, 47(7/8):583–595, Jul./Aug. 1999. (Cited on pages 66 and 67)

[KN04] A. W. H. Khong and P. A. Naylor. Reducing Inter-Channel Coherence in Stereophonic Acoustic Echo Cancellation Using Partial Update Adaptive Filters. In *Proc. EURASIP European Signal Processing Conference (EUSIPCO)*, pages 405–408, Vienna, Austria, Sep 2004. (Cited on page 33)

[KNHOb98] O. Kirkeby, P. A. Nelson, H. Hamada, and F. Orduna-bustamante. Fast Deconvolution of Multichannel Systems using Regularization. *IEEE Trans. on Speech and Audio Processing*, 6(2):189–194, March 1998. (Cited on pages 66, 67, and 110)

[KNKP94] I.S. Kim, H.S. Na, K.J. Kim, and Y. Park. Constraint Filtered-X and Filtered-U Least-Mean-Square Algorithms for the Active Control of Noise in Ducts. *Journal of the Acoustical Society of America (JASA)*, 95(6):3379–3389, June 1994. (Cited on page 107)

[KRF99] O. Kirkeby, P. Rubak, and A. Farina. Analysis of Ill-Conditioning of Multi-Channel Deconvolution Problems. In *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pages 155–158, New Paltz, New York, USA, October 1999. (Cited on page 67)

[Kut00] H. Kuttruff. *Room Acoustics*. Spoon Press, London, 4. edition, 2000. (Cited on pages 9, 16, 61, 64, 93, 161, 162, 163, and 164)

[LCC10] Y. Litvin, I. Cohen, and D. Chazan. Monaural Speech/Music Source Separation using Discrete Energy Separation Algorithm. *Signal Processing*, 90(12):3147–3163, December 2010. (Cited on page 31)

[LE06] K. Lee and D. P. W. Ellis. Voice Activity Detection in Personal Audio Recordings using Autocorrelogram Compensation. In *Proc. Int. Conf. on Spoken Language Processing (ICSLP)*, pages 1970–1973, Pittsburgh, PA, USA, October 2006. (Cited on page 31)

[LGN06] X.S. Lin, N. D. Gaubitch, and P. A. Naylor. Two-Stage Blind Identification of SIMO Systems with Common Zeros. In *Proc. EURASIP European Signal Processing Conference (EUSIPCO)*, Florcence, Italy, September 2006. (Cited on page 66)

[Loi07] P.C. Loizou. *Speech Enhancement: Theory and Practice*. CRC Press Inc., Boca Raton, USA, 2007. (Cited on pages 71, 169, 170, 171, 189, and 190)

[LV08] H. Löllmann and P. Vary. Estimation of the Reverberation Time in Noisy Environments. In *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, Seattla, USA, September 2008. (Cited on page 92)

[LZTZ02] Q. Li, J. Zheng, A. Tsai, and Q. Zhou. Robust Endpoint Detection and Energy Normalization for Real-Time Speech and Speaker Recognition. *IEEE Transactions on Speech and Audio Processing*, 10(3):146–157, 2002. (Cited on page 31)

[MAG95] E. Moulines, O. Ait Amrane, and Y. Grenier. The Generalized Multi Delay Filter: Structure and Convergence Analysis. *IEEE Trans. on Signal Processing*, 43(1):14–28, Jan 1995. (Cited on pages 5, 19, 26, 29, and 110)

[Mar01] R. Martin. Noise Power Spectral Density Estimation Based on Optimal Smoothing and Minimum Statistics. *IEEE Trans. on Speech and Audio Processing*, 9, July 2001. (Cited on page 30)

[MCH82] J. N. Mourjopoulos, P. M. Clarkson, and J.K. Hammond. A Comparative Study of Least-Quares and Homomorphic Techniques for the Inversion of Mixed Phase Signals. In *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 1858–1861, 1982. (Cited on page 64)

[MDEJ03] R.K. Martin, M. Ding, B.L. Evans, and C.R. Johnson. Efficient Channel Shortening Equalizer Design. *EURASIP Journal on Applied Signal Processing*, 2003(13):1279–1290, December 2003. (Cited on page 122)

[Mer99] A. Mertins. Memory Truncation and Crosstalk Cancellation in Transmultiplexers. *IEEE Communications Letters*, 3(6):180–182, June 1999. (Cited on page 66)

[Mer01] A. Mertins. Memory Truncation and Crosstalk Cancellation for Efficient Viterbi Detection in fdma Systems. *Journal of Telecommunications and Information Technology*, 2(3):29–35, December 2001. (Cited on page 66)

[MG82] D. Mansour and A. Gray. Unconstrained Frequency Domain Adaptive Filter. *IEEE Trans. on Acoustics, Speech and Signal Processing*, 30(5):726–734, October 1982. (Cited on pages 18 and 29)

[MGA11] N. Moritz, S. Goetze, and J.-E. Appell. Ambient Voice Control for a Personal Activity and Household Assistant. In R. Wichert and B. Eberhardt, editors, *Ambient Assisted Living - Advanced Technologies and Societal Change, Springer Lecture Notes in Computer Science (LNCS)*, number 978-3-642-18166-5, pages 63–74. Springer Science, January 2011. (Cited on page 5)

[MGK06a] V. Mildner, S. Goetze, and K.-D. Kammeyer. Multi-Channel Noise-Reduction-Systems for Speaker Identification in an Automotive Acoustic Environment. In *Audio Engineering Society (AES), 120th Convention*, Paris, France, May 2006. (Cited on page 5)

[MGK06b] V. Mildner, S. Goetze, and K.-D. Kammeyer. Multi-Channel Speech Enhancement using a Psychoacoustic Approach for a Post-Filter. In *German ITG-Symposium on Speech Communication*, Kiel, Germany, 26.-28. April 2006. (Cited on pages 5 and 30)

[MGK06c] V. Mildner, S. Goetze, and K.-D. Kammeyer. Performance of Text-Independent Speaker Identification considering In-Car Acoustics. In *German 32. Deutsche Jahrestagung für Akustik (DAGA'06)*, pages 223–224, Braunschweig, Germany, 20.-23. March 2006. (Cited on page 5)

[MGKM07] V. Mildner, S. Goetze, K.-D. Kammeyer, and A. Mertins. Optimization of Garbor Features for Text-Independent Speaker Identification. In *Proc. IEEE Int. Symposium on Circuits and Systems (ISCAS)*, pages 3932–3935, New Orleans, USA, 27.-30. May 2007. (Cited on page 5)

[MHB01] D. R. Morgan, J. L. Hall, and J. Benesty. Investigation of Several Types of Nonlinearities for Use in Stereo Acoustic Echo Cancellation. *IEEE Trans. on Speech and Audio Processing*, 9(6):686–696, Sep 2001. (Cited on page 33)

[MJM12] R. Mazur, Jan Ole Jungmann, and A. Mertins. Optimized gradient calculation for room impulse response reshaping algorithm based on p-norm optimization. In *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, Kyoto, Japan, Mar. 2012. (Cited on page 126)

[MK86] M. Miyoshi and Y. Kaneda. Inverse Control of Room Acoustics Using Multiple Loudspeakers and/or Microphones. *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 917–920, April 1986. (Cited on pages 65 and 97)

[MK88] M. Miyoshi and Y. Kaneda. Inverse Filtering of Room Acoustics. *IEEE Trans. on Acoustics, Speech and Signal Processing*, 36(2):145–152, February 1988. (Cited on pages 66, 96, and 97)

[MK02] M. Marzinzik and B. Kollmeier. Speech Pause Detection for Noise Spectrum Estimation by Tracking Power Envelope Dynamics. *IEEE Trans. on Speech and Audio Processing*, 10(2):109–118, February 2002. (Cited on page 31)

[MKM09a] T. Mei, M. Kallinger, and A. Mertins. Room Impulse Response Reshaping/Shortening Based on Least Mean Squares Optimization with Infinity Norm Constraint. In *Proc. 16th International Conference on Digital Signal Processing (DSP 2009)*, pages 1–6, Santorini, Greece, July 2009. (Cited on page 125)

[MKM09b] T. Mei, M. Kallinger, and A. Mertins. Room Impulse Response Shortening with Infinity-Norm Optimization. In *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 3745–3748, Taipei, Taiwan, April 2009. (Cited on pages 66 and 125)

[MM80] R. A. Monzingo and T. W. Miller. *Introduction to adaptive Arrays*. John Wiley and Sons, New York, 1980. (Cited on page 67)

[MM01] S. Müller and P. Massarani. Transfer-Function Measurement with Sweeps. *Journal of the Audio Engineering Society*, 49(6):443–471, June 2001. (Cited on pages 10, 17, 89, and 133)

[MMGP07] H. K. Maganti, P. Motlicek, and D. Gatica-Perez. Proc. ieee int. conf. on acoustics, speech, and signal processing (icassp). In *Unsupervised Speech/Non-Speech Detection for Automatic Speech Recognition in Meeting Rooms*, pages IV–1037 – IV–1040, Honolulu, HI, USA, April 2007. (Cited on page 31)

[MMK10] A. Mertins, T. Mei, and M. Kallinger. Room Impulse Response Shortening/Reshaping with Infinity- and p-Norm Optimization. *IEEE Trans. on Audio, Speech and Language Processing*, 18(2):249–259, February 2010. DOI:10.1109/TASL.2009.2025789. (Cited on pages 66, 67, 74, 119, 125, and 126)

[Moo79] J.A. Moorer. About this Reverberation Business. *Computer Music Journal*, 3(2):13–28, 1979. (Cited on page 13)

[Moo97] B.C.J. Moore. *An Introduction to the Psychology of Hearing*. Academic Press, New York, USA, 4 edition, 1997. (Cited on page 119)

[Moo03] B.C.J. Moore. *An Introduction to the Psychology of Hearing*. Academic Press, Boston, USA, 5 edition, 2003. (Cited on page 118)

[Mou85] J. N. Mourjopoulos. On the Variation and Invertibility of Room Impulse Response Functions. *Journal of Sound and Vibration*, 102(2):217–228, 1985. (Cited on pages 65 and 96)

[Mou94] J. N. Mourjopoulos. Digital Equalization of Room Acoustics. *Journal of the Audio Engineering Society*, 42(11):884–900, November 1994. (Cited on pages 62, 63, 64, 65, and 165)

[MP91] J. N. Mourjopoulos and M. Paraskevas. Pole and Zero Modeling of Room Transfer Functions. *Journal of Sound and Vibration*, 146:281–302, 1991. (Cited on page 65)

[MPS00] A. Mader, H. Puder, and G. Schmidt. Step-Size Control for Acousatic Echo Cancellation Filters – an Overview. *Elsevier Signal Processing*, 80(9):1697–1719, September 2000. (Cited on pages 30, 31, 32, 56, 137, 138, and 141)

[MS97] J. Meyer (Bitzer) and K. U. Simmer. Multi-channel speech enhancement in a car environment using Wiener filtering and spectral subtraction. In *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 1167–1170, Munich, Germany, April 1997. (Cited on page 52)

[MSA+13] N. Moritz, M.R. Schädler, K. Adiloglu, B.T. Meyer, T. Jürgens, T. Gerkmann, B. Kollmeier, S. Doclo, and S. Goetze. Noise Robust Distant Automatic Speech Recognition Utilizing NMF based Source Separation and Auditory Feature Extraction. In *Proc. 2nd International Workshop on Machine Listening in Multisource Environments (CHiME 2013)*, pages 1–6, Vancouver, Canada, June 2013. (Cited on page 5)

[MSS+97] S. Makino, K. Strauss, S. Shimauchi, Y. Haneda, and A. Nakagawa. Subband Stereo Echo Canceller Using the Projection Algorithm with Fast Convergence to the True Echo Path. In *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 299–302, Munich, Germany, Apr 1997. (Cited on page 32)

[MV93] R. Martin and P. Vary. Combined Acoustic Echo Cancellation, Dereverberation and Noise Reduction: A Two Channel Approach. In *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, pages 429–438, Lannion, France, Sep. 1993. (Cited on pages 30 and 68)

[MV94] R. Martin and P. Vary. Combined Acoustic Echo Cancellation, Dereverberation and Noise Reduction: A Two Channel Approach. *Annalesd des Télécommunications*, 49(7-8):429–438, Jul.-Aug. 1994. (Cited on page 68)

[MV96] R. Martin and P. Vary. Combined Acoustic Echo Control and Noise Reduction for Hands–Free Telephony – State of the Art and Perspectives. In *Proc. EURASIP European Signal Processing Conference (EUSIPCO)*, Trieste, Italy, Sep. 1996. (Cited on pages 30 and 51)

[MV08] M. Myllymäki and T. Virtanen. Voice Activity Detection in the Presence of Breathing Noise using Neural Network and Hidden Markov Model. In *Proc. EURASIP European Signal Processing Conference (EUSIPCO)*, Lausanne, Switzerland, August 2008. (Cited on page 31)

[MYR96] P.J.W. Melsa, R.C. Younce, and C.E. Rohrs. Impulse Response Shortening for Discrete Multitone Transceivers. *IEEE Trans. on Communications*, 44(12):1662–1672, December 1996. (Cited on pages 63 and 122)

[NA79] S. T. Neely and J. B. Allen. Invertibility of a Room Impulse Response. *Journal of the Acoustical Society of America (JASA)*, 66:165–169, July 1979. (Cited on pages 18, 64, 65, and 93)

[NCB06] P.A. Naylor, J. Cui, and M. Brookes. Adaptive Algorithms for Sparse Echo Cancellation. *ITSP*, 86(6):1182–1192, 2006. (Cited on pages 39 and 43)

[NG03] D.B. Ward N.D. Gaubitch, P.A. Naylor. On the Use of Linear Prediction for Dereverberation of Speech. In *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, pages 99–102, Kyoto, Japan, September 2003. (Cited on page 68)

[NG05] P.A. Naylor and N.D. Gaubitch. Speech Dereverberation. In *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, Eindhoven, The Netherlands, September 2005. (Cited on pages 61, 71, 88, and 168)

[NGH10] P.A. Naylor, N.D. Gaubitch, and E.A.P. Habets. Signal-Based Performance Evaluation of Dereverberation Algorithms. *Journal of Electrical and Computer Engineering, Article ID 127513*, 2010. (Cited on page 72)

[NGM01] E. Nemer, R. Goubran, and S. Mahmoud. Robust voice activity detection using higher-order statistics inthe lpc residual domain. *IEEE Transactions on Speech and Audio Processing*, 9(3):217–231, 2001. (Cited on page 31)

[Nit00] B. Nitsch. A Frequency-Selective Stepfactor Control for an Adaptive Algorithm working in the Frequency Domain. *Elsevier Signal Processing*, 80(9):1733–1745, September 2000. (Cited on pages 29 and 31)

[NM03] T. Nakatani and M. Miyoshi. Blind Dereverberation of Single Channel Speech Signal Based on Harmonic Structure. In *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 92–95, Hong Kong, China, April 2003. (Cited on page 68)

[NN67] J.I. Nagumo and A. Noda. A Learning Method for System Identification. *IEEE Trans. Autim. Control*, AC-12:282–287, 1967. (Cited on page 39)

[NOBH95] P. A. Nelson, F. Orduna-Bustamante, and H. Hamada. Inverse Filter Design and Equalization Zones in Multichannel Sound Reproduction. *IEEE Trans. on Speech and Audio Processing*, 3(3):185–192, May 1995. (Cited on page 66)

[NTSS04] Y. Nagata, Y. Tatekura, H. Saruwatari, and K. Shikano. Iterative Inverse Filter Relaxion Algorithm for Adaptation to Acoustic

Fluctuations in Sound Reproduction System. *Electronics and Communications in Japan*, 87(7), July 2004. (Cited on page 65)

[OAO97] K. Ochiai, T. Araseki, and T. Ogihara. Echo Canceller with Two Echo Path Models. *IEEE Trans. on Communications*, 25(6):589–595, June 1997. (Cited on page 31)

[OS89] A. V. Oppenheim and R. W. Schafer. *Discrete-Time Signal Processing*. Prentice Hall, second edition, 1989. (Cited on page 64)

[OT89] S. E. Olive and F. E. Toole. The Detection of Reflections in Typical Rooms. *Journal of the Audio Engineering Society*, 37(7/8):539–553, July 1989. (Cited on pages 67 and 119)

[OU84] K. Ozeki and T. Umeda. An Adaptive Filtering Algorithm Using Orthonormal Projection to an Affine Subspace and it's Properties. *Electronic and Communications in Japan*, 67-A(5):126–132, Feb 1984. (Cited on page 29)

[PAG95] R. D. Patterson, M. Allerhand, and C. Giguere. Time-domain modelling of peripheral auditory processing: A modular architecture and software platform. *Journal of the Acoustical Society of America,*, 98:1890–1894, 1995. (Cited on page 186)

[Pee04] G. Peeters. A Large Set of Audio Features for Sound Description (Similarity and Classification) in the CUIDADO Project. CUIDADO IST Project Report, April 2004. (Cited on page 32)

[PMV+06] R. V. Prasad, R. Muralishankar, S. Vijay, H. N. Shankar, P. Pawelczak, and I. Niemegeers. Voice Activity Detection for VoIP-An Information Theoretic Approach. In *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)*, pages 1 – 6, San Francisco, CA, USA, November 2006. (Cited on page 31)

[PN10] N.D. Gaubitch (editors) P.A. Naylor. *Speech Dereverberation*. Springer, London, 2010. (Cited on pages 62, 63, and 68)

[Pol88] R.D. Poltmann. Stochastic Gradient Algorithm for System Identification Using Adaptive FIR-Filters with too Low Number of Coefficients. *IEEE Trans. on Circuits and Systems*, 35(2):247–250, Feb 1988. (Cited on page 28)

[PRLN92] J. G. Proakis, C. M. Rader, F. Ling, and C. L. Nikias. *Advanced Digital Signal Processing*. Macmillan, New York, 1992. (Cited on page 14)

[PS01] R. Pintelon and J. Schoukens. *Identifikation of Linear Systems - A Practical Guideline to Accurate Modeling*. IEEE Press, New York, 2001. (Cited on page 10)

[QBC88] S. R. Quakenbusch, T. P. Barnwell, and B.A. Clemens. *Objective Measures of Speech Quality*. Prentice-Hall, Englewood Cliffs, NJ, 1988. (Cited on pages 169, 170, and 171)

[Raa61] D.H. Raab. Forward and Backward Masking between Acoustic Clicks. *Journal of the Acoustical Society of America (JASA)*, 33:137–139, February 1961. (Cited on page 119)

[Rad79] C. M. Rader. An Improved Algorithm for High Speed Autocorrelation with Application to Spectral Estimation. *IEEE Trans. on Audio and Electroacoustics*, 18, Dec 1979. (Cited on page 22)

[RAGA10] J. Rennies, E. Albertin, S. Goetze, and J.-E. Appell. Automatic Live Monitoring of Communication Quality for Normal-hearing and Hearing-impaired Listeners. In K. Miesenberger, J. Klaus, W. Zagler, and A. Karshmer, editors, *Computers Helping People with Special Needs, ICCHP 2010, Part II, LNCS 6180, Proc. 12th International Conference on Computers Helping People with Special Needs (ICCHP), Vienna, Austria*, pages 568–575. Springer-Verlag: Berlin Heidelberg New York, July 2010. (Cited on page 5)

[RF94] M. Rupp and R. Frenzel. Analysis of LMS and NLMS Algorithms with Delayed Coefficient Update under the Presence of Spherically Invariant Processes. *IEEE Trans. on Signal Processing*, 42(3):668–672, March 1994. (Cited on page 107)

[RG12] M. Ruhland and S. Goetze. Computational Efficient Noise Reduction for Dialogue Systems in Car Environments based on Binary Time-Frequency Masking and Autoregressive Interpolation. In *Workshop on Dialog systems that think along - Do they really understand me*, Saarbrücken, Germany, Sep. 2012. (Cited on page 5)

[RGA09] J. Rennies, S. Goetze, and J.E. Appell. How can audio technology improve working conditions? In J. Eschenbächer, D. Wewetzer, and P. Hoffmann, editors, *Change 2009 - Ambient Assisted Working Accessible and assistive ICT in Enterprise Environments, Emden, Germany, September 10-11, 2009*, 2009. (Cited on page 5)

[RGA11] J. Rennies, S. Goetze, and J.-E. Appell. Considering Hearing Deficiencies in Human-Computer Interaction. In M. Ziefle and C.Röcker, editors, *Human-Centered Design of E-Health Technologies: Concepts, Methods and Applications*, chapter 8, pages 180–207. IGI Global, 2011. In press. (Cited on page 5)

[RGB+12] M. Ruhland, S. Goetze, M. Brandt, J. Bitzer, and S. Doclo. A New Approach for Reduction of Supergaussian Noise using Autoregressive Interpolation and Time-Frequency Masking. In *Interna-*

tional Workshop on Acoustic Signal Enhancement (IWAENC 2012), Aachen, Germany, Sep. 2012. (Cited on page 5)

[RGH+08a] T. Rohdenburg, S. Goetze, V. Hohmann, K.-D. Kammeyer, and B. Kollmeier. Objective Perceptual Quality Assessment for Self-Steering Binaural Hearing Aid Microphone Arrays. In *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 2449–2452, Las Vegas, USA, March 30 - April 4 2008. (Cited on pages 5 and 67)

[RGH+08b] T. Rohdenburg, S. Goetze, V. Hohmann, B. Kollmeier, and K.-D. Kammeyer. Combined Source Tracking and Noise Reduction for Application in Hearing Aids. In *8. ITG-Fachtagung Sprachkommunikation*, Aachen, Germany, October 2008. (Cited on page 5)

[RGH+11] R. Rehr, S. Goetze, D. Hollosi, J.-E. Appell, and J. Bitzer. Speech / Non-Speech Discrimination for Acoustic Monitoring Considering Privacy Issues. In *Proc. 37th Annual Convention for Acoustics (DAGA)*, pages 879–880, Düsseldorf, Germany, March 2011. (Cited on pages 5, 32, and 138)

[RGS07] J. Ramírez, J. M. Gorriz, and J. C. Segura. Voice Activity Detection. Fundamentals and Speech Recognition System Robustness. In M. Grimm and K. Kroschel, editors, *Robust Speech Recognition and Understanding*, number ISBN: 978-3-902613-08-0, chapter 1, pages 1–22. I-Tech Education and Publishing, Vienna, Austria, June 2007. (Cited on page 31)

[RHK06] T. Rohdenburg, V. Hohmann, and B. Kollmeier. Subband-based parameter optimization in noise reduction schemes by means of objective perceptual quality measures. In *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, Paris, France, September 2006. (Cited on pages 70 and 71)

[RJW+03] R. Ratnam, D.L. Jones, B.C. Wheeler, J.W.D. O'Brien, C.R. Lansing, and A.S. Feng. Blind Estimation of Reverberation Time. *Journal of the Acoustical Society of America (JASA)*, 114(5):2877–2892, 2003. (Cited on page 92)

[RK00] B. D. Radlović and R. A. Kennedy. Nonminimum-Phase Equalization and its Subjective Importance in Room Acoustics. *IEEE Trans. on Speech and Audio Processing*, 8(6):728–737, November 2000. (Cited on pages 61 and 64)

[RN88] J.L. Rodgers and W.A. Nicewander. Thirteen Ways to Look at the Correlation Coefficient. *The American Statistician*, 42(1):59–66, Feb. 1988. (Cited on page 71)

[Roh08] T. Rohdenburg. *Development and Objective Perceptual Quality Assessment of Monaural and Binaural Noise Reduction Schemes for Hearing Aids*. PhD thesis, University of Oldenburg, Medical Physics, Oldenburg, 2008. (Cited on pages 67, 68, and 71)

[RS75] L. R. Rabiner and M. R. Sambur. An Algorithm for Determining the Endpoints of Isolated Utterances. *The Bell System Technical Journal*, 54(2):297–315, 1975. (Cited on page 31)

[RSB+04] J. Ramírez, J. C. Segura, C. Benitez, A. de la Torre, and A. Rubio. Voice Activity Detection with Noise Reduction and Long-Term Spectral Divergence Estimation. In *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, pages II–1093 – II–1096, Montreal, Quebec, Canada, May 2004. (Cited on page 31)

[RSB+05] J. Ramírez, J. C. Segura, C. Benítez, Á. de la Torre, and A. Rubio. An Effective Subband OSF-based VAD with Noise Reduction for Robust Speech Recognition. *IEEE Transactions on Speech and Audio Processing*, 13(6):1119, 2005. (Cited on page 31)

[RV89a] D. Rife and J. Vanderkooy. *Transfer-function measurement with maximum-length sequences*. J. Audio Eng. Soc. 37, 419-444, 1989. (Cited on page 10)

[RV89b] D. D. Rife and J. Vanderkooy. Transfer-Function Measurement with Maximum-Length Sequences. *Journal of the Audio Engineering Society*, 37:419–444, June 1989. (Cited on page 65)

[RWK99] B. D. Radlović, R. C. Williamson, and R. A. Kennedy. On the Poor Robustness of Sound Equalization in Reverberant Environments. In *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 881 – 884, March 1999. (Cited on pages 65 and 96)

[RWK00] B. D. Radlović, R. C. Williamson, and R. A. Kennedy. Equalization in an Acoustic Reverberant Environment: Robustness Results. *IEEE Trans. on Speech and Audio Processing*, 8(3):311–319, May 2000. (Cited on pages 65, 93, 96, 103, 105, and 141)

[SA02] S. Sukittanon and L. Atlas. Modulation Frequency Features for Audio Fingerprinting. In *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, pages II 1773–II 1776, Minneapolis, MN, USA, May 2002. (Cited on page 31)

[SAP04] S. Sukittanon, L. E. Atlas, and J. W. Pitton. Modulation-scale analysis for content identification. *IEEE Transactions on Signal Processing*, 52(10):3023–3035, 2004. (Cited on page 31)

[SAVJ09] S. Shafiee, F. Almasganj, B. Vazirnezhad, and A. Jafari. A Two-Stage Speech Activity Detection System Considering Fractal Aspects of Prosody. *Pattern Recognition Letters*, 31(9):936–948, 2009. (Cited on page 31)

[SB96] P. Scalart and A. Benamar. A System for Speech Enhancement in the Context of Hands–Free Radiotelephony with Combined Noise Reduction and Acoustic Echo Cancellation. *Speech Communication*, 20(1):203–214, Dec 1996. (Cited on page 30)

[SBM01] K. U. Simmer, J. Bitzer, and C. Marro. Post-filtering techniques. In M. S. Brandstein and D. Ward, editors, *Microphone Arrays: Signal Processing Techniques and Applications*, chapter 3, pages 39–60. Springer, 2001. (Cited on pages 30, 67, and 68)

[SBR04a] S. Spors, H. Buchner, and Rabenstein R. A Novel Approach to Active Listening Room Compensation for Wave Field Synthesis Using Wave-Domain Adaptive Filtering. In *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, Montreal, Canada, May 2004. (Cited on page 69)

[SBR04b] S. Spors, H. Buchner, and Rabenstein R. Active Listening Room Compensation for Spatial Audio Systems. In *Proc. EURASIP European Signal Processing Conference (EUSIPCO)*, 2004. (Cited on page 69)

[SBR04c] S. Spors, H. Buchner, and Rabenstein R. Efficient Active Listening Room Compensation for Wave Field Synthesis. In *Proc. AES Convention (Audio Engineering Society)*, volume 116, Berlin, Germany, May 2004. (Cited on page 69)

[Sch65] J.M. Schroeder. New Method of Measuring Reverberation Time. *Journal of the Acoustical Society of America (JASA)*, 37(3):409–412, 1965. (Cited on page 15)

[Sch01] H. Schmidt. *OFDM für die drahtlose Datenuebertragung innerhalb von Gebaeuden*. PhD thesis, Universität Bremen, Arbeitsbereich Nachrichtentechnik, 2001. (Cited on page 66)

[SCS+13] J. Schröder, B. Cauchi, M.R. Schädler, N. Moritz, K. Adiloglu, J. Anemüller, S. Doclo, B. Kollmeier, and S. Goetze. Acoustic Event Detection Using Signal Enhancement and Spectro-temporal Feature Extraction. In *IEEE AASP Challenge: Detection and Classification of Acoustic Scenes and Events*, New Paltz, NY, USA, Oct. 2013. (Cited on page 5)

[SGR+11] J. Schröder, S. Goetze, J. Rennies, F. Xiong, and J. Anemüller. Real-time Room Reverberation Estimation for Online Speech Intelligibility Monitoring. In *Proc. 37th Annual Convention for Acoustics (DAGA)*, pages 869–870, Düsseldorf, Germany, March 2011. (Cited on pages 5 and 92)

[SH84] A. Sekey and B.A. Hanson. Improved 1-Bark Bandwidth Auditory Filter. *Journal of the Acoustical Society of America (JASA)*, 75(6):1902–1904, June 1984. (Cited on page 175)

[SH02] H. Steeneken and T. Houtgast. Basics of the STI-Measuring Method. *Past, Present, and Future of the Speech Transmission Index, International Symposium on STI, The Netherlands*, pages 13–44, October 2002. (Cited on page 12)

[Shy92] J. J. Shynk. Frequency-Domain and Multirate Adaptive Filtering. *IEEE Signal Processing Magazine*, pages 14–34, January 1992. (Cited on pages 18, 29, 110, 134, and 138)

[SK91] D. Slock and T. Kailath. Numerically Stable Fast Transversal Filters for Recursive Least Squares Adaptive Filtering. *IEEE Trans. on Signal Processing*, 39(1):92–114, Jan 1991. (Cited on page 29)

[SK00] A. Stenger and W. Kellermann. Adaptation of a Memoryless Preprocessor for Nonlinear Acoustic Echo Cancelling. *EURASIP Signal Processing*, 80:1747–1760, September 2000. (Cited on pages 32 and 66)

[SKR03a] S. Spors, A. Kuntz, and R. Rabenstein. An Approach to Listening Room Compensation with Wave Field Synthesis. In *Proc. of the AES International Conference*, volume 24, pages 70–82, Banff, Alberta, Canada, June 2003. (Cited on page 69)

[SKR03b] S. Spors, A. Kuntz, and R. Rabenstein. Listening Room Compensation for Wave Field Synthesis. In *Int. Conf. on Multimedia and Expo, Baltimore, Maryland, USA*, pages 725–728, July 2003. (Cited on page 69)

[SKW92] K. U. Simmer, P. Kuczynski, and A. Wasiljeff. Time delay compensation for adaptive multichannel speech enhancement systems. In *Proc. Int. Symp. on Signals, Systems and Electronics ISSSE-92*, pages 660–663, Paris, France, September 1992. (Cited on page 31)

[SL61] M.R. Schroeder and B.F. Logan. "colorless" artificial reverberation. *Audio, IRE Transactions on*, 9(6):209 – 214, nov. 1961. (Cited on page 11)

[Sla93] M. Slaney. An efficient implementation of the patterson-holdsworth auditory filter bank. Technical Report 35, Apple Computer, Inc., 1993. (Cited on page 186)

[SLS01] N. Suditu, J. van de Laar, and P.C.W. Sommen. Evaluation of Dereverberation Capabilities of Broadband Beamformers. In *12th Annual Workshop on Circuits, Systems and Signal Processing (ProRISC 2001), Veldhoven, The Netherlands*, pages 656–661, 28-30 November 2001. (Cited on page 67)

[SM95] S. Shimauchi and S. Makino. Stereo Echo Canceller with True Echo Path Estimation. In *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 3059–3062, 1995. (Cited on page 32)

[SMH95] M.M. Sondhi, D. Morgan, and J. Hall. Stereophonic Acoustic Echo Cancellation - an Overview of the Fundamental Problem. *IEEE Signal Processing Letters*, 2(8):148–151, Aug 1995. (Cited on page 33)

[SMS+13] J. Schröder, N. Moritz, M.R. Schädler, B. Cauchi, K. Adiloglu, J. Anemüller, S. Doclo, B. Kollmeier, and S. Goetze. On the Use of Spectro-Temporal Features for the IEEE AASP Challenge 'Detection and Classification of Acoustic Scenes and Events'. In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New Paltz, NY, USA, Oct. 2013. (Cited on page 5)

[SMZ08] K. Shi, X. Ma, and G.T. Zhou. A Double-Talk Detector Based on Generalized Mutual Information for Stereophonic Acoustic Echo Cancellation in the Presence of Nonlinearity. In *Proc. Asilomar Conf. on Signals, Systems, and Computers*, Pacific Grove, CA, USA, October 2008. (Cited on pages 31 and 33)

[Som90] P. Sommen. On the Convergence Properties of a Partitioned Block Frequency Domain Adaptive Filter (PBFDAF). In *Proc. EURASIP European Signal Processing Conference (EUSIPCO)*, pages 201–204, Barcelona , Spain, September 1990. (Cited on page 19)

[Son67] M.M. Sondhi. An Adaptive Echo Canceller. *Bell Syst. Tech. J*, 46(3):497–511, 1967. (Cited on pages 4, 28, 31, and 38)

[SP90] J.-S. Soo and K. Pang. Multidelay Block Frequency Domain Adaptive Filter. *IEEE Trans. on Acoustics Speech and Signal Processing*, 38(2):373–376, Feb 1990. (Cited on pages 19, 29, 110, and 113)

[SRHE09] J. Schroeder, T. Rohdenburg, V. Hohmann, and S.D. Ewert. Classification of Reverberant Acoustic Situations. In *Int. Conf. on Acoustics (NAG/DAGA'09)*, Rotterdam, The Netherlands, March 2009. (Cited on page 92)

[SRR05] S. Spors, R. Renk, and Rabenstein R. Limiting effects of active room compensation using wave field synthesis. In *Proc. AES Convention (Audio Engineering Society)*, Barcelona, Spain, May 2005. (Cited on page 69)

[SS97] E. Scheirer and M. Slaney. Construction and Evaluation of a Robust Multifeature Speech/Musicdiscriminator. In *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 1331 – 1334 vol.2, Munich, Germany, April 1997. (Cited on page 32)

[SS98] J. Sohn and W. Sung. A Voice Activity Detector Employing Soft Decision Based Noise Spectrum Adaptation. In *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 365 – 368, Seattle, WA , USA, May 1998. (Cited on page 31)

[TBA10] L. N. Tan, B. J. Borgstrom, and A. Alwan. Voice Activity Detection using Harmonic Frequency Components in Likelihood Ratio Test. In *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 4466 – 4469, Dallas, TX, USA, March 2010. (Cited on page 31)

[TGS97a] V. Turbin, A. Gilloire, and P. Scalart. Comparison of Three Post–Filtering Algorithms for Residual Acoustic Echo Reduction. In *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Proc. (ICASSP)*, volume 1, pages 307–310, Munich, Germany, April 1997. (Cited on pages 30 and 51)

[TGS97b] V. Turbin, A. Gilloire, and P. Scalart. Using Psychoacoustic Criteria in Acoustic Echo Cancellation Algorithms. In *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, pages 53–56, London, September 1997. (Cited on page 31)

[TNK03] M. Miyoshi T. Nakatani and K. Kinoshita. Implementation and Effects of Single Channel Dereverberation Based on the Harmonic Structure of Speech. In *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, pages 91–94, Kyoto, Japan, September 2003. (Cited on page 68)

[TNM06] K. Kinoshita T. Nakatani and M. Miyoshi. Harmonicity-Based Blind Dereverberation for Single-Channel Speech Signals. *IEEE Trans. on Speech and Audio Processing*, 15(1):80 – 95, January 2006. (Cited on page 68)

[TO88] F. E. Toole and S. E. Olive. The Modification of Timbre by Resonances: Perception and Measurement. *Journal of the Audio Engineering Society*, 36:122–142, March 1988. (Cited on pages 67 and 80)

[TPM93] D. Tsoukalas, M. Paraskevas, and J. Mourjopoulos. Speech Enhancement Using Psychoacoustic Criteria. *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 359 – 362, April 1993. (Cited on page 31)

[TRB+06] Á. Torre, J. Ramírez, C. Benítez, J. C. Segura, L. García, and A. J. Rubio. Noise Robust Model-Based Voice Activity Detection. In *Proc. Annual Conference of the International Speech Communication Association (INTERSPEECH)*, pages 1954–1957, Pittsburgh, PA, USA, September 2006. (Cited on page 31)

[TS06] M. Triki and D.T.M. Slock. Iterated Delay and Predict Equalization for Blind Speech Dereverberation. In *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, Paris, France, September 2006. (Cited on page 165)

[Van94] J. Vanderkooy. Aspects of MLS Measuring Systems. *Journal of the Audio Engineering Society*, 42(4):219–231, April 1994. (Cited on pages 10 and 65)

[VHH98] P. Vary, U. Heute, and W. Hess. *Digitale Sprachsignalverarbeitung*. Teubner, Stuttgart, first edition, 1998. (Cited on pages 172 and 173)

[Vir99] N. Virag. Single channel speech enhancement based on masking properties of the human auditory system. *IEEE Trans. on Speech and Audio Processing*, 7(2):126–137, March 1999. (Cited on page 31)

[VM06] P. Vary and R. Martin. *Digital Speech Transmission - Enhancement, Coding, and Error Concealment*. Wiley & Sons, 1 edition, 2006. (Cited on pages 12, 18, 22, 28, 31, 33, 35, 36, 67, 119, 172, and 174)

[Wag03] K. Wagner. *Factors Influencing Sentence Intelligibility in Noise*. PhD thesis, University of Oldenburg, Oldenburg, Germany, 2003. (Cited on page 69)

[Wan95] H. Wang. Multi-Channel Deconvolution using Padé Approximation. In *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 3007 – 3010, May 1995. (Cited on page 66)

[Wan96] E.A. Wan. Adjoint-LMS: An Efficient Alternative to the Filtered-X LMS and Multiple Error LMS Algorithms. In *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 1842–1845, Atlanta, USA, May 1996. (Cited on page 107)

[Wüb06] D. Wübben. *Effiziente Detektionsverfahren für Multilayer-MIMO-Systeme*. PhD thesis, Universität Bremen, Arbeitsbereich Nachrichtentechnik, 2006. (Cited on pages 66 and 67)

[WG00] P. J. Wolfe and S. J. Godsill. The application of psychoacoustic criteria to the restauration of musical recordings. In *108th Convention of the Audio Engineering Society (AES)*, Paris, France, 2000. (Cited on page 30)

[WGH+06] J.Y.C. Wen, N.D. Gaubitch, E.A.P. Habets, T. Myatt, and P.A. Naylor. Evaluation of Speech Dereverberation Algorithms using the MARDY Database. In *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, Paris, France, September 2006. (Cited on pages 71 and 89)

[WGH+11] S. Wilksen, S. Goetze, D. Hollosi, J.-E. Appell, and J. Bitzer. Speech Activity Detection for Activity Monitoring using an Embedded Platform. In *Proc. 37th Annual Convention for Acoustics (DAGA)*, pages 875–876, Düsseldorf, Germany, March 2011. (Cited on pages 5 and 32)

[WH60] B. Widrow and M.E. Hoff. Adaptive Switching Circuits. In *IRE Western Electric Show and Convention Record, Part 4*, pages 96–104, August 1960. (Cited on page 38)

[Wie49] N. Wiener. *Extrapolation, Interpolation, and Smoothing of Stationary Time Series, with Engineering Applications*. Wiley, New York, 1949. (Cited on page 30)

[WK03] D. Wübben and K.-D. Kammeyer. Impulse Shortening and Equalization of Frequency-Selective MIMO Channels with Respect to Layered Space-Time Architectures. *EURASIP Signal Processing Magazine*, 83(8):1643–1659, August 2003. (Cited on pages 63 and 66)

[WMB75] B. Widrow, J.M. McCool, and M. Ball. The Complex LMS Alorithms. In *Proc. IEEE*, volume 63, pages 719–720, 1975. (Cited on page 38)

[WN06] J.Y.C. Wen and P.A. Naylor. An Evaluation Measure for Reverberant Speech using Decay Tail Modeling. In *Proc. EURASIP European Signal Processing Conference (EUSIPCO)*, Florence, Italy, September 2006. (Cited on pages 71, 80, 179, 180, 181, 182, and 183)

[WN07] J.Y.C. Wen and P.A. Naylor. Objective Measurement of Colouration in Reverberation. In *Proc. EURASIP European Signal Processing Conference (EUSIPCO)*, pages 1615–1619, Poznan, Poland, September 2007. (Cited on pages 11, 71, 80, 184, and 185)

[WQW11] S. Wu, X. Qiu, and M. Wu. Stereo Acoustic Echo Cancellation Employing Frequency Domain Preprocessing and Adaptive Filter. *IEEE Trans. on Audio, Speech and Language Processing*, 19(3):614–621, March 2011. (Cited on page 33)

[WS85] B. Widrow and S.D̃. Stearns. *Adaptive Signal Processing*. Englewood Cliffs, 1985. (Cited on pages 18, 28, 38, 105, and 106)

[WSG92] S. Wang, A. Sekey, and A. Gersho. An Objective Measure for Predicting Subjective Quality of Speech Coders. *IEEE J. Selected Areas of Communications*, 10(5):819–829, June 1992. (Cited on pages 88, 172, and 177)

[WSS81] B. Widrow, B. Shur, and S. Shaffer. On Adaptive Inverse Control. In *Proc. Asilomar Conf. on Signals, Systems, and Computers*, pages 185–189, Pacific Grove, USA, 1981. (Cited on page 105)

[WWB07] A. Warzybok, K.C. Wagener, and T. Brand. Intelligibility of German Digit Triplets Test by Non-Native Listeners. In *Proc. 8th EFAS Congress / 10th Congress of the German Society of Audiology*, Heidelberg, Germany, 2007. (Cited on page 69)

[XAG12] F. Xiong, J.-E. Appell, and S. Goetze. System Identification for Listening-Room Compensation by means of Acoustic Echo Cancellation and Acoustic Echo Suppression Filters. In *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, Kyoto, Japan, March 2012. (Cited on pages 5, 28, and 141)

[XGM13] F. Xiong, S. Goetze, and B.T. Meyer. Blind Estimation of Reverberation Time based on Spectro-Temporal Modulation Filtering. In *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 443–447, Vancouver, Canada, May 2013. (Cited on page 5)

[Yan99] W. Yang. *Enhanced Modified Bark Sectral Distortion (EMBSD): A Objective Speech Quality Measure Based on Audible Distortion and Cognition Model*. PhD thesis, Temple University, Philadelphia, USA, May 1999. (Cited on page 172)

[YHC05] J. Benesty Y. Huang and J. Chen. A Blind Channel Identification-Based Two-Stage Approach to Separation and Dereverberation of Speech Signals in a Reverberant Environment. *IEEE Trans. on Speech and Audio Processing*, 13(5):882–895, September 2005. (Cited on pages 62, 97, and 132)

[YHC07] J. Benesty Y. Huang and J. Chen. On Crosstalk Cancellation and Equalization With Multiple Loudspeakers for 3-d Sound Reproduction. *IEEE Signal Processing Letters*, 14(10):649 – 652, October 2007. (Cited on page 84)

[YK82] S. Yamamoto and S. Kitayama. An Adaptive Echo Canceller with Variable Step Gain Method. In *Trans. IECE, Japan*, volume E65, pages 1–8, January 1982. (Cited on page 28)

[YM00] B. Yegnanarayana and P.S. Murthy. Enhancement of Reverberant Speech Using LP Residual Signal. *IEEE Trans. on Speech and Audio Processing*, 8(3):267–280, May 2000. (Cited on pages 68 and 72)

[YY09] I. C. Yoo and D. Yook. Robust Voice Activity Detection using the Spectral Peaks of Vowel Sounds. *ETRI journal*, 31(4):451–453, 2009. (Cited on page 31)

[ZF99] E. Zwicker and H. Fastl. *Psychoacoustics: Facts and Models*. Springer, Berlin, second edition, 1999. (Cited on pages 67, 173, and 176)

[ZGN08] W. Zhang, N.D. Gaubitch, and P.A. Naylor. Computationally Efficient Equalization of Room Impulse Responses Robust to System Estimation Errors. In *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 4025–4028, Las Vegas, NV, USA, March 2008. (Cited on page 96)

[ZKN08] W. Zhang, A.W.H. Khong, and P.A. Naylor. Adaptive Inverse Filtering of Room Acoustics. In *Proc. Asilomar Conf. on Signals, Systems, and Computers*, pages 26–29, Pacific Grove, USA, October 2008. (Cited on page 65)

Index