Acoustic Echo Cancellation for Multiple Reproduction Channels: From First Principles to Real-Time Solutions

Herbert Buchner

Deutsche Telekom Laboratories, Berlin University of Technology, Ernst-Reuter-Platz 7, 10587 Berlin, Germany
E-Mail: `herbert.buchner@telekom.de`

Abstract

Multichannel acoustic echo cancellation (MCAEC) is a key technology whenever hands-free and full-duplex communication in modern systems with multichannel sound reproduction is desired. Although the basic principle of echo cancellation has been well known for several decades, the multichannel case poses some additional and fundamentally different challenges. Moreover, there are even some notable differences between the two-channel case and the general multichannel case which has been addressed bit by bit only in recent years. The aim of this paper is twofold. On the one hand, after a brief review of the problem of multichannel acoustic echo cancellation, this paper gives an outline of how the problem may be tackled based on some fundamental principles. In this sense, the presentation in this paper brings together for the first time ideas from system theory, information theory, psychoacoustics, and also wave physics. Based on this framework, and as the other main contribution, we present in this paper some recent advances in the field of MCAEC. Thereby, important issues in the case of more than two channels are emphasized. Finally, as an outlook, we touch on our ongoing work towards MCAEC for massive multichannel sound reproduction, such as wave field synthesis.

1 Introduction

For various applications, such as home entertainment, virtual reality (e.g., games, simulations, training), or advanced teleconferencing, multimedia terminals with an increased number of audio channels for sound reproduction are highly desirable (e.g., stereo, 5.1 surround systems, or even beyond). In such applications, multichannel acoustic echo cancellation (MCAEC) is a key technology whenever hands-free and full-duplex communication is desired. Acoustic echo cancellation has already been discussed extensively for the single-channel case and for stereo sound reproduction (e.g., [1, 2, 3, 4]). Only in recent years, AEC has been realized for more than two reproduction channels [4, 5].

Figure 1 describes a typical scenario for stereo or multichannel AEC. From a transmission room, a sound source (e.g., a speaker) is picked up by P microphones ($P=2$ for stereo). The microphone signals are transmitted to a receiving room and reproduced via P loudspeakers. At the same time, a microphone in the receiving room picks up speech from a local user. In order to prevent the sound emitted from the loudspeakers coupling into the outgoing microphone signal (which is sent back to the far-end listener or some multimedia terminal), AEC attempts to cancel out any contributions of the incoming loudspeaker signals $x_{\mathrm{ref},i}(n)$ from the microphone signal by subtracting filtered versions of the loudspeaker signals from the microphone signal. This generally requires that cancellation filters (assumed to be length-L FIR filters) are dynamically adjusted by an adaptation algorithm to achieve minimum error signal $e(n)$ and thus optimum cancellation. This is the case when the adaptive cancellation filters

$$\hat{\mathbf{h}}_i(n) = \left[\hat{h}_{i,1}(n), \cdots, \hat{h}_{i,L}(n)\right]^T, \; i = 1, 2, \ldots, P \tag{1}$$

accurately model the impulse responses $\mathbf{h}_i$ from the emitting speakers to the microphone.

It has been shown for stereo AEC that a so-called *non-uniqueness problem* exists [6]: If both loudspeaker signals are strongly correlated, then the adaptive filters generally converge to a solution that does not correctly model the transfer functions between the speakers and the microphone, but merely optimizes echo cancellation for the given particular loudspeaker signals. As

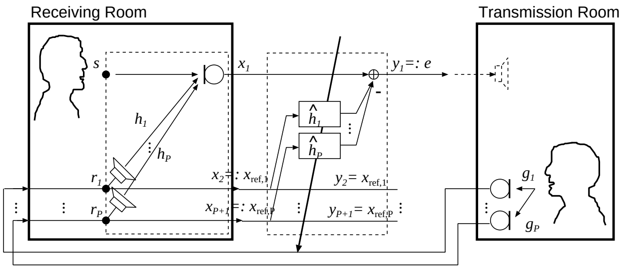

Figure 1: Scenario for multi-channel AEC.

a consequence, a change in the characteristics of the loudspeaker signals (e.g. due to a change of the geometric position of the sound source in the transmission room) results in a breakdown of the echo cancellation performance and requires a new adaptation of the cancellation filters. In practice, from a statistical point of view, the high cross-correlations between the loudspeaker signals lead to a highly ill-conditioned tap-input correlation matrix in the normal equation to be solved for the minimization of $e(n)$ [7, 8],

$$\begin{aligned}
\mathbf{R}_{\mathbf{x}_{\mathrm{ref}}\mathbf{x}_{\mathrm{ref}}}(n) &= \hat{E}\left\{\mathbf{x}_{\mathrm{ref}}(n)\mathbf{x}_{\mathrm{ref}}^T(n)\right\} \qquad (2)\\
&= \begin{bmatrix} \mathbf{R}_{\mathbf{x}_{\mathrm{ref},1}\mathbf{x}_{\mathrm{ref},1}}(n) & \cdots & \mathbf{R}_{\mathbf{x}_{\mathrm{ref},1}\mathbf{x}_{\mathrm{ref},P}}(n) \\ \vdots & \ddots & \vdots \\ \mathbf{R}_{\mathbf{x}_{\mathrm{ref},P}\mathbf{x}_{\mathrm{ref},1}}(n) & \cdots & \mathbf{R}_{\mathbf{x}_{\mathrm{ref},P}\mathbf{x}_{\mathrm{ref},P}}(n) \end{bmatrix},\\
\mathbf{x}_{\mathrm{ref}}(n) &= \left[\mathbf{x}_{\mathrm{ref},1}(n), \ldots, \mathbf{x}_{\mathrm{ref},P}(n)\right],\\
\mathbf{x}_{\mathrm{ref},i}(n) &= \left[x_{\mathrm{ref},i}(n), \ldots, x_{\mathrm{ref},i}(n-L+1)\right]^T.
\end{aligned}$$

To tackle this challenging problem of ill conditioning, various techniques have been proposed mainly in the stereo context so far. They can be distinguished into two different classes representing separate system components, e.g., [2]:

(a) Application of a robust and fast converging adaptation algorithm taking all cross-correlations into account.

(b) Preprocessing of the signals transmitted from the transmission room prior to their reproduction in the receiving room in order to *partially decorrelate* all channels relative to each other.

Due to the conflicting key requirements that on the one hand the preprocessing must not introduce any objectionable artifacts into the reproduced audio signals, and on the other hand for the convergence enhancement, a systematic design for MCAEC based on first principles of coefficient estimation and optimization together with a complete stochastic signal description, and human audio perception is highly desirable. The structure of this paper is motivated by a step-by-step incorporation of these first principles. Within this framework, we place recent advances in MCAEC with emphasis on more than two reproduction channels, and deduce various new insights and practical results.

2 Elements from System Theory

In general, to optimally exploit the information contained in the involved signals, the coefficient estimation process should take into account all their fundamental stochastic properties: *nongaussianity, nonwhiteness, nonstationarity*. A suitable broadband signal formulation for this purpose was developed within the so-called TRINICON ('TRIple-N Independent Component Analysis for CONvolutive mixtures') framework for adaptive multiple-input and multiple-output (MIMO) filtering [9, 10, 11]. For an overview, see also [12] in the present conference.

In [13] the AEC problem was linked to the more general *MIMO system identification and signal separation problem* as addressed by TRINICON, and as illustrated by the two dashed boxes in Fig. 1. The left and right dashed box correspond to a sparse MIMO mixing system, and a corresponding MIMO demixing system, respectively. The demixing system follows rigorously from the ideal MIMO separation solution derived in [14, 15]. This formal connection facilitates the introduction of the general formulation of stochastic signal models as multivariate probability densities which capture the temporal structure by multiple time lags and the nonstationarity by the respective time-varying stochastic parameters, such as correlation matrices.

3 Elements from Information Theory and Optimization

The TRINICON optimization criterion for the case of separation (and system identification) problems is based on minimizing the information-theoretic quantity of *mutual information* between the output channels of the demixing MIMO system using the multivarate densities mentioned above, i.e., in the special case of AEC, we separate the contributions of the loudspeaker signals from the error signal $e(n)$ at the AEC output (Fig. 1 and [13]). In the special case of Gaussian signals, this separation process corresponds to a simultaneous block-diagonalization of the output correlation matrix for multiple time instants since the local speech $s(n)$ is assumed to be uncorrelated from the loudspeaker signals [11, 13].

In this paper we generalize the information-theoretic approach in [13] to *multichannel* AEC with typically highly correlated loudspeaker signals. In other words, the output channels $x_{\mathrm{ref},1}(n),\ldots,x_{\mathrm{ref},P}(n)$ of the demixing system in Fig. 1 should *not* be separated from each other. Figure 2 illustrates the output correlation matrix with time-lags and the corresponding *desired* structure of it after convergence for the special case of Gaussian signals and $P=2$. It can be shown that the approach in [13] generalizes straightforwardly to the MCAEC case by just using this modified matrix partitioning.

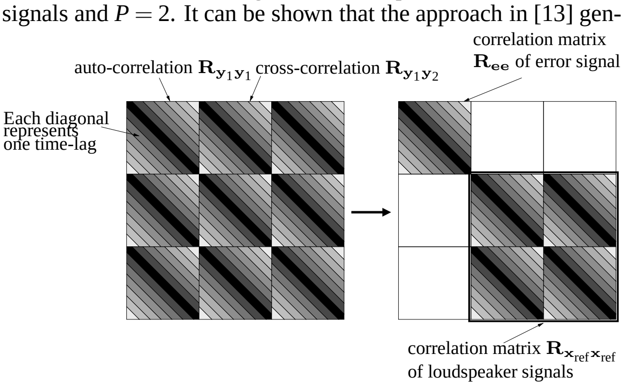

Figure 2: Illustration of the AEC process for second-order statistics and $P=2$.

In [13] the update equations for TRINICON-based coefficient adaptation in AEC have been presented for the simple case of gradient-based optimization. However, it is known that gradient-descent algorithms (e.g., LMS/NLMS [7]) generally exhibit very slow convergence for highly correlated input signals such as in the multichannel case [7].

The so-called Newton-Raphson-type optimization procedure is known as the canonical method for more challenging optimization problems. A TRINICON-based Newton update can be derived in a way analogous to [16]. The Newton algorithm contains virtually all of the well-known adaptation schemes as special cases, most notably the recursive least-squares (RLS) algorithm. The important feature of Newton-type/RLS-type algorithms is that they explicitly take all input correlations (2) into account within their Hessian matrix [7, 16] which makes them very attractive for the MCAEC application [2].

In addition to this desirable property of RLS-type algorithms, the more general TRINICON-based approach inherently leads to a *multivariate* error *nonlinearity* to take both the nongaussianity and the nonwhiteness of the near-end signal into account [13]. This provides an inherent double-talk handling and a link to the powerful concept of robust statistics, e.g., [17, 18]. Moreover, the block online adaptation and block averaging obtained in [13] further speeds up the convergence (especially in MCAEC).

Note also that the general TRINICON-based approach also leads to important insights in the case of AEC for multiple microphone channels in the receiving room, as explained further in Sect. 5.

Finally, another aspect in the design of a real-time solution to the MCAEC problem is its computationaly complexity. Unfortunately, straightforward implementations of RLS-type algorithms are computationally very expensive due to the required (implicit or explicit) inversion of the Hessian matrix. A very efficient practical solution to this problem is to formulate the above-mentioned broadband algorithm in a mathematically rigorous way in the frequency domain, as shown, e.g., in [5, 11, 16], followed by the introduction of *carefully selected* approximations. The most important features of this concept of *frequency-domain adaptive filtering* (FDAF) is that in addition to the efficient use of the FFT (gains for both, adaptation and filtering), all the sub-matrices of the input correlation matrix (2) are approximately diagonalized by the DFT. In this way, it is possible to efficiently take into account all cross-correlations [5]. This is possible for both, second-order and higher-order statistics. A first MCAEC system for 5-channel surround sound applications, based on the multichannel FDAF algorithm has been presented in [4, 5]. This real-time implementation also utilizes the concept of robust statistics [16].

4 Elements from Psychoacoustics

As mentioned in Sect. 1, among the key requirements for the techniques to preprocess the signals transmitted from the transmission room prior to their reproduction in the receiving room is the subjective sound quality. While several of the known preprocessing techniques provide enough decorrelation to achieve proper AEC convergence in the stereo case, considerations of sound quality have frequently not been addressed adequately. In this section we first give a brief overview of the known two-channel preprocessing approaches. We then describe a recently introduced novel approach [19], based on perceptual considerations. It easily generalizes to the multi-channel case and has been demonstrated to be effective in surround sound echo cancellation.

4.1 Known two-channel preprocessing approaches

A first simple preprocessing method for stereo AEC was proposed by Benesty et al. [8, 20] and achieves signal decorrelation by adding non-linear distortions to the signals. While this approach features extremely low complexity, the introduced distortion products can become quite audible and objectionable, especially for high-quality applications using music signals. Moreover, the generalization of this approach to an arbitrary number of channels is not straightforward.

A second well-known approach consists of adding uncorrelated noise to the signals. In [21], this is achieved by perceptual audio coding / decoding of the signal which introduces uncorrelated quantization distortion that is masked due to the noise shaping according to the coder's psychoacoustic model. The use of an explicit psychoacoustic model plus analysis / synthesis filterbanks is able to prevent audible distortions for arbitrary types of audio signals and may be easily generalized to more than two channels. However, the associated implementation complexity and the introduced delay render this approach unattractive for most applications.

Other approaches employ switched / time-varying time-delays [3] or variable all-pass filtering [22] to produce a time-varying phase shift / signal delay between the two channels of a stereo AEC and thus "decorrelate" both signals. Specifically, [3] describes a preprocessing system in which the output signal switches between the original signal and a time-delayed / filtered version of it. As a disadvantage, this switching process may introduce unintended artifacts into the audio signal. [22] describes a system in which an allpass preprocessor is randomly modulating its allpass filter variable. In [23], it was proposed to apply this allpass preprocessor only to the low frequency range up to 1 kHz due to convergence requirements.

4.2 Psychoacoustically motivated method suitable for the multichannel case

In order to obtain a preprocessing method offering both good decorrelation properties for the enhancement of AEC convergence *and* minimal alteration of the perceived stereo image, the method proposed in [19] is based on several considerations. From the previously discussed approaches time-varying modulation of the phase of the audio signal, as proposed in [3, 22], is an effective method which is generally unobtrusive in its perceptual effects on audio signals as compared to other methods while avoiding computationally expensive masking models. Nonetheless, it is difficult to achieve maximum decorrelation while guaranteeing that introducing a time / phase difference between left and right channels does not result in an alteration of the perceived stereo image. Several aspects must be accounted for:

- Interaural phase / time difference is a relevant perceptual parameter for subjective perception of a sound stage [24] and has been used extensively in synthesis of stereo images (e.g. [25]). Consequently, a change in the perceived stereo image can only be avoided if the introduced time / phase difference stays below the threshold of perception, as it applies to audio signals that are reproduced via loudspeakers.
- Optimal AEC convergence enhancement can be achieved if the preprocessing introduces time / phase differences just at the threshold of perception, i.e., applies the full amount of tolerable change.
- As is known from psychoacoustics, the human sensitivity to phase differences is high at low frequencies, and gradually reduces for increasing frequencies, until it fully vanishes for frequencies above ca. 4 kHz.
- Neither a simple time delay modulation nor a low-order time-varying allpass filtering offer the flexibility to tailor the amount of time / phase shifting as a function of frequency, such that the full potential of perceptually tolerable change is exploited.

Hence, in contrast to the earlier phase modulation approaches, the recently proposed novel method in [19] is designed to allow a perceptually motivated frequency-selective choice of phase modulation parameters (modulation frequency, modulation amplitude, and modulation waveform) by employing analysis / synthesis filterbanks. The input audio signal is decomposed into subband signals by means of an analysis filterbank. Then, the subband phases are modified based on a set of frequency-dependent modulating signals. According to the above considerations, subbands belonging to the low frequency part of an audio signal should be left largely untouched, while subbands corresponding to frequencies above 4 kHz may be modulated heavily. As detailed in [19], the frequency-selective phase modulation amplitude was optimized by a listening procedure. Finally, the modified spectral coefficients are converted back into a time-domain representation by a synthesis filterbank. To allow easy access to the signal's phase, a complex-valued filterbank [26] is used, and a phase modification is implemented by a complex multiplication of the subband coefficient with $e^{j\varphi(t,\nu)}$ where $\varphi(t,\nu)$ denotes the intended time varying phase shift in subband ν. It is preferable to choose a smooth modulating function $\varphi(t,\nu)$, such as a sine wave at a relatively low frequency. Moreover, to account for the symmetry of typical multi-channel speaker setups, such as 5.1 or 7.1, the modulation of channel pairs is carried out in a complex conjugate fashion. The modulation frequencies for pairs are chosen such that they provide "orthogonal" modulation activity, as detailed in [19].

Figure 3 shows a summary of the results of a standardized subjective listening test carried out with 10 experienced listeners in a typical surround sound listening setup. The sound quality was quantified on a scale from 0 to 100 for 5 critical music excerpts and one speech excerpt (see [19] for further details). The different preprocessing types are the original reference and a

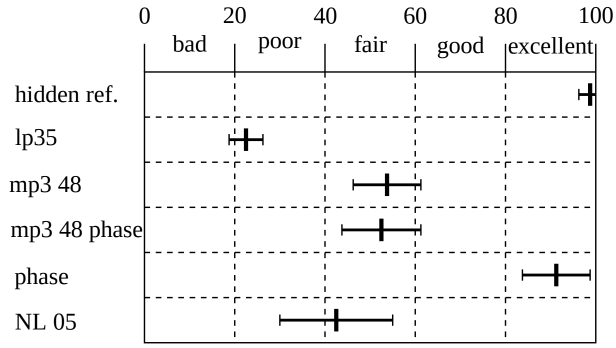

Figure 3: Results of the MUSHRA listening test (average and 95% confidence intervals).

3.5 kHz band-limited version thereof (both included as required by the MUSHRA ('MUlti Stimulus test with Hidden Reference and Anchor') listening test), individual channel mp3 en/decoding at 48 kbit/s ('mp3 48'), the novel perceptual phase modulation method ('phase'), a combination of mp3 encoding/decoding and phase modulation ('mp3 48 phase') and the conventional nonlinear processing ('NL' after [8, 20]). It is visible from the graph that the phase modulation method emerges as the clear winner in terms of sound quality. Note that the latter four methods were tuned for comparable coefficient convergence speeds.

5 MIMO Processing and Elements from Wave Physics

5.1 MIMO case for multiple microphones

So far in this paper, we have focused on the case of multiple reproduction channels but only one microphone in the receiving room. The more general case of a full MIMO loudspeaker-room-microphone system appears when combining MCAEC with a microphone array, e.g., [5]. Traditionally, in this case several parallel multiple-input and single-output (MISO) systems are independently applied, which has been shown to be optimal in terms of least-squares-based coefficient estimation.

As explained in Sect. 3, TRINICON-based AEC is generally able to exploit the nonwhiteness of the signals in the receiving room (upper left sub-matrix in Fig. 2). By further generalizing the TRINICON-based AEC to the case of MIMO loudspeaker-room-microphone systems it is also able to exploit the *spatial* nonwhiteness in the receiving room by simultaneously taking into account all microphone signals for the adaptation process. In other words, the performance may be improved with multiple microphones.

5.2 Massive multichannel systems and wave physics

Current loudspeaker setups, such as the 5.1 format, still rely on a restrained listening area ('sweet spot'). A high-quality volume solution for a large listening space is offered by the wave field synthesis (WFS) method which is based on wave physics [27]. The so-called *Kirchhoff-Helmholtz integrals* which can be derived from the acoustic wave equation state that at any point within a source-free listening area, the sound pressure field can be calculated if both the sound pressure and its gradient are known on the *contour* enclosing this area. Thus, in WFS, closely spaced arrays of a large number P of individually driven loudspeakers generate a prespecified sound field. P may lie between 20 and several hundred. An analogous approach is possible for wave field analysis (WFA) using microphone arrays.

Building a full-duplex system with this massive multichannel setup for unrestricted audio content might be considered as the supreme discipline of MCAEC research since in this case even the $P \times P$ frequency bin-wise correlation matrices of the loudspeaker driving signals are generally still large and ill-conditioned after the approximate blockwise diagonalization of

(2) within the frequency-domain adaptive filtering (FDAF) coefficient update (cf. Sect. 3).

The basic idea of *wave-domain adaptive filtering (WDAF)*, e.g., [28, 29], is to replace the point-to-point MIMO system model by a more detailed spatial consideration exploiting wave-physics foundations as in WFS/WFA. In particular, WDAF extends the conventional FDAF approach by a suitable *spatio-temporal* transform domain for efficiency. Figure 4 illustrates this two-step transformation approach from the RLS via FDAF towards WDAF in terms of the loudspeaker correlation matrix and its approximate temporal and spatio-temporal diagonalizations.

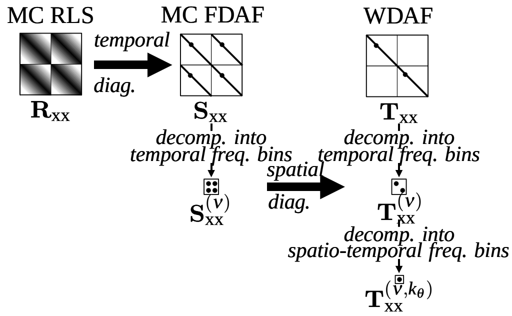

Figure 4: Illustration of the WDAF concept and its relation to conventional algorithms.

Requirements for the spatio-temporal basis functions are that they should be orthogonal and must fulfill the acoustic wave equation (e.g., circular harmonics). Moreover, since the transducers are only placed on the contour enclosing the listening area, corresponding transformations taking into account the Kirchhoff-Helmholtz Integrals are necessary. These transformations depend on the array geometries, and for certain setups, e.g., circular arrays [28, 29], they can in fact be formulated in a compact form.

Advantages of the approximate MIMO decoupling due to the spatio-temporal transformation are both an improved convergence and a significant complexity reduction, as shown, e.g., in [28, 29]. Note also that the WDAF concept can be well applied to the general TRINICON approach. Since all microphone signals are jointly taken into account by the spatio-temporal transformation, WDAF also facilitates an efficient exploitation of the spatial nonwhiteness mentioned in the previous subsection.

6 Conclusions

Although acoustic echo cancellation has been a well established topic in acoustic signal processing for many years, the multichannel case is still an active and interesting area of research. Recently significant progress for more than two reproduction channels has been made. As illustrated in this paper, the research area of MCAEC is a very good example for the necessity to bring together fundamentals from various different disciplines.

References

[1] S. Shimauchi and S. Makino, “Stereo projection echo canceller with true echo path estimation,” *Proc. IEEE ICASSP*, pp. 3059-3062, May 1995.

[2] T. Gänsler and J. Benesty, “Stereophonic acoustic echo cancellation and two-channel adaptive filtering: an overview,” *Int. Journal of Adaptive Control and Signal Processing,* vol. 14, pp. 565-586, 2000.

[3] A. Sugiyama, Y. Joncour, and A. Hirano, “A stereo echo canceller with correct echo path identification based on an input-sliding technique,” *IEEE Trans. Signal Processing,* 49(1), pp. 2577-2587, 2001.

[4] H. Buchner and W. Kellermann, “Acoustic echo cancellation for two and more reproduction channels,” *Proc. Int. Workshop on Acoustic Echo and Noise Control,* pp. 99-102, Sept. 2001.

[5] H. Buchner, J. Benesty, and W. Kellermann, “Generalized multichannel frequency-domain adaptive filtering: efficient realization and application to hands-free speech communication,” *Signal Processing,* vol. 85, no. 3, pp. 549-570, March 2005.

[6] M. M. Sondhi, D. R. Morgan, and J. L. Hall, “Stereophonic acoustic echo cancellation- An overview of the fundamental problem,” *IEEE Signal Processing Lett.,* vol. 2, pp. 148-151, Aug. 1995.

[7] S. Haykin, *Adaptive Filter Theory*, 4th ed., Prentice-Hall, 2002.

[8] 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. Speech Audio Processing,* vol. 6, pp. 156-165, Mar. 1998.

[9] H. Buchner, R. Aichner, and W. Kellermann, “Blind source separation for convolutive mixtures exploiting nongaussianity, nonwhiteness, and nonstationarity,” in *Proc. Int. Workshop Acoustic Echo and Noise Control,* Kyoto, pp. 223-226, Sept. 2003.

[10] H. Buchner, R. Aichner, and W. Kellermann, “TRINICON: A versatile framework for multichannel blind signal processing,” in *Proc. IEEE ICASSP*, Montreal, vol. 3, pp. 889-892, May 2004.

[11] H. Buchner, R. Aichner, and W. Kellermann, “Blind source separation for convolutive mixtures: A unified treatment,” in Y. Huang and J. Benesty (eds.), *Audio Signal Processing for Next-Generation Multimedia Communication Systems*, Kluwer Academic Publishers, Boston, pp. 255-293, Feb. 2004.

[12] H. Buchner, R. Aichner, and W. Kellermann, “The TRINICON framework for adaptive MIMO signal processing with focus on the generic Sylvester constraint,” in *ITG Fachtagung Sprachkommunikation*, Aachen, Oct. 2008.

[13] H. Buchner and W. Kellermann, “A fundamental relation between blind and supervised adaptive filtering illustrated for blind source separation and acoustic echo cancellation,” in *Conf. Rec. Joint Workshop on Hands-Free Speech Communication and Microphone Arrays (HSCMA)*, Trento, Italy, May 2008.

[14] H. Buchner, R. Aichner, and W. Kellermann, “Relation between blind system identification and convolutive blind source separation,” in *Proc. Joint Workshop Hands-Free Speech Communication and Microphone Arrays (HSCMA)*, Piscataway, NJ, USA, Mar. 2005.

[15] H. Buchner, R. Aichner, and W. Kellermann, “TRINICON-based blind system identification with application to multiple-source localization and separation,” in S. Makino, T.-W. Lee, and S. Sawada (eds.), Blind Speech Separation, Springer-Verlag, Berlin/Heidelberg, pp. 101-147, Sept. 2007.

[16] H. Buchner, J. Benesty, T. Gänsler, and W. Kellermann, “Robust extended multidelay filter and double-talk detector for acoustic echo cancellation,” in *IEEE Trans. Acoustics, Speech, and Language Processing*, Sept. 2006.

[17] P.J. Huber, *Robust Statistics*, Wiley, New York, 1981.

[18] T. Gänsler, S.L. Gay, M.M. Sondhi, and J. Benesty, “Double-talk robust fast converging algorithms for network echo cancellation,” *IEEE Trans. Speech Audio Processing*, vol. 8, pp. 656-663, Nov. 2000.

[19] J. Herre, H. Buchner, and W. Kellermann, “Acoustic echo cancellation for surround sound using perceptually motivated convergence enhancement,” in *Proc. IEEE ICASSP* , Honolulu, HI, USA, April 2007.

[20] D.R. Morgan, J.L. Hall, and J. Benesty, “Investigation of Several Types of Nonlinearities for Use in Stereo Acoustic Echo Cancellation,” *IEEE Trans. Speech Audio Processing*, vol. 5, no. 6, pp. 686-696, Sept. 2001.

[21] T. Gänsler and P. Eneroth, “Influence of audio coding on stereophonic acoustic echo cancellation,” *Proc. IEEE ICASSP,* pp. 3649-3652, 1998.

[22] M. Ali, “Stereophonic acoustic echo cancellation system using time-varying all-pass filtering for signal decorrelation,” *Proc. IEEE ICASSP,* pp. 3689-3692, 1998.

[23] T. Hoya, J.A. Chambers, and P.A. Naylor, “Low complexity ε-NLMS algorithms and subband structures for stereophonic acoustic echo cancellation,” *Proc. Int. Workshop on Acoustic Echo and Noise Control,* pp. 36-39, 1999.

[24] J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization,* revised edition, MIT Press, 1997.

[25] C. Faller and F. Baumgarte, “Binaural Cue Coding - Part II: Schemes and applications,” *IEEE Trans. on Speech and Audio Proc.,* vol. 11, no. 6, Nov. 2003.

[26] H.S. Malvar, “A modulated complex lapped transform and its application to audio processing,” *Proc. IEEE ICASSP,* pp. 1421-1424, 1999.

[27] A.J. Berkhout, D. de Vries, and P. Vogel, “Acoustic control by wave field synthesis,” *Journal of the Acoustic Society of America*, vol. 93, no. 5, pp. 2764–2778, May 1993.

[28] H. Buchner, S. Spors, W. Herbordt, and W. Kellermann, “Wave-Domain Adaptive Filtering for Acoustic Human-Machine Interfaces based on Wavefield Analysis and Synthesis,” in *Proc. EUSIPCO* , Vienna, Sept. 2004.

[29] S. Spors, H. Buchner, R. Rabenstein, and W. Herbordt, “Active listening room compensation for massive multichannel sound reproduction systems using wave-domain adaptive filtering,” *Journal of the Acoustical Society of America*, vol. 122, no. 1, pp. 354-369, Jul. 2007.