MULTICHANNEL ACOUSTIC ECHO CANCELLATION: WHAT'S NEW?

Tomas Gänsler[1] and Jacob Benesty[2]

[1]Agere Systems, Research
[2]Bell Laboratories, Lucent Technologies
600 Mountain Avenue, Murray Hill, New Jersey 07974-0636
gaensler@agere.com, jbenesty@bell-labs.com

ABSTRACT

Since the groundbreaking paper by Sondhi and Morgan in 1991, stereophonic (as well as multichannel) acoustic echo cancellation has been a field for active research. What problems have been studied and solved so far? What are the future directions? This paper presents some history of stereophonic echo cancellation research, newly developed theory for improved description of the misalignment problem, and views on interesting problems that could be pursued in the future.

1. INTRODUCTION

The specific problems of stereophonic acoustic echo cancellation have intrigued researchers over the past decade. The reason for this interest has mainly been the so-called "nonuniqueness problem" which exists in two or multichannel echo cancellation, but not in the single-channel case. The nonuniqueness problem arises if the (multiple) transmission audio streams originate from the same source. This means that the normal equation to be solved by the adaptive filter is singular. In this situation, *the echo canceler cannot provide a unique echo path solution*, and it can be shown that all solutions found by the echo canceler depend on the reverberation paths of the transmission room (Fig. 1). This was first presented in [26] and later thoroughly described in [27]. While the fact of nonuniqueness can be discouraging to many people, it has posed a great challenge to researchers in the field of adaptive filtering, and has inspired many to search for effective solutions to the problem.

Early discoveries showed that the normal equations to be solved are in fact not singular in practice because of unmodeled tails of the transmission room echo paths, but still are severely ill-conditioned [3], [2]. The non-stationarity of the reverberation paths in the transmission room also improves the condition number of the normal equations [25], [22]. The most extensive results regarding the nonuniqueness problem, the problem of solving the ill-conditioned normal equations, and an appropriate solution to these problems were given in [9], [10]. Another important result reported in these papers is the fact that the *only* solution to the nonuniqueness and misalignment problems is to reduce the correlation between the transmitted signals. An effective means of decorrelation by nonlinear processing was also proposed. However, what was really interesting to the research community was that there are many ways to reduce the channel correlation. This inspired many researchers to look for other preprocessing methods e.g. time-

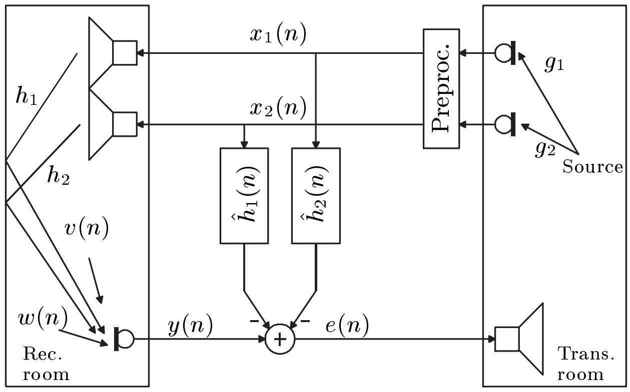

Figure 1: Block diagram of a generic two-channel acoustic echo canceler. For simplicity, only one return channel is shown.

varying filters [20], [1], noise shaping [18], [15], complementary comb filtering [8], and a variant of nonlinear processing [24]. Some reports of decorrelating filters can also be found in [27] among others. However, this approach has been less successful since the main issue of preprocessing is how to achieve significant decorrelation without adversely affecting stereo perception.

Another issue that remains after dealing with the nonuniqueness problem is slow convergence of the adaptive filter since the signals are still highly correlated. This is also a point where multichannel echo cancellation differs from the single-channel case. For the single-channel case, the common normalized least mean square algorithm (NLMS) is sufficient from a tracking and convergence point of view [19]. For the multichannel case however, we are in general forced to use more sophisticated algorithms than NLMS. Most common algorithm choices are two-channel subband versions of the fast affine projection (FAP) algorithm [16], [22], fast recursive least-squares (FRLS) [12], and lately two-channel frequency-domain adaptive algorithms [6], [13]. The frequency-domain approach has also been found very interesting from a performance and convergence point of view for systems with more than two channels [11].

Overviews of the preprocessing approaches and choices of adaptive algorithms can be found be found in [17], [21]. Following a summary of the essential knowledge about the stereo echo cancellation problem, we will present some new theoretical results explaining the effects of nonlinear processing and its influence on misalignment. Then we will conclude by discussing some relevant challenges that might be interesting to study.

2. BACKGROUND – WHAT WE KNOW

In this section, we will present what is known about stereophonic echo cancellation with respect to nonuniqueness, misalignment problems, and methods for solving these. In this discussion, we distinguish between the length (M) of the impulse responses (reverberation) in the transmission room, the length (L) of the modeling filters, and the length (N) of the impulse responses in the receiving room.

If we assume that the transmission room is linear and time invariant, we have the following important relation [3]:

$$\mathbf{x}_{1,M}^T(n)\mathbf{g}_{2,M} = \mathbf{x}_{2,M}^T(n)\mathbf{g}_{1,M} \qquad (1)$$

where we define the signal vectors of arbitrary length P as

$$\begin{aligned} \mathbf{x}_{i,P}(n) &= \begin{bmatrix} x_i(n) & x_i(n-1) & \cdots & x_i(n-P+1) \end{bmatrix}^T, \\ i &= 1,2. \end{aligned}$$

The length of these vectors can be either $P = M,\ L$, or N depending on the situation we analyze. The superscript ^{T} denotes the transpose of a vector or a matrix, and the impulse response vectors are defined as

$$\mathbf{g}_{i,M} = \begin{bmatrix} g_{i,0} & g_{i,1} & \cdots & g_{i,M-1} \end{bmatrix}^T,\ i = 1,2.$$

Let the receiving room response vectors of length N be denoted by $\mathbf{h}_{i,N},\ i = 1,2$, and the estimated responses by

$$\hat{\mathbf{h}}_{i,P}(n) = \begin{bmatrix} \hat{h}_{i,0}(n) & \hat{h}_{i,1}(n) & \cdots & \hat{h}_{i,P-1}(n) \end{bmatrix}^T$$

where also here $P = M,\ L$, or N depending again on the situation we analyze. For simplicity of notation, define the concatenated vectors as $\mathbf{h}_{2N} = \left[\mathbf{h}_{1,N}^T\ \mathbf{h}_{2,N}^T\right]^T$ and $\hat{\mathbf{h}}_{2P} = \left[\hat{\mathbf{h}}_{1,P}^T\ \hat{\mathbf{h}}_{2,P}^T\right]^T$ (analogously for the regressor vector $\mathbf{x}_{2P} = \left[\mathbf{x}_{1,P}^T\ \mathbf{x}_{2,P}^T\right]^T$).

The objective of the echo canceler is to estimate the receiving room echo paths by minimizing a criterion based on the error signal $e(n)$. Thus, let us define the recursive least-squares error criterion with respect to the modeling filters:

$$J(n) = \sum_{p=1}^{n} \lambda^{n-p} e^2(p) \qquad (2)$$

where λ ($0 < \lambda \leq 1$) is an exponential forgetting factor,

$$e(n) = y(n) - \hat{\mathbf{h}}_{2L}^T(n)\mathbf{x}_{2L}(n) \qquad (3)$$

is the error signal at time n between the microphone[1] output

$$y(n) = \mathbf{h}_{2N}^T\mathbf{x}_{2N}(n) \qquad (4)$$

and its estimate. The minimization of (2) leads to the normal equation:

$$\mathbf{R}(n) \begin{bmatrix} \hat{\mathbf{h}}_{1,L}(n) \\ \hat{\mathbf{h}}_{2,L}(n) \end{bmatrix} = \mathbf{r}(n) \qquad (5)$$

where

$$\mathbf{R}(n) = \sum_{p=1}^{n} \lambda^{n-p} \begin{bmatrix} \mathbf{x}_{1,L}(p) \\ \mathbf{x}_{2,L}(p) \end{bmatrix} \begin{bmatrix} \mathbf{x}_{1,L}^T(p) & \mathbf{x}_{2,L}^T(p) \end{bmatrix} \qquad (6)$$

[1] In reality, $y(n)$ may contain ambient noise, $w(n)$, and a near-end talker $v(n)$.

is an estimate of the input signal covariance matrix and

$$\mathbf{r}(n) = \sum_{p=1}^{n} \lambda^{n-p} y(p) \begin{bmatrix} \mathbf{x}_{1,L}(p) \\ \mathbf{x}_{2,L}(p) \end{bmatrix} \qquad (7)$$

is an estimate of the cross-correlation vector between the input and output signals.

With all of these formal definitions, we can now summarize the following about the solution of the normal equation (5).

2.1. Nonuniqueness of the solution for $L \geq M$

For $L \geq M$ case, any algorithm finds the (non-unique) solution:

$$\hat{\mathbf{h}}_{2L}(n) \quad = \quad \mathbf{h}_{2N}(n) + \zeta\mathbf{u} \qquad (8)$$

where ζ is an arbitrary constant and

$$\mathbf{u} = \begin{bmatrix} \mathbf{g}_{2,M}^T & 0 & \cdots & 0 & -\mathbf{g}_{1,M}^T & 0 & \cdots & 0 \end{bmatrix}^T \qquad (9)$$

is a vector in the nullspace of the obviously singular matrix $\mathbf{R}(n)$ since $\mathbf{x}_{2L}^T(n)\mathbf{u} = 0$ in (1). In the realistic case $L < M$ we find that the normal equations theoretically provide a unique solution but are severely ill-conditioned.

2.2. Misalignment of the solution for $L < N$

Assume the signal covariance matrix is nonsingular. The misalignment ($\|\boldsymbol{\varepsilon}\|$) is a proper way to measure how close to the true solution our estimate is. It is defined as

$$\|\boldsymbol{\varepsilon}\| \quad = \quad \|\mathbf{h}_{2L} - \hat{\mathbf{h}}_{2L}(n)\|. \qquad (10)$$

In the case where $L < N$, it is can be shown [10] that

$$\hat{\mathbf{h}}_{2L}(n) \quad = \quad \mathbf{h}_{2L} + \mathbf{Q}_{\mathrm{t}}^{1/2}\mathbf{h}_{2\mathrm{t}}(n) \qquad (11)$$

where $\mathbf{h}_{2\mathrm{t}}(n) = \left[\mathbf{h}_{1,\mathrm{t}}^T\ \mathbf{h}_{2,\mathrm{t}}^T\right]^T$ and $\mathbf{h}_{i,\mathrm{t}},\ i = 1,2$, each of length $N-L$, are the truncated (unmodeled) tails of the receiving room impulse responses. The matrix $\mathbf{Q}_{\mathrm{t}}^{1/2}$ is of full rank. Hence, we see from (10) and (11) that the misalignment is always nonzero if the truncated tails are nonzero.

2.3. Link between coherence and nonunique/ ill-conditioned solution

The coherence between two random signals x_1 and x_2 is defined in the frequency domain as

$$\gamma(f) = \frac{S_{x_1x_2}(f)}{\sqrt{S_{x_1x_1}(f)S_{x_2x_2}(f)}} \qquad (12)$$

where $S_{x_px_q}(f),\ p,q = 1,2$ are the cross- and auto-spectra (of the corresponding signals). The significance of the coherence function is that it can be shown to relate to the conditioning of the covariance matrix (6), and therefore determines the sensitivity of the normal equation solution to noise. It has been shown that the eigenvalues of the covariance matrix are lower bounded by a factor $[1 - |\gamma(f)|^2]$ [10]. Therefore, a magnitude-squared coherence of 0.999 at some frequency f would mean that the solution would be sensitive to noise at the –30 dB level. In the case where $|\gamma(f)| = 1$, there is of course no unique solution because the normal equation is singular.

2.4. Nonlinear processing

An effective method for decreasing the correlation between the signals is to pass one channel through a positive half-wave rectifier and the other channel through a negative half-wave rectifier [7] i.e.,

$$x_1'(n) = x_1(n) + \frac{\alpha}{2}\left[x_1(n) + |x_1(n)|\right], \quad (13a)$$

$$x_2'(n) = x_2(n) + \frac{\alpha}{2}\left[x_2(n) - |x_2(n)|\right], \quad (13b)$$

where $\alpha \geq 0$ is the "level of nonlinearity." Experiments have shown that, for speech, this function does not affect the stereo perception and as long as $\alpha \leq 0.5$, the introduced distortion is hardly audible. Another study has also shown that for audio, the distortion introduced by $\alpha = 0.3$ is comparable to the distortion experienced after MPEG level III compression [28].

As mentioned in the introduction, there are other ways of reducing the correlation between the signals than nonlinear processing. However, not only does the half-wave rectifier approach work well in practice, but as we will see in Sect. 3, it is actually possible to theoretically analyze its impact on the performance of the echo canceler.

Before proceeding, we would like to clarify an erroneous assumption of how nonlinear processing can be applied in a multichannel system to "solve" the nonuniqueness problem.

2.5. A common misconception

A common misunderstanding is that it is enough to process the input to the adaptive filter only in order to solve the nonuniqueness problem. That is, let the (linear or nonlinear) preprocessing (here denoted by the general function $\Phi[\cdot]$) of the signals only affect the regressor to the adaptive filter while unprocessed signals are transmitted to the receiving room. This would of course solve the problems of audible distortion, but does it solve the nonuniqueness problem? The answer is no and it can be shown by rewriting the normal equation (5), which can be expressed as the recursive average of the regressor $\mathbf{x}_{2L}(n)$ and residual error $e(n)$ as

$$\sum_{p=1}^{n} \lambda^{n-p} \Phi\left[\mathbf{x}_{2L}(n)\right] \mathbf{x}_{2L}^T(n) \mathbf{h}_{2L} = \sum_{p=1}^{n} \lambda^{n-p} \Phi\left[\mathbf{x}_{2m}(n)\right] y(n). \quad (14)$$

Unfortunately, by using (1) and (9), we find that $(L \geq M)$

$$\left[\sum_{p=1}^{n} \lambda^{n-p} \Phi\left[\mathbf{x}_{2L}(n)\right] \mathbf{x}_{2L}^T(n)\right] \mathbf{u} = \mathbf{0}. \quad (15)$$

Hence, the normal equation is still singular and there are still nonunique solutions.

3. NEW THEORY FOR THE MISALIGNMENT PROBLEM

Since the fundamental issues of stereo as well as multichannel echo cancellation have been summarized, we will now move on to present new theory that gives more detailed insight into the misalignment caused by residual correlation between the channels after nonlinear preprocessing. This is achieved by extending the existing link between the coherence function and the condition of the covariance matrix (6). Moreover, we will present an implicit relation between the coherence value after nonlinear processing (13) (γ_α) and the level of nonlinearity and the actual coherence value before processing (γ). In other words, we would like to find answers to the questions:

(i) How does the level of nonlinearity (α) influence the level of coherence (γ_α) after processing?

(ii) How does the level of coherence influence the misalignment $\|\boldsymbol{\epsilon}\|$ (10)?

Since the derivations of the desired relations are somewhat lengthy, we refer to [14] for details.

To get manageable relations, we choose to model the transmission signals x_1 and x_2 as constant spectrum (white) Gaussian signals. The coherence between the signals is also constant $\gamma(f) = \gamma \geq 0$, and the signals are band-limited in frequency between $\pm f_s/2$ with variance σ_x^2. (The sampling frequency is denoted by f_s.) We refer to this model as the anechoic model. This may not look like the best model for a real-life situation. However, for speech in an office environment, it works fairly well as is shown in [14].

3.1. Link between nonlinearity and coherence

Question (i) can be answered by looking at a theoretical model (in the same fashion that was done in [23]) for the coherence between the two signals that have passed through a nonlinearity. However, a modification of the ideas in [23] is necessary for the positive and negative half-wave rectifier. The resulting magnitude coherence between the signals after nonlinear processing (x_1' and x_2'), denoted by γ_α, is given by

$$\gamma_\alpha = |\gamma_\alpha(f)| = \frac{|S_{x_1'x_2'}(f)|}{\sqrt{S_{x_1'x_1'}(f) S_{x_2'x_2'}(f)}},$$

and it is shown in [14] that this magnitude coherence can be explicitly expressed for $f \neq 0$ as

$$\gamma_\alpha = \frac{\gamma + \frac{\beta}{2}\left\{\gamma - \frac{1}{\pi}\left[\gamma \cos^{-1}(-\gamma) + \sqrt{1-\gamma^2} - 1\right]\right\}}{1 + \frac{\beta}{2}\left(1 - \frac{1}{\pi}\right)}, \quad (16)$$

where $\beta = \alpha^2/(1+\alpha)$.

In Fig. 2 we compare (16), i.e., γ_α, with estimates of the coherence function of the signals generated according to the anechoic model after they have passed through the nonlinear function given in (13). The theory and simulation completely agree over our choice of range of the parameters: $0 \leq \alpha \leq 2$ and $\gamma = 1,\ 0.9,\ 0.8$.

3.2. Link between coherence and misalignment

Question (ii) is given by the misalignment formula for the two-channel recursive least-squares (RLS) or frequency-domain algorithms. For these algorithms, it can be shown that the expected normalized (by $\|\mathbf{h}_{2L}\|^2$) misalignment energy is given by [5, Chap. 8],

$$\frac{E\{\|\boldsymbol{\epsilon}(n)\|^2\}}{\|\mathbf{h}_{2L}\|^2} = \frac{(1-\lambda)}{2} \frac{\sigma_b^2}{\|\mathbf{h}_{2L}\|^2} \mathrm{tr}\left\{\mathbf{R}^{-1}\right\}, \quad (17)$$

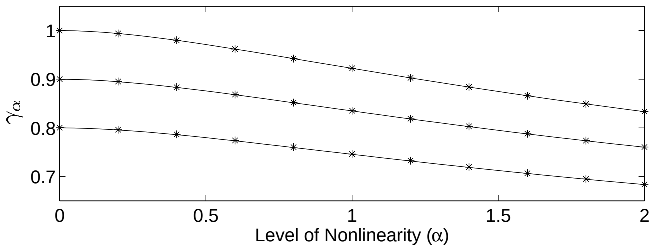

Figure 2: Comparison of theoretical (solid) and estimated (∗) magnitude coherence after nonlinearly processing the transmission signals with a positive and a negative half-wave rectifier (13). The magnitude coherence of the unprocessed signals (before the rectifiers) corresponds to $\alpha = 0$.

where $\sigma_b^2 = \sigma_v^2 + \sigma_w^2$ (variance of ambient noise and near-end talker) is the background noise power, and $\mathbf{R}$ is defined as

$$\mathbf{R} \quad = \quad (1-\lambda)E\left[\mathbf{R}(n)\right]. \tag{18}$$

$E\left[\cdot\right]$ denotes expectation and $\mathbf{R}(n)$ is given in (6).

By Fourier transforming to diagonalize $\mathbf{R}$, we can explicitly compute the trace in (17). Furthermore, to be somewhat more general than the anechoic model, assume a model where $\gamma(l),\ l = 0 \ldots L-1$, is *not* constant with frequency ($f = l/L$), but the signal autospectra $S_{x_p x_p} = \sigma_x^2,\ p = 1,2$ are still constant with frequency. We then find

$$\frac{E\{\|\boldsymbol{\varepsilon}(m)\|^2\}}{\|\mathbf{h}_{2L}\|^2} \quad = \quad \frac{1-\lambda}{\mathrm{EBR}}\left[\frac{1}{L}\sum_{l=0}^{L-1}\frac{1}{1-|\gamma(l)|^2}\right], \tag{19}$$

where EBR is the echo-to-background ratio

$$\mathrm{EBR} = \frac{||\mathbf{h}_{2L}||^2\sigma_x^2}{\sigma_b^2}. \tag{20}$$

For expression (19), we have also assumed that $L \geq N$, $\|\mathbf{h}_{1,N}\| = \|\mathbf{h}_{2,N}\|$, and $\sigma_{x_1}^2 = \sigma_{x_2}^2$.

Figure 3 shows the theoretical misalignment, given by expression (19), and the estimated misalignment that results from using the frequency-domain adaptive algorithm in [5, Chapt. 8]. Here, the magnitude coherence ($|\gamma|$) is varied between 0 and 1. The estimated misalignment is found by averaging instantaneous normalized misalignment estimates ($\|\mathbf{h} - \hat{\mathbf{h}}(n)\|^2/\|\mathbf{h}\|^2$) over 32000 samples after the echo canceler has converged. The ambient noise level, $\mathrm{EBR} = \sigma_{y_e}^2/\sigma_b^2 \approx 1000$ (30 dB). For the adaptive filter, we have chosen $L = N = 1024$, and $\lambda = [1 - 1/(3 \cdot 2L)]^L$.

The factor within brackets of (19) quantifies: (1) the increase in misalignment due to coherence, e.g. if $\gamma = 0.95$ the misalignment will be increased by 10 dB; and (2) the mean square error echo perturbation that occurs after a transmission room echo path change, i.e. a 10 dB louder echo will be returned before the echo canceler reconverges if $\gamma = 0.95$. Furthermore, the two expressions (16) and (19) can also be used for adjusting the level of nonlinearity such that the misalignment is kept at a desired level [14].

4. CHALLENGES – WHAT WE WOULD LIKE TO KNOW

In this paper, the theory for multichannel echo cancellation problems has been presented. What are the next crucial

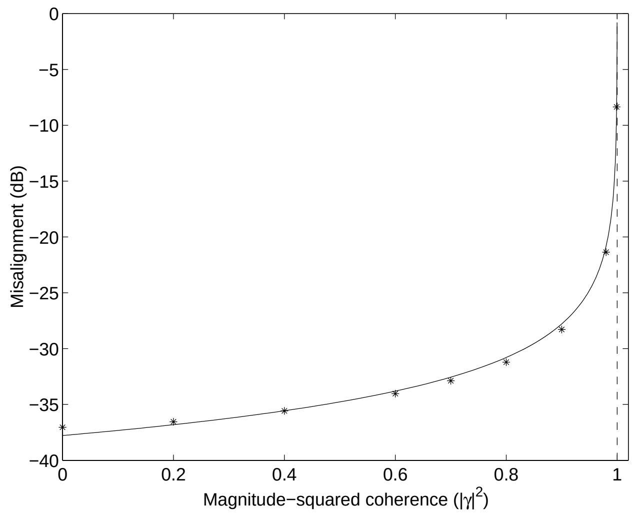

Figure 3: Comparison of theoretical (solid) and estimated misalignment (∗) as a function of magnitude-squared coherence.

steps in this field of research? The authors are of the opinion that the best way to proceed is to acquire experience with real-time systems like the ones presented in [12], [13]. Problems that arise from implementations like these inspire and fuel the much needed research in order to make multichannel acoustic echo cancellation a successful product in the telecommunication industry. The following list of problems could be considered as "hot topics" for the research community in the near future:

1. Echo path imbalance problems. If directional microphones are used in the receiving room, the two echo paths will be very different in magnitude ($\|\mathbf{h}_{1,L}\| \neq \|\mathbf{h}_{2,L}\|$). This leads to high sensitivity of the estimate of the path with low magnitude. How should the adaptive algorithm handle this situation?

2. The echo path imbalance becomes especially problematic during double-talk. Will independent path control of the adaptive algorithm help?

3. Optimize the general double-talk detection structure presented in [4] for the echo path imbalance problem.

4. Improved preprocessing techniques is required for audio, not only speech, especially for a multichannel (e.g. 5.1 channels surround sound) system.

5. Analysis of other preprocessing methods like that presented in Sect. 3 are required for optimization.

6. Fair comparisons of preprocessing methods that take into account perceived distortion (noise and audio image).

7. Higher convergence rate of the adaptive algorithm is desirable. There is a noticeable penalty in convergence rate in two-channel systems compared to the single-channel case.

8. Complexity reduction is of major interest for the multichannel algorithms.

Acknowledgments

The authors would like to thank D. R. Morgan for inspiring discussions and useful comments, G. Elko and once again D. R. Morgan for proof reading this manuscript. Also, we would especially like to thank our recently retired colleagues, M. M. Sondhi, J. L. Hall, and B.-H. Juang for the excellent support and productive collaboration we have enjoyed over the years working together.

5. REFERENCES

[1] Ali M.: *Stereophonic echo cancellation system using time-varying all-pass filtering for signal decorrelation.* In *Proc. IEEE ICASSP*, pages 3689–3692, 1998.

[2] Amand F.: *Étude de l'annulation d'écho multi-voie et application á la téléconférence de haute qualité.* PhD thesis, Université de Rennes 1, 1996.

[3] Benesty J., Amand F., Gilloire A., and Grenier Y.: *Adaptive filtering algorithms for stereophonic acoustic echo cancellation.* In *Proc. IEEE ICASSP*, pages 3099–3102, 1995.

[4] Benesty J. and Gänsler T.: *A multichannel acoustic echo canceler double-talk detector based on a normalized cross-correlation matrix.* In *Proc. IWAENC*, 2001.

[5] Benesty J., Gänsler T., Morgan D. R., Sondhi M. M., and Gay S. L.: *Advances in Network and Acoustic Echo cancellation.* Springer-Verlag, Berlin, 2001.

[6] Benesty J., Gilloire A., and Grenier Y.: *A frequency domain stereophonic acoustic echo canceler exploiting the coherence between the channels. J. Acoust. Soc. Am.*, 106:L30–L35, Sept. 1999.

[7] Benesty J., Morgan D. R., Hall J. L., and Sondhi M. M.: *Synthesized stereo combined with acoustic echo cancellation for desktop conferencing. Bell Labs Tech. J.*, 3:148–158, July-Sept. 1998.

[8] Benesty J., Morgan D. R., Hall J. L., and Sondhi M. M.: *Stereophonic acoustic echo cancellation using nonlinear transformations and comb filtering.* In *Proc. IEEE ICASSP*, pages 3673–3676, 1998.

[9] Benesty J., Morgan D. R., and Sondhi M. M.: *A better understanding and an improved solution to the problems of stereophonic acoustic echo cancellation.* In *Proc. IEEE ICASSP*, pages 303–306, 1997.

[10] Benesty J., Morgan D. R., and Sondhi M. M.: *A better understanding and an improved solution to the specific problems of stereophonic acoustic echo cancellation. IEEE Trans. Speech Audio Processing*, 6:156–165, Mar. 1998.

[11] Buchner H. and Kellermann W.: *Acoustic echo cancellation for two and more reproduction channels.* In *Proc. IWAENC*, 2001.

[12] Eneroth P., Gay S. L., Gänsler T., and Benesty J.: *A real-time stereophonic acoustic subband echo canceler.* In S. L. Gay and J. Benesty, editors, *Acoustic Signal Processing for Telecommunications*, chapter 8, pages 135–152. Kluwer Academic Publishers, 2000.

[13] Fischer V., Gänsler T., Benesty J., and Diethorn E. J.: *A software stereo acoustic echo canceler under Microsoft Windows™.* In *Proc. IWAENC*, 2001.

[14] Gänsler T. and Benesty J.: *New insights to the stereophonic acoustic echo cancellation problem and an adaptive nonlinearity solution. IEEE Trans. Speech Audio Processing.* Submitted.

[15] Gänsler T. and Eneroth P.: *Influence of audio coding on stereophonic acoustic echo cancellation.* In *Proc. IEEE ICASSP*, pages 3649–3652, 1998.

[16] Gay S. L. and Tavathia S.: *The fast affine projection algorithm.* In *Proc. IEEE ICASSP*, pages 3023–3026, 1995.

[17] Gilloire A.: *Current issues in stereophonic and multichannel acoustic echo cancellation.* In *Proc. IWAENC*, pages K–5–8, 1997.

[18] Gilloire A. and Turbin V.: *Using auditory properties to improve the behavior of stereophonic acoustic echo cancellers.* In *Proc. IEEE ICASSP*, pages 3681–3684, 1998.

[19] Hänsler E.: *From algorithms to systems—it's a rocky road.* In *Proc. IWAENC*, pages K1–K4, 1997.

[20] Y. Joncour and Sugiyama A.: *A stereo echo canceller with pre-processing for correct echo path identification.* In *Proc. IEEE ICASSP*, pages 3677–3680, 1998.

[21] Makino S. and Shimauchi S.: *Stereophonic acoustic echo cancellation - an overview and recent solutions.* In *Proc. IWAENC*, pages 12–19, 1999.

[22] Makino S., Strauss K., Shimauchi S., Haneda Y., and Nakagawa A.: *Subband stereo echo canceller using the projection algorithm with fast convergence to the true echo path.* In *Proc. IEEE ICASSP*, pages 299–302, 1997.

[23] Morgan D. R., Hall J. L., and J. Benesty J.: *Investigation of several types of nonlinearities for use in stereo acoustic echo cancellation. IEEE Trans. Speech Audio Processing.* To appear.

[24] Shimauchi S., Haneda Y., Makino S., and Kaneda Y.: *New configuration for a stereo echo canceller with nonlinear pre-processing.* In *Proc. IEEE ICASSP*, pages 3685–3688, 1998.

[25] Shimauchi S. and Makino S.: *Stereo projection echo canceller with true echo path estimation.* In *Proc. IEEE ICASSP*, pages 3059–3062, 1995.

[26] Sondhi M. M. and Morgan D. R.: *Acoustic echo cancellation for stereophonic teleconferencing.* In *Proc. IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics*, 1991.

[27] Sondhi M. M., Morgan D. R., Hall J. L.: *Stereophonic acoustic echo cancellation – an overview of the fundamental problem. IEEE Signal Processing Lett.*, pages 148–151, Aug. 1995.

[28] Till O., Rodger B., and Johansson N.: *Informal listening test of stereophonic perception after nonlinear transformation.* Unpublished work, May 1997.