ACOUSTIC ECHO AND NOISE CONTROL – A LONG LASTING CHALLENGE

Pia Dreiseitel *Eberhard Hänsler* *Henning Puder*
Signaltheorie
Darmstadt University of Technology, D–64283 Darmstadt, Germany
{dreiseit,haensler,hpuder}@nesi.tu-darmstadt.de

ABSTRACT

Hands–free operation of telephones, incorporating echo cancellation and noise reduction, has been discussed for over a decade. This paper presents an overview of the wide range of algorithms which are applicable to echo cancellers and noise reduction. Practical problems associated with implementation and overall system control are also discussed.

1 INTRODUCTION

When telecommunications started about a century ago users had their two hands busy [1]. They had to hold a microphone close to their mouth and a loudspeaker close to one ear. It did not take long to get one hand free: microphone and loudspeaker were assembled in a handset. However, the aim of hands–free operation has not yet been attained.

In early years of telecommunication the lack of efficient electro acoustic devices and amplifiers justified the inconvenience to the customer. At the same time two problems were solved:

- acoustic echos transmitted back to the remote user were reduced by providing sufficient attenuation,
- operation in a noisy environment was possible by an improved signal to noise ratio.

For non–experts it is still difficult to understand that it takes all the signal processing capabilities available today to achieve at least "some" solution of these easily explained problems of hands–free operation. A large number of papers on the topic under consideration have been published within the last few years including bibliographies [2, 3, 4, 5] and reports on the state of the art [6, 7]. Adaptive algorithms for acoustic echo compensation and noise control gained special attention in [8, 9].

2 BASICS

At the most general level, there are two sources that make the solution of the hands–free problem difficult: first the physical properties of loudspeaker–enclosure–microphone systems (LEMS's) and speech signals and secondly the fulfillment of the regulations of the International Telecommunications Union (ITU). Although the latter may seem arbitrary, it is essential for the equipment to be licensed by telecommunication authorities.

2.1 Physics

Audio communication systems include at least one loudspeaker and one microphone housed within the same enclosure. Consequently, the microphone picks up not only locally generated signals like speech and environmental noise but also the signal radiated by the loudspeaker as well as its echos caused by reflections at the boundaries of the enclosure. Assuming linearity, the audio characteristics of the LEMS may be modeled by an impulse response. The duration of the response depends on the reverberation time of the enclosure. In case of an office room this time is in the order of several hundred milliseconds, in case of a passenger car it is in the order of fifty to one hundred milliseconds. Furthermore, the response of the LEMS is extremely sensitive to any movements within the enclosure. Finally, the system is driven by audio signals, typically a mixture of speech and noise, where speech itself is comprised by periodic and aperiodic components with highly fluctuating magnitudes and pauses. Briefly, from a signal processing point of view the system and the signals involved are extremely unpleasant.

2.2 Regulations

The ITU–T Recommendations [10] put very stringent conditions on hands–free telephone systems. For "ordinary" telephones the echo attenuation has to be at least 45 dB in the case of single talk. In double talk situations this value can be reduced by 15 dB. Beyond that, only a negligible delay may be introduced into the signal path by the hands–free facility.

2.3 Systems for Acoustic Echo and Noise Control

A number of systems have been proposed for acoustic echo and noise control. They all use three units (or a subset of them) (Fig. 1): The first unit is a loss control that attenuates the incoming and/or the outgoing signal. Early hands–free communication systems used this unit only reducing conversations to half duplex. Because of the ITU regulations loss control still remains the most important function because it has to guarantee the required attenuation. The second unit consists of an adaptive filter functioning as a replica of the LEMS. If perfect adaptation could be achieved loudspeaker and microphone would be decoupled entirely without any impact on locally generated signals. The third – most modern – "ingredient" of an echo and noise control system consists of a Wiener filter within the outgoing signal path. In contrast to the loss control unit this filter provides a frequency dependent attenuation of the outgoing signal. Its aim is twofold: to suppress residual echos not covered by the adaptive filter and to enhance speech quality by suppressing noise components.

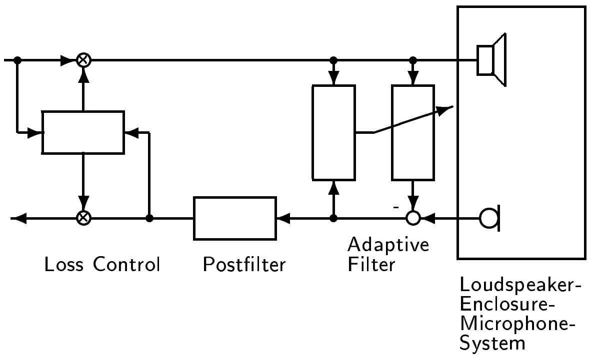

Figure 1: Basic structure for acoustic echo and noise control

Design considerations and results achieved by these three units will be given within the following sections.

3 ADAPTIVE ALGORITHMS

Several adaptation algorithms have been applied to acoustic echo cancellers. Each of these algorithms minimises the mean square error of signal $e(k)$ (s. Fig. 2). The algorithms discussed in this section are dealt with in order of increasing complexity.

3.1 NLMS

The least mean square (NLMS) algorithm is the most easily and frequently implemented algorithm. It is described by the following relations:

$$e(k) \quad = \quad d(k) - \underline{c}^T(k)\,\underline{x}(k), \tag{1}$$

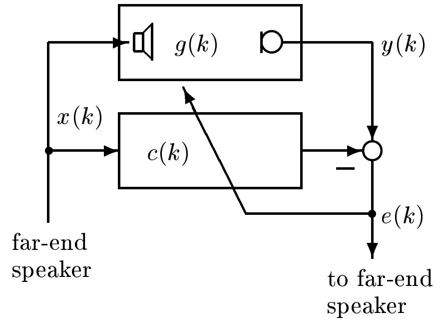

Figure 2: Adaptive System

$$\underline{c}(k+1) \quad = \quad \underline{c}(k) + \alpha\,\frac{e(k)\,\underline{x}(k)}{\|\underline{x}(k)\|^2}, \tag{2}$$

where $e(k)$ denotes the adaptation error, $d(k)$ the desired signal, $\underline{c}(k)$ the coefficient vector of the adaptive filter, $\underline{x}(k)$ the excitation vector and finally α a variable step size factor.

The computational requirements of the NLMS algorithms are low. This is important since the application considered here requires a large number of coefficients. The disadvantage is its slow speed of convergence especially in case of correlated inputs.

3.2 NLMS algorithm with pre–whitening filters

A simple approach to overcome this problem is pre–whitening of both input and reference signal. This can be achieved by a linear prediction error filter. Restricting oneself to a time invariant filter, a filter of order one to four proved to be sufficient. Pre–whitening filters of higher order have to be adaptive. They can be designed using the Levinson–Durbin–algorithm.

3.3 Step size Control

As mentioned before, the NLMS algorithm uses an adaptation factor α called step size which is responsible for both stability and speed of convergence. Controlling the step size becomes especially important in case of noisy microphone signals like those in car environments or in double talk situations [16].

It can be shown [17] that an optimal step size exists for adaptation in a noisy environment which is also suitable for non stationary excitation signals:

$$\alpha_{opt} = \frac{\mathrm{E}\{\epsilon(k)^2\}}{\mathrm{E}\{e(k)^2\}}, \tag{3}$$

where E{...} denotes expectation, $e(k)$ the current error value and $\epsilon(k)$ the residual echo signal. The application of (3), however, requires a reliable double talk detector [18].

3.4 Affine Projection Algorithm

Looking closely at the affine projection algorithm (APA), it can be considered as an extension of the NLMS algorithm, taking into account the P last excitation vectors.

$$\begin{aligned} e(k) &= d(k) - \underline{c}^T(k)\underline{x}(k), && (4) \\ \underline{e}(k) &= (e(k), e(k-1), \ldots, e(k-P+1))^T, && (5) \\ \underline{X}(k) &= (\underline{x}(k), \underline{x}(k-1), \ldots, \underline{x}(k-P+1))^T, && (6) \\ \underline{c}(k+1) &= \underline{c}(k) + \underline{X}(k)\,(\underline{X}^T(k)\,\underline{X}(k))^{-1}\,\underline{e}^T(k). && (7) \end{aligned}$$

Usually P is small compared to the total number of filter coefficients. In contrast to the NLMS algorithm, the matrix $\underline{X}^T(k)\,\underline{X}(k)$ has to be inverted. This can be carried out recursively. A fast version of the APA – called FAP (fast affine projection [11]) – has been developed for an efficient implementation. This algorithm is therefore suitable for acoustic echo cancellers. However, numerical instabilities occur because of recursively calculated correlation matrices. One can overcome these problems by regularising the correlation matrix by adding a constant value to the values of the main diagonal. Furthermore, the algorithm has to be reinitialised whenever divergence is detected.

If an affine projection of second order is applied, the inverse of the matrix can be calculated directly requiring only small computational load. Compared to the NLMS algorithm, even a second order APA increases the speed of convergence remarkably.

3.5 RLS and FTF

The recursive least squares algorithm (RLS) is known as a very fast converging recursive algorithm. A straight forward notation of this algorithm is given here:

$$\begin{aligned} \underline{w}(k) &= \lambda^{-1} R_{xx}^{-1}(k)\,\underline{x}(k), && (8) \\ R_{xx}^{-1}(k+1) &= \lambda^{-1} R_{xx}^{-1}(k) - \frac{\underline{w}(k)\,\underline{w}^T(k)}{1 + \underline{w}^T(k)\,\underline{x}(k)}, && (9) \\ e(k) &= d(k) - \underline{c}^T(k)\,\underline{x}(k), && (10) \\ \underline{c}(k+1) &= \underline{c}(k) + e(k)\,\underline{w}(k), && (11) \end{aligned}$$

where $R_{xx}(k)$ denotes an estimate of the autocorrelation matrix of the excitation signal, λ an exponential forgetting factor ($0 < \lambda < 1$) and $\underline{w}(k)$ the gain vector. The convergence of the RLS algorithm is superior to the NLMS algorithm. However, there is the problem of locking when λ is chosen close to one. The tracking performance of the RLS algorithm is therefore not as satisfying as the initial convergence.

If the algorithm is implemented with finite–precision, it can become unstable for the numerical round–off error increases. A QR–Decomposition based inversion of the autocorrelation matrix does not show this behaviour [12].

If one has to deal with a large number of coefficients, the direct implementation of the RLS algorithm is not feasible since its computational complexity of order M^2.

Several approaches for a fast version of the RLS algorithm are known, principally based on pre–windowing techniques which reduce the computational load to order M. A fast implementation of the RLS algorithm – called Fast Transversal Filter algorithm (FTF) – is organised in four steps:

- Recursive forward linear prediction.
- Recursive backward linear prediction.
- Recursive computation of the gain vector.
- Recursive estimation of the desired response.

Unfortunately, the FTF algorithm is numerically unstable and tends to diverge. In fact, stabilising the RLS algorithms is a topic in its own right [13, 15]. One basic idea is to extend the algorithm by accumulating the round–off errors and to perform corrections when the numerical error becomes significant.

3.6 Fast Newton

Whereas the APA may be considered as an extended version of the NLMS algorithm, the Fast Newton algorithm can be seen as a simplified version of the fast RLS algorithm [14]. In fast implementations of the RLS algorithm, linear predictions of the order M are required, where M is the size of the coefficient vector. When the order of correlation of the excitation signal is small, there is actually no need to calculate the full prediction vector of order M. Reducing the size of the prediction vector to a size P appropriate to the excitation signal leads to the Fast Newton algorithm. The convergence performance is comparable to the RLS algorithm, whereas the numerical complexity is of order MP.

3.7 Fullband – Subband – Block–processing

Until now, our discussion of adaptive filters has dealt only with fullband signals, since this is the most straight forward method of implementation. However, straightforward does not necessarily mean most efficient. Both sub–band and block processing enable implementations resulting in less computational cost.

If a signal is split up into subbands, one can subsample the resulting signals leading to shorter adaptive echo cancellers. All of the adaptive algorithms mentioned above are suitable for subband implementation. The processing power saved may be used for more complex adaptation. However, subband realisations do have one substantial disadvantage that may prohibit their application: they introduce delay into the system [10]. This delay is caused by the filter–banks for analysis (decomposition) and synthesis of the excitation and error signals. These filter–banks have to be designed with respect to the special demands of an adaptive echo canceller. The aliasing terms for example have to be minimised [19]. There is a substantial body of literature concerned with the design of polyphase filter banks used for echo cancellation (e.g. [20, 21]).

In block processing, the impulse response of the adaptive filter is split up into blocks. Using fast convolution techniques, the calculation of the output signal can be carried out very efficiently [22]. Again, there is a trade–off between efficiency of processing and delay. However, block–processing offers the advantage of optimising delay versus processing power. Small block sizes keep the delay low but increase the processing power required.

4 STEREOPHONIC ECHO CANCELLATION

Recently stereophonic acoustic echo cancellation became more and more important for applications such as teleconferencing or video games [23].

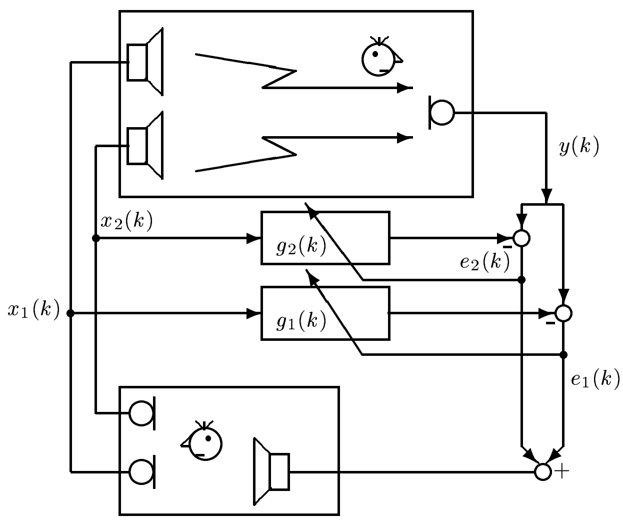

Figure 3: Stereophonic echo cancellation

As the excitation signals of the two channels are correlated (Fig. 3), there is no unique solution for identifying the two impulse responses. Furthermore, an extended correlation matrix of the two input signals has to be inverted. In case of high correlation, this causes numerical instabilities due to ill–conditioning which, in turn, leads to divergence. However, there are a number of approaches to overcome the correlation of the two excitation signals. One technique applies a nonlinear function to one of the excitation signals [23]. In a second approach the correlation matrix is regularised by introducing leakage into the update of the coefficient vector [24].

5 NOISE REDUCTION

With the increasing number of mobile telephones, more and more people use them in cars. This generates a demand for hands–free telephone sets for cars that not only increase the comfort to the user but also allow the driver to keep his hands on the steering wheel.

To enhance the speech signal outgoing to the far–end user, noise reduction methods are desirable.

We describe one channel methods for two reasons: first the cost for installing a second channel may be prohibitive, and secondly single channel procedures can also be extended to multi–channel methods.

5.1 Basic architecture

Most noise reduction procedures are based on the Wiener solution [25]:

$$G_{opt}(kB,n) = \begin{cases} 1 - \left(\kappa \frac{N_{PSD}(kB,n)}{X_{PSD}(kB,n)}\right)^p : G_{opt} > \beta_f \\ \beta_f \quad : \text{otherwise}, \end{cases} \tag{12}$$

where $N_{PSD}(kB,n)$ and $X_{PSD}(kB,n)$ denote the PSD of the noise and the distorted input signal respectively and B is equal to the block size. The frequency index is given by n. Compared to the well–known Wiener filter an overestimation factor κ, a variable power p, and a spectral floor β_f are introduced.

Unfortunately, there is a conflict between the ratio of the noise reduction and the quality of the resulting speech signal. The parameters suggested above have to be chosen such that a subjective obtimum is achieved.

To preserve natural sounding speech the spectral floor is introduced which in turn limits the SNR–improvement to $-20\log(\beta_f)$dB. The imprecision associated with estimation of the time–varying PSDs causes an unpleasant tonal noise. The so–called musical–tones can be attenuated by tailoring the transfer function adequately with the additional parameters.

Modifications of the filter (12) are given by the MMSE–STSA estimator (Minimum Mean Square Error Short–Time Spectral Amplitude) and its derivation the MMSE–LSA estimator (Minimum Mean Square Error Logarithmic Spectral Amplitude) [26, 27]. To derive the algorithms the time–varying property of the distorted input signal has been taken into account. For these algorithms an ‘a priori‘ and an ‘a posteriori‘ signal to noise ratio (SNR) are estimated:

$$SNR_{post}(kB,n) = \frac{|X(kB,n)|^2}{|N_{PSD}(kB,n)|^2} - 1, \tag{13}$$

$$SNR_{prio}(kB,n) = (1-\gamma)\max(SNR_{post}(kB,n),0)$$

$$+\gamma \frac{|G_{opt}((k-1)B,n)\, X(kB,n)|^2}{|N_{PSD}(kB,n)|^2}. \quad (14)$$

$X(kB,n)$ describes the STFT of the input signal $x(k)$ at block and kB. The weighting rules for the algorithms are given by:

1) MMSE–STSA:

$$G_{opt}(kB,n) \quad = \quad (15)$$

$$\frac{\sqrt{\pi}}{2}\sqrt{\frac{1}{1+SNR_{post}(kB,n)}\frac{SNR_{prio}(kB,n)}{1+SNR_{prio}(kB,n)}}$$

$$* M_1\left[(1+SNR_{post}(kB,n))\frac{SNR_{prio}(kB,n)}{1+SNR_{prio}(kB,n)}\right]$$

with: $M_1[u] = \exp(-\frac{u}{2})\left[(1+u)I_0(\frac{u}{2}) + I_1(\frac{u}{2})\right]$

2) MMSE – LSA:

$$G_{opt}(kB,n) \quad = \quad \frac{SNR_{prio}(kB,n)}{1+SNR_{prio}(kB,n)} \quad (16)$$

$$* M_2\left[(1+SNR_{post}(kB,n))\frac{SNR_{prio}(kB,n)}{1+SNR_{prio}(kB,n)}\right]$$

with: $M_2[u] = \exp\left\{\frac{1}{2}\int_u^\infty \frac{e^{-t}}{t}dt\right\}$ and I_0, I_1 the modified Bessel functions of first and second order.

5.2 Frequency Decomposition

As shown above the noise reduction filter is defined in the frequency domain. Therefore a frequency analysis of the non–stationary input signal is required. One method achieving this is to use the STFT (Short Time Fourier Transform) which needs the multiplication of the input signal by a time–window $\gamma(m)$:

$$X(kB,n) = \sum_{m=0}^{N-1} x(m)\gamma(m-kB)e^{-j\frac{2\pi}{N}nm} \quad (17)$$

Subband decomposition provides a second class of methods. The sample values of the subband signals can produce a set of spectral coefficients for the noise reduction algorithm (Fig. 4).

After noise reduction the subband signals are upsampled, passed through anti–aliasing filters, and synthesised to obtain the enhanced output signal. The filterbanks shown in Fig. 4 split the input signal into uniformly spaced frequency bands [20] comparable to the STFT. Modifications include non–uniformly spaced frequency resolutions [28] offering the possibility of modeling the human perception system (ear and brain) (s. 5.4). Alternatively non–uniformly distributed resolution can be obtained by cascaded filter banks (Fig. 5) including the special case of the discrete wavelet transform (Fig. 6) [29, 30].

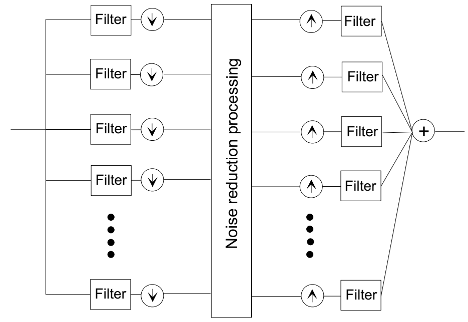

Figure 4: Filterbank

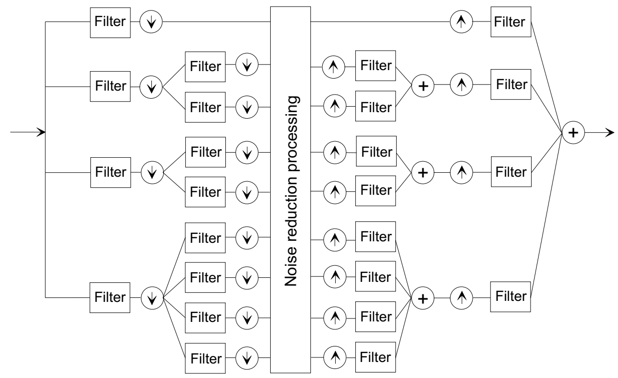

Figure 5: Cascaded filterbank

With these cascaded structures also different time–resolutions are obtained as subsampling is performed after each filter stage. Fast varying high frequency components can be treated with a higher resolution in time whereas low frequency components show a more detailed frequency resolution.

5.3 Estimation of the Power Spectral Densities

The time–frequency analysed input signal can be used to estimate $X_{PSD}(kB,n)$ and $N_{PSD}(kB,n)$.

To determine $X_{PSD}(kB,n)$ a recursively smoothed periodogramm is sufficient. However, only slight smoothing is tolerable to avoid echo–reverberation effects in the enhanced signal.

The estimation of $N_{PSD}(kB,n)$ has to be based on $X(kB,n)$ also. To distinguish between noise compo-

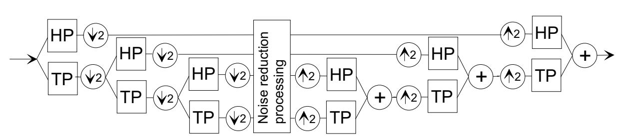

Figure 6: Wavelet filterbank

nents and speech components, either voice activity detection or the so–called Minimum–Statistics [31, 32] are necessary. In the latter, the minima of the smoothed input spectrum are observed for each frequency bin over a time window. The length of this time window is chosen according to the duration of speech components.

5.4 Psycho Acoustics

Recent studies use psycho acoustics to improve noise reduction algorithms. Two approaches are followed: the first is to adapt the signal analysis to the human ear, the second is to use the so–called masking thresholds.

It is known that the human ear performs a non–uniform frequency analysis on a logarithmic scale (Bark–scale) [33]. Methods presented in section 5.2 allow a frequency analysis adapted to the human perception.

Masking means that weak tones are covered by stronger neighbouring tones in time or frequency. Noise reduction filter design takes advantage of these properties [34, 35].

5.5 Combinations

When acoustic echo control is applied in a noisy environment such as in cars, a combination of noise reduction and echo control is desirable. As far as the succession is concerned echo control should precede noise reduction [36] so that parts of the echo not compensated by the adaptive filter may be considered as additional noise [37, 38].

The stationarity assumption for the background noise does not hold for the residual echo. Therefore different estimation methods have to be applied. The power spectral density (PSD) of the residual echo ($E_{PSD}(kB,n)$) is estimated by a power transfer factor $\gamma(kB)$ [38], $E_{PSD}(kB,n) = \gamma(kB) * F_{PSD}(kB,n)$, or a transfer function $\Gamma(kB,n)$ [37], $E_{PSD}(kB,n) = \Gamma(kB,n) * F_{PSD}(kB,n)$, specifying the ratio between the far–end signal and the residual echo, where $F_{PSD}(kB,n)$ denotes the PSD of the far–end signal.

The distortion signal is given by the sum of the estimated noise and the residual echo

$$N'_{PSD}(kB,n) = N_{PSD}(kB,n) + E_{PSD}(kB,n). \quad (18)$$

Consequently $N'_{PSD}(kB,n)$ replaces $N_{PSD}(kB,n)$ in the noise reduction methods presented above.

Separating the two problems of cancelling the residual echo and noise suppression by applying two filters in series offers additional degrees of freedom [35].

5.6 Multi–Microphone–Solutions

Microphone arrays offer further improvements of noise reduction.

A simple approach consists of delay and sum beamforming, where a control system adapts the direction of maximum sensitivity towards the near–end speaker.

Assuming that the different microphone signals are comprised by correlated speech and uncorrelated noise one yields an improved estimate of the noise power spectral density [39]. The following formulas illustrate the procedure:

$$X_1(kB,n) = \alpha_1\, S(kB,n) + N_1(kB,n), \quad (19)$$

$$X_2(kB,n) = \alpha_2\, S(kB,n) + N_2(kB,n), \quad (20)$$

$$\begin{aligned} C(n) &= \frac{E^2\left\{X_1(kB,n)\, X_2(kB,n)\right\}}{E\left\{{X_1}^2(kB,n)\right\}\, E\left\{{X_1}^2(kB,n)\right\}} \quad (21) \\ &= \frac{\left(\alpha_1\, \alpha_2\, \sigma_s^2(n)\right)^2}{[\alpha_1^2\, \sigma_s^2(n) + \sigma_{n1}^2(n)]\, [\alpha_2^2\, \sigma_s^2(n) + \sigma_{n2}^2(n)]}, \end{aligned}$$

with $S(kB,n)$ being the STFT of the speech signal $s(k)$, $N_1(kB,n)$, $N_2(kB,n)$ the STFTs of the uncorrelated noise signals $\sigma_s^2(n)$, $\sigma_{n1}^2(n)$, and $\sigma_{n2}^2(n)$ the corresponding power spectra.

The assumption that the noise signals are uncorrelated is more valid in higher frequency bands and for microphones located further apart. The correlation coefficients $C(n)$ may also influence the transfer function $G_{opt}(n)$ of the noise reduction filter.

6 LOSS CONTROL

Loss control is required to guarantee a prescribed echo suppression level (e.g. by the ITU–T). The total attenuation introduced by the loss control is distributed between the loudspeaker and microphone paths respectively in such that the communication is disturbed as less as possible.

In combination with the acoustic echo cancellation, loss control has to insert only the difference between the attenuation reached by the acoustic echo canceller and the level required by ITU–T. This requires that means have to be provided to estimate the performance of the echo canceller.

7 SYSTEM CONTROL

So far we have not discussed the importance of controlling the hands–free telephone systems. In a realistic scenario the adaptive filter does not achieve more than 30 dB ERLE (echo return loss enhancement) and may achieve less, if the processing power does not allow a large number of coefficients. A loss control is therefore required. Since the LEM–System is time–variant, the adaptation has to be performed whenever possible to track system changes. Situations, however, may occur (e.g. double talk, low SNR) where only small step sizes for the adaptation are permissible. An exact observation of the total system is therefore important for the overall performance [16].

8 IMPLEMENTATIONS

At this time low–cost, real–time processing means that the algorithms have to be implemented on fixed–point signal processors with only 16–24 bit word–length. The step from mathematical notation to implementation is therefore not trivial. Restrictions of computational cost are still limiting the performance of the system. However, this problem may be resolved in the future.

9 QUALITY TESTING

Although international test standards are desirable, no definition has yet been made. Besides transmission quality, in case of echo cancellation the system should be judged with regard to conversation ability. For noise reduction the naturalism of the outgoing speech signal is most important.

As an example of objective testing of echo cancellers the composite source signal should be mentioned [40]. It consists of three sections with different signal characteristics (tonal, random noise, silence) which enables both signal detection and adaptation. For testing the double talk ability composite source signals with increasing respectively decreasing envelops are superimposed [41]. However, it is not sufficient to judge echo cancelling and noise reduction merely on objective measures, since they cannot be brought into line with human perception. Subjective judgments by system users are therefore still necessary. These tests are based on the mean opinion score (MOS) which marks the signal on a scale of 1 to 5 (very bad to very good).

10 OUTLOOK

Voice communication systems with hands–free facilities are on the market. Double talk capability and noise reduction are provided at least to a certain extent. Further improvements, however, are necessary. These may result from a better understanding of the problem and more powerful, yet affordable, hardware. As far as the problem understanding is concerned, an improved joint control of the various algorithms comprising an echo and noise procedure are most promising. Therefore, even after more than one decade of intensive research and development, the challenge still remains.

References

[1] The new Bell telephone. *Scientific American 37 (1877)*, p. 1.

[2] E. Hänsler: The hands–free telephone problem – An annotated bibliography. *Signal Processing 27 (1992) 3*, pp. 259–271.

[3] E. Hänsler: The hands–free telephone problem – An annotated bibliography update. *Annales des Télécommunications 49 (1994) 7–8*, pp. 360–367.

[4] E. Hänsler: The hands–free telephone problem – A second annotated bibliography update. *Proc. of the 4th Internat. Workshop on Acoustic Echo and Noise Control (1995)*, Røros, Norway, pp. 107–114.

[5] A. Gilloire, P. Scalart et al.: Innovative speech processing for mobile terminals: An annotated bibliography. *COST 254 workshop*, Toulouse, France, July 1997.

[6] A. Gilloire: State of the art in acoustic echo cancellation. A.R. Figueiras and D. Docampo (Ed.): *Adaptive Algorithms: Applications and Non Classical Schemes.* Universidad de Vigo, 1991. pp. 20–31.

[7] A. Gilloire, E. Moulines, D. Slock and P. Duhamel: State of the art in acoustic echo cancellation. A.R. Figueiras–Vidal(Ed.): *Digital Signal Processing in Telecommunications.* Springer, London, U.K., 1996, p. 45–91.

[8] T. Pétillon, A. Gilloire and S. Théodoridis: A comparative study of efficient transversal algorithms for acoustic echo cancellation. *Proc. EUSIPCO–92 (1992)*, Brussels, Belgium, pp. 119–122.

[9] R. Martin: Algorithms for hands–free voice communication in noisy environments. *Proc. 9. Aachener Kolloquium "Signaltheorie" (1997)*, Institut für Technische Elektronik der RWTH, Aachen, pp. 233–238.

[10] International Telecommunication Union: General characteristics of international telephone connections and international telephone circuits – acoustic echo controllers. *ITU–T Recommendation G.167*, 1993.

[11] S. Gay and S. Tavathia: The fast affine projection algorithm. *Proc. ICASSP–95*, 1995.

[12] S. Haykin: *Adaptive Filter Theory.* 2nd ed., Prentice–Hall International, New Jersey, 1991.

[13] J. Cioffi and T. Kailath: Fast RLS Transversal Filter for adaptive filtering. *IEEE Trans. on ASSP*, Vol. 32, 1984, pp. 304–337.

[14] G.V. Moustakides and S. Theodoridis: Fast Newton Transversal Filters – A New Class of Adaptive Estimation Algorithms. *IEEE Trans. on Signal Processing*, Vol 39, No. 10, 1991, pp. 2184–193.

[15] B. Nitsch and S. Binde: Stabilizing the LFTF algorithm by Leakage control. *Proc. EUSIPCO–96*, 1996.

[16] C. Breining: Control of a Hands–free Telephone Set. *Signal Processing*, Elsevier, Holland Vol. 61, No. 2, 1997.

[17] U. Schultheiss: *Über die Adaption eines Kompensators für akustischer Echos.* Fortschrittsberichte VDI, Reihe 10, Nr. 90, Düsseldorf, 1988.

[18] P. Heitkämper: An adaption control for acoustic echo cancellers. *IEEE Signal Processing Letters*, 4(6), June 1997.

[19] G. Schmidt: Acoustic echo control in subbands – an application of multirate systems. *Proc. of the EUSIPCO–98*, 1998.

[20] R. Crochiere, L. Rabiner: *Multirate Digital Signal Processing.* Prentice–Hall, London, 1983

[21] G. Wackersreuter: On the Design of Filters for Ideal QMF and Polyphase Filter Banks. *AEÜ*, Vol.39, No. 2, 1985, pp. 123–130.

[22] B. Nitsch: Implementation of a block adaptive filter working in the frequency domain combined with a robust adaptation control. *Proc. EUSIPCO–98*, 1998.

[23] J. Benesty, D.R. Morgan, and M.M. Sondhi: A better understanding and an improved solution to the problems of stereophonic acoustic echo cancellation. *Proc. ICASSP–97*, 1997.

[24] Y. Loke and J. Chambers: Leakage factor: Its application in stereophonic acoustic echo cancellation. *Signal Processing*, 1998.

[25] G. Doblinger: Computationally efficient Speech enhancement by spectral Minima Tracking in Subbands. *Proc. EUROSPEECH–95*, Madrid, Spain, 1995.

[26] Y. Ephraim, D. Malah: Speech Enhancement Using a Minimum Mean–Square Error Short–Time Spectral Amplitude Estimator. *IEEE Trans. on ASSP*, Vol. ASSP–32, No.6, 1984, pp. 1109–1121.

[27] Y. Ephraim, D. Malah: Speech Enhancement Using a Minimum Mean–Square Error Log–Spectral Amplitude Estimator. *IEEE Trans. on ASSP*, Vol. ASSP–33, No.2, 1985, pp. 443–445.

[28] M. Kappelan, B. Strauß, P. Vary: Flexible Nonuniform Filter Banks Using Allpass Transformation of Multiple Order. *Proc. EUSIPCO–96*, Tieste, Italy, 1996.

[29] T. Gülzow, A. Engelsberg: Comparison of a Discrete Wavelet Transform and a Nonuniform Polyphase Filterbank Applied for Spectral Substraction. *Proc. IWAENC–97*, Imperial College London, U.K., 1997.

[30] P. Dreiseitel and H. Puder: Speech Enhancement for Mobile Telephony based on Nonuniformly Spaced Frequency Resolution. *Proc. EUSIPCO–98*, Rhodos, Greece, 1998.

[31] R. Martin: Spectral Subtraction Based on Minimum Statistics. *Proc. EUSIPCO–94*, 1994, pp. 1182–1185.

[32] R. Martin: An Efficient Algorithm to Estimate the Instaneous SNR of Speech Signals. *Proc. EUROSPEECH–93*, Berlin, 1993, pp. 1093–1096.

[33] E. Zwicker: *Psychoakustik.* Springer–Verlag, Berlin Heidelberg New York, 1982.

[34] H. Reininger, C. Kuhn: Signalverbesserung durch gehörgerechte spektrale Subtraktion. *Proc. 9. Aachener Kolloquium*, Aachen, 1997.

[35] S. Gustafsson, P. Jax: Combined Residual Echo and noise reduction: A novel Psychoacoustically Motivated Algorithm. *Proc. EUSIPCO–98*, Rhodos, Greece, 1998.

[36] R. Martin, P. Vary: Combined Acoustic Echo Control and Noise Reduction for Hands–Free Telephony – State of the Art and Perspectives. *Proc. EUSIPCO–96*, Trieste, Italy, 1996.

[37] S. Gustafsson, R. Martin: Combined Acoustic Echo Control and Noise Reduction Based on Residual Echo Estimation. *Proc. IWAENC–97*, Imperial College London, U.K., 1997.

[38] P. Dreiseitel, H. Puder: A combination of Noise Reduction and Improved Echo Cancellation. *Proc. IWAENC–97*, Imperial College London, U.K., 1997.

[39] D. Mahmoudi, A. Drygajlo: Wavelet Transform Based Coherence Fonction for Multi–Channel Speech Enhancement. *Proc. EUSIPCO–98*, Rhodos, Greece, 1998.

[40] H.W. Gierlich: Binaural analysis methods and their relationship to quality evaluation of hands–free telecommunication equipment. *Proc. EUSIPCO–96*, Trieste, Italy, 1996.

[41] H.W. Gierlich and F. Kettler: Speech Quality Evaluation of Hands–Free Telephones During Double Talk: New Evaluation Methodologies. *Proc. EUSIPCO–98*, Rhodos, Greece, 1998.