pp. 429-438

Combined acoustic echo cancellation, dereverberation and noise reduction : a two microphone approach

Rainer MARTIN *
Peter VARY *

Abstract

This paper presents new algorithms for acoustic echo cancellation and noise reduction which use two (or possibly more) microphone signals. In contrast to the single microphone method the multimicrophone approach can exploit the spatial coherence properties of sound fields which arise from noise and reverberated speech. Besides the standard FIR *echo canceller the proposed algorithms comprise an adaptive filter to eliminate non coherent signal components. The combined system achieves better Erle than the* FIR *echo canceller alone, attenuates ambient noise, dereverberates near end speech, and possibly leads to implementations with reduced complexity. The paper analyzes the acoustical properties of typical environments, presents the algorithms and experimental results.*

Key words : Telephone, Echo, Acoustic signal, Echo canceller, Acoustic noise, Ambient noise, Noise reduction, Spatial coherence, Diffuse reflection.

ANNULATION D'ÉCHO ACOUSTIQUE, DÉRÉVERBÉRATION ET RÉDUCTION DU BRUIT COMBINÉES : UNE APPROCHE AVEC DEUX MICROPHONES

Résumé

Cet article présente de nouveaux algorithmes pour l'annulation d'écho et la réduction des bruits qui utilisent deux signaux (ou plus) provenant de microphones. Au contraire des méthodes utilisant un seul microphone, les approches à plusieurs microphones peuvent tirer parti des propriétés de cohérence spatiale des champs sonores qui sont générées par des bruits ou de la parole réverbérée. En plus de l'annuleur d'écho standard FIR*, les algorithmes proposés se caractérisent par la présence d'un filtre adaptatif dédié à l'élimination des composantes non cohérentes des signaux. Le système combiné permet d'obtenir un meilleur facteur Erle que l'annuleur d'écho* FIR *seul. De plus, il atténue les bruits ambiants, déréverbère la parole proche et mène probablement à une réalisation dont la complexité serait réduite. L'article analyse les propriétés acoustiques d'environnements sonores typiques, présente les algorithmes et les résultats expérimentaux.*

Mots clés : Téléphone, Echo, Signal acoustique, Annuleur d'écho, Bruit acoustique, Bruit ambiant, Réduction bruit, Cohérence spatiale, Réflexion diffuse.

Contents

I. INTRODUCTION

A hands-free telephone interface provides more comfort and flexibility to the user of telecommunication equipment than its hand-held counterpart. With the advent and wide dissemination of mobile telephone systems and the recent advance in digital signal processing

* Institute for Communication Systems and Data Processing (IND), Aachen University of Technology, Templergraben 55, D–52056 Aachen, Germany.

technology the hands-free telephone with true *double talk* and noise reduction features becomes of increasing interest [1, 2].

Any realization of a hands-free telephone has to deal with two major problems. Firstly, due to the coupling between loudspeaker and microphone the loudspeaker signal is echoed back to the microphone making an effective echo cancellation device inevitable. Secondly, due to the relatively large distance from the speakers mouth to the microphone the microphone signal might be severely degraded by ambient noise. The relatively high noise levels found for example in mobile telecommunication applications call for a noise reduction device, not only to enhance speech quality, but also to improve the echo cancellers performance. It is the objective of this paper to investigate algorithmic structures and microphone-loudspeaker configurations which effectively combine the echo cancellation and noise reduction tasks.

The multimicrophone approach presented here allows us to exploit the spatial acoustic properties of noise and reverberation and to combine acoustic echo cancellation and noise reduction in a mutual synergy. Our approach is based on the observation that ambient noise as well as reverberated speech is spatially uncorrelated. The basic scenario for the application of our multimicrophone echo cancellation and noise reduction system can be described as follows. The near end speaker is typically situated in a noisy and reverberant room relatively close to the microphones such that the near end speech is highly correlated on all microphone channels. Without an echo cancellation device the far end speech would be echoed back to the far end terminal via the loudspeaker-room-microphone-array system (LRMA system). The cancellation of echoes is accomplished by means of an adaptive FIR filter which estimates the room impulse response. Our algorithms supplement the standard FIR echo canceller by an adaptive filter which reduces uncorrelated signal components, i.e. noise and reverberated components of speech. The echo cancellation task is now distributed onto two filters : the standard FIR canceller which takes care of direct sound and early reflections and the additional filter which attenuates the reverberation components of the microphone signals. As a consequence, the number of taps used to represent the impulse response of the LRMA system within the echo canceller might be reduced and the intelligibility of the near end speech is improved.

Only few authors have yet explored algorithms and systems that reduce acoustic echoes as well as noise. The idea to use several microphones to reduce reverberation can be traced back to the paper of Allen *et al.* [3]. They devised a frequency domain method to eliminate reverberation by computing an adaptive filter based on the cross power spectral density of two microphone signals. Zelinski successfully applied the dereverberation principle of Allen *et al.* to noise reduction. He proposed a time domain [4] and a frequency domain [5] algorithm which exploit the spatial coherence of speech and noise using four microphones arranged in a rectangular array. Very recently his approach was combined with acoustic echo cancellers [6]. Hands-free interfaces based on a larger number of microphones and the adaptive antenna algorithm due to Frost [7] were investigated by Xu and Grenier [8]. Finally, a two microphone echo cancellation system which uses a noise only reference microphone [9] was proposed in [10]. Our adaptive algorithms are based on the dereverberation principle of Allen *et al.* [3]. Since a large number of microphones is deemed impractical for mobile communications we focus on methods that use a linear array of two or three microphones.

The remainder of this paper is organized as follows : after a brief presentation of the echo cancellation problem we explore in Section III some of the acoustic issues which are specific for the multimicrophone approach. In Section IV we present our multimicrophone noise reduction and dereverberation approach and characterize it in terms of improvement in SNR. In Section V we merge the echo cancellation and noise reduction algorithms to introduce two distinct combined systems and an algorithm with reduced complexity. Finally, we analyze the performance of the combined systems and outline a route for further research. Throughout the paper we will emphasize the close interaction between the electro-acoustic interface and the performance of the multimicrophone algorithms.

II. THE ECHO CANCELLATION PROBLEM

The identification of the room impulse response by means of an adaptive FIR filter is theoretically the most appealing approach to acoustic echo cancellation. The cancellation approach does not degrade the near end speech and has the advantage to allow true *double talk*. However, the implementation of an FIR acoustic echo cancellation device is challenged by three major problems : computational complexity, speed of convergence, and robustness against noise. The computational complexity is governed by the length of the FIR filter which should ideally match a significant portion of the room impulse response. The number of filter taps necessary to achieve a sufficient amount of echo attenuation (see e.g. CCITT Recommendation P.30 [11]) is therefore proportional to the reverberation time T_{60}, which is defined as the time in which the sound energy, after shut off of the sound source, has decayed to a -60 dB level. At a sampling rate of 8 kHz and a reverberation time $T_{60} = 0.5$ s several thousand filter taps are necessary to achieve sufficient echo attenuation. When using the popular LMS algorithm (or one of its numerous variants) the speed of convergence is inversely proportional to the number of filter taps and must be reduced for low signal to noise ratios (SNR). Thus, the reverberation time and the SNR of the environment are at the heart of the problem. For adverse environments there are no easy solutions available. Besides a possible feedback of his echo the far end listener might be disturbed by the reverberation of near end speech, which is due to the relatively large distance from the speakers mouth to the microphone.

As it will be shown in Section V, the combination of state-of-the-art echo canceller with multimicrophone noise reduction systems does promise to ease these problems. To facilitate the discussion of the combined systems we will now briefly describe the echo cancellation algorithm which we employed for our experiments.

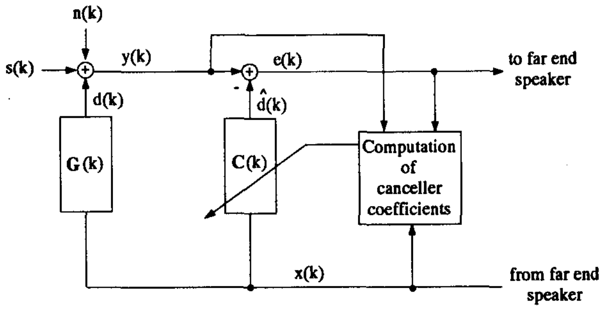

FIG. 1. — Discrete time model of electro-acoustic path and acoustic echo canceller.

Modèle en temps discret du chemin électroacoustique et de l'annuleur d'écho acoustique.

Figure 1 shows a discrete time model of acoustic echo cancellation. The N-tap FIR filter $C(k)$ models the impulse response $G(k)$ of the loudspeaker-room-microphone system (LRM system). The estimated reverberated far end signal $\widehat{d}(k)$ is subtracted from the microphone signal $y(k)$ to yield the improved near end signal $e(k)$. The near end speaker and the ambient noise are modelled by signals $s(k)$ and $n(k)$, respectively. The attenuation of the far end signal by the LRM system is characterized by the echo return loss (ERL) :

$$\text{(1)} \qquad \mathrm{ERL}(k) = 10 \log_{10} \left(\frac{E\{x^2(k)\}}{E\{d^2(k)\}} \right).$$

To compute the canceller coefficients we use an adaptation algorithm due to Antweiler [12] and Schultheiss [13] which is based on the LMS algorithm. The key features of this algorithm are linear predictors to whiten the signals used by the LMS and an adaptive step size controller originally proposed by Yamamoto [14]. The step size controller estimates the system distance $||G(k) - C(k)||$ and adjusts the step size according to the estimated system distance and the SNR.

In our experiments we used predictors of order 2 to prewhiten the far end speech $x(k)$ and the error signal $e(k)$. The predictors were updated every 128 samples. During an initialization phase the echo canceller is adapted using a small fixed step size. Experiments were carried out with coefficient vectors of length $N = 128$, 512, and 1536.

The performance of the echo canceller is usually measured by computing the echo return loss enhancement (Erle). The Erle is defined as :

$$\text{(2)} \qquad Erle(k) = 10 \log_{10} \left(\frac{E\{d^2(k)\}}{E\{(d(k) - \widehat{d}(k))^2\}} \right).$$

In real world experiments only the measurable quantity $Erle_m(k)$ is available. $Erle_m(k)$ computes the power ratio of the microphone signal before and after the cancellation :

$$\text{(3)} \qquad Erle_m(k) = 10 \log_{10} \left(\frac{E\{y^2(k)\}}{E\{e^2(k)\}} \right).$$

III. COHERENCE OF REVERBERATED SOUND FIELDS

We now turn to issues which strongly influence the design of multimicrophone echo cancellation and noise reduction algorithms. While the reverberation time T_{60} is one of the most important design parameters for acoustic echo cancellers, the coherence properties of the acoustic sound field are of central importance for the multimicrophone approach. It is the latter feature of sound fields which we will examine in some more detail in this section.

To demonstrate the impact of the acoustic interface on algorithm performance we present experimental results for two distinct acoustic environments and different loudspeaker-microphone configurations. The first environment is an office room with relatively large reverberation time and low noise levels. The second is a moving car characterized by a relatively short reverberation time but high noise levels. To investigate the effect of microphone directivity and the relative position of loudspeaker and microphones we examined two configurations, LRMA1 and LRMA2. Loudspeaker-microphone configuration LRMA1 uses omnidirectional microphones. The microphones are 0.4 m apart and the loudspeaker is placed between the microphones. Loudspeaker and microphones are facing the near end speaker. Loudspeaker-microphone configuration LRMA2 uses hypercardioid microphones also 0.4 m apart. The loudspeaker faces the ceiling and is placed 0.4 m behind the microphones, approximately in the minimum of the microphone directivity pattern. The amplification of the microphone signals was adjusted to yield the same near end speech power, regardless of microphone directivity. However, the two loudspeaker-microphone configurations have different attenuations of the far end echo signal, measured in terms of echo return loss (ERL), as well as slightly different reverberation times. In the car we used just one loudspeaker-microphone configuration with hypercardioid microphones. The car LRMA system will be denoted by LRMA3. We summarize the properties of all three configurations in Table I. The reverberation time T_{60} was measured by backwards integration of the squared impulse response.

III.1. The coherence function.

The (magnitude squared) coherence function is a frequency domain measure of correlation between two signals x and y and is defined as follows :

TABLE I. — Properties of loudspeaker-room-microphone-array configurations LRMA$_1$, LRMA$_2$, and LRMA$_3$.

Propriétés des configurations haut-parleur-salle-réseaux de microphones LRMA$_1$, LRMA$_2$ and LRMA$_3$.

LRMA system	Environment	Microphone directivity	Loudspeaker position	ERL (dB)	T_{60} (s)
LRMA$_1$	office	omnidirectional	between microphones	– 6.44	0.7
LRMA$_2$	office	hypercardioid	0.4 m behind microphones	9.8	0.65
LRMA$_3$	car	hypercardioid	on the dashboard	2.3	0.063

$$C_{xy}(\Omega) = \frac{|S_{xy}(\Omega)|^2}{S_{xx}(\Omega)S_{yy}(\Omega)}, \tag{4}$$

$\Omega = 2\pi f/f_s$ denotes the frequency normalized by the sampling rate f_s. $S_{xy}(\Omega)$, $S_{xx}(\Omega)$, and $S_{yy}(\Omega)$ are the cross spectral power densities and the auto spectral power densities of signals x and y, respectively. With the help of the Cauchy-Schwarz inequality it can be shown that :

$$0 \leq C_{xy}(\Omega) \leq 1. \tag{5}$$

The coherence function attains its maximum whenever frequency components of the two signals are completely correlated. The coherence function is zero for uncorrelated signals.

III.2. The coherence of the ideal diffuse sound field.

The coherence of two bandlimited and sampled signals recorded with omnidirectional microphones in an ideal diffuse sound field is given by [15] :

$$C_{id}(\Omega) = \frac{\sin^2(\Omega d_{mic} f_s/c)}{(\Omega d_{mic} f_s/c)^2}, \quad 0 \leq \Omega \leq \pi, \tag{6}$$

where d_{mic} denotes the microphone distance and c the speed of sound. $C_{id}(\Omega)$ attains its minimum values at frequencies :

$$\Omega_{min,l} = \frac{l\pi c}{d_{min} f_s}, \quad l = 1, 2, 3, \cdots \tag{7}$$

III.3. Coherence of real noise fields.

Uncorrelated noise signals are an essential requirement for our noise reduction system. In this section we show that real noise fields exhibit similar correlation properties as the ideal diffuse sound field in a reverberant enclosure. Figure 2 plots the ideal coherence (equation 6) and the coherence of noise measured in an office for different microphone distances. It is obvious that a sufficiently large microphone distance will guarantee almost uncorrelated noise signals in the frequency band used for speech transmission. We measured the coherence in different offices as well as cars and other noisy environments. It was found that the coherence is always closely approximated by equation 6, the microphone directivity having a larger impact on the coherence than the environment itself.

III.4. Coherence of reverberation.

Reverberation is a non stationary phenomenon. Thus, coherence evaluations of reverberated sound fields must be based on time intervals which are short compared with the reverberation time of the sound field. We estimated the short time coherence of reverberated sound fields using recursive FFT-based short time power spectrum estimates of the measured impulse responses of the LRMA-system. To achieve sufficient time resolution the frequency resolution of the FFT was restricted to 32 frequency bins. The spectral density estimates were obtained by smoothing of the FFT data blocks by means of a first order recursive network with a pole at $\alpha = 0.92$. Since the loudspeaker of a hands-free telephone is typically relatively close to the microphones, the sound

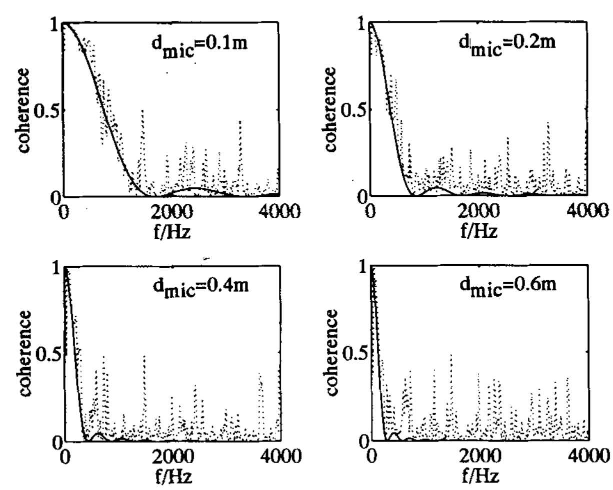

FIG. 2. — Coherence of the ideal diffuse sound field (solid) and coherence of noise recorded inside an office (dotted) with omnidirectional microphones ; microphone distances : d_{mic} = 0.1 ; 0.2 ; 0.4 ; 0.6 m.

Cohérence du champ sonore idéalement diffus (trait continu) et cohérence du bruit enregistré à l'intérieur d'un bureau (trait pointillé) à l'aide de microphones omnidirectionnels. Distances entre les deux microphones : d_{mic} = 0,1 ; 0,2 ; 0,4 ; 0,6 m.

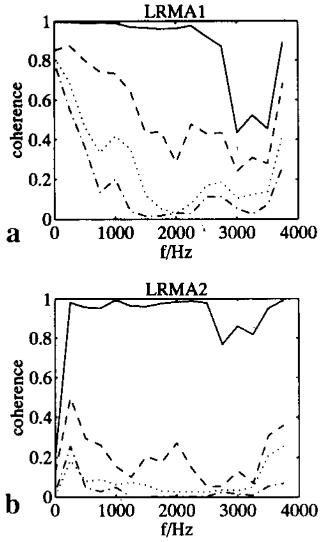

FIG. 3. — Time evolution of coherence of the impulse responses recorded with omnidirectional microphones (LRMA$_1$, a) and hypercardioid microphones (LRMA$_2$, b) ; solid line : 0.001 s, dashed line : 0.05 s, dotted line : 0.1 s, dash/dotted line : 0.15 s.

Evolution temporelle de la cohérence des réponses impulsionelles enregistrées à l'aide de microphones omnidirectionels. (LRMA$_1$, a) et de microphones hypercardiodes (LRMA$_2$, b) ; trait continu : 0,001 s, trait interrompu : 0,05 s, trait pointillé : 0,1 s, trait et points alternés : 0,15 s.

field produced by far end speech at the location of the microphones is not diffuse and the positioning of the loudspeaker and the microphones has a major impact on the coherence. Figure 3 plots the time evolution of the estimated short time coherence for two different microphone loudspeaker configurations, LRMA$_1$ and LRMA$_2$. The microphone distance is in both cases 0.4 m.

Since the directional microphones in configuration LRMA$_2$ pick up less direct and more reverberated far end sound than the omnidirectional microphones of configuration LRMA$_1$ the coherence of the microphone signals decays much faster for configuration LRMA$_2$ than for configuration LRMA$_1$. Thus, it might be suspected and it is actually shown in Section V that configuration LRMA$_2$ is principally better suited to support our multimicrophone noise reduction and dereverberation filter.

To conclude this section we note that the reverberation of the far end echo signal has similar spatial coherence properties as stationary ambient noise. Thus, noise and the reverberated far end signal components can be reduced by our multimicrophone system alike. A large microphone distance and a microphone loudspeaker configuration that decouples the loudspeaker and the microphones support this approach. However, the maximum microphone distance is usually limited by geometric restrictions of the environment and by the requirement that the signal of the near end speaker must be highly correlated on all microphone channels. In the next section we will present the noise and reverberation reduction algorithm and its dependence on the coherence of noise and speech.

IV. MULTIMICROPHONE NOISE REDUCTION

During the last years multimicrophone noise reduction techniques have attracted considerable interest in the speech enhancement community and there are by now a number of different methods available. In reverberant rooms the noise cancelling principle with a noise only reference microphone [9] is of limited applicability [15]. The multimicrophone methods of Allen [3] *et al.* and of Zelinski [4, 5], however, are more easily adapted to realistic situations. The latter methods estimate a Wiener filter (or some variant thereof) which attenuates incoherent signals, leaving the coherent signal components intact. As it was shown in the previous section the coherence of ambient noise is low if the microphone distance is sufficiently large. Thus, if the near end speaker is close enough to the microphones to produce highly correlated signals the adaptive filter will separate near end speech from noise and reverberated speech components. Since a large number of microphones and rectangular arrays are impractical for mobile communication purposes, we restricted our investigations to linear arrays with two or three microphones. Since a three microphone system is described in [16] we here present a two microphone system.

Figure 4 shows a block diagram of our two microphone time domain system. The microphone signals x

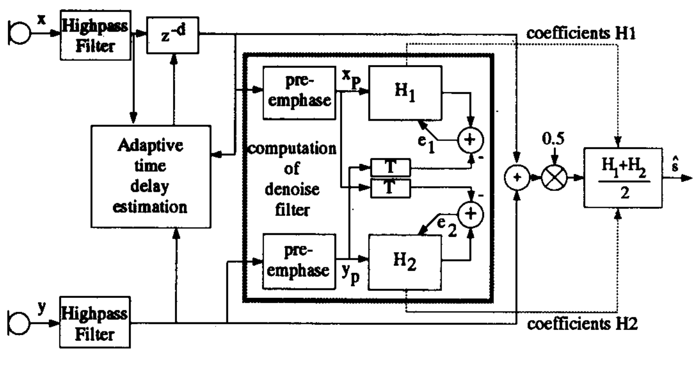

FIG. 4. — Two microphone LMS-adapted speech enhancement system.

Système d'amélioration des signaux de la parole avec deux microphones et adaptation par les moindres carrés.

and y are considered to be a sum of near end speech signals $s_{1/2}$ and noise signals $sn_{1/2}$, i.e. $x = s_1 + n_1$ and $y = s_2 + n_2$. The microphone signals are A/D converted and highpass filtered to cut off low frequency noise below $f_c = 240$ Hz. An adaptive time delay estimation algorithm tracks the position of the near end speaker and compensates the delay difference of the microphone signals which might arise from a non symmetrical position of the near end speaker. The time delay estimation is implemented by means of a cross correlation computation, smoothing of the correlation function, and a search for the maximum of the cross correlation function [17]. After the time delay compensation the two signals are added and filtered with an adaptive FIR filter. The coefficients of this filter are computed as the mean of coefficients of two linear phase FIR filters H_1 and H_2. The filters H_1 and H_2 are adapted using a linear phase version of the LMS algorithm [18]. The i-th coefficient $h_i(k)$ of the linear phase filter H_1 is adjusted according to equation (8) :

$$(8) \quad h_i(k+1) = h_i(k) + \mu_1(k)\, e_1(k)[x_p(k-i) + x_p(k-M+i+1)], \quad i = 0, \cdots, (M-1)/2,$$

and the symmetry condition $h_i(k) = h_{M-i-1}(k)$ where $M = 65$ denotes the number of filter taps. $x_p(x)$ is the input signal of the adaptive filter, $e_1(k)$ is the adaptation error, and $\mu_1(k)$ is a variable step size parameter. The three-tap FIR preemphasis filters decorrelate the input signals and were optimized in an off line experiment to produce best subjective speech quality. The complexity of this two microphone system (including time delay compensation) is roughly equivalent to the complexity of 250 echo canceller taps when using the algorithm described in Section II.

It was shown in [19] that the minimum mean square distortion of the estimated speech signal at the output of the speech enhancement system might be approximated in terms of coherence functions as :

$$(9) \quad E\{d_{min}^2\} = \frac{1}{2\pi}\int_{-\pi}^{\pi} S_{s_2 s_2}(\Omega) \left(1 - C_{s_1 s_2}(\Omega)\frac{\text{SNR}_x(\Omega)}{1+\text{SNR}_x(\Omega)}\right) d\Omega + \frac{1}{2\pi}\int_{-\pi}^{\pi} S_{n_2 n_2}(\Omega) C_{n_1 n_2}(\Omega)\frac{1}{1+\text{SNR}_x(\Omega)}\, d\Omega,$$

where $\text{SNR}_x(\Omega) = S_{s_1 s_1}(\Omega)/S_{n_1 n_1}(\Omega)$ denotes the signal to noise ratio of signal x. From equation (9) we find that a small distortion requires strongly correlated speech signals and high SNR input signals. Low coherence of noise signals is important, especially under low SNR conditions.

Similar to other noise reduction algorithms artificial sounding residual noise *(musical tones)* becomes a problem at low input SNR. The main source of musical tones is the residual correlation of noise at low frequencies (see Fig. 2). Musical tones can be avoided if the noise reduction system processes only frequency components above $\Omega_{min,2}$ (see equation (7)). This constraint requires either a large and hence not realizable microphone distance or a limitation of bandwidth by means of highpass filters. Frequencies above 240 Hz and below the cutoff frequency of such a highpass filter must be processed by some other method or just bypassed without further processing.

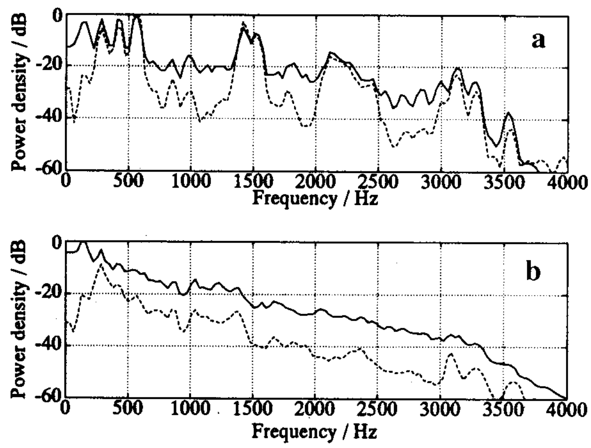

FIG. 5. — Power spectral densities at system input (solid) and output (dashed) during speech activity (a) and speech pause (b), microphone distance $d_{mic} = 0.4$ m, car noise.

Densités spectrales de puissance à l'entrée du système (trait continu) et à la sortie (trait interrompu) avec parole (a) et sans parole (b). Distance des microphones $d_{mic} = 0{,}4$ m, bruit de voiture.

Figure 5 demonstrates the performance of the two microphone system using the power spectral density of the microphone signals and of the system output. The upper graph was measured during speech activity and the lower graph during speech pause. During speech activity the formant frequencies of the speech signal are well reproduced. During speech pause we achieve up to 15 dB attenuation of noise.

To summarize we note that the two microphone noise reduction system separates coherent from non coherent signal components. As a result an FIR echo canceller placed at the output needs to cancel only the direct sound and early reflection part of the far end echo. It is this effect that we exploit in the combined system which we will present in the next section.

V. COMBINED SYSTEMS

The basic ingredients of our approach were introduced in Sections II and IV. We will now present the combined systems with the aim to develop a true symbiosis of echo cancellation and noise reduction. We will introduce two basic structures and a computationally attractive version and discuss their respective advantages and limitations. In what follows we will not consider the time delay compensation which is necessary whenever the near end speaker is not in a symmetric position with respect to the microphones.

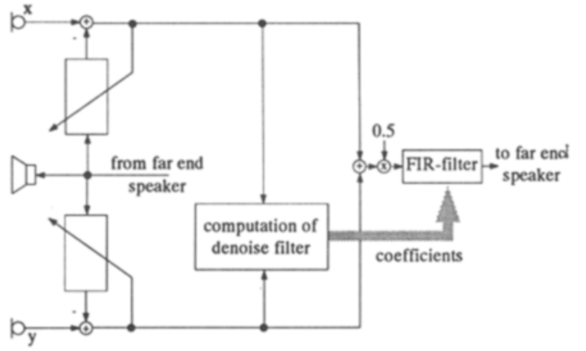

FIG. 6. — Block diagram of combined system cs_1.

Schéma bloc du système combiné cs_1.

V.1. Combined system cs_1 : echo cancellation before noise reduction.

Figure 6 depicts a block diagram of the first combined algorithm. Each microphone channel has its own echo canceller. The compensated signals are processed by a two channel noise reduction system as presented in Section IV. The advantage of this combination is that the coherent components of the far end speech are removed before the signals are passed to the noise reduction. The noise reduction has to deal only with the reverberated sound which is mainly not coherent. Thus, it can be expected that the noise cancelling filter will also reduce much of the reverberated far end speech. The obvious disadvantage of this design is that we need an echo canceller for each microphone channel. Hence this approach is impractical for more than two microphones. However, as the experimental results will show we might reduce the number of filter taps for each of the echo cancellers and achieve thus a complexity reduction.

V.2. Combined system cs_2 : noise reduction before echo cancellation.

This algorithm first processes the two microphone noise reduction before passing the combined denoised signal to a single echo canceller which adapts on the sum of the two microphone channels. Figure 7 shows a block diagram of combined system cs_2. The noise and reverberation reduction now introduces a time varying filter into the echo path. A copy of that filter is placed before the far end input of the echo canceller. Because of the time variant behavior we have to recompute the echo cancellers filter states each time the coefficients of the noise reduction filter change. This recomputation increases complexity and makes this version less attractive. Also, it is not easy to devise a time delay adjustment scheme for this algorithm since a potential time delay between the microphone channels must be compensated before the summation of the two microphone signals. Any time shift of one microphone signal with respect to the other will change the impulse response of the overall system making it more difficult for the echo canceller to track the LRMA system response.

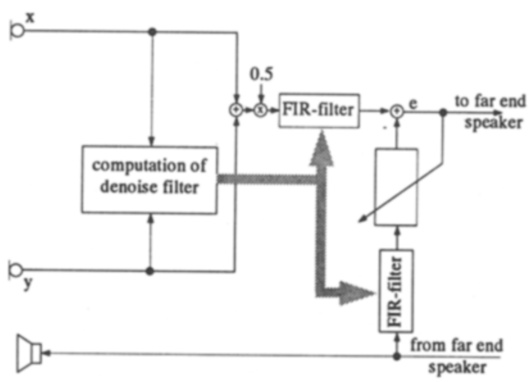

FIG. 7. — Block diagram of combined system cs_2.

Schéma bloc du système combiné cs_2.

V.3. Combined system cs_3 : reduced complexity.

The combined system shown in Figure 8 combines to a certain extent the advantages of combined system cs_1 and cs_2. We here use just one echo canceller to cancel the mean of the microphone signals. Due to the summation of two microphone signals with mostly uncorrelated noise the SNR at the echo cancellers input is increased by a maximum of 3 dB. The cancellation signal is also used to cancel the individual microphone signals. The cancellation of the individual microphone channels will be not perfect since the room impulse response of the two microphone channels will be somewhat different. However, since the direct sound and early reflection parts of the room impulse responses of both microphone channels are similar and the (imperfectly) compensated signals are used only for the computation of the noise reduction filter, the loss in Erle is tolerable.

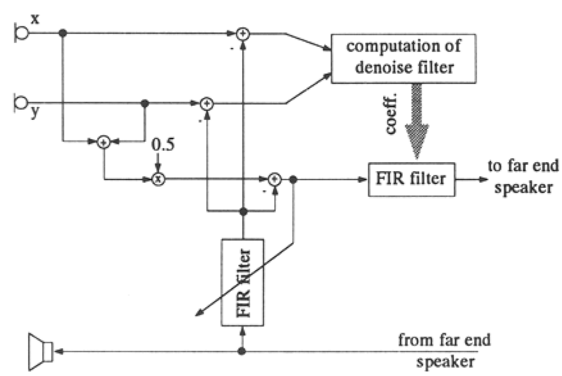

FIG. 8. — Block diagram of combined system cs_3.

Schéma bloc du système combiné cs_3.

V.4. Experimental results.

Reverberated speech samples were recorded in an office room ($LRMA_1$, $LRMA_2$) and in a car ($LRMA_3$) using different microphones and loudspeaker positions (see Table I). Our speech data base consists of 16 phonetically balanced speech samples, produced by two female and two male speakers. Each speech sample is four seconds long and contains two short sentences with a small pause in between. All Erle plots and figures were computed using averages of these 16 speech samples. Car noise was recorded separately from the speech signals and was added at various SNR values for the simulations. All signals were highpass filtered with a cutoff frequency of $f_c = 240$ Hz. Whenever noise was added, the Erle was computed using the original speech signals only.

Before we discuss detailed results we will summarize our general results :

1) Informal listening test suggest that the optimal microphone distance for noise reduction as well as reverberation reduction is about 40 cm. Larger microphone distances will lead to degraded near end speech. Smaller microphone distances will limit the maximum noise and reverberation reduction and will induce *musical tones* at low frequencies.

2) The lower the coupling of the loudspeaker and microphones the better the performance of the overall system. Therefore directional microphones and the loudspeaker placed in the minimum of the microphone directivity gives the best results in office and car.

3) Combined system CS_1 performes best, followed by CS_3, and CS_2.

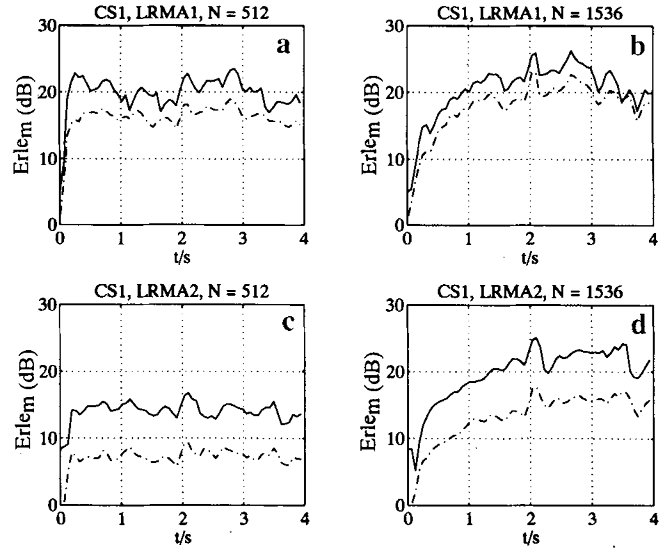

FIG. 9. — Office environment : Erle_m vs. time of echo canceller (dashed) and combined system (solid) for 512 canceller taps (a) and c)) and 1536 canceller taps (b) and d)). a) and b) : Omnidirectional microphones (LRMA_1) ; c) and d) : hypercardioid microphones (LRMA_2).

Environnement de bureau : Erle_m en fonction du temps de l'annuleur d'écho (trait interrompu) et du système combiné (trait continu) avec 512 coefficients (a) et c)) et 1536 coefficients (b) et d)). a) et b) microphones omnidirectionels (LRMA_1), c) et d) microphones hypercardiodes (LRMA_2).

V.4.1. Office environment : large reverberation time, low noise levels (LRMA_1, LRMA_2).

Figure 9 shows single talk Erle_m vs. time plots for the office LRMA system (LRMA_1 and LRMA_2, low noise level, no additional noise). These plots were computed using system CS_1. The SNR for the received far end speech depends on the microphone directivity and the loudspeaker microphone coupling (see Table I for ERL values). The SNR is approximately 35 dB for configuration LRMA_1 and about 15 dB for configuration LRMA_2.

Due to the relatively large reverberation time we need a fairly large number of canceller coefficients to achieve sufficient attenuation of the echo signal. Since the SNR is much better for configuration LRMA_1 than for LRMA_2 the echo canceller converges much better in configuration LRMA_1. However, in configuration LRMA_2 the microphone signals contain relatively more reverberated signal. Therefore the dereverberation filter works much better for LRMA_2 than for LRMA_1. In configuration LRMA_2 the noise reduction filter contributes an additional Erle_m of 7 dB. Since also the ERL of LRMA_2 is significantly larger than the ERL of LRMA_1 (see Table I) the overall best performance was achieved using LRMA_2.

The same dependencies can be observed for the combined systems CS_2 and CS_3. The mean Erle values are summarized in Table II. The mean was computed using a time period from 1 s to 4 s, i.e. the time after initial convergence. For two reasons the absolute Erle values are much smaller for system CS_2 than for the other combined systems. Firstly, the input signals of the dereverberation filter contain relatively large amount of far end direct sound which is coherent and cannot be removed. Secondly, the adaptation of the echo canceller

TABLE II. — Mean Erle_m of systems CS_1, CS_2, and CS_3 in office environment (LRMA_1 and LRMA_2).

Erle_m moyen des systèmes CS_1, CS_2 et CS_3 en environnement de bureau (LRMA_1 et LRMA_2).

	LRMA_1 $N = 512$	LRMA_1 $N = 1536$	LRMA_2 $N = 512$	LRMA_2 $N = 1536$
CS_1, Erle of echo canceller only (dB)	16.5	19.6	7.4	14.9
CS_1, Erle of combined system (dB)	19.9	22.2	14.2	21.7
CS_2, Erle of echo canceller only (dB)	13.8	14.5	5.6	8.6
CS_2, Erle of combined system (dB)	15.2	15.8	7.2	10.1
CS_3, Erle of echo canceller only (dB)	15.6	19.4	8.0	14.5
CS_3, Erle of combined system (dB)	19.8	22.9	14.1	18.8

is disturbed by the time varying dereverberation filter in the echo path. Hence, increasing the number of canceller taps N does not yield as much additional Erle for system CS_2 as for the other systems. Combined system CS_3 performes almost as well as combined system CS_1.

V.4.2. Car environment : small reverberation time, high noise levels (LRMA3).

High noise levels are characteristic for mobile communications environments. In the car we also have a relatively short reverberation time such that an echo canceller with 256-512 taps is sufficient. Figure 10 plots the time evolution of Erle_m for the speech samples without additional noise (car at standstill, engine off). Erle_m plots for combined systems CS_1 show that the dereverberation filter improves system performance by approximately 5 dB, independent of the number of canceller taps. For $N = 512$ a sufficiently large attenuation of the echo signal is achieved. For system CS_3 the gain of the dereverberation filter is less pronounced (2-3 dB).

The upper and the lower graph in Figure 11 plot the echo canceller performance and filter gain versus SNR for systems CS_1 and CS_3, respectively. The echo canceller has 512 taps, car noise was added to the speech signals. The Erle was time averaged over a time period from 1 s to 4 s. In the presence of noise the echo canceller converges rather slowly. In consequence, the Erle is significantly decreased in the time period which was considered here.

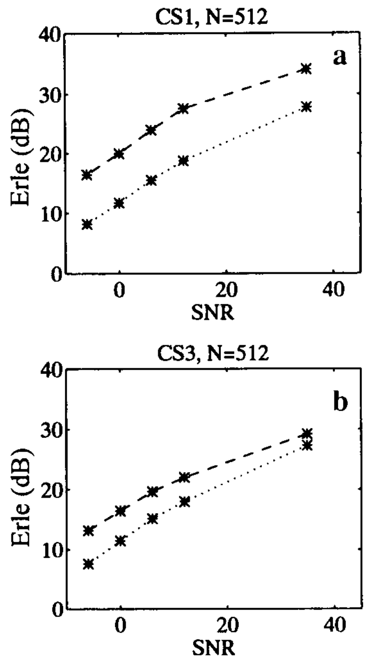

FIG. 11. — Car environment (LRMA3) : mean Erle vs. SNR for CS_1 (a) and CS_3 (b) ; dashed line : combined system, dotted line : echo canceller only.

Environnement de voiture (LRMA3) : Erle moyen en fonction du rapport signal/bruit pour le CS_1 (a) et le CS_3 (b) ; trait interrompu : système combiné, trait pointillé : annuleur d'écho seul.

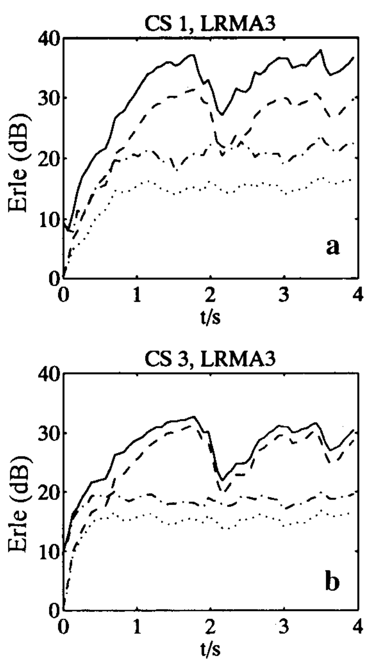

FIG. 10. — Car environment (LRMA3, engine off) : Erle_m vs. time for system CS_1 (a) and CS_3 (b). Solid line : combined system and $N = 512$, dashed line : echo canceller only and $N = 512$, dash/dotted line : combined system and $N = 128$, dotted line : echo canceller only and $N = 128$.

Environnement de voiture (LRMA3, moteur arrêté) : Erle_m en fonction du temps pour les systèmes CS_1 (a) et CS_3 (b). Trait continu : systèmes combinés avec $N = 512$. Trait interrompu : annuleur d'écho seul avec $N = 512$. Traits et points alternés : système combiné avec $N = 128$. Trait pointillé : annuleur d'écho avec $N = 128$.

The gain of the noise reduction and dereverberation filter is almost 10 dB. It should be noticed however, that noise and far end echo are attenuated alike. Thus, the relative level of noise with respect to the far end echo is maintained and the improvement in Erle is subjectively not immediately obvious to the far end listener.

As in the previous experiments the gain of combined system CS_3 is not as large as the gain of CS_1. Nevertheless, the prospect of reduced complexity makes it worthwhile to further pursue this approach. System CS_2 performed significantly worse than CS_3 and is not documented here.

VI. SUMMARY AND CONCLUSIONS

In this paper we introduced a class of algorithms which combines a multimicrophone noise and reverberation reduction system with FIR echo cancellers. The noise reduction system significantly increases the Erle and dereverberates the near end speaker, thus improving intelligibility at the far end terminal. We showed that there is a close interaction between the acoustics of the reverberated sound field, the design of the acoustic interface, and the proposed multimicrophone algorithms. Two distinct experimental environments were considered : office room (large reverberation time, low noise levels) and moving car (short reverberation time, high noise levels). In both cases best results were achieved with a highly decoupled loudspeaker-microphone configuration and an algorithm that provides individual echo cancellers to all microphone channels (CS_1). A computationally less expensive solution was developed (CS_3) without sacrificing too much of the performance. The incorporation of time delay compensation algorithms into the systems is currently investigated. We furthermore work on improvements speech quality at low SNR.

ACKNOWLEDGMENT

The authors thank the reviewer for suggestions to improve this paper.

Manuscrit reçu le 25 mai 1994.

REFERENCES

[1] HÄNSLER (E.). The hands-free telephone problem. An annotated bibliography. *Signal Processing* (1992), **27**, pp. 259-271.
[2] NAYLOR (P.). Enhancement of hands-free telecommunications. *Proc. Third Int. Workshop on Acoustic Echo Control*, Lannion (Sep. 7-8, 1993).
[3] ALLEN (J.), BERKLEY (D.), BLAUERT (J.). Multimicrophone signal-processing technique to remove room reverberation from speech signals. *J. Acoust. Soc. Am.* (Oct. 1977), **62**, n° 4.
[4] ZELINSKI (R.). Noise reduction based on microphone array with LMS adaptive post-filtering. *Electronics Letters* (Nov. 1990), **26**, n° 24, pp. 2036-2037.
[5] ZELINSKI (R.). A microphone array with adaptive post-filtering for noise reduction in reverberant rooms. *Proc. ICASSP*, **88**, pp. 2578-2581.
[6] HUHN (T.), JENTSCHEL (H.-J.). Kombination von Geräuschreduktion und Echokompensation beim Freisprechen. *Nachrichtentechnik, Elektronik* (1993), **43**, n° 6, pp. 274-280.
[7] FROST (O.). An algorithm for linearly constrained adaptive array processing. *Proc. IEEE* (Aug. 1972), **60**, n° 8, pp. 926-935.
[8] XU (M.), GRENIER (Y.). Acoustic echo cancellation by adaptive antenna. *1st Workshop on Acoustic Echo Control*, Berlin (1989).
[9] WIDROW (B.), GLOVER (J.), MCCOOL (J.), KAUNITZ (J.), WILLIAMS (C.), HEARN (R.), ZEIDLER (J.), DONG (E.), GOODLIN (R.). Adaptive noise cancelling : principles and applications. *Proc. IEEE* (Dec. 1975), **63**, n° 12, pp. 1692-1716.
[10] YASUKAWA (H.). Acoustic echo canceller with subband noise cancelling. *Electronics Letters* (July 1992), **28**, n° 15, pp. 1403-1404.
[11] ***. Transmission performance of group audio terminals (GATs). *CCITT Recommendation P.30*, Melbourne (1988).
[12] ACKER (ANTWEILER) (C.), VARY (P.). Combined implementation of predictive speech coding and acoustic echo cancellation. *Signal Processing*, VI : Theories and applications, VANDEWALLE (J.) *et al.* (eds.), pp. 1641-1644, *Eurasip, Elsevier* (1992).
[13] SCHULTHEISS (U.). Über die Adaption eines Kompensators für akustische Echos. *VDI Verlag* (1988), **10**, n° 90.
[14] YAMAMOTO (S.), KITAYAMA (S.). An adaptive echo canceller with variable step gain methods. *Trans. IECE*, Jap (1982), **65**, pp. 1-8.
[15] ARMBRÜSTER (W.), CZARNACH (R.), VARY (P.). Adaptive noise cancellation with reference input. Possible applications and theoretical limits. *Signal Processing III* : theories and applications. YOUNG (I. T.), *et al.* (editors), *Elsevier Science Publishers B.V.*, North Holland (1986), pp. 391-394.
[16] JANSE (K.), KELLERMANN (W.), KNIBBELER (C.), MARTIN (R.), VARY (P.). A noise reduction system for automobiles. *Proc. Third Int. Workshop on Acoustic Echo Control*, Lannion (Sep. 7-8, 1993).
[17] CARTER (G.). Coherence and time delay estimation. *Proc. IEEE* (Feb. 1987), **75**, n° 2, pp. 236-255.
[18] COWAN (C.), GRANT (P.). Adaptive filters. *Prentice Hall* (1985).
[19] MARTIN (R.), VARY (P.). A symmetric two microphone speech enhancement system. Theoretical limits and application in a car environnement. *Processings of the fifth IEEE Signal Processing Workshop, Starved Rock State Park,* Illinois (Sep. 13, 1992).

BIOGRAPHIES

Rainer MARTIN was born in 1962. He received the Dipl.-Ing. degree in Electrical Engineering from Aachen University of Technology in 1988 and the MSEE degree from Georgia Institute of Technology in 1989. Since 1990 he is a research employee at Aachen University of Technology, Germany.

Peter VARY was born in 1948. He received his Ph.D. degree from the University of Erlangen-Nürnberg in 1978. From 1980 to 1988 he was head of the advanced development group *Digital Signal Processing* at Philips Kommunikations Industrie, Nürnberg. Since 1988 he is a full professor at Aachen University of Technology.