LOW-DELAY SUBBAND ECHO CONTROL IN AN AUTOMOTIVE ENVIRONMENT

Kai Steinert[1], Martin Schönle[1], Christophe Beaugeant[2], Tim Fingscheidt[3]

[1]Siemens AG, Corporate Technology, Munich, Germany
[2]Nokia Siemens Networks, Munich, Germany
[3]Institute for Communications Technology, Braunschweig Technical University, Germany
E-Mail: {kai.steinert.ext,martin.schoenle}@siemens.com, christophe.beaugeant@nsn.com, t.fingscheidt@tu-bs.de

ABSTRACT

Acoustic echoes are usually attenuated by an adaptive filter with a subsequent residual echo suppression postfilter. Subband adaptive filtering approaches may lead to a reduced computational complexity and an increased filter convergence speed relative to time-domain algorithms. However, a disadvantage is the processing delay caused by the analysis and synthesis filtering. To achieve a considerable reduction of the signal delay, we employ a delayless subband adaptive filter where the echo path model coefficients are calculated in the subband domain, transformed into a fullband impulse response, and the actual filtering is performed in the time domain. Likewise the postfilter weights are calculated in the time-frequency domain, but applied via an FIR filter of low group delay in the time domain. The adaptive filter and postfilter are controlled by a novel cross correlation approach that does not require a separate VAD.

1. INTRODUCTION

In hands-free systems applied in challenging acoustic environments like cars, the acoustic echo is typically reduced with an adaptive FIR filter and an additional postfilter [1]. The adaptive filter estimates the linear part of the echo path up to a certain order. Due to several influences such as nonlinearities, a possibly long and highly time-varying echo path, or local speech and noise activity, a residual echo component remains in the output signal of the adaptive filter. Therefore spectral weighting is applied by a postfilter for a further residual echo reduction.

An adaptive filter can easily be implemented with a time-domain NLMS algorithm. However, in case of acoustic echo cancellation in cars with model filter lengths of up to 40-50 ms, the algorithm may become computationally complex and lead to slow convergence. These drawbacks may be circumvented by a subband adaptive filter [2] which also offers a new degree of freedom: parameters such as the filter lengths can be chosen subband-dependent according to the frequency characteristics of the echo path and the signals involved. Yet subband implementations are at the expense of a large signal delay caused by the analysis and synthesis filter bank. However, many conversational systems have tight constraints with respect to delay, e.g. 30 ms in the uplink path according to [3]. Delayless subband adaptive filtering [4] is possible if the adaptation takes place in the time-frequency domain while the actual filtering is performed in the time domain with a broadband impulse response obtained by transforming the subband responses. Furthermore aliasing in the signal path is prevented.

A postfilter for spectral weighting can conveniently be included into the delayless structure if, similar to the adaptive filter, the weights are calculated in the subband domain, but applied in the time domain with a low-delay FIR filter [5].

For the control of both the adaptive filter and the postfilter we propose a cross correlation approach that does not require a separate double-talk detector. Besides it implicitly takes into account echo path variations without the requirement of an echo path change detector as proposed in [1] or an additional term modeling these impulse response variations according to [6]. Results are presented for a simulation with a time-varying car impulse response and car background noise.

This paper does not consider noise reduction. However, an appropriate weighting rule may easily be combined with the residual echo suppression postfilter we proposed.

2. SYSTEM OVERVIEW

The echo cancellation system under discussion is depicted in Fig. 1. The far-end speaker signal $x(n)$ and the microphone signal $y(n)$ —composed of the echo part $d(n)$ and the local speech and noise signal $v(n)$ —are transformed into the subband signals $X_k(m)$ and $Y_k(m)$, respectively, where k indicates the subband and m the subband time index. In the subband domain an adaptive filter impulse response $\hat{\mathbf{H}}_k(m) = [\hat{H}_{k,0}(m), ..., \hat{H}_{k,N_k-1}(m)]^{\mathrm{T}}$ of length N_k in subband k is calculated such as to model the subband responses corresponding to the time-domain loudspeaker-enclosure-microphone (LEM) system $\mathbf{h}(n) = [h_0(n), ..., h_{N-1}(n)]^{\mathrm{T}}$ assumed to be of length N. A time-domain echo path estimate $\hat{\mathbf{h}}(n) = [\hat{h}_0(n), ..., \hat{h}_{N-1}(n)]^{\mathrm{T}}$ is then calculated from all $\hat{\mathbf{H}}_k(m)$ via a weight transform [7]. For a further attenuation of the remaining echo $e_u(n)$, residual echo

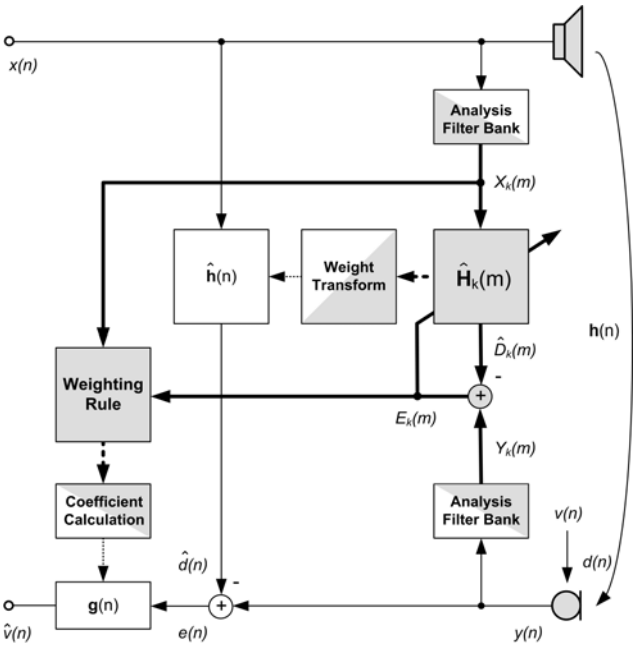

Fig. 1 Delayless echo cancellation combined with low-delay postfiltering. Thin lines stand for time-domain, thick lines for subband-domain signal paths, and dashed lines indicate a coefficient or weight transfer. The shading in the boxes indicates an operation in the subband domain.

suppression postfilter weights $G_k(m)$ are calculated in the subbands and transformed to a low-delay time-domain moving-average filter $\mathbf{g}(n)$ [5]. Thus in the time domain the echo $d(n)$ is cancelled without any additional delay. Postfiltering is performed with a filter of lower delay compared to an analysis-synthesis subband system. The output signal is then calculated as

$$\hat{v}(n) = \left(y(n) - x(n) * \hat{\mathbf{h}}(n) \right) * \mathbf{g}(n) \,. \tag{1}$$

2.1. Subband decomposition

The subband decomposition is accomplished with a uniform DFT-modulated polyphase filter bank [1, chap. 9]. We chose 32 channels at a subsampling rate of 16 and a prototype analysis filter of length 128 designed according to [1, app. B].

2.2. Delayless subband adaptive filter

We have implemented a subband NLMS adaptive filter with the step size control according to [1, chap. 13] in the open-loop scheme [4] described by

$$E_k(m) = Y_k(m) - \hat{\mathbf{H}}_k^{\mathrm{H}}(m) \cdot \mathbf{X}_k(m) \tag{2}$$

$$\mu_k(m) = \frac{\mathrm{E}\{|E_{u,k}(m)|^2\}}{\mathrm{E}\{|E_k(m)|^2\}} \tag{3}$$

$$\hat{H}_k(m+1) = \hat{H}_k(m) + \mu_k(m) \frac{\mathbf{X}_k(m) E_k^*(m)}{\|\mathbf{X}_k(m)\|^2} \tag{4}$$

with the subband excitation signal vector $\mathbf{X}_k(m) = [X_k(m), X_k(m-1), ..., X_k(m-N_k+1)]^{\mathrm{T}}$, the step size $\mu_k(m)$, and the undisturbed error $E_{u,k}(m)$. The expression $\mathrm{E}\{\cdot\}$ stands for the expectation value, $|\cdot|$ is the magnitude, $\|\cdot\|$ the Euclidean vector norm, $(\cdot)^*$ the complex conjugate, and $(\cdot)^{\mathrm{H}}$ the Hermitian vector. The undisturbed error is the adaptive filter output signal without the local signal part $V_k(m)$, i.e. $E_{u,k}(m) = E_k(m) - V_k(m)$. The estimation of its power will be discussed below. To save memory, the subband-dependent adaptive filter length N_k can be chosen shorter for higher-frequency subbands where the impulse response of real rooms decays faster.

The analysis filter bank group delay results in non-causal taps in the optimum subband impulse responses. An additional artificial delay of one subband sample is introduced in the microphone signal path to partly model these. As has already been mentioned, the fullband impulse response is calculated by transforming the subband coefficients into the time domain with a weight transform [7].

It has to be noted that the adaptive filter convergence and tracking is delayed relative to fullband adaptation. This is because, contrary to the filtering operation, the adaptation takes place in the subsampled subbands delayed by the analysis filter bank.

2.3. Low-delay postfilter

An adaptive FIR echo canceller in a real system still leaves residual echo components in the error signal. Therefore a residual echo suppression Wiener postfilter is used [8]. Its spectral weights $G_k(m)$ are calculated according to

$$G_k(m) = \frac{\xi_k(m)}{1 + \xi_k(m)} \tag{5}$$

$$\text{with } \xi_k(m) = \frac{\mathrm{E}\{|V_k(m)|^2\}}{\mathrm{E}\{|E_{u,k}(m)|^2\}} \tag{6}$$

the a priori signal-to-residual-echo ratio. The estimate of $\xi_k(m)$, $\hat{\xi}_k(m)$, is calculated with a decision-directed approach [8] as follows

$$\hat{\xi}_k(m) = \alpha \cdot \frac{|\hat{V}_k(m-1)|^2}{\mathrm{E}\{|E_{u,k}(m-1)|^2\}} + (1-\alpha) \cdot \max\{\gamma_k(m) - 1, 0\} \tag{7}$$

where $\gamma_k(m)$ is the a posteriori signal-to-residual-echo ratio defined as

$$\gamma_k(m) = \frac{|E_k(m)|^2}{\mathrm{E}\{|E_{u,k}(m)|^2\}} \tag{8}$$

and $0 < \alpha < 1$. The instantaneous local signal power of the last iteration is calculated according to $|\hat{V}_k(m-1)|^2 = G_k^2(m) \cdot |E_k(m-1)|^2$. The estimation of the residual echo power $\mathrm{E}\{|E_{u,k}(m)|^2\}$ will be discussed in the following subsection.

The weights $G_k(m)$ are transformed into a low-delay time-domain postfilter $\mathbf{g}(n)$ [5]. This broadband postfilter approximates an analysis-synthesis subband system with spectral weighting in the subbands. In order to avoid synthesis filtering and the additional delay involved, no subsampling takes place and the prototype analysis filter is realized by an M-th band filter of length 128 designed with the window method. The synthesis stage consists of a mere addition of all subbands. In our case $\mathbf{g}(n)$ was an FIR filter truncated to a length of 32 samples.

Similar to the case of the delayless adaptive filter, the calculation and the application of the postfilter coefficients does not take place temporally synchronized anymore. However, for our system we were not able to perceive a degraded signal quality.

2.4. Control of adaptive filter and postfilter

Both the adaptive filter and the residual echo postfilter need an estimate of the undisturbed echo power $\mathrm{E}\{|E_{u,k}(m)|^2\}$. It can approximately be written as [1]

$$\mathrm{E}\{|E_{u,k}(m)|^2\} \approx \mathrm{E}\{|X_k(m)|^2\} \cdot \beta_k(m) \qquad (9)$$

where $\beta_k(m) = \mathbf{E}\{\|\mathbf{H}_k(m) - \hat{\mathbf{H}}_k(m)\|^2\}$ stands for the system distance. The true echo path impulse response $\mathbf{H}_k(m)$ is defined similarly to $\hat{\mathbf{H}}_k(m)$. We calculate $\beta_k(m)$ with the cross spectral method [9]. In this context, however, we aim at estimating the squared system distance (and not an accurate echo path). Therefore we define the vectors

$$\breve{\mathbf{X}}_k(m) = \mathrm{FFT}\{[X_k(m - N_{Cr} + 1), ..., X_k(m)]^{\mathrm{T}}\} \qquad (10)$$

$$\breve{\mathbf{E}}_k(m) = \mathrm{FFT}\{[E_k(m - N_{Cr} + 1), ..., E_k(m)]^{\mathrm{T}}\} \qquad (11)$$

of FFT-transformed subband sample vectors of length N_{Cr}. For improved results Hann windowing can be applied before taking the FFT. We assume that the excitation signal is not correlated with the local speech and noise signal. The correlation estimates are given by

$$\breve{\mathbf{\Gamma}}_{XX,k}(m) = \mathrm{E}\{\breve{\mathbf{X}}_k^*(m) \otimes \breve{\mathbf{X}}_k(m)\} \qquad (12)$$

$$\breve{\mathbf{\Gamma}}_{XE,k}(m) = \mathrm{E}\{\breve{\mathbf{X}}_k^*(m) \otimes \breve{\mathbf{E}}_k(m)\} \qquad (13)$$

so that, similar to [9], the squared expected system distance can be formulated as

$$\beta_k(m) = \| \breve{\mathbf{\Gamma}}_{XE,k}(m) \div \breve{\mathbf{\Gamma}}_{XX,k}(m) \|^2 \qquad (14)$$

where $\otimes$ and $\div$ stand for elementwise multiplication and division, respectively. The correlation measures have to be smoothed in order to keep the influence of uncorrelated signals low. However, too strong smoothing may impair the result for fast time variations in the system distance and/or the echo path. We chose a smoothing constant of 0.99 (first order IIR smoothing) and a correlation length of $N_{Cr} = 8$ and performed this measurement every 8 subband samples.

An adaptive filter mismatch due to an LEM system variation is implicitly considered by taking the cross correlation (13). Thus an additional echo path change detector [1] or an additional term modeling the impulse response variations [6] is not needed.

3. SIMULATION RESULTS

To assess the performance of the algorithm presented, a simulated car environment was used. The far-end and near-end signals were male and female speech, respectively, of approximately the same signal level. The echo signal was created by convolving the far-end signal with an impulse response of length 400 varying over time. This echo path variation was accomplished by linearly interpolating between different impulse responses measured in a real car environment such that after each 400 ms a new impulse response was reached. Weak car noise was added to the local speech signal with an SNR of 20 dB. A sampling rate of 8 kHz was chosen.

The results are given in terms of the disturbed echo return loss enhancement $\mathrm{ERLE}_d(n)$ of the adaptive filter with postfilter defined as

$$\mathrm{ERLE}_d(n) = 10 \log_{10} \frac{\mathrm{E}\{y^2(n)\}}{\mathrm{E}\{\hat{v}^2(n)\}}. \qquad (15)$$

For comparison, results are also given for a state-of-the-art hands-free system [10] consisting of a time-domain NLMS of length 400 with residual echo suppression postfiltering in the FFT bins. The adaptive filter was controlled with a double-talk detector.

4. DISCUSSION AND OUTLOOK

The waveform of the input and the enhanced output signal is given in Fig. 2 (first and second plot). The third plot shows the $\mathrm{ERLE}_d(n)$ in dB for the algorithm described and the reference hands-free system.

The adaptive filter by itself performed slightly worse than the reference system adaptive filter. A possible reason could be that, due to the adaptation of the time-varying echo path delayed relative to the time-domain filtering, the delayless filter will for this case always produce a certain residual error. However, for an activated residual echo suppression postfilter, the presented system outperforms the state-of-the-art system, especially during

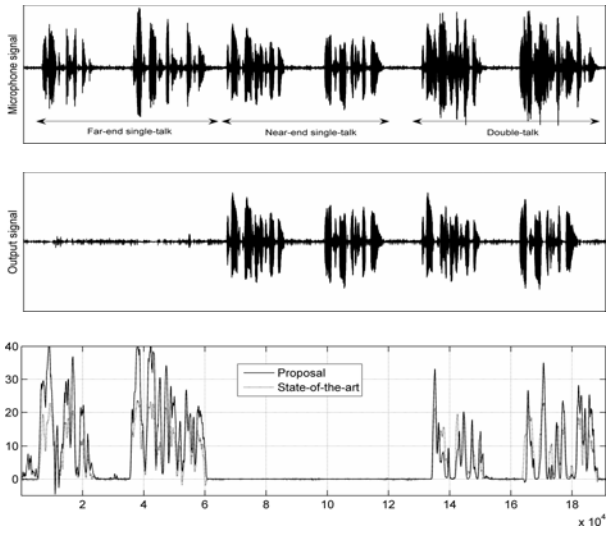

Fig. 2 The first plot shows the microphone signal, the second shows the enhanced (postfilter) output signal, and the ERLE (in dB) of the presented system and the state-of-the-art reference is depicted in the third plot.

far-end single-talk. But also during the double-talk period a higher residual echo suppression (at the postfilter output) could be observed.

As has been described earlier, the adaptive filter does not introduce a delay into the signal path. Only the low-delay postfilter produces a delay which can, however, be more than halved compared to an analysis-synthesis subband system. In our case the whole system has a delay of 16 samples compared to 143 samples for a system with a synthesis filter bank (127 samples filter bank group delay and 16 samples for considering the non-causal subband impulse response).

A higher frequency resolution (i.e. a larger number of subbands) for spectral weighting can be realized with an appropriately-designed prototype filter which should, on the one hand, exhibit a high stopband attenuation for a satisfactory adaptive filter performance [1, chap. 9], and, on the other hand, have a small delay to avoid too big an adaptation time lag.

5. CONCLUSION

In this paper we have introduced a combined subband delayless adaptive filter and low-delay postfilter with a control algorithm based on a cross correlation estimate. A separate VAD or echo path change detection is not required. With the structure presented the delay of an analysis-synthesis system can be more than halved compared to a system with a synthesis filter bank. The residual echo suppression postfilter allows for an easy integration of a noise reduction weighting rule. In a simulation with a time-varying car impulse response we compared the algorithm with a state-of-the-art hands-free system.

6. REFERENCES

[1] E. Hänsler and G. Schmidt, *Acoustic Echo and Noise Control. A Practical Approach*, Wiley, Hoboken NJ, 2004.

[2] W. Kellermann, "Analysis and design of multirate systems for cancellation of acoustical echoes," in *Proc. ICASSP*, New York NY, USA, pp. 2570–2573 Apr. 1988.

[3] VDA Specification for Car Hands-free Terminals, Version 1.5, Draft, Verband der Automobilindustrie, 12/2004.

[4] D. R. Morgan and J. C. Thi, "A delayless subband adaptive filter architecture," *IEEE Signal Processing Mag.*, vol. 43, no. 8, pp. 1819–1830, Aug. 1995.

[5] H. W. Löllmann and P. Vary, "A warped low delay filter for speech enhancement," in *Proc. IWAENC*, Paris, France, Sept. 2006.

[6] G. Enzner and P. Vary, „Robust and elegant, purely statistical adaptation of acoustic echo canceler and postfilter," in *Proc. IWAENC*, Kyoto, Japan, Sept. 2003.

[7] J. M. de Haan, *Filter Bank Design for Digital Speech Signal Processing. Methods and Applications.*, Ph.D. thesis, Blekinge Institute of Technology, Ronneby, Sweden, 2004.

[8] C. Beaugeant, V. Turbin, P. Scalart, and A. Gilloire, "New optimal filtering approaches for hands-free telecommunication terminals," *Signal Processing (Elsevier)*, vol. 64, pp. 33–47, 1998.

[9] T. Okuno, M. Fukushima, and M. Tohyama, "Adaptive cross-spectral technique for acoustic echo cancellation," *IEICE Trans. Fundamentals*, vol. E82-A, no. 4, pp. 634–639, Apr. 1999.

[10] M. Schönle, C. Beaugeant, K. Steinert, H. W. Löllmann, B. Sauert, and P. Vary, „Hands-Free Audio and its Application to Telecommunication Terminals," *Proc. AES, 29th International Conference*, Seoul, South Korea, Sept. 2006.