Realization of an Adaptive Algorithm with Subband Filtering Approach for Acoustic Echo Cancellation in Telecommunication Applications

Malte Sandrock, Stefan Schmitt
DSPecialists GmbH, D-10245 Berlin, Germany
http://www.dspecialists.de

Abstract

This work presents a model and its realization on digital signal processors (DSP) for suppressing annoying acoustic echoes in telephone and mobile communication applications. For cost sensitive subscriber devices and high quality voice communications the aims are 1) minimum computational complexity at 2) maximum convergence speed and 3) maximum echo attenuation. The chosen method for the echo cancellation involves a subband decomposition performing the adaptive filtering in subbands at a down-scaled sampling rate. The algorithm is described theoretically, modeled under MATLAB/Simulink and implemented on a fixed-point DSP.

1. Introduction

Figure 1 depicts a system where a Far End Speaker (FES) is impaired by the echo of his own voice which superposes the Near End Speaker´s (NES) voice. The system composed of loudspeaker, room and microphone forms the impulse response $\vec{g}(n)$ of the LEM-system (i.e. loudspeaker-enclosure-microphone system). The total electroacoustic circuit may become unstable, if the FES also is in a LEM system. Without an impact to universal validity we focus on a model with the NES being in a LEM-system and the FES not.

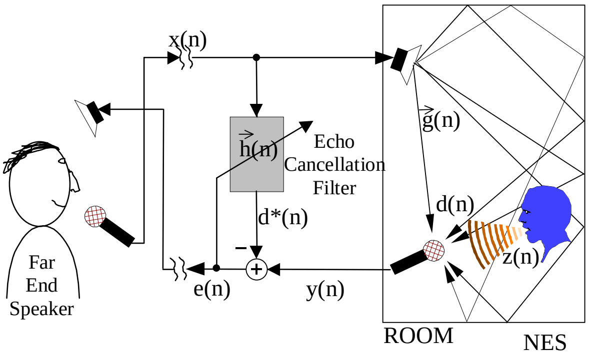

Figure 1: Acoustic echo cancellation with adaptive filtering

The FES signal *x(n)* and $\vec{g}(n)$ determine the echo component *d(n)*. Thus, the microphone signal *y(n)* is a superposition of the NES speech component *z(n)* and the LEM-system echo component *d(n)*. By introducing an echo cancellation filter (ECF) $\vec{h}(n)$ incl. subtraction in parallel to the LEM-system the FES loudspeaker signal *e(n)* is expected to contain only the NES speech *z(n)*, i.e. $\vec{h}(n) = \vec{g}(n)$ leads to *e(n)* = *z(n)*. Since the real LEM-system is supposed to be time-variant, $\vec{h}(n)$ must be realized as an adaptive filter. Its coefficients are computed recursively using an approximation algorithm.

Strong acoustic echo cancellation is characterized by fast convergence and a good tracking behavior. The convergence speed characterizes the time the system needs to reach a certain degree of echo-attenuation after initialization. Tracking behavior determines the ability of the echo canceller to re-adapt when the echo path changes (e.g. a change in the speaker’s position), taking into consideration that minimal echo path changes cause big influences on the impulse response.

“Real world” acoustic echo cancellation (AEC) controllers typically include in addition to the ECF non-linear filters and a double talk detector for freezing filter coefficients during the presence of NES speech. Our work, however, focuses on a subband decomposition approach for the ECF $\vec{h}(n)$.

2. System Identification with Adaptive Filters

A widely implemented technique of digital ECF is based on modeling the impulse response of the echo path as an adaptive filter. To achieve a sufficiently high echo attenuation, the replica of the acoustic echo path must be very long (e.g. 1000 taps, depending on the reverberation time of the LEM-System and the sample-rate chosen). Recursive filters implementations are not advantageous compared to non-recursive filter implementations [1][2]. The main reason for taking FIR filters is their guaranteed stability. This advantage is especially important in case of adaptive filter structures implemented on fixed-point DSPs. Hence, using high order transversal adaptive filter structures is widespread in ECF. However, the consequences are 1) a high computational complexity of filtering and of filter coefficient actualization, 2) a low convergence speed of the adaptive filter arises.

For filter coefficient adaptation the Least Mean Square Algorithm (LMS) by Widrow et al. has been used widely. Because of the instationarity of speech signals with regard to short-time variance we apply its normalized version NLMS. In comparison to other algorithms, NLMS is robust and quite simple regarding the computational complexity. Its disadvantage in fullband applications is that the convergence speed decreases significantly when correlated colored signals like speech are used as excitation signals [1][3].

Other algorithms like the Affine Projection (AP) show a better convergence behavior but the computational complexity increases with the factor S in relation to NLMS, where S denotes the order of the AP algorithm The Recursive-Least-Squares (RLS) Algorithm is suitable for colored, stationary excitation, but involves possible instability in fixed-point DSP implementations.

For real time implementations on low cost fixed-point DSPs our ECF approach focuses on NLMS and adaptive FIR filters.

3. Subband Filtering

The basic idea of a subband decomposition approach is its increase in convergence speed in comparison to a fullband solution. This is due to a reduced spectral magnitude range, i.e. subband filtering has a decorrelating effect because colored input signals are decomposed into subbands with “whiter” sub-spectra.

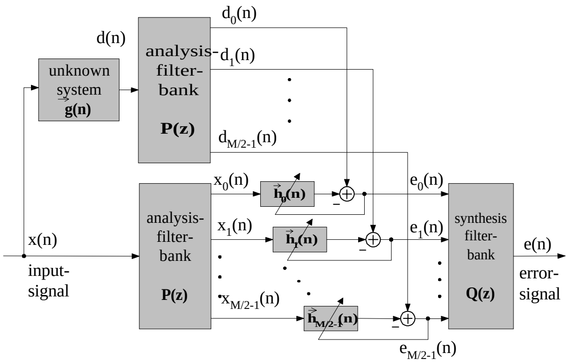

Figure 2: Subband adaptive filtering (SAF)

Figure 2 depicts the subband adaptive filtering (SAF). Using analysis filter banks *P(z)* the original signal from FES and the NES microphone signal are decomposed by subdividing their spectra into smaller intervals ($x_0(n)$, $x_1(n)$, ...). Adaptive filtering is then performed in these subbands by a set of independent filters ($\vec{h}_0(n), \vec{h}_1(n), ...$). The outputs of these filters are subsequently combined using a synthesis filter bank *Q(z)* to reconstruct the fullband output.

The bandwidth in each subband is reduced. Thus the sampling frequency for each subband filter can be lowered. Consequently adaptive filters need fewer taps in comparison to fullband solutions to cover the same time interval and are updated at a lower rate. This leads to a significant reduction of computational complexity.

3.1 Characterization of Subband Decomposition

Because the quality of subband separation is highly significant for the obtained decimation rate and for the convergence behavior of the adaptive filters in subbands, the design of analysis and synthesis filter banks is the determining factor for the quality and efficiency of the overall system [3].

Because linear group delays are required for subband adaptive filtering (SAF), only non-recursive filters are allowed for the filter banks.

Apart from the advantages, one considerable disadvantage of subband adaptive filtering must be accepted: it is the additional signal delay depending on the length of the filters used in the filter banks. The necessary filter length is directly dependent of the number of subbands and the quality of the filterbank, i.e. the flatness in passband, the transition bandwidth in frequency domain and the stop-band rejection. These factors in turn correspond directly to the obtainable decimation factor (and therefore the reduction of computational complexity), the convergence speed and the maximum echo attenuation of SAF. Non-ideal filter banks themselves color the subband signals significantly and in that way they reduce the whitening effect of spectral division. Moreover, they cause alias in subbands which results in lower SAF convergence speed as well (Perfect reconstruction of output full-band signal is not enough to reach good SAF convergence behavior!).

With regard to the filter length a tradeoff between the contradictory requirements short signal delay, which is associated with low computational overhead resulting from subband decomposition, and quality of filter banks must be found.

One further condition is the need of complementary filter banks in order to achieve insignificant distortions at output of SAF. Consequently, spectral overlap among neighboring filters must be permitted to avoid spectral loss when using non-ideal filters [3][4]. On the other, hand distortion caused by aliasing should be avoided.

3.2 Filter Bank Structure

The tradeoff is achieved by using an oversampled filter bank, depicted in figure 3. Whereas in case of critical decimation the down-sampling ratio *L* equals the number of uniform subbands *M*, oversampling is characterized by a ratio $L < M$ and leads to an introduction of spectral redundancy. With respect to computational efficiency it is advantageous to chose a decimation factor *L* close to the critical one *M*, which results in non-integer oversampling rates ${}^{M}\!/_{L}$.

In case of decimating real valued subband signals the choice of non integer oversampling factors is restricted in order to avoid aliasing. However, complex modulated subband signals can be decimated at any integer decimation rate below the critical one [5]. Hence, we prefer complex valued subband processing.

For subband separation and recombination we use DFT filter banks. To subdivide the sequence *x(n)*, (apart from the low-frequency part) any part of the spectrum centered around the frequencies $\omega = \omega_m$ (for $m = 0, 1, \ldots, M$) are shifted into the base-band by multiplying *x(n)* with the complex sinusoid $e^{-j\omega_m n}$ (with $\omega_m = 2\pi \frac{m}{M}$). After this modulation, the low-pass filter *P(z)* is used to extract the desired frequency band. Since the separated subband signals are in base-band and have a smaller bandwidth, they may be decimated before further processing.

The prototype filter *P(z)* and the down-sampler are used repeatedly, which has the advantage of leading to efficient DSP implementation. Moreover, to reduce the computational complexity, this combination can be realized as polyphase filter. In this case, down-sampling is performed before the filtering and the lowpass filter. Finally, the filter decomposed into polyphase components is working at the lower sample-rate [4][6].

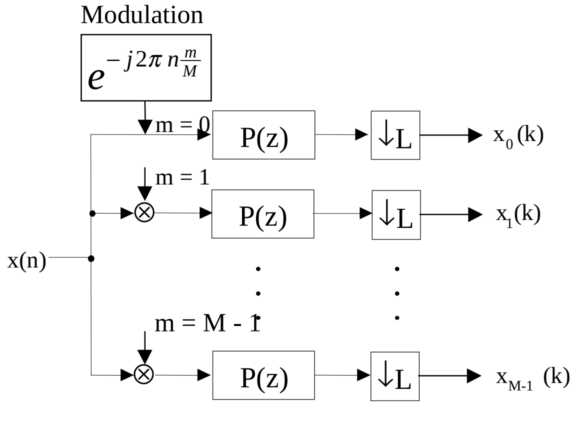

Figure 3: Analysis Filter Bank

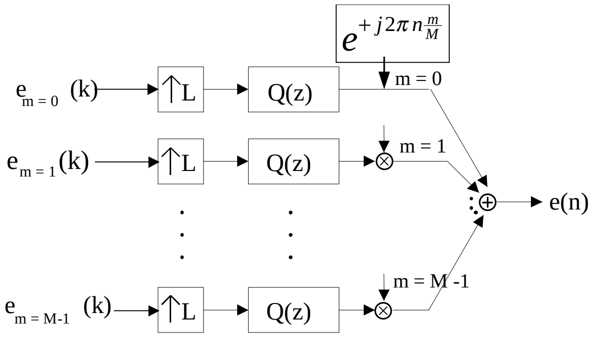

Figure 4: Synthesis Filter Bank

The structure of the synthesis filter in figure 4 is complementary to the analysis filter bank. Again, upsampling and interpolation are realized for all subbands with the same polyphase lowpass filter.

3.4 Design of Prototype Filters

Our aim in prototype filter design is to build complementary analysis and synthesis filter banks, so that the reconstructed output signal contains very low distortion. We choose different prototype filters for analysis and synthesis filterbank using the freedom that is offered by oversampled decimation [3]

The bandwidth of the synthesis filter $\frac{\pi}{M}$ is determined by the number of subbands M whereas the analysis filter bandwidth is larger and only limited by the decimation rate *L*. Our analysis and synthesis prototype filters are selected as in figure 5.

The analysis prototype filter is chosen such that it has a flat magnitude response between zero and a frequency larger than ω_{ss}, where ω_{ss} is the beginning of the stop band of the synthesis prototype filter. Therefore it prevents the subband spectra from varying and decaying over the frequency range which is not cut off by the synthesis filter. Over this range matching between the impulse response of each subband adaptive filter to the desired (room) impulse response from the associated band should be achieved. Thus this range is of prime importance for the convergence behavior of each SAF. Hence, choosing prototypes as in figure 5, all subband filters are well excited over their respective bands of interest which leads to good convergence behavior of the overall filter.

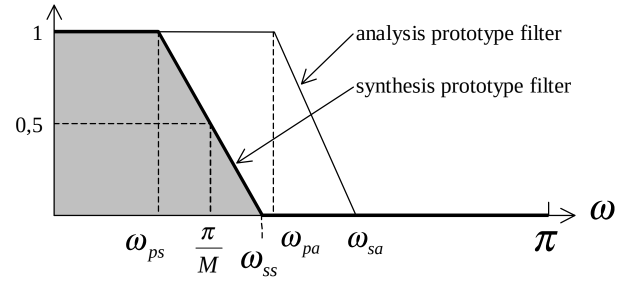

Figure 5: Possible choice of analysis and synthesis prototype filters

4. Simulation Model

We built an ECF model using MATLAB/Simulink with a 16-band complex modulated filter bank and NLMS adaptive filters in subbands. We chose analysis and synthesis prototype filters with length S_P and S_Q of 190 taps. For filter design we use a method from [3]. The decimation factor *L* is 10. All calculations and the later DSP implementation are based on a sampling rate of 8 kHz.

The echo-path is modeled by FIR-filter. The FIR impulse response has a reverberation time of 72 ms and was measured in a large car. As excitation signal we use colored noise, which is generated by first order IIR filtering of white noise, and speech. For comparison means we built an equivalent fullband ECF model.

In comparison to the fullband system, the convergence speed of the subband realization is significantly higher, as depicted in figure 6. The improvement depends on the correlation of the excitation signal. With a colored noise excitation (IIR feedback coefficient *b*=0.9) SAF achieves an ERLE of appr. 30 dB one second after initialization, in contrast to 15 dB of the fullband system. The maximum echo-attenuation is better than 30 dB. With speech signals the improvement is considerable as well.

Concerning the requirements of computational power in case of a DSP implementation, with a system whose parameters are given in table 1, the subband system only needs 30 % DSP operations (real multiplications or multiply-and-accumulate operations, including overhead associated with the filterbanks) of the fullband system. For this calculation it must be taken into account that *M*/2-1 subbands are redundant because of spectral symmetry and complex modulation. Moreover, the signals in the base band and in the band around ◊/2 are real.

5. DSP Implementation

After completing the verification phase of the simulation model, the modular model was transferred into a code library on a fixed-point Motorola DSP56301. The code modules are implemented in assembly. All modules are realized object oriented. Thus, multi-instances of them can be used during run-time. The assembly implementation and the simulation model show exactly the same quantitative behavior with regard to *ERLE*.

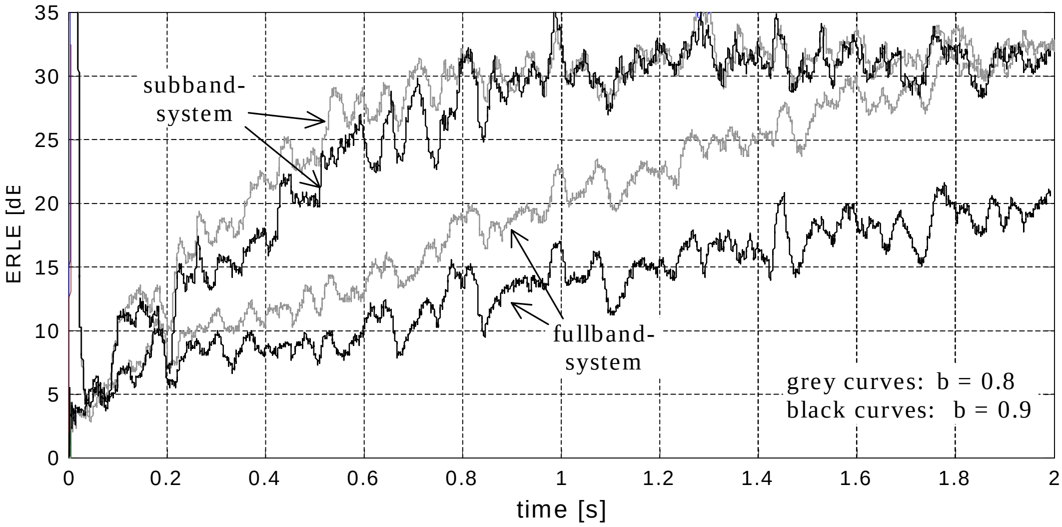

Figure 6: Simulation results: Echo Return Loss Enhancement (*ERLE*) of the subband system in comparison to fullband system when colored noise is used for excitation (two different degrees of input signal correlation: the feedback coefficient *b* set to 0.8 and 0.9). The chosen filter-length for the adaptive filters is 1000 in full-band system and 100 in each decimated subband channel.

Parameters	
Sample Rate	$f_T = 8000$ Hz
Number of Subbands	$M = 16$
Decimation Rate	$L = 10$
Filter length of adaptive Filters (subband, fullband)	$S_{hSB} = 100$ $S_{hFB} = 1000$
Filter length of Prototype Filters	$S_P = S_Q = 190$
Necessary Operations in Time	
Subband System	
Filterbanks 2x Analysis 1x Synthesis	$\approx 3\cdot\frac{M}{L}\cdot S_{P/Q}\cdot f_T$ = 2,4 MOPS
Adaptive Filtering	$\approx\left[12\cdot\left(\frac{M}{2}-1\right)+8\right]\cdot S_{hSB}\cdot\frac{f_T}{L}$ = 7,36 MOPS
Fullband System	
Adaptive Filtering	$4\cdot S_{hFB}\cdot f_T$ = 32 MOPS

Table 1: Model and Implementation Parameters

6. Conclusion

Acoustic echo cancellation is characterized by the need of high order adaptive filters requiring much computational power. Adaptive filtering suffers from slow convergence when handling with speech signals. Subband decomposition is a suitable means to improve both, to accelerate convergence and to reduce the computational complexity of the overall system.

The longer the reverberation time of the echo to be cancelled is, the higher the relative win of computational complexity is. The design of the filterbank and especially of its prototype filters is the decisive factor to the successful implementation of a SAF. The created DSP-algorithm behaves like the simulation model and confirms our theoretical considerations.

Future work will be done to optimize the implementation with regard to computational complexity of the filterbank. Additionally it will be extended by components needed for a “real world” AEC.

7. References

[1] Breining, Chr. et al: “Acoustic Echo Control“, IEEE Sign. Proc. July 1999, p. 42-69
[2] Breining, Chr. et al: “Freisprechen – die Jagd nach akustischen Echos“, Frequenz 51, 1997, p. 245-253
[3] Farhang-Boroujeny, B.: “Adaptive Filters, Theory and Applications”, Wiley, 1999
[4] Fliege, N. J.: “Multiraten-Signalverarbeitung”, Teubner, Stuttgart 1993
[5] Weiss, S., Stewart, R. W.: “On Adaptive Filtering in Oversampled Subbands”, Shaker Verlag, Aachen, 1998
[6] W. J. Hess: “Digitale Filter”, Teubner 1993