

DESIGN OF OVERSAMPLED DFT MODULATED FILTER BANKS OPTIMIZED FOR ACOUSTIC ECHO CANCELLATION

Qin Li, Wei-Ge Chen, Chao He, and Henrique S. Malvar

Microsoft Research,
One Microsoft Way, Redmond, WA 98052, USA

ABSTRACT

This paper describes a method for designing oversampled DFT filter banks (FB) optimized for subband acoustic echo cancellation (AEC). For this application, the design requirements typically are good echo cancellation quality, low delay, small reconstruction error, and low computation complexity. Our method explicitly includes a model for echo return loss enhancement (ERLE) as part of the optimization criteria. Convergence of the high-dimensional, nonlinear optimization problem is facilitated by decorrelating the prototype filter impulse response via a discrete cosine transform (DCT), discarding many insignificant coefficients and thus reducing the dimensionality of the search. The experimental results demonstrate the effectiveness of our design and the effects on the performance of AEC, with ELRE improvements on the order of 3 dB or better. The method is flexible and could also be extended to other application domains.

Index Terms – oversampled filter banks, modulated filter banks, acoustic echo cancellation, adaptive filters.

1. INTRODUCTION

In communication systems, the presence of acoustic echo is highly undesirable as it degrades the perceived voice quality. AEC is an effective approach to mitigate this problem and is thus widely used in modern communication devices. Among different techniques for AEC, subband adaptive filtering has the advantage of high computational efficiency and fast rate of convergence [1].

Multirate signal processing has been extensively studied and remarkable theories have been developed thoroughly [2], [3]. Due to the nature of subband adaptive filtering, critically-sampled FBs are generally considered unsuitable, despite of their attractiveness and successful applications in compression. In [4] it was shown that if critically sampled FBs are to be used correctly for adaptive filtering, one must consider cross-band filters, at the cost of significantly increased computation complexity. It is also commonly recognized that the relaxation of the perfect reconstruction (PR) constraint can lead to better prototype filter designs, because it allows for further optimization of the filter characteristics. Moreover, because the adaptive filters alter the subband signals, the loss of PR is less of an issue. Near-perfect reconstruction (NPR) FBs can keep the signal distortion to a low enough level, while providing more flexible design choices. For this reason, oversampled NPR FBs are often chosen in practice, and nu-

merous filter design choices have been proposed (see [5]–[10] and references therein). In this paper, we focus on the design of NPR oversampled uniform filter banks using discrete Fourier transform (DFT) modulation.

We propose a design mechanism that directly optimizes a cost function which includes a quantity that approximately models the ERLE of the overall system. Such explicit optimization is expected to boost the overall performance of AEC in comparison to other design methods that are indirect in this regard. In [11] a similar attempt was made to quantify the performance of subband adaptive filters. Furthermore, we introduce a regularization procedure that increases the prospect of reaching a global minimum of the cost function. Instead of directly optimizing the coefficients of the prototype filter, the actual minimization takes place in the DCT domain, where the components are decorrelated and the energy is concentrated in a few low frequencies.

2. UNIFORM OVERSAMPLED DFT FILTER BANKS

When viewed from the bandpass filter interpretation, the FBs first process the input signal $x(n)$, by K bandpass filters denoted by $h_k(n)$. The resulting signals are decimated by a factor of M and become $X_k(m)$. With AEC, each of the K channels has an adaptive filter that minimizes the echo energy, resulting in echo cancellation [1]. After processing, the subband signals are upsampled by M , filtered by the synthesis filter $f_k(n)$, and added to produce the reconstructed signal $\hat{x}(n)$. We consider oversampled DFT modulated filter banks, where $M < K$ and the channel filters are related to each other by DFT modulation, $h_k(n) = h_0(n) e^{j2\pi kn/K}$. Furthermore, as usual, we choose $f(n) = h(n)$ so that the design problem is simplified to a single prototype one. Note that the prototype filter responses without any subscript are used interchangeably as the ones with subscript zero throughout the rest of the paper. The modulated FBs can also be interpreted as a poly-phase structure, resulting a connection to lapped transforms and efficient implementations [3]. With NPR FBs we relax exact reconstruction, and instead seek $\hat{x}(n) \approx x(n)$.

Most existing approaches optimize a selection of design criteria that are usually chosen ad hoc as noted in [8]. For example, [6] optimizes the combination of the stop band energy and the power complementary condition, while the recent work in [10] optimizes the peak amplitude distortion. In [8] the authors derived explicit bounds for different design criteria and the globally optimal solution can be obtained via

a convex optimization problem. The method is simplified and extended to design two prototypes filter banks in [9].

When designing FBs for AEC, the requirements are typically specified as audio quality, delay and computation complexity as dictated by the overall system design. Correlating common filter characteristics to the system specifications is not trivial, and in practice requires significant heuristics. It is thus desirable to design the FBs by directly targeting the system performance specifications, such as ERLE.

3. FREQUENCY-DOMAIN CONSIDERATIONS

Let the analysis filters $h_k(n)$ and the synthesis filters $f_k(n)$ have frequency responses $H_k(e^{j\omega})$ and $F_k(e^{j\omega})$ respectively. If the subband signals are not modified, the input-output relationship can be written in the frequency domain as [12]

$$\begin{aligned} X(e^{j\omega}) &= \frac{1}{M} \sum_{k=0}^{K-1} F_k(e^{j\omega}) \sum_{l=0}^{M-1} H_k(e^{j(\omega-2\pi l/M)}) X(e^{j(\omega-2\pi l/M)}) \\ &= \frac{1}{M} \sum_{k=0}^{K-1} F_k(e^{j\omega}) H_k(e^{j\omega}) X(e^{j\omega}) + \\ &\quad \frac{1}{M} \sum_{k=0}^{K-1} F_k(e^{j\omega}) \sum_{l=1}^{M-1} H_k(e^{j(\omega-2\pi l/M)}) X(e^{j(\omega-2\pi l/M)}) \end{aligned} \quad (1)$$

The first term in (1) is typically referred to as the linear component and the second part as the aliasing component. In a PR FB, the first term would be a simple delay while the second term would equal zero for all frequencies. In a NPR FB, neither of above needs to be completely true and aliasing is largely a function of the stopband characteristics of the channel filter. Indeed, many existing design methods use various forms of the stopband characteristics as the design criteria. One way to measure the subband aliasing considers the total aliasing energy from subsampling:

$$E_a = \int_{-\pi/M}^{\pi/M} \sum_{l=1}^{M-1} \left| H_0(e^{j(\omega-2\pi l/M)}) \right|^2 d\omega \quad (2)$$

We note that it suffices to look at the case of $k=0$ because of the modulation relationship of the channel filters. For a well-behaved prototype filter, E_a should be very low.

Minimizing E_a alone, however, would not guarantee good performance of AEC. In fact, since ERLE is an effective measure of the echo cancellation quality, it will be used directly as a design criterion as shown in the following. For each subband k , let us assume that the echo signal in the passbands $[2\pi k/K - \pi/M, 2\pi k/K + \pi/M]$ of the analysis filters can be completely cancelled. Although this may seem a rather bold assumption, it is largely justified as the passband components remain linear after decimation. Under ideal conditions, the adaptive filters will converge to the unknown system and thus drive the residue errors to a negligible level [11]. Note that the presence of ‘‘doubletalk’’ is irrelevant to the validity of this assumption. Thus, we can incorporate the ‘‘ideal’’ echo canceller into the analysis filters:

$$H'_k(e^{j\omega}) = \begin{cases} 0, & \omega \in \left[\frac{2\pi k}{K} - \frac{\pi}{M}, \frac{2\pi k}{K} + \frac{\pi}{M} \right] \\ H_k(e^{j\omega}), & \text{elsewhere} \end{cases} \quad (3)$$

Replacing $H_k(e^{j\omega})$ in (1) with (3) yields a model for the echo residual signal:

$$\hat{X}(e^{j\omega}) = \frac{1}{M} \sum_{k=0}^{K-1} F_k(e^{j\omega}) \sum_{l=0}^{M-1} H'_k(e^{j(\omega-2\pi l/M)}) X(e^{j(\omega-2\pi l/M)}) \quad (4)$$

While (4) looks similar to the aliasing term in (1), it is actually a different quantity. Obviously, the smaller $\hat{X}_k(e^{j\omega})$ is, the higher the echo cancellation quality or ERLE will be. Hence a reasonable design criterion is to minimize

$$E_r = \int_{-\pi/M}^{\pi/M} \left| \sum_{k=0}^{K-1} F_k(e^{j\omega}) \sum_{l=0}^{M-1} H'_k(e^{j(\omega-2\pi l/M)}) \right|^2 d\omega \quad (5)$$

which represents the total echo residual energy of the ‘‘ideal’’ echo canceller when the input signal $X(e^{j\omega})$ is white. While real signals are most likely not white noise, this simplification allows us to construct a signal independent design criterion. Note that minimizing E_r is not the same as enforcing an NPR condition via the minimization of the aliasing term in (1).

Another frequency domain characteristic is the linear component of (1). While the stopband behavior is measured by E_a , we have not developed any constraint on the passband. Considering that the prototype filter should not deviate much from an ideal lowpass filter, we seek to minimize the passband ripple energy of the analysis filter given by

$$E_p = \int_{-\pi/M}^{\pi/M} \left(\left| H(e^{j\omega}) \right| - \sqrt{N/2} \right)^2 d\omega \quad (6)$$

Experiments confirm that minimizing E_p helps to stabilize the optimization procedure.

4. TIME-DOMAIN CONSIDERATIONS

While the frequency domain characteristics are necessary for designing a good prototype filter, they are not sufficient. For example, one may satisfy the frequency domain criteria by relying mostly on the time domain aliasing cancellation. However, because the gain factors for each subband are subject to independent changes over time as a result of AEC, time domain cancellation becomes much less effective. It is thus reasonable that a good solution should strike a balance between time- and frequency-domain criteria.

Similarly to the frequency domain version, the input-output relationship of the FB can be written in time as [12]

$$\hat{x}(n) = \sum_{m=-\infty}^{\infty} \sum_{s=-\infty}^{\infty} x(n-sK) f^m(n) h^m(sK-n) \quad (7)$$

where $f^m(n) = f_0(n-mM)$ and $h^m(n) = h_0(n-mM)$. From (7) it follows that, assuming the FB has a delay of M samples,

PR can be accomplished by satisfying the following conditions:

$$\sum_{m=-\infty}^{\infty} q^m(n,1) = 1, \quad \text{for all } n \quad (8)$$

$$\sum_{m=-\infty}^{\infty} \sum_{s=-\infty}^{\infty} q^m(n,s) = 0, \quad \text{for all } s \neq 1 \text{ and all } n \quad (9)$$

where $q^m(n,s) = f^m(n) h^m(sK - n)$. These are generalized conditions compared to the approach taken in [5], where the time domain reconstruction error is measured by an ‘‘impulse response’’ (we use quotes because strictly speaking, the overall FB, even without AEC subband processing, is not a linear time-invariant system; it’s a periodically time-varying system with period M).

Following (8) and (9), the two time-domain criteria are then defined as the direct term:

$$\epsilon_p = \sum_{n=-\infty}^{\infty} \left| \sum_{m=-\infty}^{\infty} q_m(n,1) - 1 \right|^2 \quad (10)$$

and the aliasing term:

$$\epsilon_a = \sum_{n=-\infty}^{\infty} \sum_{s=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} |w(m,s) q_m(n,s)|^2 \quad (11)$$

where $w(m,s)$ is a weighting function determined according to a human psychoacoustic model [13].

5. COMPUTATIONAL ISSUES

In general, there is no guarantee that simultaneous minimization of E_r , E_a , E_p , ϵ_p and ϵ_a , according to (2)–(11) would be possible. Thus, a practical approach is to use a cost function that linearly combines the error metrics, in the form

$$\epsilon \triangleq \alpha E_r + \beta E_a + \gamma E_p + \lambda \epsilon_a + \eta \epsilon_p$$

where α , β , γ , λ , and η are positive weights. Empirically, we determine these constants such that E_r is the dominant cost, about an order of magnitude higher than the aliasing constraints E_a and ϵ_a , while the ‘‘in-band’’ metrics E_p and ϵ_p receive half the weight of E_a and ϵ_a .

In typical applications the prototype filter has hundreds of taps, so it lies in a relatively high dimension for straight use of common optimization routines to be effective. Because $h(n)$ is low-pass in nature, its DCT spectrum decays quickly with frequency. Thus, we can efficiently approximate $h(n)$ by using only a small number of its DCT coefficients, which become the free variables, significantly reducing the dimensionality of the optimization problem.

The frequency domain error metrics, E_r , E_a , and E_p , require the integration from $-\pi$ to π , we approximate that via evaluation on a dense grid of frequencies. The time domain metric ϵ_p and ϵ_a are evaluated in a finite range of n due to periodicity. It can be shown that in practice computation of all

these metrics can be done efficiently; we do not provide the straightforward details due to space limitations.

6. RESULTS

In the following, we use a particular design example to demonstrate the effectiveness of the proposed method, although it should be understood that our method is versatile enough to handle general design conditions. For the multi-variable minimization, we use the MATLAB `fminsearch` function, a simplex search method. The implementation runs in MATLAB and typical takes a few minutes to converge on a typical desktop computer. For this example, the system requirements on the FB are the following: audio sampling rate = 16 kHz, delay ≤ 4 ms, FFT size ≤ 128 (due to the complexity consideration), and ERLE ≥ 30 dB for white noise input. With these in mind, the following dimensions of the FB are now fixed: $M = 64$, $K = 128$.

With these parameters, we first design a prototype filter with the windowing approach, using a Kaiser window. The outcome of this first step is used as the initial condition for the optimization loop. With the Kaiser window, let $\delta_p = \delta_s$, the passband and stopband deviations, range from 0.1 to 0.01. The transition bandwidth is $\Delta F = \pi/2M$. Following [12], the length N of the prototype filter is estimated at the range [220, 576]. For ease of implementation, we choose $N = 3M$ which stands in the middle of the range. Note that while the above method of estimating N is useful, it does not imply a limitation of our design method. It is certainly possible to design the prototype filters at different lengths and optimize for N .

Tab. 1 shows the values of the different design criteria and the overall cost function at different iterations. We see that at the end of optimization, E_r is improved by more than 7dB, E_a by 5dB, and ϵ_a by 3dB, while other metrics decrease to or hold at the desirable levels. Figures 1 and 2 show the impulse and frequency responses of the initial filter and the optimized filter, respectively.

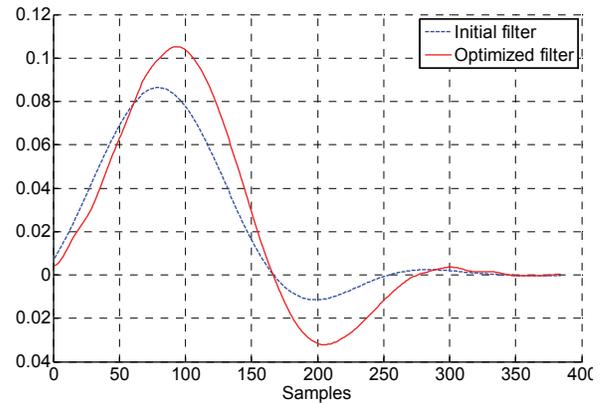


Figure 1. Impulse responses of the initial and optimized filters.

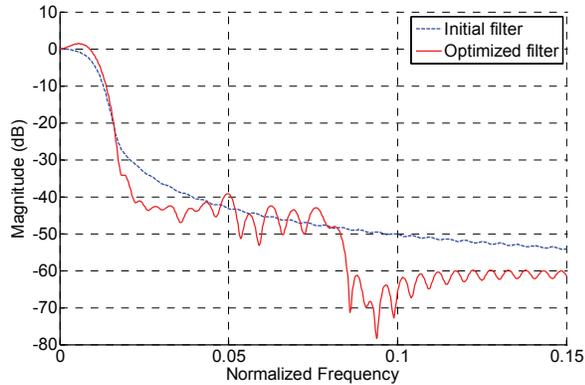


Figure 2. Frequency responses of the initial and optimized filters.

The top line in Fig. 3 shows the ERLE of the full-band output from AEC as a function of time in seconds, using the optimized filter. Here the input to AEC is a white noise signal convolved by a room impulse response filter. We can see that after about 15 seconds, the ERLE converges to around 34.5dB. In the steady state, the optimized filter outperforms the initial Kaiser window design (middle line) by 3.5dB.

By contrast, when we remove the ERLE metric E_r from the cost function, the resulting filter is not as effective. At the end of the iterations, we are still able to achieve superior or comparable levels for all design criteria except for E_r (see the last row in Tab. 1). However, as shown in Fig. 3, the steady state ERLE of this experimental filter (bottom line) actually regresses from the initial filter, dropping by a margin as predicted by E_r .

iter.	E_r (dB)	E_a (dB)	E_p (dB)	ε_a (dB)	ε_p (dB)	ε
1	-27.06	-25.80	-29.27	-15.28	-33.49	0.1293
401	-29.16	-27.65	-30.98	-15.49	-34.16	0.0829
801	-31.75	-31.35	-29.37	-16.03	-32.36	0.0536
1201	-33.68	-31.21	-28.48	-16.30	-32.47	0.0441
6001	-34.17	-30.84	-29.21	-17.77	-35.36	0.0386
end	-34.10	-30.58	-29.55	-17.88	-36.25	0.0382
alt	-30.98	-31.49	-30.64	-18.39	-29.95	0.0294

Table 1. Design criteria and cost functions at different iterations.

7. CONCLUSION

We present a technique for designing oversampled DFT filter banks optimized for AEC. Our design technique directly targets the system performance (ERLE) jointly with other desirable criteria and takes the guesswork out of the designers' hands. The effectiveness of our proposed method is demonstrated by a design example as well as experimental results performed with subband AEC. It is worth noting that our approach can be generalized to design other types of filter banks, such as those based on two prototypes. Applications other than AEC may also benefit from the same design philosophy by exploiting domain-specific error metrics.

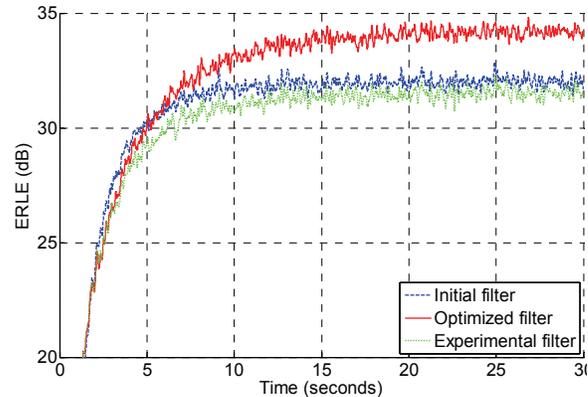


Figure 3. ERLEs of the initial, optimized, and experimental filters.

REFERENCES

- [1] E. Hansler and G. Schmidt, *Acoustic Echo and Noise Control: A Practical Approach*, Chapter 9, New York: Wiley-IEEE, 2004.
- [2] P. P. Vaidyanathan, *Multirate Systems and Filter Banks*. Englewood Cliffs, NJ: Prentice Hall, 1993.
- [3] H. S. Malvar, *Signal Processing with Lapped Transforms*. Norwood, MA: Artech House, 1992.
- [4] A. Gilloire and M. Vetterli, "Adaptive filtering in subbands with critical sampling: analysis, experiments, and application to acoustic echo cancellation," *IEEE Trans. on Signal Processing*, vol. 40, no. 8, pp. 1862–1875, Aug. 1992.
- [5] M. Harteneck, S. Weiss, and R.W. Stewart, "Design of near perfect reconstruction oversampled filter banks for subband adaptive filters," *IEEE Trans. on Circuits and Systems II: Analog and Digital Signal Processing*, vol. 46, no. 8, pp. 1081–1085, Aug. 1999.
- [6] Q-G. Liu, B. Champagne, and D. K. C. Ho, "Simple design of oversampled uniform DFT filter banks with applications to subband acoustic echo cancellation," *Signal Processing*, vol. 80, pp. 831–847, 2000.
- [7] K. Eneman and M. Moonen, "DFT modulated filter bank design for oversampled subband systems," *Signal Processing*, vol. 81, pp. 1947–1973, 2001.
- [8] M.R. Wilbur, T.N. Davidson, and J.P. Reilly, "Efficient design of oversampled NPR GDFT filterbanks," *IEEE Trans. on Signal Processing*, vol. 52, no. 7, pp. 1947–1963, 2004.
- [9] B. Dumitrescu, R. Bregovic, and T. Saramaki, "Simplified design of low-delay oversampled NPR GDFT filterbanks," *EURASIP Journal on Applied Signal Processing*, vol. 2006, pp. 1–11, 2006.
- [10] D. Hermann, E. Chau, R.D. Dony, and S.M. Areibi, "Window based prototype filter design for highly oversampled banks in audio applications," in Proc. ICASSP, Hawaii, vol. 2, pp. 405–408, Apr. 2007.
- [11] S. Weise, R.W. Stewart, A. Stenger, and R. Rabenstein, "Steady-state performance limitations of subband adaptive filters," *IEEE Trans. on Signal Processing*, vol. 49, no. 9, pp. 1982–1991, Sept. 2001.
- [12] R. E. Crochiere and L. R. Rabiner, *Multirate Digital Signal Processing*. Englewood Cliffs, NJ: Prentice Hall, 1983.
- [13] J. D. Johnston, "Transform coding of audio signals using perceptual noise criteria," *IEEE J. Selected Areas Communications*, vol. 6, pp. 314–323, Feb. 1988.