A GENERALIZED COMB FILTERING TECHNIQUE FOR SPEECH ENHANCEMENT

D. Malah*

Electrical Engineering Dept.
Technion - Israel Institute of Technology
Technion City, Haifa 32000, Israel

R.V. Cox

Acoustics Research Dept.
Bell Laboratories
Murray Hill, N.J. 07974, USA.

ABSTRACT

Because of speech signals nonstationarity, usual comb filtering of noisy speech signals results only in a modest improvement in signal to noise ratio, and only in a small perceptual reduction of structured noise or interference.

A generalized comb filtering technique, which applies a time-varying weighting to each pitch period, is mathematically analyzed and shown to be capable of breaking up the noise structure, in addition to comb filtering. This is found to provide a meaningful perceptual improvement when the noise or interference are structured.

The mathematical analysis is facilitated by using a polyphase network model of the generalized comb filter. Design constraints and rules are developed and several filter families are proposed. Computer simulation results are discussed.

INTRODUCTION

The harmonic structure of voiced speech has been the basis for applying comb filtering techniques to enhance speech degraded by noise or interference [1]. In the usual comb filter implementation, an output speech segment of pitch period duration is produced by weighting several adjacent pitch periods of the corrupted speech. In principal, one can improve the performance of the filter by increasing the number of signal periods used in the weighting process. However, since speech is only quasi-stationary, the number of pitch periods that can be weighted is limited to the quasi-stationary interval, which is typically only 20-40 msec in duration. The small number of pitch periods that can be used limits the performance of the comb filter. The attained improvement in signal to noise ratio (SNR) is therefore quite small [1].

If the noise or interference is structured (e.g., frame-rate noise in block processing coders [2], periodic interference, and a competing speaker [3]) one typically finds that the modest noise attenuation achieved by comb filtering does not harm much the noise structure. As a result, the perceptual reduction by the comb filter of structured noise or interference is quite small.

In this work we develop a mathematical model and analyze a generalized comb filtering (GCF) technique which in addition to usual comb filtering (CF) also breaks up the interference spectral structure.

The GCF technique, similar to the CF method, is based on weighting several adjacent pitch periods to produce the filtered signal. However, unlike CF, the weight given to the samples within each pitch period is not fixed.

MATHEMATICAL MODELING AND ANALYSIS

Let $x(n)$ be the noisy input signal. The output signal $y(n)$ from a comb filter is given by [1]

$$y(n) = \sum_{k=-L}^{L} a_k x(n-kP) \tag{1}$$

where, a_k, $-L \leq k \leq L$, are the 2L+1 comb filter coefficients (assumed here to be symmetrical). Its transfer function $H_c(z)$ is given by

$$H_c(z) = \sum_{k=-L}^{L} a_k z^{-kP} = A(z^P) \tag{2}$$

where $A(z)$ is the z-transform of the sequence $\{a_k\}$. Evaluation of $H_c(z)$ on the unit circle gives the frequency response $H_c(e^{j\theta}) = A(e^{jP\theta})$, which is periodic in θ with period $2\pi/P$.

One way of implementing the CF operation described in (1) is in block form [2]. This is done by stack-adding 2L+1 signal segments of P samples duration each, weighted by the 2L+1 filter coefficients $\{a_k\}$. This is equivalent to using a (2L+1)P point data window $w_c(n)$ of the form

$$w_c(n) = a_{-k}, \quad kP \leq n < (k+1)P, \quad -L \leq k \leq L , \tag{3}$$

by which the current 2L+1 input signal segments are weighted prior to the stack-add operation [2].

The generalization described in [2] corresponds to applying a non-fixed weighting to the samples within a given segment of duration P, as illustrated in Fig. 4. We denote the (2L+1)P point generalized data window by $w_g(n)$ and express it as follows :

$$w_g(n) = \alpha_{-k}(n-kP), \quad kP \leq n < (k+1)P \quad ; -L \leq k \leq L \tag{4}$$

where $\alpha_k(n)$ is the time dependent weighting (P samples) in the k-th segment (assumed here to be symmetrical in k). In analogy to (1), the output signal $y(n)$ is now

$$y(n) = \sum_{k=-L}^{L} \alpha_k(n \bmod P)x(n-kP) \tag{5}$$

* This work was carried out at Bell Labs while on leave from the Technion.

To facilitate the mathematical analysis, we use a polyphase network representation [4] of the generalized comb filter, as illustrated in Fig. 2. According to this description, the m-th sample in each signal segment (m=0,1,...P-1) is directed to the m-th channel by the commutating switch. It is then filtered by the time-invariant polyphase filter $p_m(k)$, where (with $\alpha_k = \alpha_{-k}$)

$$p_m(k) = w_g(kP+m) = \alpha_k(m), \quad -L \le k \le L, \quad m=0,1,\dots,P-1 \quad (6)$$

For the usual comb filter all the P polyphase filters are identical and are given by $p_m(k) = a_k$. Using results presented in [4] - in the context of decimation and interpolation - a mathematically equivalent scheme is given by the signal flow graph in Fig. 3. In this scheme, the block denoted by $\downarrow P$ is a compressor [4], i.e. it performs a P:1 decimation of its input, whereas the block denoted by $\uparrow P$ is a 1:P expander [4] which inserts P-1 zeros between adjacent input samples. z^{-1} and z^{+1} correspond to a unit delay and a unit advance, respectively. From Fig. 3 we find

$$Y(z) = Z\{\, y(n) \,\} = \frac{1}{P}\sum_{s=0}^{P-1} X(ze^{-j2\pi s/P})G_s(z^P) \quad (7)$$

$$\text{where} \quad G_x(z^P) = \sum_{m=0}^{P-1} P_m(z^P)e^{-j2\pi ms/P}, \quad (8)$$

with $X(z)$ and $P_m(z)$ being the z-transform of $x(n)$ and $p_m(k)$, respectively. Note that for the usual comb filter $P_m(z^P)=A(z^P)$, and hence $G_s(z^P)=PA(z^P)\delta(s)$, where $\delta(s)$ is the unit impulse function, so as expected, $Y(z) = X(z)\,A(z^P)$.

We continue the analysis by explicitely expressing $P_m(z^P)$ in (8) in terms of $p_m(k)$, and evaluating $G_s(z^P)$ and $Y(z)$ at a discrete dense set of frequencies $z = e^{j2\pi r/P^2}$, $r=0,1,\dots,P^2-1$. The results are

$$G_s(e^{j2\pi r/P}) = \sum_{m=0}^{P-1}\sum_{k=-L}^{L} p_m(k)e^{-j2\pi(rk+ms)/P}, \quad (9)$$
$$s=0,1,\dots,P-1, \quad r=0,1,\dots,P^2-1.$$

$$Y(e^{j2\pi r/P^2}) = \frac{1}{P}\sum_{s=0}^{P-1} X(e^{j2\pi(r-sP)/P^2})G_s(e^{j2\pi r/P}). \quad (10)$$

The output signal frequency response is seen from (10) to consist of two major spectral contributions. One is the weighted input spectral component, corresponding to the term for s=0 in (10). The other is the sum of aliasing components corresponding to the terms for s=1,...,P-1. Since $G_s(e^{j2\pi r/P})$ is periodic in r with period P, it is sufficient to examine this function for r=0,1,...,P-1, which we denote by G(s,r). The frequency weighting of the original signal in each of the P signal bands (having each a bandwidth of $2\pi/P$) is given by G(0,r)/P, and is denoted the "inband" weighting. For $s\neq 0$, G(s,r)/P is the weight given to aliasing components into each band from signal components s bands away. Thus, the two-dimensional sequence G(s,r) characterizes the generalized comb filter. From (9) we see that G(s,r), s,r=0,1,...,P-1, is the two-dimensional discrete Fourier transform (2D-DFT) of the sequence $p(m,k) = p_m(k)$, m=0,1,...,P-1, $-L \le k \le L$, i.e., G(s,r) = 2D-DFT { p(m,k)} . For illustration, we show in Fig. 4 the frequency characteristics of the generalized comb filter in Fig. 1. Note that for a given value of s, G(s,r) is obtained by sampling the corresponding curve at the normalized frequency points $r2\pi/P^2$, r=0,1,...,P-1.

DESIGN CONSIDERATIONS

In this section we consider the design problem of generalized comb filters. We first derive several constraints which are based on commonly desired properties of such filters.

One basic requirement is to make the inband frequency weighting, specified by G(0,r)/P, equal to the frequency response of a comb filter with given coefficients a_k. From (9) one finds $G(0,r)/P = \text{DFT}\,\{\bar{\alpha}_k\}$, where $\bar{\alpha}_k$ is the average value of the window segment $\alpha_k(m)$, m=0,1,...,P-1. Therefore, the inband frequency weighting can be made equal to that of a given comb filter by letting

$$\bar{\alpha}_k = a_k, \quad -L \le k \le L. \quad (11)$$

Another requirement is that the aliasing from other bands should not harm the periodic signal harmonics. This corresponds to setting G(s,0) = 0, s=1,2,...,P-1, $(s\neq 0)$. For unity gain of the desired harmonics we require that G(0,0) = P. Thus, $G(s,0) = P\delta(s)$. Using this result in (9) with $p_m(k) = \alpha_k(m)$, we get the constraint

$$\sum_{k=-L}^{L} \alpha_k(m) = 1, \quad m=0,1,\dots,P-1. \quad (12)$$

which can be shown to include the constraint that the sum over k of $\bar{\alpha}_k$, $-L \le k \le L$ is 1. This is a commonly used design constraint for a_k as well [1].

A third constraint is due to a time domain consideration. To avoid discontinuities in the output signal at segment boundaries, due to speech nonstationarity or pitch measurement errors, we introduce the following continuity constraint

$$\alpha_k(0) = \alpha_k(P-1) = \begin{cases} 0 & k\neq 0 \\ 1 & k=0 \end{cases} \quad (13)$$

If this constraint is satisfied, the first sample in a given output segment and the last sample in the previous segment are adjacent samples in the input signal so that continuity is ensured.

The above three constraints can be satisfied by many window functions. To narrow down the possibilities and to simplify the design we first limit our attention to a class of windows which have the property that the time variation within the different window segments, $\alpha_k(n)$, is governed by a single function, q(n), as follows

$$\alpha_k(n) = a_k q(n), \quad \begin{matrix} k = -L,\dots,L, & k\neq 0 \\ n = 0,1,\dots,P-1 \end{matrix}, \quad (14)$$

and $\alpha_0(n)$ is set so that (12) is satisfied. The constraint in (11) is satisfied if we require that the average value of q(n), n=0,1,...,P-1, is 1. We then find $\alpha_0(n) = 1 + (a_0-1)q(n)$. Finally, the continuity constraint in (13) holds if q(0) = q(P-1) = 0.

The frequency weighting function G(s,r) is now given by $G(s,r) = P\delta(s) + [A(r)-1]Q(s)$ where $A(r) = A(e^{j2\pi r/P})$ is found by computing the DFT of $\{a_k\}$ (since usually $P>2L+1$, zeros can be appended to the sequence $\{a_k\}$ before computing a P-point DFT), and $Q(s)$ is the DFT of $q(n)$. Note that because the coefficients a_k sum up to 1, $A(0)=1$, and because the average value of $q(n)$ is 1, $Q(0)=P$, resulting in $G(0,r) = P\,A(r)$; $G(s,0) = P\delta(s)$, as required earlier.

The design process consists therefore of selecting the coefficients a_k (as in a comb filter design [1]), and a modulation function $q(n)$ to control the aliasing. Some specific modulation functions and their aliasing characteristics are presented below.

Let $q(n)$ be a symmetric sequence, i.e., $q(n)=q(P-1-n)$, $n=0,1,\ldots,P-1$, and consider the following parametric representation

$$q(n) = c(\mu)[1-w(n)^{\mu}] \quad , \quad 0\le n\le M \qquad (15)$$

In (15) μ is a parameter; $c(\mu)$ is a constant chosen so that the average value of $q(n)$ is 1; $w(n)$ is a prototype window function satisfying $w(0)=1$; and $M=(P-1)/2$ if P is odd, $M=(P/2)-1$ if P is even. Note that for $M<n\le P-1$: $q(n)=q(P-1-n)$. We have particularly examined the following two choices for $w(n)$: a triangular window $w_T(n)$ and a Hanning window $w_H(n)$, namely

$$w_T(n) = 1 - n/M, \quad w_H(n) = \tfrac{1}{2}[1+\cos(\pi n/M)], \quad 0\le n\le M \qquad (16)$$

Fig. 5 shows $q(n)$ for $w(n)=w_T(n)$ and different values of μ. Because of lack of space we do not show here the corresponding $Q(s)$. It is found that by varying the parameter μ, one can control the relative aliasing from different bands.

It is noted th for $\mu=1$ and $w(n) = w_T(n)$, $q(n)$ becomes a triangular window ($n=0,1,\ldots,P-1$), and if $w(n) = w_H(n)$, it becomes a Hanning window. It is also noted that the generalized comb filter shown in Fig. 1 corresponds to using $q(n)$ which is obtained from (15) with $\mu=2$ and $w(n) = w_T(n)$.

We have found in simulations that the amount of aliasing obtained with the above families of modulating functions is insufficient in certain applications (e.g. a competing speaker). To further increase the aliasing we have modified $\alpha_k(n)$ in (14) to the following form

$$\alpha_k(n) = [a_k + v(n)]q(n), \quad k\ne 0 \qquad (17)$$

and $\alpha_0(n)$ is set to satisfy (12). To satisfy the earlier discussed three constraints, the above constraints on $q(n)$ must hold, and in addition the average value of $v(n)q(n)$ must be 0. We then find $\alpha_0(n) = 1 + [a_0 - 2Lv(n)]q(n)$. The sequence $v(n)$ is used here to increase the aliasing. We found it useful to derive $v(n)$ from a zero mean, unit variance, random sequence $e(n)$, in the following way

$$v(n) = G[e(n) + \varepsilon] \qquad (18)$$

where G is a gain factor and ε is a bias constant chosen to satisfy the above requirement on $q(n)v(n)$. Fig. 6 shows a three-segment generalized comb filter constructed according to (17).

SIMULATION RESULTS

The GCF technique has already proven itself useful in the perceptual reduction of frame-rate noise in an adaptive transform speech coder [2]. The results of a subjective test, reported in [2], have shown an improvement which is equivalent to an increase in coder rate of 2.4 to 3 Kb/s for speech coded at 7.2 to 12 Kb/s.

Following the above mathematical analysis we have further examined the capability of the GCF technique to break up the spectral structure of interfering signals. We have also done some preliminary work in filtering and perceptually reducing speech of a competing speaker.

To illustrate the structure destruction caused by the GCF technique to an interfering wide band periodic signal, we show in Fig. 7 computer simulation results. Part a shows the spectrum of the interfering signal. Part b shows the spectrum of the processed signal by a usual comb filter (with $a_{-1}=a_0=a_1=0.33$) matched to a pitch period of 51 samples (the interference period is 136 samples). An inability of the usual comb filter to break the spectral structure of the interference signal is observed. Part c shows the result of applying a generalized comb filter of the form shown in Fig. 6 with the following parameters $G = 0.22$, $\varepsilon = -0.12$ and $\bar{\alpha}_k = a_k = 0.33$, as before. The effect on the interference structure is evident.

For further illustration, the above interference signal was added to a synthetic vowel at 7dB SNR. The spectrum of the synthetic vowel is shown in Fig. 8 Part a. The combined noisy signal is shown in Part b. The result of filtering with a fixed comb filter (with weights $a_{-1}=a_1= 0.25$; $a_0= 0.50$) is shown in Part c. When we applied the generalized comb filter we obtained the spectrum shown in Part d. While the structure of the interference is indeed destroyed, the overall noise reduction (SNR) is seen to be less than with the fixed comb filter because of the aliasing components. To improve the performance we found it useful to apply both filters (the fixed and generalized comb filters) in cascade thus achieving both the filtering and structure destruction of the interfering signal, as shown in Part e.

The last approach was also applied to the filtering and perceptual reduction of the speech of a competing speaker. In a preliminary experiment which involved combining the speech of a male speaker with an interfering female speaker (at 6.6dB SNR) we found that the above combination of a fixed comb filter and a generalized comb filter was more effective in filtering the competing speaker than a cascade of two fixed comb filters. This is a potential application which needs further examination and subjective evaluation.

REFERENCES

[1] J.S. Lim et al., IEEE Trans. ASSP-26, pp.354-358 Aug. 1978.

[2] R.V. Cox and D. Malah, ICASSP-81, pp.1089-1092, 1981.

[3] Y.M. Perlmutter et al., ICASSP-77, pp.212-215, 1977.

[4] R.E. Crochiere and L.R. Rabiner, Proc. IEEE, Vol. 69, pp.300-331, March 1981.

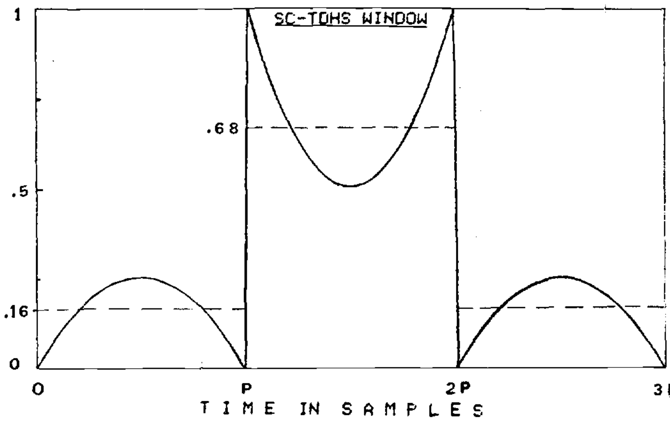

Fig. 1 Generalized comb filter window used in [2].

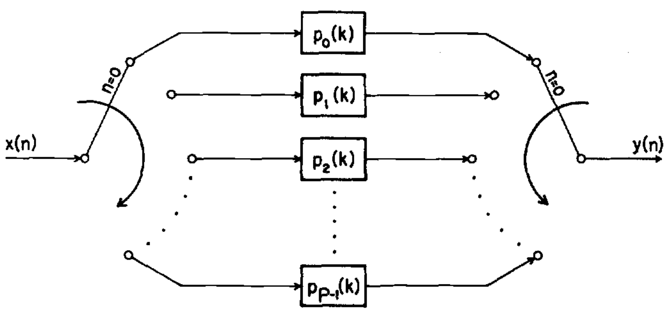

Fig. 2 Polyphase network representation of GCF

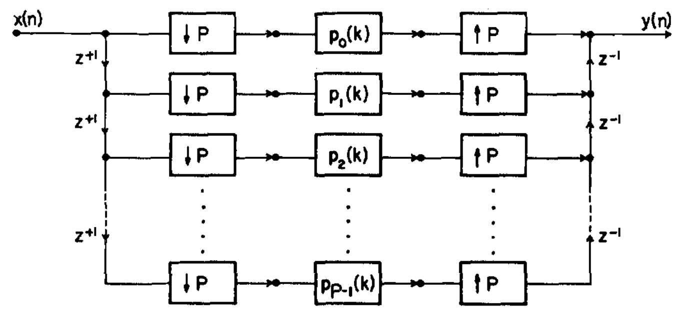

Fig. 3 Signal flow graph of GCF.

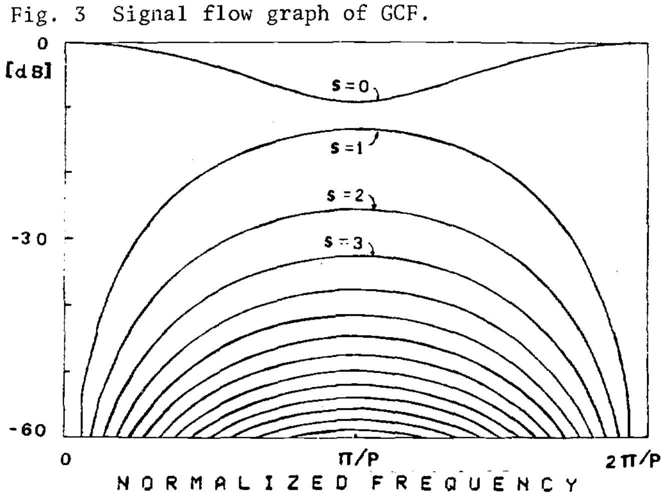

Fig. 4 Frequency weighting function G(s,r) of the generalized comb filter in Fig. 1.

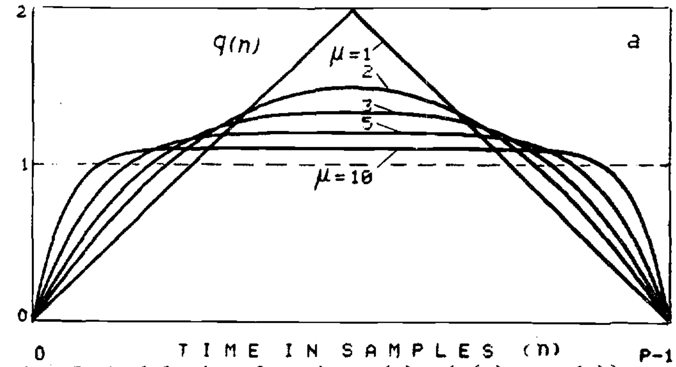

Fig. 5 Modulating function q(n), ($w(n) = w_T(n)$).

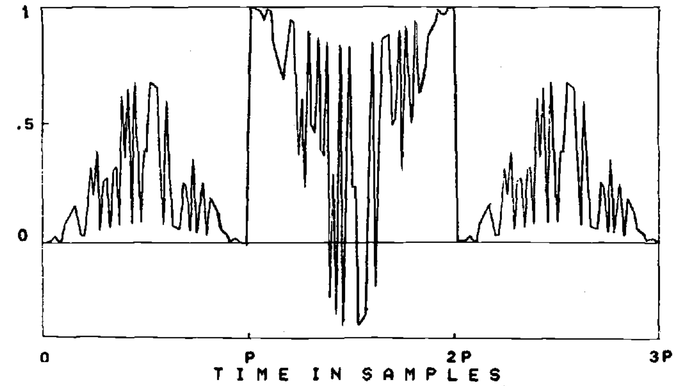

Fig. 6 Generalized comb filter window designed from (17) with v(n) according to (18).

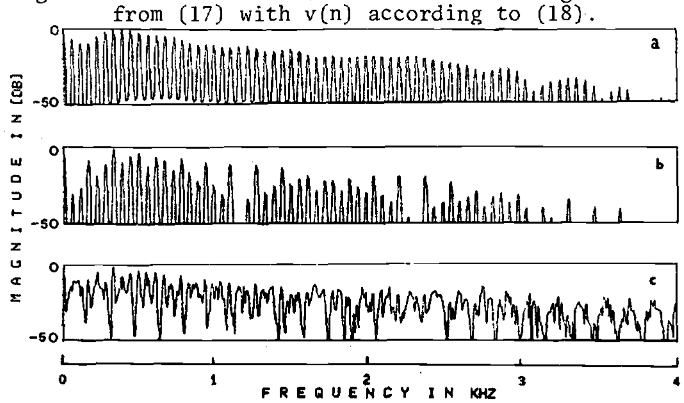

Fig. 7 Comb filtering of periodic interference.

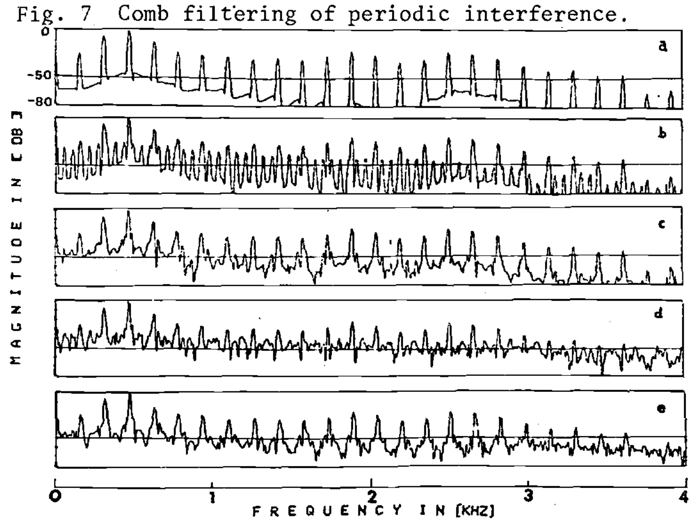

Fig. 8 Processing a synthetic vowel with periodic interference.