THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 54, No. 3, March 1975
Printed in U.S.A.

FIR Digital Filter Banks for Speech Analysis

By R. W. SCHAFER, L. R. RABINER, and O. HERRMANN

(Manuscript received August 9, 1974)

In using filter banks for processing speech signals, it is often important that the sum of the individual frequency responses of the bandpass filters (composite response) be flat with linear phase. This paper presents a technique for achieving flat composite response using linear-phase FIR digital filters. The design method is based on some special properties of FIR filters designed by the windowing method. Excellent response characteristics can be achieved with complete flexibility in choosing the center frequencies and bandwidths of the individual filters.

I. INTRODUCTION

Filter banks are used to perform short-time spectrum analysis in a variety of speech processing systems.[1–4] Typically, a set of bandpass filters is designed so that a desired portion of the speech band is entirely covered by the combined passbands of the filters composing the filter bank. The outputs of the bandpass filters therefore are considered to be a time-varying spectrum representation of the speech signal. If special care is taken in the design of the bandpass filters, it is possible to reconstruct a very good approximation to the input speech by simply adding together the outputs of the bandpass filters.[5] This is the basic principle of a variety of vocoder systems.

Since the bandpass filters are linear systems, we can characterize the behavior of such filter banks by considering the composite frequency response when all the outputs are added together. Since, ideally, the output should be equal to the input, then we desire that the composite frequency response have constant magnitude and linear phase in the desired band of frequencies. This criterion, together with specifications on the desired bandwidths of the individual frequency channels, forms a meaningful basis for the design of filter banks for speech analysis.

An earlier paper[5] showed that careful attention to the relative phases between channels is important in achieving a flat composite frequency response. That paper, which was concerned primarily with filter banks composed of infinite impulse response (IIR) digital filters, described a method of obtaining flat composite frequency response by a relatively simple adjustment of the relative phases of the channels. This method was later applied to the design of a speech analysis/synthesis scheme in which finite impulse response (FIR) digital filters were used.[3] Using this method, excellent overall response can be obtained for both IIR and FIR digital filters in filter banks in which the center frequencies are *uniformly* spaced. However, the method is not easily extended to nonuniformly spaced filter banks.

In the present paper, we describe a different approach that is not limited to the design of uniformly spaced filter banks. The method exploits some special properties of linear-phase FIR filters and thus cannot be applied very successfully to the design of IIR filter banks. We first discuss the basic design principles, and then show some design examples. We conclude with a discussion of some computational considerations of FIR digital filter banks.

II. DESIGN METHOD

FIR digital filters are attractive for design of speech filter banks for several reasons. First, such filters can be designed to have precisely linear phase simply by imposing the constraint

$$h(n) = h(N-1-n) \qquad 0 \leqq n \leqq N-1 \tag{1}$$

(on each individual filter band*), where $h(n)$ is the impulse response of the filter and N is its length in samples. This means that the criterion of linear phase for the composite filter bank response is trivially met if the individual filters have identical linear-phase characteristics. Therefore, it is possible to focus attention on achieving arbitrary frequency selective properties for the individual filters and on obtaining the desired flat response for the composite filter bank. The second great advantage of FIR filters is that a variety of design methods exist ranging from the straightforward windowing method[6,7] to iterative approximation methods that allow great flexibility in realizing complicated design specifications.[8]

2.1 FIR bandpass filters

The bandpass filters that we shall consider have impulse responses of the form

$$\begin{aligned} h_k(n) &= h_{lk}(n) \cos(\omega_{ck} nT) \qquad & 0 \leqq n \leqq N-1 \\ &= 0, & \text{otherwise,} \end{aligned} \tag{2}$$

where $h_{lk}(n)$ is the impulse response of the kth linear-phase low-pass

* It is assumed, for simplicity, that the impulse response of each bandpass filter is of duration N samples, although it is trivial to remove this restriction by adding appropriate delays for each channel.

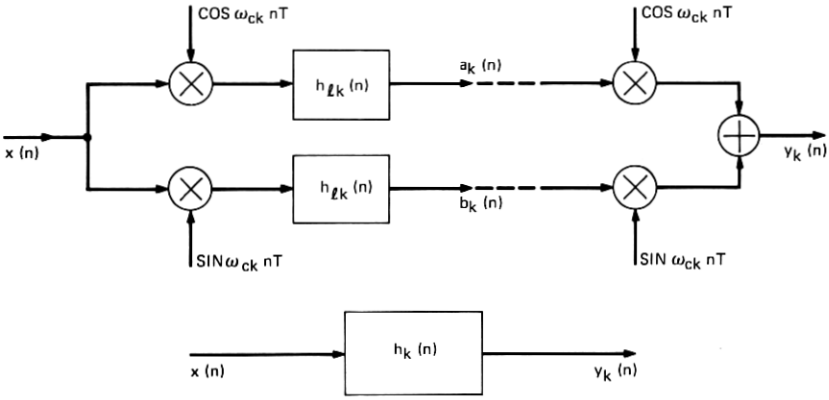


Fig. 1—Implementation of a typical bandpass channel.

filter. This particular form for the impulse response is motivated by the fact that, in some vocoder applications,[2,4] each bandpass channel is implemented as shown in Fig. 1. The overall impulse response of the system of Fig. 1 from input $x(n)$ to output $y(n)$ is easily shown to be given by eq. (2).

The spacing of the individual channels of the filter bank is determined by the choice of the set of center frequencies, ω_{ck}, which is in turn determined by the desired frequency resolution of the filter bank. The frequency selectivity of each channel is determined by the frequency response characteristics of the prototype low-pass filters $h_{lk}(n)$. Since phase considerations can be simply avoided by designing all the bandpass filters to have the same linear phase, we can focus our attention entirely on designing a set of prototype low-pass filters that have the desired individual frequency selective properties and that give the flattest amplitude response for the composite set of bandpass filters.

2.2 Low-pass filter design

The window design method appears to have a number of advantages for design of the prototype low-pass FIR filters. This method is depicted in Fig. 2. First, a desired ideal low-pass filter of the form

$$\begin{aligned} H_{dk}(e^{j\omega T}) &= e^{-j\omega n_0 T} && |\omega| \leqq \omega_{pk} \\ &= 0, && \text{otherwise,} \end{aligned} \tag{3}$$

is defined by choosing the cutoff frequency ω_{pk}. Note that, for simplicity, we have omitted in the figure the linear phase term $\exp(-j\omega n_0 T)$ corresponding to a delay of n_0 samples. The value of n_0 required is $n_0 = (N - 1)/2$. This means that, if N is even, the

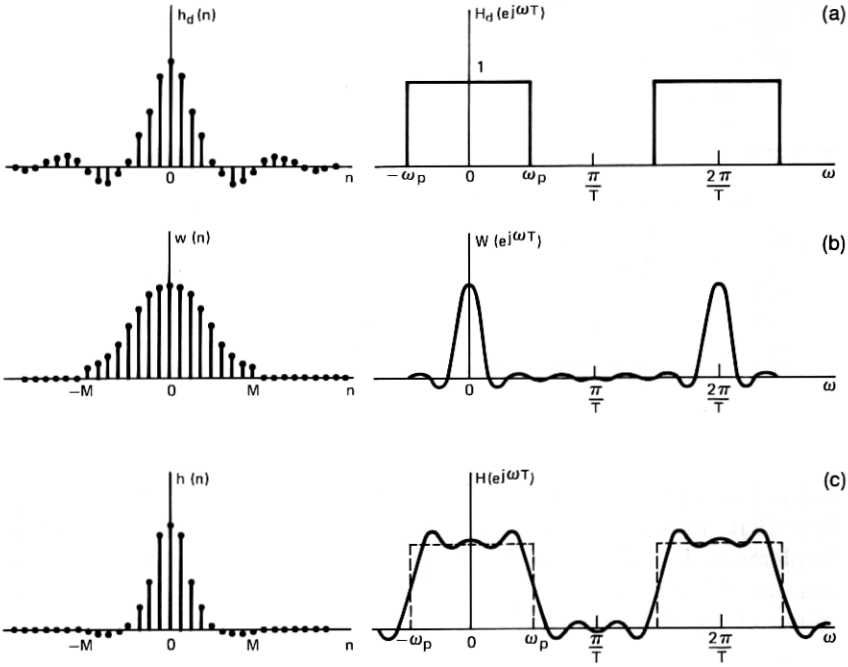


Fig. 2—Windowing technique for a low-pass design.

delay corresponds to a noninteger number of samples. The ideal impulse response for the kth channel is, therefore,

$$h_{dk}(n) = \frac{1}{2\pi}\int_{-\omega_{pk}}^{\omega_{pk}} e^{-j\omega n_0 T}\, e^{j\omega nT} d\omega = \frac{\sin\left[\omega_{pk}(nT - n_0T)\right]}{\pi(n - n_0)}. \tag{4}$$

Of course, this impulse response is infinite in extent and must be truncated to obtain an FIR filter. This is done by defining

$$h_{lk}(n) = w(n - n_0)h_{dk}(n), \tag{5}$$

where $w(n)$ is a window function and $h_{lk}(n)$ is the impulse response of the kth prototype low-pass filter. The length of the window, denoted by N, can be either an even integer ($N = 2M$) or an odd integer ($N = 2M + 1$). Figure 2 shows the case when N is odd.

The result of multiplying the ideal low-pass impulse response by the window corresponds to a convolution in the frequency domain of the ideal frequency response and the Fourier transform, $W(e^{j\omega T})$, of the window; i.e.,

$$H_k(e^{j\omega T}) = \frac{T}{2\pi}\int_{-\pi/T}^{\pi/T} H_{dk}(e^{j\theta T})W(e^{j(\omega-\theta)T})d\theta. \tag{6}$$

The result of this convolution is depicted in Fig. 2. It can be seen that the main effects are the introduction of a smooth transition between the passband and the stopband and the introduction of ripples in the passband and stopband regions. The properties of this approximation are depicted in Fig. 3. If ω_p is larger than the width of the "main lobe" of $W(e^{j\omega T})$, then the following set of properties are generally true:

(*i*) The transition region, $\Delta\omega$, is inversely proportional to N.
(*ii*) The function $H(e^{j\omega T})$ is very nearly antisymmetric about the point $(\omega_p, 0.5)$.
(*iii*) The peak approximation errors in the passband and stopband are very nearly equal.
(*iv*) The approximation error is greatest in the vicinity of ω_p, and it decreases for values of ω away from ω_p.

The above properties of the windowing design method are true of all the commonly used windows. However, Kaiser has proposed a family of window functions that are very flexible and nearly optimum for filter design purposes.[6] Specifically, the Kaiser window is

$$\begin{aligned} w(n) &= \frac{I_0[\alpha\sqrt{1-(n/n_0)^2}]}{I_0(\alpha)} \qquad & |n| \leqq n_0 \\ &= 0, & \text{otherwise}, \end{aligned} \tag{7}$$

where $n_0 = (N-1)/2$ and $I_0[\cdot]$ is the modified zeroth-order Bessel function of the first kind. By adjusting the parameter α, one can trade off between transition width and peak approximation error. Furthermore, Kaiser[7] has formalized the window design procedure by giving

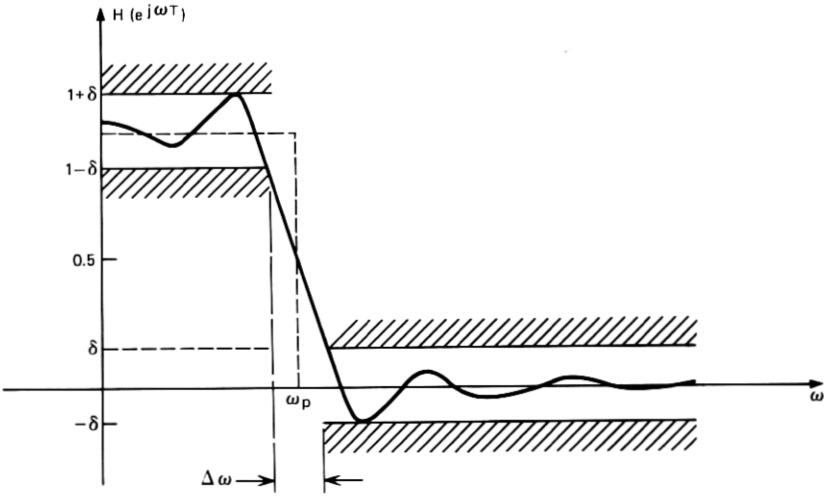


Fig. 3—Resulting low-pass design from windowing.

the empirical design formula

$$N = \frac{-20 \log_{10} \delta - 7.95}{14.36 \Delta f} + 1, \tag{8a}$$

where N is the filter order, δ is the peak approximation error, and Δf is the normalized transition width

$$\Delta f = \frac{\Delta \omega T}{2\pi}. \tag{8b}$$

To use this formula, we fix δ and Δf at values that provide the desired frequency selectivity. Then eq. (8a) can be used to compute N, and the parameter α can be computed from the equation[7]

$$\begin{aligned} \alpha &= 0.1102(-20 \log_{10} \delta - 8.7), && -20 \log_{10} \delta > 50 \\ &= 0.5842(-20 \log_{10} \delta - 21)^{0.4} \\ &\quad + 0.07886(-20 \log_{10} \delta - 21), && 21 < -20 \log_{10} \delta < 50. \end{aligned} \tag{9}$$

In the present application of this design method, the choice of δ and Δf depends upon the specifications of the bandpass filters that constitute the filter bank.

2.3 Filter bank design

To design a filter bank using FIR filters, we must first determine the range of frequencies to be covered by the composite response. Let us assume that these are denoted $\omega_{\min}$ and $\omega_{\max}$, where $\omega_{\max} \leqq \pi/T$. Now, if there are a total of N_f filters, we must choose the bandwidths and center frequencies so that the entire range of frequencies $\omega_{\min} \leqq \omega \leqq \omega_{\max}$ is covered. This is depicted in Fig. 4 for the case $N_f = 3$. This figure shows the ideal responses for each bandpass filter;

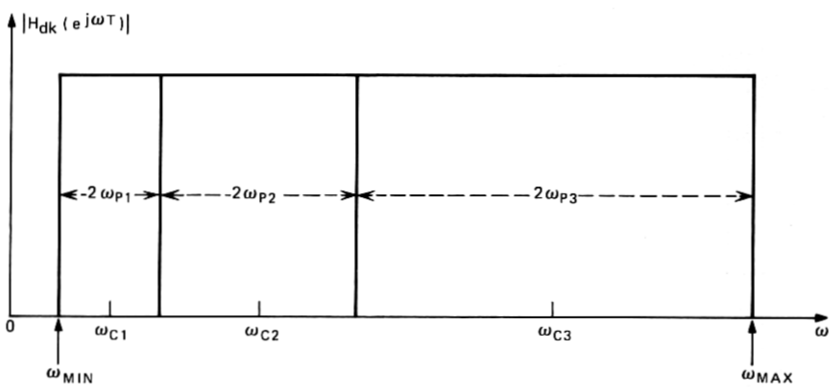


Fig. 4—A typical nonuniform filter bank.

i.e., as would be obtained if windowing were not required. In general, it is clear that

$$\omega_{\max} - \omega_{\min} = \sum_{k=1}^{N_f} 2\omega_{pk} \tag{10}$$

and

$$\begin{aligned} \omega_{ck} &= \omega_{\min} + \sum_{m=1}^{k-1} 2\omega_{pm} + \omega_{pk} \qquad && k \geqq 2 \\ &= \omega_{\min} + \omega_{p1} && k = 1. \end{aligned} \tag{11}$$

If all the filters have the same bandwidth, i.e., $\omega_{pk} = \omega_0$, then it is easily seen that

$$\omega_0 = \frac{\omega_{\max} - \omega_{\min}}{2N_f}. \tag{12}$$

Alternatively, if the bandwidths are to increase exponentially; e.g., $\omega_{pk} = 2^{k-1}\omega_0$, then

$$\omega_0 = \frac{\omega_{\max} - \omega_{\min}}{2(2^{N_f} - 1)}. \tag{13}$$

The center frequencies can be found in either case by using eq. (11).

The choice of peak approximation error depends upon how much stopband attenuation is deemed necessary in a given application. Typical values of $-20 \log_{10} \delta$ would most likely be between 40 and 60 dB. Using eq. (9), the appropriate value of α can be computed. Finally, the normalized transition width Δf must be fixed to compute N from eq. (8a). Again, the choice of $\Delta\omega$ (or Δf) is governed by consideration of the desired frequency selectivity for the individual filters. Clearly, the transition width $\Delta\omega_k$ should not be more than $2\omega_{pk}$.

In the filter bank context, we shall require that $\Delta\omega$ be the same for all filters so that we can take advantage of property (*ii*) of Section 2.2. That is, if all the filters have identical transition regions and, furthermore, if these transitions are antisymmetric about the crossover points, then we can expect that the sum of the frequency responses will be very close to unity. This is illustrated in Section III.

III. DESIGN EXAMPLES

In this section, we illustrate the use of the principles established in Section II with examples of both uniform and nonuniform filter banks. For all the examples, the sampling rate is assumed to be 9.6 kHz.

Example 1

Suppose that we wish to design a bank of 15 equally spaced filters that covers the range 200 to 3200 Hz. Then, using eq. (12), we find that the cutoff frequency for all the low-pass filters* is

$$f_0 = \frac{\omega_0}{2\pi} = 100 \text{ Hz}.$$

Using eq. (11), the center frequencies are

$$f_{ck} = \frac{\omega_{ck}}{2\pi} = 100(2k+1) \text{ Hz} \qquad k = 1, 2, \cdots, 15.$$

If we assume that 60-dB attenuation is required outside the transition regions of each channel, we find from eq. (9) that $\alpha = 5.65326$. Since the cutoff frequency is 100 Hz for all the prototype low-pass filters, the widest transition band that is reasonable is 200 Hz. Using this value and $-20 \log_{10} \delta = 60$ in eq. (8a), we obtain $N = 175$ as the lowest reasonable value for N. Note that, if lower attenuation is acceptable, then N can be smaller for the same Δf.

The filter bank designed with the above parameters is shown in Fig. 5. Figure 5a shows the individual bandpass filters. Note how the fall-off in the upper transition band of a given filter complements the ascent of the next filter. Also note that adjacent channels cross at an amplitude value of 0.5. Figure 5b shows the composite response of the filter bank. It is clear that the filters merge together very well at the edges of the frequency bands. Indeed, the deviation from unity is less than or equal to the peak approximation error, $\delta = 0.001$, that was used in designing the prototype low-pass filters.

Example 2

A nonuniform spacing of the filters is often used to exploit the ear's decreasing frequency resolution with increasing frequency. Suppose that we wish to cover the same range 200 to 3200 Hz as in Example 1, but we wish to use only four octave band filters. That is, each successive filter will have a bandwidth twice the bandwidth of the previous filter. Using eq. (13), we find that the lowest frequency channel has cutoff frequency

$$f_0 = \frac{\omega_0}{2\pi} = \frac{3200 - 200}{2(2^4 - 1)} = 100 \text{ Hz}.$$

In general, the cutoff frequencies of the prototype low-pass filters are

$$f_{pk} = \frac{\omega_{pk}}{2\pi} = 2^{k-1} f_0 \qquad k = 1, 2, 3, 4,$$

---

* For the actual low-pass filter, the response will be approximately 0.5 at $\omega = \omega_p$, the cutoff frequency of the ideal low-pass filter.

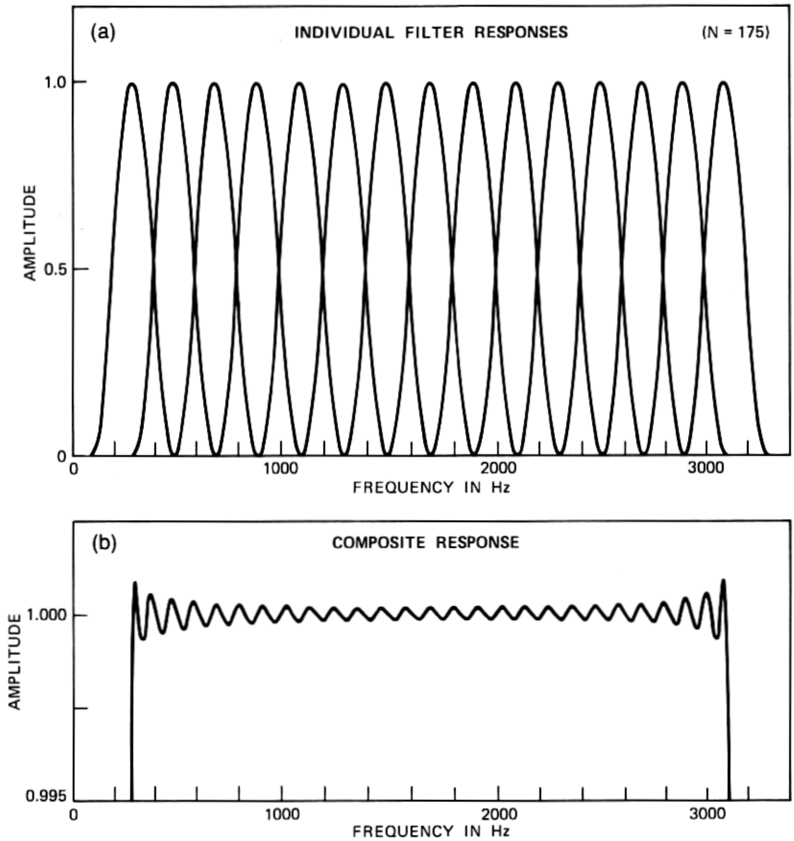


Fig. 5—Individual and composite frequency responses of a bank of 15 uniform bandpass filters for $N = 175$.

or the bandwidths of the bandpass filters are 200, 400, 800, and 1600 Hz, respectively. The center frequencies are found, from eq. (11), to be 300, 600, 1200, and 2400 Hz, respectively. Again requiring 60-dB attenuation, we note that the narrowest bandwidth is 100 Hz, so that the smallest reasonable transition width is 200 Hz. This leads again to a minimum value of $N = 175$. The filter bank corresponding to these design parameters is shown in Fig. 6. In Fig. 6a, again note the relationship between the ascending and descending transitions between adjacent filters. Particularly note that, since N and α are the same for each of the prototype low-pass designs, the shape of the curves in the transition region is independent of the bandwidth. Figure 6b shows the composite response where the deviation from unity is again less than 0.001.

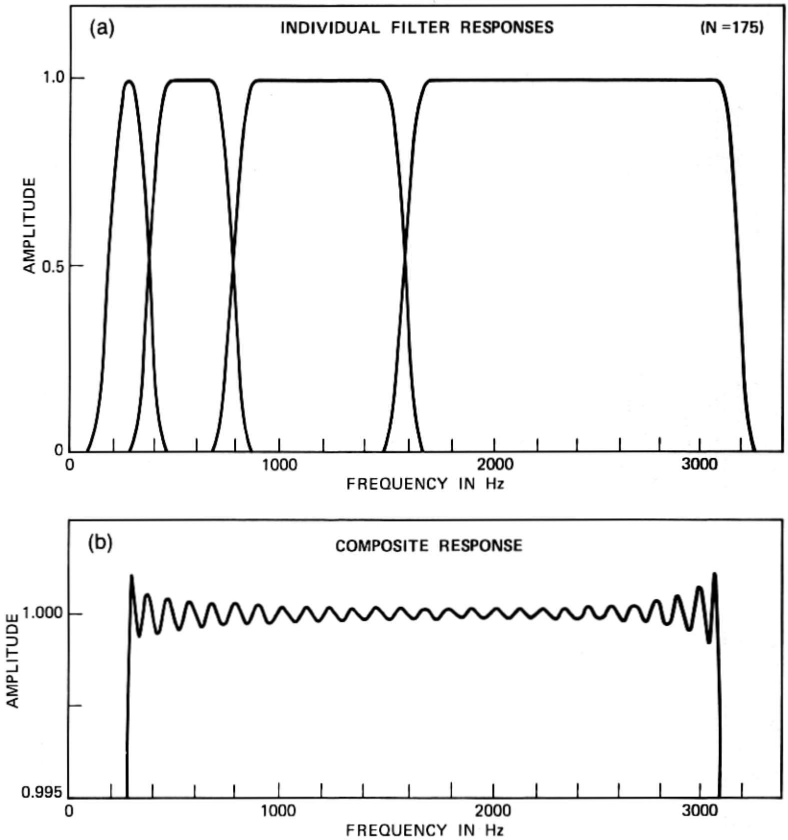


Fig. 6—Individual and composite frequency responses of a bank of 4 nonuniform bandpass filters for $N = 175$.

It is interesting to note that the composite frequency response of the filter bank is *independent* of the number and distribution of the individual filters, so long as the same window is used to design all the individual filters in the bank. This result can be verified by writing the overall frequency response of the filter bank, $H(e^{j\omega T})$, as

$$H(e^{j\omega T}) = \sum_{k=1}^{N_f} H_k(e^{j\omega T}), \tag{14}$$

which, from eq. (6), can be written as

$$H(e^{j\omega T}) = \sum_{k=1}^{N_f} \frac{T}{2\pi} \int_{-\pi/T}^{\pi/T} H_{dk}(e^{j\theta T}) W(e^{j(\omega-\theta)T}) d\theta. \tag{15}$$

Interchanging the order of summation and integration, eq. (15)

can be written as:

$$H(e^{j\omega T}) = \frac{T}{2\pi}\int_{-\pi/T}^{\pi/T}\left[\sum_{k=1}^{N_f} H_{dk}(e^{j\theta T})\right]W(e^{j(\omega-\theta)T})d\theta \tag{16}$$

$$= W(e^{j\omega T})⊛H_T(e^{j\omega T}), \tag{17}$$

where

$$H_T(e^{j\omega T}) = \sum_{k=1}^{N_f} H_{dk}(e^{j\omega T}). \tag{18}$$

Equations (17) and (18) show that the overall frequency response of the filter bank is the circular convolution of the frequency response of the window with the frequency response of the combined *ideal* bandpass filters. As seen in Fig. 4, the combined ideal frequency response of the bandpass filters is an ideal bandpass filter from $\omega = \omega_{\min}$ to $\omega = \omega_{\max}$, independent of the number and distribution of the individual filters. Thus, the composite filter bank frequency responses for the examples in Figs. 5 and 6 are identical because the same window was used in both cases and the filters spanned the identical frequency ranges.

Example 3

Suppose that all the parameters remain the same as in Example 2 except that we require narrower transition regions. This means that a larger value of N is required. In fact, Eq. (8a) shows that N and Δf are roughly inversely proportional. Figure 7 shows the filter bands corresponding to the parameters of Example 2 except that $N = 301$ and $\Delta f = 0.012082$ (transition width is 116 Hz). The sharper transitions are apparent in Fig. 7a, and Fig. 7b shows that the composite response remains very flat.

Example 4

We have assumed throughout that the transition width was less than twice the smallest low-pass cutoff frequency. In our examples, this constraint required that N be at least 175. The result of reducing N below this value is illustrated in Fig. 8. In this case, all the parameters were the same as in Examples 2 and 3, except in the case of $N = 101$ and $\Delta f = 0.0362465$. The transition width is 348 Hz, which is much greater than twice the cutoff frequency of the first low-pass filter. This is clearly in evidence in Fig. 8a. It is clear that reasonable filters are obtained for the wider bandwidth filters; however, the lowest filter does not attain unity response anywhere in its passband.

The preceding examples make it abundantly clear that, for sufficiently long impulse responses, the composite filter-bank response can

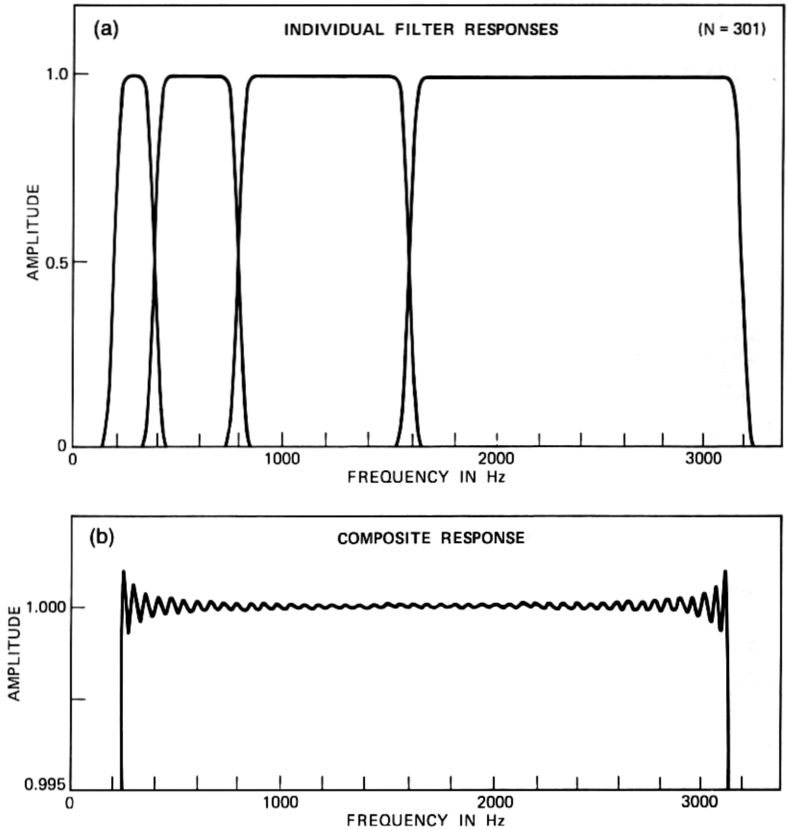
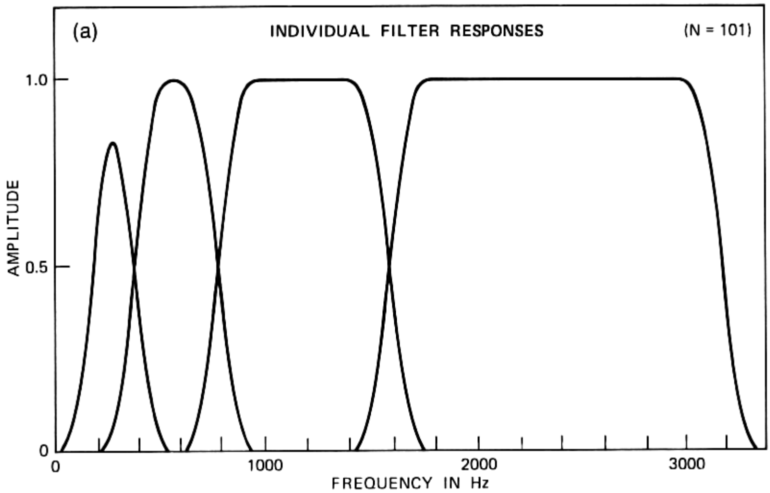


Fig. 7—Individual and composite frequency responses of a bank of 4 nonuniform bandpass filters for $N = 301$.

be very flat. In Ref. 5, where design techniques for IIR filter banks were discussed, the best results achieved for the composite response were approximately 1-dB peak-to-peak ripple for uniform bandwidths and about 2.5-dB peak-to-peak ripple for nonuniform bandwidths. This is in contrast to the results of the examples of this section, where the peak-to-peak ripple in the composite response was about 0.0274 dB for all the filter banks independent of how the bandwidths were chosen. This, together with the precise linear phase that is easily achieved, makes the FIR filter banks superior to what can be achieved for IIR filter banks. The price that is paid for this is that rather large values of N are required to achieve sharp transitions. However, the values of N used in the previous examples are certainly not unreasonable if the filters are implemented by FFT convolution methods or in special-purpose hardware.



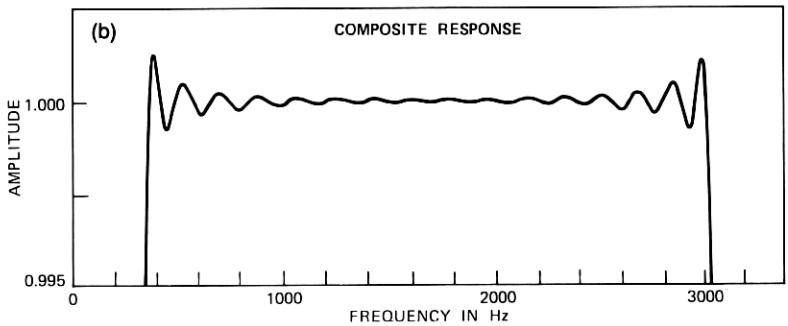


Fig. 8—Individual and composite frequency responses of a bank of 4 nonuniform bandpass filters for $N = 101$.

IV. SUMMARY

We have discussed a design method for filter banks composed of FIR digital filters. The method exploits the linear-phase properties obtainable for such filters, as well as the symmetry of the transition region that results from the windowing method of design. We summarized this method of design for the Kaiser window and illustrated the filter-bank design method with several examples. These examples show that the proposed design method has a great deal of flexibility and that excellent response characteristics can be achieved.

REFERENCES

1. J. L. Flanagan, *Speech Analysis, Synthesis and Perception,* Second Ed., New York: Springer-Verlag, 1972.
2. J. L. Flanagan and R. M. Golden, "Phase Vocoder," B.S.T.J., *45,* No. 9 (November 1966), pp. 1493–1509.

3. R. W. Schafer and L. R. Rabiner, "Design and Simulation of a Speech Analysis-Synthesis System Based on Short-Time Fourier Analysis," IEEE Trans. Audio and Electroacoustics, *AU-21*, No. 3 (June 1973), pp. 165–174.
4. J. L. Flanagan and R. C. Lummis, "Signal Processing to Reduce Multipath Distortion in Small Rooms," J. Acoust. Soc. Am., *47*, part 1, June 1970, pp. 1475–1481.
5. R. W. Schafer and L. R. Rabiner, "Design of Digital Filter Banks for Speech Analysis," B.S.T.J., *50*, No. 10 (December 1971), pp. 3097–3115.
6. J. F. Kaiser, "Digital Filters," *System Analysis by Digital Computer*, Ch. 7, F. F. Kuo and J. F. Kaiser, eds., New York: John Wiley, 1966, pp. 228–243.
7. J. F. Kaiser, "Nonrecursive Digital Filter Design using the I_0-Sinh Window Function," Proc. 1974 IEEE Int. Symp. on Cir. and Syst., San Francisco, 1974, pp. 20–23.
8. L. R. Rabiner, J. H. McClellan, and T. W. Parks, "FIR Digital Filter Design Techniques using Weighted Chebyshev Approximation," to appear in Proc. IEEE, April 1975.