THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 50, No. 10, December, 1971
Printed in U.S.A.

Design of Digital Filter Banks for Speech Analysis

By R. W. SCHAFER and L. R. RABINER

(Manuscript received June 22, 1971)

A bank of bandpass filters is often used in performing short-time spectrum analysis of speech signals. This paper is concerned with the analysis and design of digital filter banks composed of equally spaced bandpass filters. It is shown that significant improvement in the composite filter bank response can be achieved by proper choice of the relative phases of the bandpass filters. The results are extended to more general filter bank configurations.

I. INTRODUCTION

Many speech processing systems are based on the concept of short-time spectrum analysis.[1,2] Spectrum analyzers for such systems often consist of a set of bandpass filters whose combined passbands cover a desired frequency range. Although continuous-time filters have traditionally been used in filter banks for speech analysis, hardware realizations of digital filters are now available,[3] and the advantages which digital filters offer should be exploited in filter bank design. These advantages include: flexibility of design of the individual bandpass filters, precision of realization, stability of digital hardware, and the efficiency of realization of the filter bank afforded by the possibility of multiplexing the digital hardware. Thus it is important to consider design techniques for filter banks composed of digital bandpass filters.

To focus on the basic concepts in filter bank design, it is useful to define an *ideal filter bank* spectrum analyzer. Figure 1 depicts such a filter bank composed of digital filters whose impulse responses are denoted by $h_k(nT)$, $k = 0, 1, \cdots, M$, where $1/T$ is the sampling frequency of the input signal.* Such a filter bank constitutes an ideal spectrum analyzer if the input $x(nT)$ (with possibly further band limiting) can be synthesized exactly (within some fixed delay) by a linear

* The filter $h_0(nT)$ is a lowpass filter which is included for completeness although this band is usually not analyzed in practical speech analysis systems.

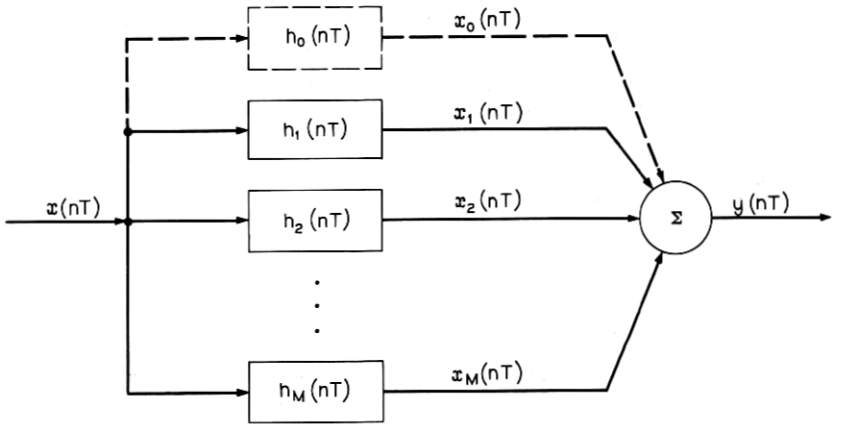

Fig. 1—Bank of digital bandpass filters.

combination of the bandpass filter outputs $x_k(nT)$. An example of such a system would be one in which the filters are ideal rectangular bandpass filters with the same constant gain and linear phase in their passbands and zero gain outside. If the filter bandwidths are such that the frequency range

$$\frac{-\pi}{T} \leqq \omega \leqq \frac{\pi}{T}$$

is completely covered without overlap, then the input can be synthesized exactly by adding together the outputs of the bandpass filters.

The essential characteristics of the ideal spectrum analyzer are that the frequency response of the combined outputs must exhibit a flat magnitude response and a linear phase, and therefore the combined impulse response must be a delayed digital impulse (unit sample). Causal digital filters (filters whose impulse responses are zero for $n < 0$) cannot have the desired ideal gain characteristics and may not have linear phase.* Therefore a filter bank composed of such filters cannot achieve the ideal characteristics of flat magnitude response and linear phase. In this paper we will describe an approach to the design of filter banks that approximate the ideal spectrum analyzer. First we present a detailed analysis of a filter bank configuration in which equally spaced, equal-bandwidth digital filters are used. This analysis suggests a technique for optimizing the filter bank characteristics and

* Finite duration impulse response digital filters can have precisely linear phase.

also suggests how the results can be used where nonuniform bandwidths are desired. The results are illustrated with examples.

II. ANALYSIS OF UNIFORM FILTER BANKS

Assume that the bandpass filters in Fig. 1 have impulse responses of the form

$$\begin{aligned} h_k(nT) &= 2 \mid D_k \mid h(nT) \cos(\omega_k nT + \Phi_k) \qquad k = 1, 2, \cdots, M \\ h_0(nT) &= \mid D_0 \mid h(nT) \end{aligned} \tag{1}$$

where $\omega_k = \omega_1 + (k-1)\,\Delta\omega$, and $h(nT)$ is the impulse response of a prototype lowpass filter. (When the $k = 0$ lowpass filter is used, $\omega_1 = \Delta\omega$ and $\omega_0 = 0$.) The system functions for this set of bandpass filters are

$$\begin{aligned} H_k(z) &= \mid D_k \mid e^{j\Phi_k} H(ze^{-j\omega_k T}) + \mid D_k \mid e^{-j\Phi_k} H(ze^{j\omega_k T}) \quad k = 1, 2, \cdots, M \\ H_0(z) &= \mid D_0 \mid H(z). \end{aligned} \tag{2}$$

The frequency response of these filters is obtained after substituting $z = e^{j\omega T}$ in (2) as

$$\begin{aligned} H_k(e^{j\omega T}) &= \mid D_k \mid e^{j\Phi_k} H(e^{j(\omega-\omega_k)T}) + \mid D_k \mid e^{-j\Phi_k} H(e^{j(\omega+\omega_k)T}), \\ &\qquad\qquad k = 1, 2, \cdots, M, \\ H_0(e^{j\omega T}) &= \mid D_0 \mid H(e^{j\omega T}). \end{aligned} \tag{3}$$

If the frequency response of prototype lowpass filter, $H(e^{j\omega T})$, drops off sharply, then it can be seen from (3) that

$$\begin{aligned} \mid H_k(e^{j\omega T}) \mid &\approx \mid D_k \mid \mid H(e^{j(\omega-\omega_k)T}) \mid \qquad 0 \leqq \omega \leqq \pi/T \\ &\approx \mid D_k \mid \mid H(e^{j(\omega+\omega_k)T}) \mid \quad -\pi/T \leqq \omega \leqq 0. \end{aligned}$$

In this case the filter bank consists of a set of $(M+1)$ equally spaced bandpass filters with identical magnitude responses around their respective center frequencies. We have chosen this method of designing bandpass filters from lowpass prototypes for analytical convenience and because of the importance of spectrum analysis systems of this form.[1,2] The results to be discussed apply for other bandpass transformations in so far as they yield a set of uniformly spaced bandpass filters with identical frequency characteristics.

Our objective is to choose the prototype lowpass filter and the parameters $\mid D_k \mid$, ω_1, $\Delta\omega$, and Φ_k so that the filter bank will closely approximate the characteristics of the ideal spectrum analyzer. To do this we must consider the response of the composite filter bank. First,

however, it is useful to interpret the individual bandpass filter outputs in terms of spectrum analysis considerations.

The individual filter outputs are of the form

$$x_k(nT) = \sum_{r=-\infty}^{n} 2x(rT) \mid D_k \mid h(nT - rT) \cos [\omega_k(nT - rT) + \Phi_k], \quad (4)$$

$$x_0(nT) = \sum_{r=-\infty}^{n} x(rT) \mid D_0 \mid h(nT - rT)$$

which can be expressed as

$$\begin{aligned} x_k(nT) &= 2 \quad \text{Re} \{D_k X(\omega_k , n) e^{j\omega_k nT}\} \\ x_0(nT) &= \text{Re} \{D_0 X(0, n)\} \end{aligned} \quad (5)$$

where D_k is the complex constant defined by

$$\begin{aligned} D_k &= \mid D_k \mid e^{j\Phi_k} \qquad k = 1, 2, \cdots , M \\ D_0 &= \mid D_0 \mid, \end{aligned} \quad (6)$$

and

$$X(\omega_k , n) = \sum_{r=-\infty}^{n} x(rT) h(nT - rT) e^{-j\omega_k rT}. \quad (7)$$

The quantity $X(\omega_k , n)$ is the discrete-time version of the short-time Fourier transform[4] of $x(nT)$. Thus, (5) serves to relate the bandpass filter outputs $x_k(nT)$ to the short-time Fourier transform.

The frequency response and impulse response of the composite filter bank are obtained from

$$y(nT) = \sum_{k=0}^{M} x_k(nT). \quad (8)$$

After substituting (5) into (8) and noting that if $x(nT)$ is real, $X(-\omega_k , n)$ is the complex conjugate of $X(\omega_k , n)$, we obtain

$$y(nT) = \sum_{k=-M}^{M} D_k X(\omega_k , n) e^{j\omega_k nT} \quad (9)$$

where $\omega_{-k} = -\omega_k$, and D_{-k} is the complex conjugate of D_k. Substituting (7) into (9) and interchanging the order of summations results in

$$y(nT) = \sum_{r=-\infty}^{n} x(rT) \left[h(nT - rT) \cdot \sum_{k=-M}^{M} D_k e^{j\omega_k (nT - rT)} \right]. \quad (10)$$

Defining

$$d(nT) = \sum_{k=-M}^{M} D_k e^{j\omega_k nT}, \tag{11}$$

we observe from (10) that the combined impulse response of the filter bank can be expressed as

$$\tilde{h}(nT) = h(nT)\, d(nT). \tag{12}$$

Equations (11) and (12) are the basic results of the analysis. (Note that they could have been obtained directly by summing the impulse responses $h_k(nT)$, with the sacrifice of the interpretation of the filter bank outputs in terms of the short-time spectrum.)

Equation (12) shows that the filter bank impulse response is the product of the prototype lowpass filter impulse response $h(nT)$, and the sequence $d(nT)$ defined by (11). The choice of lowpass filter depends on both the desired frequency resolution and the requirement of obtaining flat magnitude and linear phase response in the composite filter bank. The sequence $d(nT)$ is independent of the prototype lowpass filter and is a function of the frequency spacing, the relative gains and phases, and the number of bandpass filters. Thus, for *a given choice of prototype lowpass filter*, the parameters of $d(nT)$ can be adjusted to achieve the best approximation to the ideal spectrum analyzer. To see how this occurs, we shall first examine in detail the characteristics of the sequence $d(nT)$.

As will be shown in the remainder of this section, a particularly useful choice of the complex coefficients D_k in (11) is

$$\begin{aligned} D_k &= e^{j\omega_k n_0 T} \qquad k = \pm 1, \pm 2, \cdots, \pm M. \\ D_0 &= 1, \end{aligned} \tag{13}$$

where n_0 is an integer. That is,

$$\Phi_k = \omega_k n_0 T, \quad \text{and} \quad |D_k| = 1.$$

(The condition $D_0 = 1$ implies that the band around $\omega = 0$ is included in the filter bank; $D_0 = 0$ implies that it is not.) It can be shown that for this choice of D_k, (11) becomes

$$d(nT) = D_0 + \frac{2 \sin [M\, \Delta\omega(nT + n_0T)/2]}{\sin [\Delta\omega(nT + n_0T)/2]} \cos [\omega_a(nT + n_0T)] \tag{14}$$

where $\omega_a = \omega_1 + (M - 1)\, \Delta\omega/2$, and D_0 is 1 or 0 depending on whether or not the lowpass channel is included.

The properties of the sequence $d(nT)$ determine the character of

the impulse response of the filter bank. Some of these properties are summarized below:

(*i*) The parameter n_0 shifts the sequence $d(nT)$ by n_0 samples with respect to $h(nT)$.

(*ii*) The sequence $d(nT)$ is even about the sample $n = -n_0$; i.e., $d(nT - n_0T) = d(-nT - n_0T)$.

(*iii*) The maximum value of $d(nT)$ occurs at $n = -n_0$. This value is $d(-n_0T) = 2M + D_0$.

(*iv*) If $2\pi/(\Delta\omega T)$ and $\omega_1/\Delta\omega$ are both integers, then the sequence $d(nT)$ is periodic with period $2\pi/\Delta\omega$. Otherwise, $d(nT)$ will be an almost periodic sequence which will peak up at time intervals of $2\pi/\Delta\omega$.

Insight into the properties of $d(nT)$ can be gained by considering a simple example. Assume that $\omega_1 = \Delta\omega$ and $2\pi/(\Delta\omega T) = N$ where N is an odd integer. That is, the entire frequency range $-\pi/T \leqq \omega \leqq \pi/T$, is divided into N equal bands. If $M = (N - 1)/2$, the entire frequency range is covered. Under these conditions (14) can be written

$$d(nT) = \frac{\sin\left[\left(\frac{2M+1}{2}\right)\Delta\omega(nT + n_0T)\right]}{\sin\left[\Delta\omega(nT + n_0T)/2\right]} - 1 \quad \text{if} \quad D_0 = 0 \tag{15a}$$

$$= \frac{\sin\left[\left(\frac{2M+1}{2}\right)\Delta\omega(nT + n_0T)\right]}{\sin\left[\Delta\omega(nT + n_0T)/2\right]} \quad \text{if} \quad D_0 = 1. \tag{15b}$$

(If N is even, the $k = 0$ filter is not used, i.e., $D_0 = 0$, and ω_1 is $\Delta\omega/2$.) It is clear from (15b) that for these conditions $d(nT)$ is a periodic sequence with period $NT = 2\pi/\Delta\omega$. In fact, $d(nT)$ may be thought of as samples of a continuous-time periodic Dirichlet kernel as shown in Fig. 2a. If $M = (N - 1)/2$ and $D_0 = 1$, $d(nT)$ is a periodic discrete-time impulse train with impulses occurring at multiples of NT. This is because the sample points on the periodic Dirichlet kernel occur at the maxima and the zero crossings, as indicated by the small circles in Fig. 2a.

The conditions for $d(nT)$ to be periodic are that both $\omega_1/\Delta\omega$ and $2\pi/(\Delta\omega T)$ be equal to integers. To see this, we must examine (14) in detail. If $2\pi/(\Delta\omega T)$ is an integer, and M is an odd integer, the sequence $2 \sin [M \Delta\omega(nT + n_0T)/2]/\sin [\Delta\omega(nT + n_0T)/2]$ is periodic with period $NT = 2\pi/\Delta\omega$. If $\omega_1/\Delta\omega$ is an integer and M is odd, the sequence $\cos [(\omega_1 + (M - 1) \Delta\omega/2)(nT + n_0T)]$ is also periodic with a period

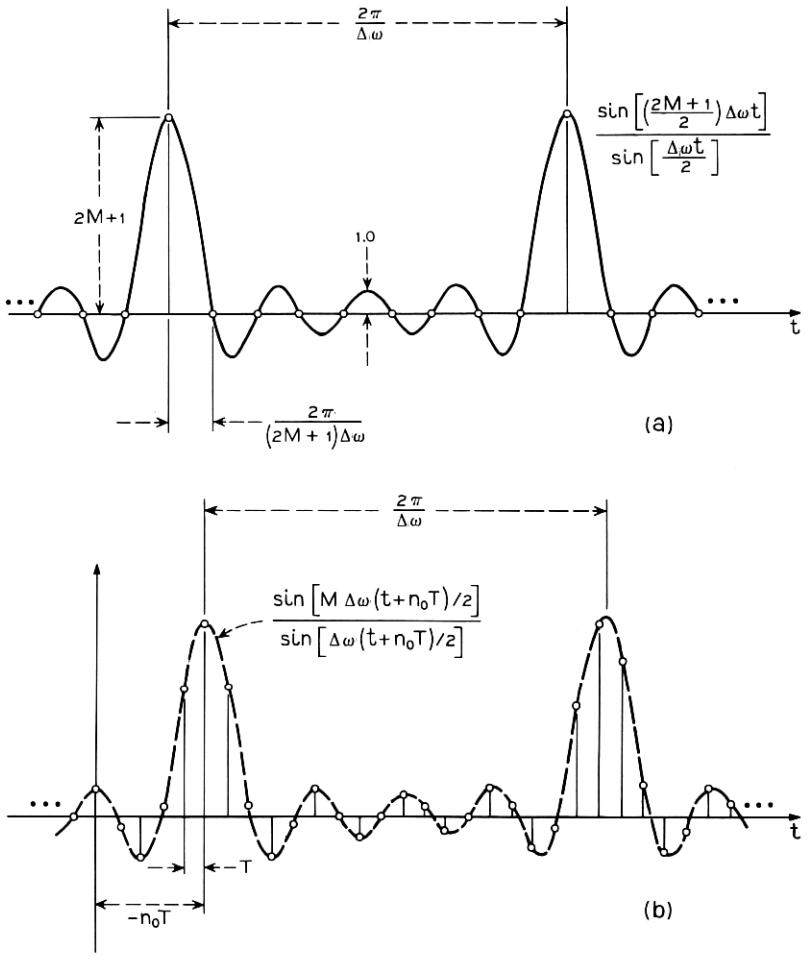

Fig. 2—(a) Periodic continuous-time Dirichlet kernel, (b) continuous-time envelope and sequence $d(nT)$ when either $\omega_1/\Delta\omega$ or $2\pi/(\Delta\omega T)$ are not integers.

that is an integer multiple of $2\pi/\Delta\omega$. Thus the product of these two sequences is periodic with period $2\pi/\Delta\omega$. The identical result holds for M an even integer and $2\pi/\Delta\omega$ and $\omega_1/\Delta\omega$ integers although the interaction between the component sequences is slightly different.

If either $2\pi/(\Delta\omega T)$ or $\omega_1/\Delta\omega$ are not integers, $d(nT)$ will not be periodic, but will still peak up at time intervals of $2\pi/\Delta\omega$. Such a case is depicted in Fig. 2b where the samples $d(nT)$ are marked by the

small circles and the dotted curve shows the factor

$$\sin [M\ \Delta\omega(nT + n_0T)/2]/\sin [\Delta\omega(nT + n_0T)/2]$$

when M is odd. As shown in Fig. 2b, $d(nT)$ will always have even symmetry about sample $n = -n_0$.

III. DESIGN OF UNIFORM FILTER BANKS USING BESSEL FILTERS

In the preceding section we presented a detailed analysis of a filter bank composed of equally spaced equal-bandwidth filters. In this section we will show how the results of that analysis can be employed in filter bank design.

The objective of flat amplitude response and linear phase is most easily achieved with bandpass filters having these same properties. For this reason, Bessel (maximally flat delay) filters are often used in filter banks.[5] In the examples shown in this paper, we have used digital filters obtained from Bessel prototype designs using impulse invariance.[6] It should be noted that the digital filters obtained this way do not have the maximally flat delay property. J. P. Thiran[7] has shown that the denominator of the system function of maximally flat delay digital filters is a Gauss hypergeometric function. It is reasonable to expect however, that for the narrow-band filters of interest here, the differences should be negligible.

As an example a digital filter derived from a sixth-order Bessel lowpass filter with asymptotic cutoff frequency of 60 Hz is shown in Fig. 3. The impulse response is shown in Fig. 3a, and the amplitude and phase responses are shown in Fig. 3b and Fig. 3c. The filter shown in Fig. 3 was used in a filter bank* with the following choice of parameters: $D_0 = 0$, $T = 10^{-4}$ sec, $\Delta\omega = 2\pi(100)$, $\omega_1 = 2\pi(100)$, $n_0 = 0$, and $M = 30$. The resulting filter bank characteristics are shown in Fig. 4. The filter bank impulse response, $\tilde{h}(nT)$, is shown in Fig. 4a along with the prototype lowpass impulse response $h(nT)$. For the above choice of parameters, $d(nT)$ is obtained from (15a) as

$$d(nT) = \frac{\sin [0.61\pi n]}{\sin [0.01\pi n]} - 1, \tag{16}$$

which is periodic with period 100 samples (10 msec), with peaks occurring at $nT = 0, \pm 10, \pm 20, \cdots$ msec. From Fig. 4a, it can be seen that in the product $h(nT) \cdot d(nT)$, the peak of $d(nT)$ at $nT = 0$ will be attenuated since $h(nT)$ is small around $nT = 0$. On the other hand,

* Note that the resulting bandpass filters are twelfth order.

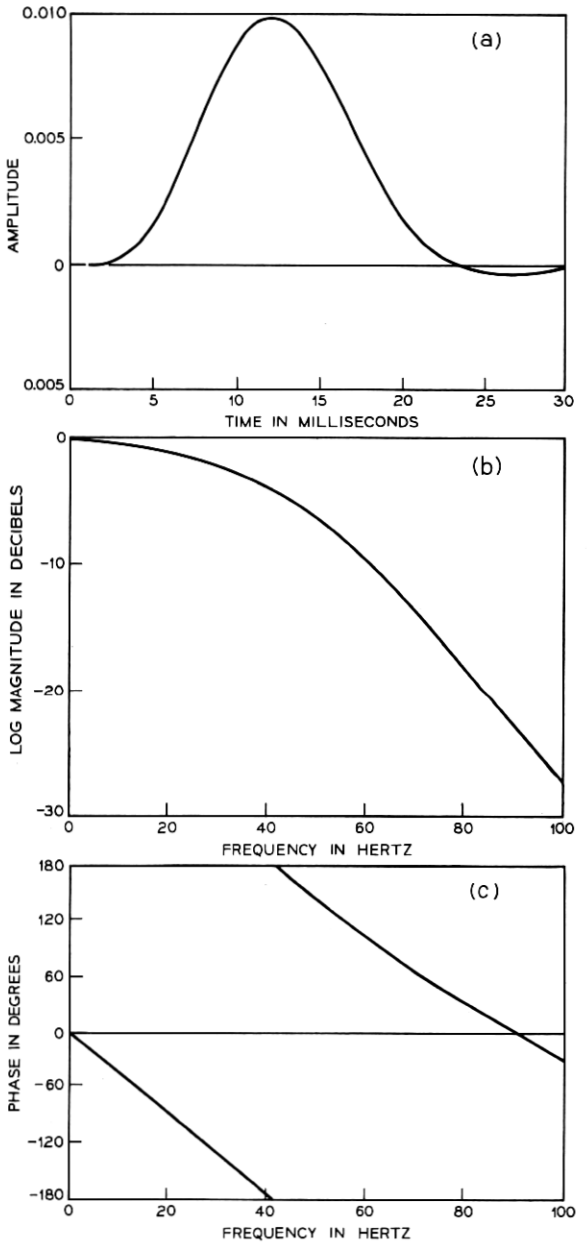

Fig. 3—Sixth-order Bessel filter characteristics. (a) Impulse response, (b) magnitude response, (c) phase response.

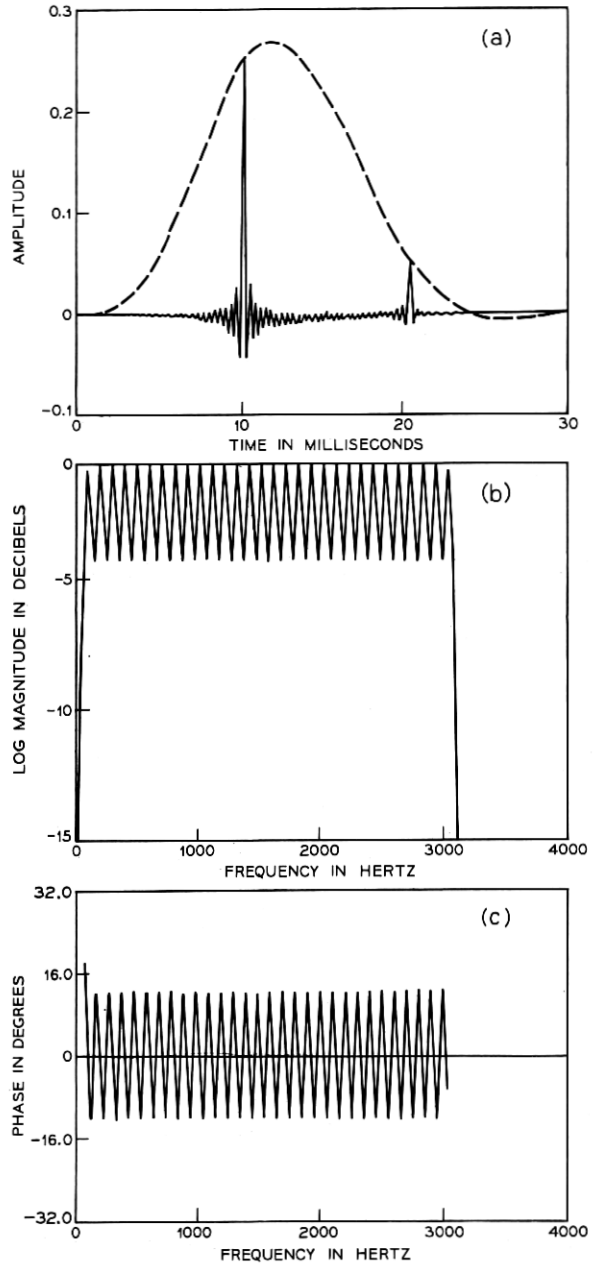

Fig. 4—Characteristics of 30-channel filter bank. (a) Impulse response (dotted curve is the impulse response of the prototype lowpass filter in Fig. 3), (b) composite magnitude response, (c) composite phase response after subtracting 10-msec delay.

the peak of $d(nT)$ at $nT = 10$ msec occurs at approximately the peak of $h(nT)$, and at $nT = 20$ msec, $h(nT)$ is large enough to produce a significant echo in the impulse response of the filter bank. As is shown in Fig. 4b and 4c, this corresponds to a 3.9-dB ripple in the amplitude response and a 25.5-degree peak-to-peak ripple in the phase response (after removing a linear phase component corresponding to a 10-msec or 100-sample delay). To decrease this amplitude and phase ripple, we should attempt to eliminate the echo in the impulse response. Furthermore, the phase ripple will be eliminated if the impulse response $h(nT)$ has even symmetry about some delay time $n_D T$. One approach is to broaden the filter bandwidths, or equivalently reduce the spacing $\Delta\omega$, so that $h(nT)$ is contracted relative to the spacing of pulses in $d(nT)$. This is generally not an acceptable solution since $h(nT)$ and $\Delta\omega$ are usually fixed by some frequency resolution criterion. However, if we refer to the properties of $d(nT)$ which were previously summarized, we note that a negative value of n_0 will shift $d(nT)$ to the right relative to $h(nT)$ so that $d(nT)$ will have even symmetry about time $n_D T = -n_0 T$. If n_0 can be chosen so than $\tilde{h}(nT) = h(nT) \cdot d(nT)$ has approximately even symmetry and consists of only one significant pulse, then the amplitude and phase ripple will be small. The manner in which this is achieved is shown in Fig. 5 where it is assumed for simplicity that $d(nT)$ is a train of digital impulses as would be the case for $M = (N - 1)/2$. Figure 5a depicts the case where $n_0 = 0$. Figure 5b shows the situation where n_0 was chosen to shift the impulse which was at $nT = 0$ in Fig. 5a to the right and into the vicinity of the peak of $h(nT)$. If it is assumed that only three impulses have nonzero amplitudes (α_1, α_2, α_3) such that $4 \mid \alpha_1 \mid \cdot \mid \alpha_3 \mid < \mid \alpha_1 + \alpha_3 \mid \cdot \mid \alpha_2 \mid$, then it can be shown (see Appendix) that the peak-to-peak amplitude ripple of the filter bank is

$$R_A = 20 \log_{10} \left[\frac{\mid \alpha_2 + \alpha_1 + \alpha_3 \mid}{\mid \alpha_2 - \alpha_1 - \alpha_3 \mid} \right]. \tag{17}$$

Similarly, if $\mid \alpha_1 + \alpha_3 \mid < \mid \alpha_2 \mid$, the peak-to-peak phase ripple about a linear phase corresponding to a delay of $-n_0 T$ is given by

$$R_P = 2 \tan^{-1} \left[\frac{\alpha_1 - \alpha_3}{(\alpha_2^2 - (\alpha_1 + \alpha_3)^2)^{1/2}} \right]. \tag{18}$$

The conditions for (17) and (18) to hold are satisfied when α_1 and α_3 are small relative to α_2, which is the normal situation. It can be seen from (18) and (17) that the phase ripple will be zero if $\alpha_1 = \alpha_3$, and the amplitude ripple will be small if $(\alpha_1 + \alpha_3)/\alpha_2$ is small.

Although these results were derived for the idealized case when $d(nT)$

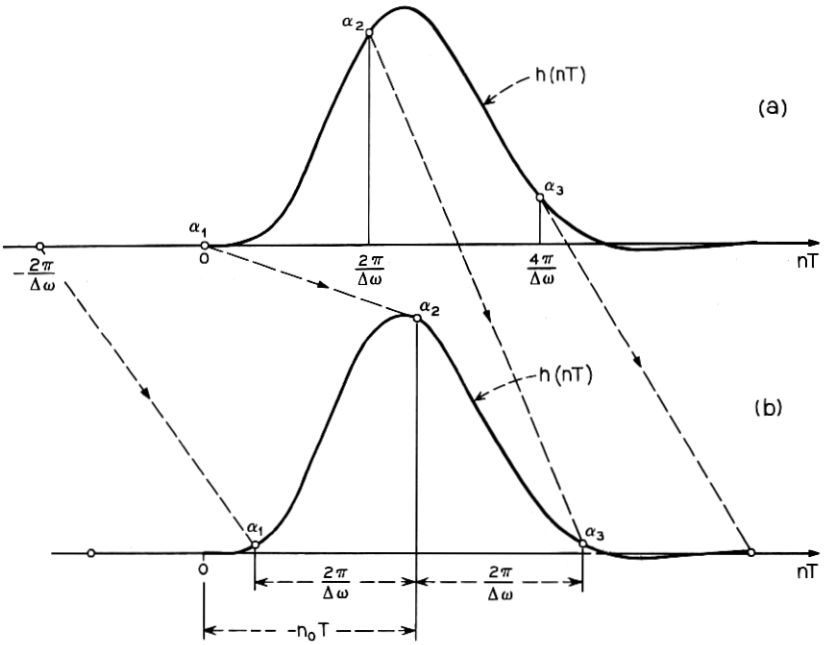

Fig. 5—Illustration of how to adjust the parameter n_0. (a) Composite impulse response for $n_0 = 0$, (b) n_0 chosen to minimize magnitude and phase ripple (dotted lines indicate movement of individual pulses in $d(nT)$).

is an impulse train, we have found that amplitude and phase ripple can be determined quite accurately using (17) and (18) in more general situations. With the foregoing principles in mind we have written an interactive computer program for filter bank design. Using this program we can design a filter bank with low amplitude and phase ripple by the following process:

(*i*) Choose ω_1 , $\Delta\omega$, and M to cover the desired analysis band and choose $h(nT)$ to provide desired frequency resolution. This results in an $h(nT)$ that has a duration of approximately $4\pi/\Delta\omega$ as shown in Fig. 5.

(*ii*) Evaluate $h(nT)$ and determine n_0 such that $\alpha_1 \approx \alpha_3$ as in Fig. 5b.

(*iii*) If the resulting filter bank is not satisfactory, steps *i* and *ii* are repeated.

In cases where $\omega_1/\Delta\omega$ is not an integer, it is important to choose n_0

so that the point of even symmetry in $d(nT)$ is shifted into the vicinity of the peak of $h(nT)$. Otherwise, it may be impossible to achieve a very good approximation to linear phase. An example of the improvement gained by proper choice of n_0 is shown in Fig. 6. In this example all the parameters were the same as in the example of Fig. 4 except a value of $n_0 = -129$ was chosen by the above process. In this case the in-band amplitude ripple is 0.8 dB and the phase ripple is 0.6 degree, as compared to 3.8 dB and 25.5 degrees when $n_0 = 0$.

R. M. Golden[5] has shown that inverting the sign of alternating channels often significantly improves the characteristics of a filter bank. This technique has a simple interpretation in terms of our results. It can be shown that inverting the sign of alternating channels is equivalent to delaying the sequence $d(nT)$ by $n_0 = -\pi/(\Delta\omega T)$ samples. This amount of delay may be nearly correct if the duration of $h(nT)$ is approximately $3\pi/\Delta\omega$; however for the situation shown in Fig. 4a, such a delay would produce a worse filter bank than no delay at all ($n_0 = 0$). Also, to achieve linear phase when $\omega_1/\Delta\omega$ is not an integer, the point of even symmetry in $d(nT)$ should be delayed to the vicinity of the peak of $h(nT)$. This does not occur when the signs of alternate channels are inverted.

IV. DESIGN OF NONUNIFORM BANDWIDTH FILTER BANKS

In speech applications it is common to take advantage of the frequency resolution characteristics of the ear[4,5] by using increasing bandwidth filters at higher frequencies. The previously discussed techniques can be applied to this situation if the filter bank consists of several sub-banks, each with different resolution. Each sub-bank can be designed as discussed above, with care being taken to ensure that the entire frequency band of interest is covered by the combination of the sub-banks. It may be necessary to equalize the delay between sub-banks by providing additional delay for all but one of the sub-banks.* This is depicted in Fig. 7 for three sub-banks with increasing-bandwidth sixth-order Bessel filters. Figure 7a shows the lowpass prototype impulse response and shifted $d(nT)$ sequence† for the first sub-bank. The lowpass asymptotic cutoff used was 78 Hz, the spacing of filters was $\Delta\omega_1 = 2\pi(125)$, the first filter was centered at $\omega_{11} = 2\pi(250)$, and a value of $n_{01} = -100$ (10-msec delay) was required to

* Golden[5] has shown that the delays can be approximately equalized by increasing the order of the lowpass prototype in direct proportion to the increase in bandwidth.

† The sequence $d(nT)$ is shown as an impulse train for convenience in plotting. The actual sequences would look like those in Fig. 4 and Fig. 6.

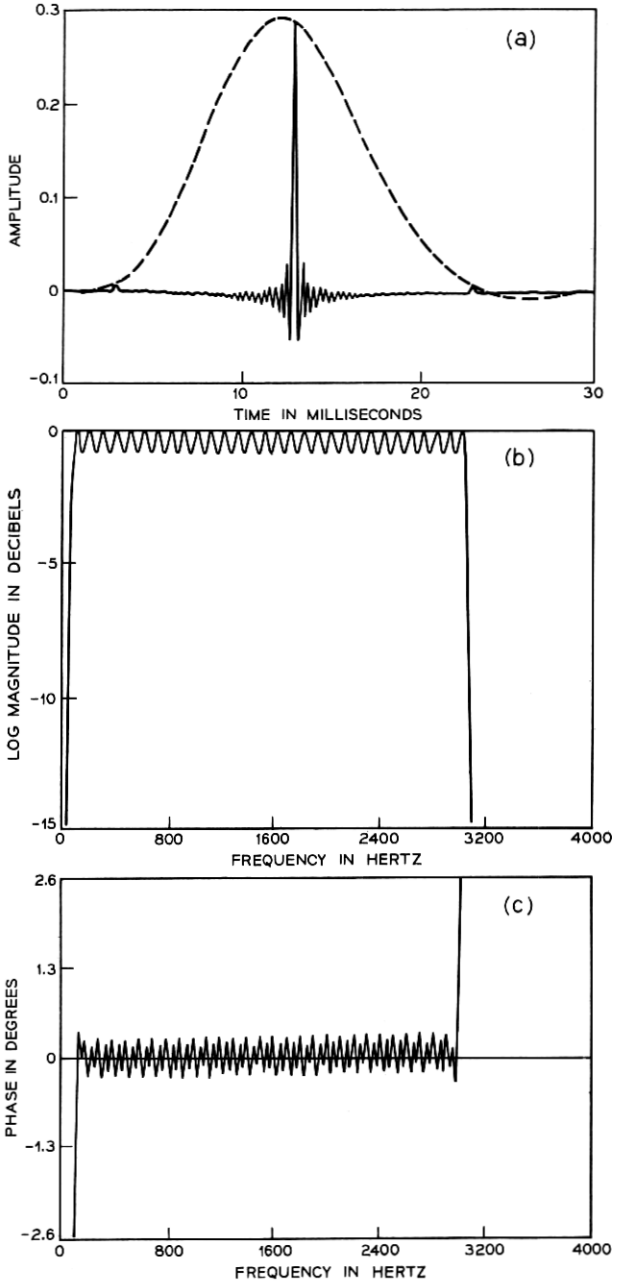

Fig. 6—Characteristics of 30-channel filter bank. (a) Impulse response for $n_0 = -129$ (dotted curve is the impulse response in Fig. 3), (b) composite magnitude response, (c) composite phase response after subtracting 12.9-msec delay.

minimize the amplitude and phase ripple. Figure 7b shows the second sub-bank in which the basic parameters were: lowpass asymptotic cutoff 136 Hz, $\Delta\omega_2 = 2\pi(218)$, $\omega_{12} = 2\pi(1296.5)$, and $n_{02} = -57$ (5.7-msec delay). To line up the central peaks, an additional delay of $n_2 = 43$ samples (4.3 msec) was required. Figure 7c shows the third sub-bank where the lowpass cutoff was 192 Hz, $\Delta\omega_3 = 2\pi(307)$, $\omega_{13} = 2\pi(2213)$, and $n_{03} = -40$ (4-msec delay). A value of $n_3 = 60$ samples (6.0 msec) is required to line up the central peak with those in Fig. 7a and 7b. The response of the combination of these three sub-banks is shown in Fig. 8. Figure 8a shows the impulse response, Fig. 8b shows the amplitude response, and Fig. 8c shows the phase after a linear phase corresponding to 10-msec delay has been subtracted. It can be

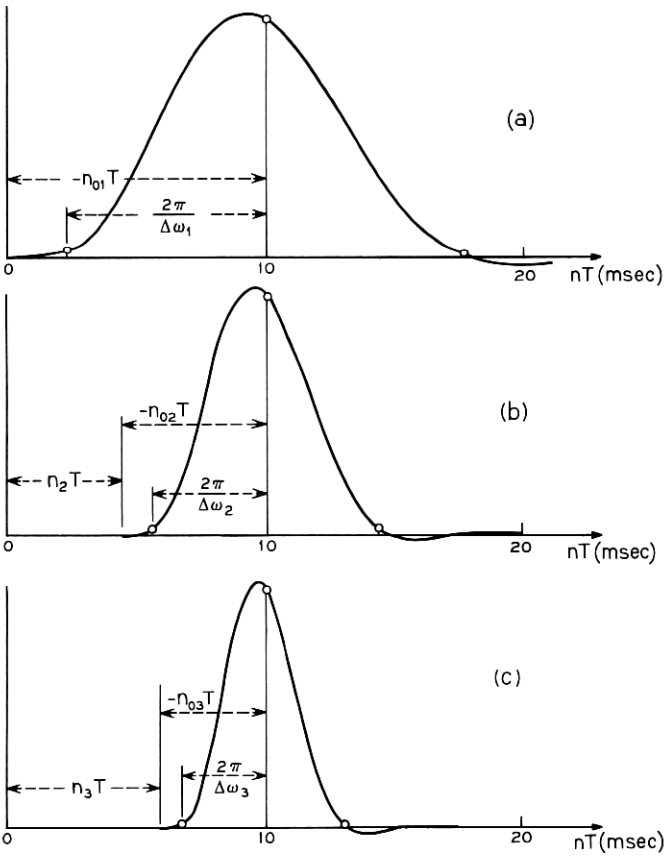

Fig. 7—Illustration of the design of nonuniform filter banks: (a) impulse response for narrow bandwidth filters, (b) impulse response for intermediate bandwidth filters, (c) impulse response for wide bandwidth filters.

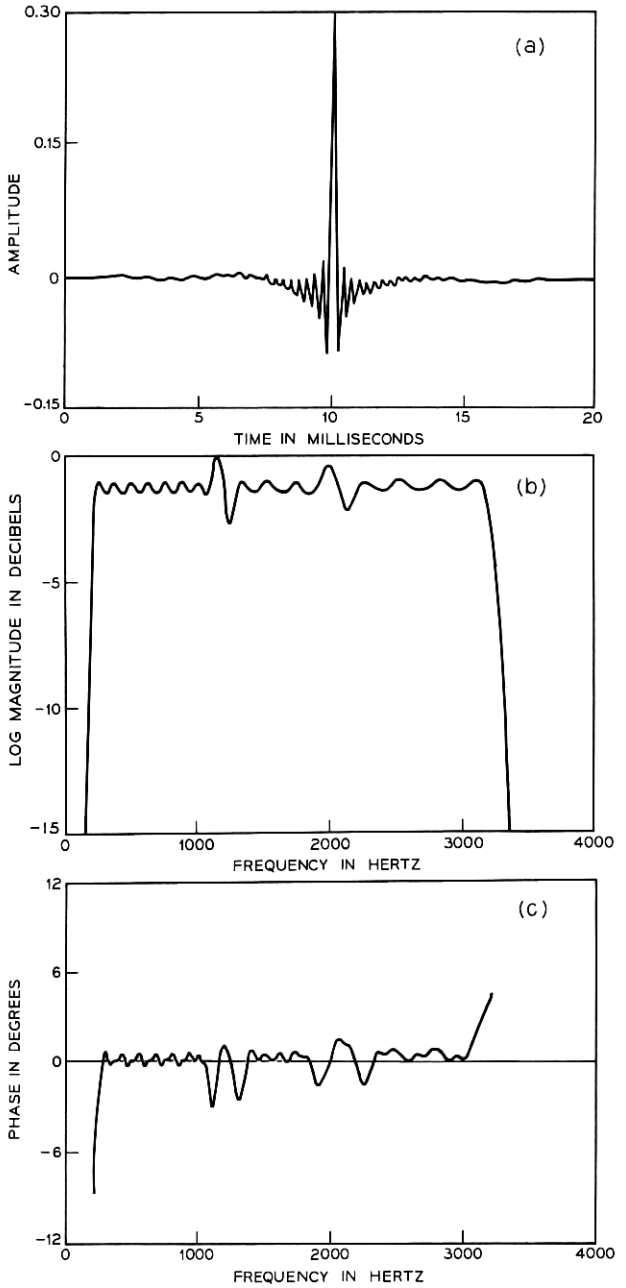

Fig. 8—Composite filter bank characteristics with three different sub-banks: (a) impulse response, (b) magnitude response, (c) phase response after subtracting 10-msec delay.

seen in Fig. 8b and 8c that the ripple in the sub-banks is quite low as would be expected from Fig. 7. At the boundary between sub-banks, however, the ripple increases significantly due to the fact that the last filter in the lower sub-bank drops off more rapidly than the first filter in the next sub-bank. This excessive variation at the boundary between sub-banks can be eliminated to some extent by using increasingly higher-order filters in the sub-banks. Alternatively, nonuniform resolution can be obtained by using equal-bandwidth filters and adding together groups of two or more of their outputs to achieve the desired bandwidth. Such an approach would require increased computation but would produce filter bank characteristics comparable to those in Fig. 6.

V. CONCLUSION

We have discussed the analysis and design of digital filter banks and have shown how the incorporation of a linearly increasing phase shift in each bandpass filter can significantly improve the overall filter bank characteristics. We also showed how the techniques can be used in nonuniform bandwidth filter banks.

The examples which we gave were based on Bessel lowpass prototypes which have impulse responses of desirable shape but rather poor amplitude response. Recent results in the design of finite duration impulse response filters[8] offer attractive possibilities for filter bank design. Such filters can have precisely linear phase and can be designed using iterative techniques with constraints on both the impulse response shape and the amplitude response. The use of such filters, together with the basic principles discussed in this paper, should yield filter banks with excellent properties.

APPENDIX

Derivation of Magnitude and Phase Ripple Formulas

Assume an impulse response sequence

$$\begin{aligned} h(n) &= \alpha_1 \qquad n = 0 \\ &= \alpha_2 \qquad n = n_p \\ &= \alpha_3 \qquad n = 2n_p \\ &= 0 \qquad \text{elsewhere.} \end{aligned} \tag{19}$$

The system function of this system is

$$H(e^{j\omega T}) = \alpha_1 + \alpha_2 e^{-j\omega n_p T} + \alpha_3 e^{-j\omega 2n_p T}. \tag{20}$$

The squared magnitude response is

$$| H(e^{j\omega T}) |^2 = [\alpha_2 + (\alpha_1 + \alpha_3) \cos (\omega n_p T)]^2 + (\alpha_1 - \alpha_3)^2 \sin^2 (\omega n_p T), \tag{21}$$

and the phase response is

$$\arg [H(e^{j\omega T})] = \tan^{-1} \left[\frac{(\alpha_1 - \alpha_3) \sin (\omega n_p T)}{\alpha_2 + (\alpha_1 + \alpha_3) \cos (\omega n_p T)} \right] \tag{22}$$

where a linear phase component $-\omega n_p T$ has been removed. Clearly, both (21) and (22) are periodic functions of ω with period $2\pi/n_p T$. To determine the amplitude and phase ripple, we must locate the maxima and minima of (21) and (22).

If we differentiate (21) with respect to ω, we find that the maxima and minima occur for values of ω satisfying

$$\sin (\omega n_p T) = 0 \tag{23a}$$

$$\cos (\omega n_p T) = -\alpha_2 \frac{(\alpha_1 + \alpha_3)}{4\alpha_1 \alpha_3}. \tag{23b}$$

The second equation is satisfied by a real value of ω if and only if

$$4 \mid \alpha_1 \mid \cdot \mid \alpha_3 \mid > \mid \alpha_1 + \alpha_2 \mid \cdot \mid \alpha_2 \mid. \tag{24}$$

In a good filter bank design, α_1 and α_3 will be positive and much smaller than α_2, and (24) will not be satisfied. Evaluating the second derivative shows that in this case the maxima and minima of $| H(e^{j\omega T}) |$ will alternate and occur at values of ω satisfying (23a); i.e., $\omega = 0, \pm\pi/n_p T, \pm 2\pi/n_p T, \cdots$. In this case the amplitude ripple in dB is given by

$$R_A = 20 \log_{10} \left[\frac{\mid \alpha_2 + \alpha_1 + \alpha_3 \mid}{\mid \alpha_2 - \alpha_1 - \alpha_3 \mid} \right]. \tag{25}$$

If (22) is differentiated with respect to ω, we find that the maxima and minima occur at values of ω satisfying

$$\cos \omega n_p T = -\left(\frac{\alpha_1 + \alpha_3}{\alpha_2} \right). \tag{26}$$

Equation (26) is satisfied by real values of ω if $\mid \alpha_1 + \alpha_3 \mid < \mid \alpha_2 \mid$. In this case the maxima and minima again alternate, and the peak-to-peak phase ripple is

$$R_p = 2 \tan^{-1} \left[\frac{\alpha_1 - \alpha_3}{(\alpha_2^2 - (\alpha_1 + \alpha_3)^2)^{1/2}} \right]. \tag{27}$$

If $|\alpha_1 + \alpha_3| > |\alpha_2|$, the phase curve will be discontinuous with a jump of 2π radians occurring at $\omega = \pm\pi/n_pT, \pm 3\pi/n_pT, \cdots$.

REFERENCES

1. Flanagan, J. L., and Golden, R. M., "Phase Vocoder," B.S.T.J., *45*, No. 9 (November 1966), pp. 1493–1509.
2. Flanagan, J. L., and Lummis, R. C., "Signal Processing to Reduce Multipath Distortion in Small Rooms," J. Acoust. Soc. Am., *47*, No. 1 (June 1970), pp. 1475–1481.
3. Jackson, L. B., Kaiser, J. F., and McDonald, H. S., "An Approach to the Implementation of Digital Filters," IEEE Trans. Audio and Electroacoust., *AU-16*, No. 3 (September 1968), pp. 413–421.
4. Flanagan, J. L., *Speech Analysis, Synthesis and Perception*, New York: Academic Press, 1965.
5. Golden, R. M., "Vocoder Filter Design: Practical Considerations," J. Acoust. Soc. Am., *43*, (December 1967) pp. 803–810.
6. Gold, B., and Rader, C. M., *Digital Processing of Signals*, New York: McGraw-Hill Book Co., 1969.
7. Thiran, J. P., "Recursive Digital Filters with Maximally Flat Group Delay," IEEE Trans. Ckt. Theory, *CT-18*, No. 4 (November 1971).
8. Rabiner, L. R., "Techniques for Designing Finite Duration Impulse Response Digital Filters," IEEE Trans. Com. Tech., *COM-19*, No. 2 (April 1971), pp. 188–195.

THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 50, No. 10, December, 1971
Printed in U.S.A.

The Preference of Slope Overload to Granularity in the Delta Modulation of Speech

By N. S. JAYANT and A. E. ROSENBERG

(Manuscript received June 18, 1971)

A preference study was made to assess the relative annoyance values of slope-overload distortion and granular noise in delta-modulated speech. A recently described adaptive delta modulator was simulated at frequencies of 20 and 40 kHz, and controlled amounts of the two types of degradation were introduced into samples of a 2-second utterance. Rankings were obtained for these samples on the basis of preference judgments of nine listeners, each of whom assessed the samples, pairwise, in a tournament-type strategy. Results indicate that the speech sample exhibiting the minimum degradation on an objective, overall-noise-power basis is not subjectively the most preferred sample. Furthermore, the subjectively optimum delta modulator exhibits greater overload and lesser granularity than the objectively optimum device.

I. INTRODUCTION

The principle of delta modulation[1] has been widely described in the literature. Briefly, delta modulation is a digital encoding strategy which uses a simple feedback mechanism to produce a "staircase" approximation to an input signal. A block diagram of the simplest form of delta modulation appears in Fig. 1. The input sequence $\{X_r\}$ is usually band-limited and suitably oversampled. The "staircase" sequence Y_r is generated according to the equations

$$C_r = \text{sgn}\,(X_r - Y_{r-1}) \tag{1}$$

$$Y_r - Y_{r-1} = m_r = \Delta_r \cdot C_r\,. \tag{2}$$

The step-size Δ_r is assumed to be a constant in conventional (linear) delta modulation. "Adaptive" delta modulation, on the other hand, allows for modifications of Δ_r in accordance with the changing slope

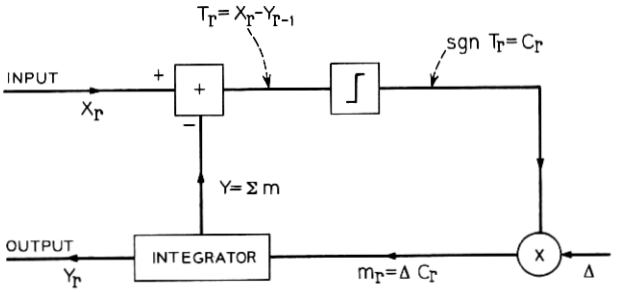

Fig. 1—Schematic diagram of a linear delta modulator.

characteristics of the input signal. Such adaptation results in better encoding, and several types of adaptive delta modulation have been described in the literature.[2,3,4]

Figure 2 illustrates the mechanism of an adaptive delta modulator and demonstrates how suitable increases and decreases of step size facilitate better encoding during steep and flat regions of the input signal waveform. Such adaptations can be effected by observations on a "recent" segment of the binary sequence $\{C_r\}$; this is illustrated by equation (5) in the sequel.

Figure 2 also brings out the distinction between two types of encoding error in delta modulation, viz., "granular noise" and "slope-overload"

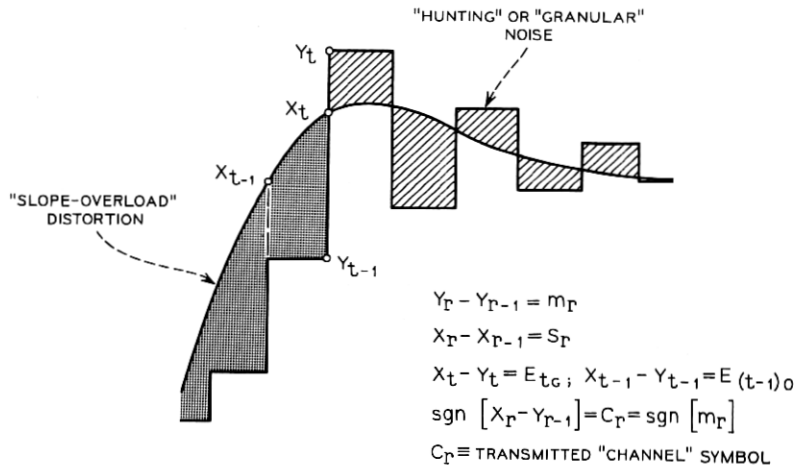

Fig. 2—Illustration of adaptive delta modulation.

distortion. A given error sample

$$E_r = X_r - Y_r \tag{3}$$

can be defined to fall into the granular or slope-overload category, depending on whether the corresponding step m_r crosses the input waveform or not. Thus, in Fig. 2, there is a 'granular' error E_{tG} at the sampling instant t, and an 'overload' error $E_{(t-1)O}$ at the sampling instant $(t - 1)$. As a matter of definition, we will note that $E_{tO} = E_{(t-1)G} = 0$.

The signal output $\{Z_r\}$ of the delta modulator is actually obtained by filtering the staircase sequence $\{Y_r\}$ to the input signal band. Let $\{X_r^F\}$ be the result of passing $\{X_r\}$ through the same lowpass filter. A perceptually relevant measure of signal degradation is accordingly defined by the encoding error

$$e_r = X_r^F - Z_r . \tag{4}$$

As with the quantity E_r in (3), one can distinguish samples of granularity and slope overload, e_{rG} and e_{rO}, in the error sequence $\{e_r\}$. Referring to Fig. 2 once more it can be seen that a physical distinction between the two types of error is suggested. Granularity can be described as a "signal-uncorrelated" random noise-type of phenomenon. It is characterized by alternation of signs and tends to be independent of signal amplitude. Slope overload, on the other hand, can be described as a "signal-correlated" distortion, since its sign and magnitude are related to the slope of the signal. This physical difference between slope overload and granularity suggests a corresponding perceptual distinction and raises the question of the relative annoyance values of the two forms of signal degradation in delta modulation. The present paper describes a study of the above question as referred to the delta modulation of a speech signal.

Earlier work in this subject is in the form of a perceptual experiment[5] in which H. Levitt, et al., characterized the perceptibility of slope-overload distortion as such. As mentioned earlier, our paper will seek to answer the complementary question of the relative perceptibilities of slope overload and granularity when they occur simultaneously in delta-modulated speech, as they usually do.

The approach we used was to vary the relative amount of slope overload and granularity introduced into samples of a test utterance, and to evaluate these samples on the basis of both objective and perceptual criteria; and then to interpret these evaluations with specific reference to the overload-granularity dichotomy.

Section II summarizes the salient features of a computer-simulated adaptive delta modulator that was utilized in the present study. This adaptive encoder has been recently described and shown to provide toll-quality speech reproduction at bit rates of practical importance.[4]

Section III defines the objective measures of speech quality used in our study, while Section IV defines a subjective measure of preference and describes an underlying perceptual experiment.

II. DESCRIPTION OF THE DELTA MODULATOR

Figure 3 is a schematic block diagram of the adaptive delta modulator utilized in the present study. This encoder is defined by the basic equations (1) and (2), and by the adaptation rule

$$\left.\begin{aligned} \Delta_r &= P \cdot \Delta_{r-1} \quad \text{if} \quad C_r = C_{r-1} \\ &= \frac{1}{P} \cdot \Delta_{r-1} \quad \text{if} \quad C_r \neq C_{r-1} \end{aligned}\right\}; \qquad P \geqq 1. \tag{5}$$

Notice that a conventional (linear) delta modulator corresponds to the special case of $P = 1$. In our study the value of P was a variable parameter; different (delta-modulated) speech samples corresponded to different suitably spaced values of P, and thereby to different mixtures of slope-overload and granularity.

The original speech sample X was a 2-second male utterance of "Have you seen Bill?" that had been band-limited to 3.3 kHz. The delta modulation was performed at sampling rates of 20 and 40 kHz. The latter frequency provides speech reproduction that approaches telephone

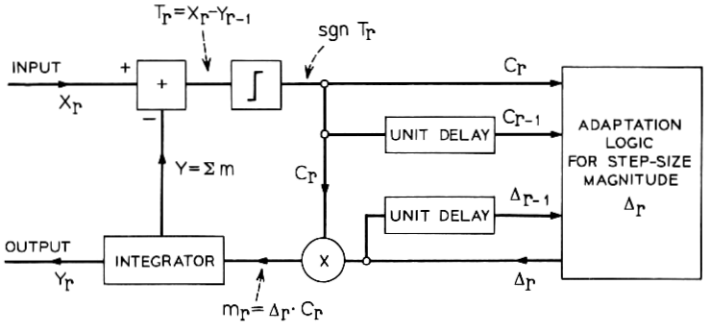

Fig. 3—Schematic diagram of an adaptive delta modulator.

THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 50, No. 10, December, 1971
Printed in U.S.A.

Design of Digital Filter Banks for Speech Analysis

By R. W. SCHAFER and L. R. RABINER

(Manuscript received June 22, 1971)

A bank of bandpass filters is often used in performing short-time spectrum analysis of speech signals. This paper is concerned with the analysis and design of digital filter banks composed of equally spaced bandpass filters. It is shown that significant improvement in the composite filter bank response can be achieved by proper choice of the relative phases of the bandpass filters. The results are extended to more general filter bank configurations.

I. INTRODUCTION

Many speech processing systems are based on the concept of short-time spectrum analysis.[1,2] Spectrum analyzers for such systems often consist of a set of bandpass filters whose combined passbands cover a desired frequency range. Although continuous-time filters have traditionally been used in filter banks for speech analysis, hardware realizations of digital filters are now available,[3] and the advantages which digital filters offer should be exploited in filter bank design. These advantages include: flexibility of design of the individual bandpass filters, precision of realization, stability of digital hardware, and the efficiency of realization of the filter bank afforded by the possibility of multiplexing the digital hardware. Thus it is important to consider design techniques for filter banks composed of digital bandpass filters.

To focus on the basic concepts in filter bank design, it is useful to define an *ideal filter bank* spectrum analyzer. Figure 1 depicts such a filter bank composed of digital filters whose impulse responses are denoted by $h_k(nT)$, $k = 0, 1, \cdots, M$, where $1/T$ is the sampling frequency of the input signal.* Such a filter bank constitutes an ideal spectrum analyzer if the input $x(nT)$ (with possibly further band limiting) can be synthesized exactly (within some fixed delay) by a linear

* The filter $h_0(nT)$ is a lowpass filter which is included for completeness although this band is usually not analyzed in practical speech analysis systems.

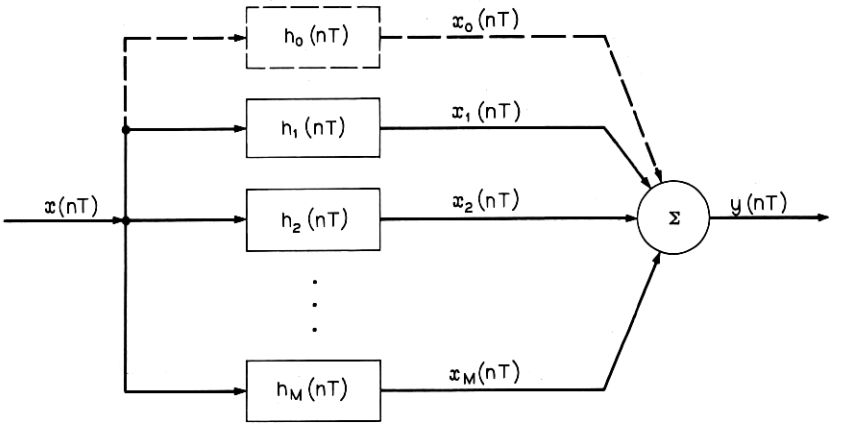

Fig. 1—Bank of digital bandpass filters.

combination of the bandpass filter outputs $x_k(nT)$. An example of such a system would be one in which the filters are ideal rectangular bandpass filters with the same constant gain and linear phase in their passbands and zero gain outside. If the filter bandwidths are such that the frequency range

$$\frac{-\pi}{T} \leqq \omega \leqq \frac{\pi}{T}$$

is completely covered without overlap, then the input can be synthesized exactly by adding together the outputs of the bandpass filters.

The essential characteristics of the ideal spectrum analyzer are that the frequency response of the combined outputs must exhibit a flat magnitude response and a linear phase, and therefore the combined impulse response must be a delayed digital impulse (unit sample). Causal digital filters (filters whose impulse responses are zero for $n < 0$) cannot have the desired ideal gain characteristics and may not have linear phase.* Therefore a filter bank composed of such filters cannot achieve the ideal characteristics of flat magnitude response and linear phase. In this paper we will describe an approach to the design of filter banks that approximate the ideal spectrum analyzer. First we present a detailed analysis of a filter bank configuration in which equally spaced, equal-bandwidth digital filters are used. This analysis suggests a technique for optimizing the filter bank characteristics and

* Finite duration impulse response digital filters can have precisely linear phase.

also suggests how the results can be used where nonuniform bandwidths are desired. The results are illustrated with examples.

II. ANALYSIS OF UNIFORM FILTER BANKS

Assume that the bandpass filters in Fig. 1 have impulse responses of the form

$$\begin{aligned} h_k(nT) &= 2 \mid D_k \mid h(nT) \cos(\omega_k nT + \Phi_k) \qquad k = 1, 2, \cdots, M \\ h_0(nT) &= \mid D_0 \mid h(nT) \end{aligned} \tag{1}$$

where $\omega_k = \omega_1 + (k-1)\,\Delta\omega$, and $h(nT)$ is the impulse response of a prototype lowpass filter. (When the $k = 0$ lowpass filter is used, $\omega_1 = \Delta\omega$ and $\omega_0 = 0$.) The system functions for this set of bandpass filters are

$$\begin{aligned} H_k(z) &= \mid D_k \mid e^{j\Phi_k} H(ze^{-j\omega_k T}) + \mid D_k \mid e^{-j\Phi_k} H(ze^{j\omega_k T}) \quad k = 1, 2, \cdots, M \\ H_0(z) &= \mid D_0 \mid H(z). \end{aligned} \tag{2}$$

The frequency response of these filters is obtained after substituting $z = e^{j\omega T}$ in (2) as

$$\begin{aligned} H_k(e^{j\omega T}) &= \mid D_k \mid e^{j\Phi_k} H(e^{j(\omega-\omega_k)T}) + \mid D_k \mid e^{-j\Phi_k} H(e^{j(\omega+\omega_k)T}), \\ &\qquad\qquad k = 1, 2, \cdots, M, \\ H_0(e^{j\omega T}) &= \mid D_0 \mid H(e^{j\omega T}). \end{aligned} \tag{3}$$

If the frequency response of prototype lowpass filter, $H(e^{j\omega T})$, drops off sharply, then it can be seen from (3) that

$$\begin{aligned} \mid H_k(e^{j\omega T}) \mid &\approx \mid D_k \mid \mid H(e^{j(\omega-\omega_k)T}) \mid \qquad 0 \leqq \omega \leqq \pi/T \\ &\approx \mid D_k \mid \mid H(e^{j(\omega+\omega_k)T}) \mid \quad -\pi/T \leqq \omega \leqq 0. \end{aligned}$$

In this case the filter bank consists of a set of $(M+1)$ equally spaced bandpass filters with identical magnitude responses around their respective center frequencies. We have chosen this method of designing bandpass filters from lowpass prototypes for analytical convenience and because of the importance of spectrum analysis systems of this form.[1,2] The results to be discussed apply for other bandpass transformations in so far as they yield a set of uniformly spaced bandpass filters with identical frequency characteristics.

Our objective is to choose the prototype lowpass filter and the parameters $\mid D_k \mid$, ω_1, $\Delta\omega$, and Φ_k so that the filter bank will closely approximate the characteristics of the ideal spectrum analyzer. To do this we must consider the response of the composite filter bank. First,

however, it is useful to interpret the individual bandpass filter outputs in terms of spectrum analysis considerations.

The individual filter outputs are of the form

$$x_k(nT) = \sum_{r=-\infty}^{n} 2x(rT) \mid D_k \mid h(nT - rT) \cos [\omega_k(nT - rT) + \Phi_k], \quad (4)$$

$$x_0(nT) = \sum_{r=-\infty}^{n} x(rT) \mid D_0 \mid h(nT - rT)$$

which can be expressed as

$$\begin{aligned} x_k(nT) &= 2 \ \ \mathrm{Re}\ \{D_k X(\omega_k , n)e^{j\omega_k nT}\} \\ x_0(nT) &= \mathrm{Re}\ \{D_0 X(0, n)\} \end{aligned} \quad (5)$$

where D_k is the complex constant defined by

$$\begin{aligned} D_k &= \mid D_k \mid e^{j\Phi_k} \qquad k = 1, 2, \cdots , M \\ D_0 &= \mid D_0 \mid, \end{aligned} \quad (6)$$

and

$$X(\omega_k , n) = \sum_{r=-\infty}^{n} x(rT)h(nT - rT)e^{-j\omega_k rT}. \quad (7)$$

The quantity $X(\omega_k , n)$ is the discrete-time version of the short-time Fourier transform[4] of $x(nT)$. Thus, (5) serves to relate the bandpass filter outputs $x_k(nT)$ to the short-time Fourier transform.

The frequency response and impulse response of the composite filter bank are obtained from

$$y(nT) = \sum_{k=0}^{M} x_k(nT). \quad (8)$$

After substituting (5) into (8) and noting that if $x(nT)$ is real, $X(-\omega_k , n)$ is the complex conjugate of $X(\omega_k , n)$, we obtain

$$y(nT) = \sum_{k=-M}^{M} D_k X(\omega_k , n)e^{j\omega_k nT} \quad (9)$$

where $\omega_{-k} = -\omega_k$, and D_{-k} is the complex conjugate of D_k. Substituting (7) into (9) and interchanging the order of summations results in

$$y(nT) = \sum_{r=-\infty}^{n} x(rT)\left[h(nT - rT) \cdot \sum_{k=-M}^{M} D_k e^{j\omega_k(nT-rT)}\right]. \quad (10)$$

Defining

$$d(nT) = \sum_{k=-M}^{M} D_k e^{j\omega_k nT}, \tag{11}$$

we observe from (10) that the combined impulse response of the filter bank can be expressed as

$$\tilde{h}(nT) = h(nT)\, d(nT). \tag{12}$$

Equations (11) and (12) are the basic results of the analysis. (Note that they could have been obtained directly by summing the impulse responses $h_k(nT)$, with the sacrifice of the interpretation of the filter bank outputs in terms of the short-time spectrum.)

Equation (12) shows that the filter bank impulse response is the product of the prototype lowpass filter impulse response $h(nT)$, and the sequence $d(nT)$ defined by (11). The choice of lowpass filter depends on both the desired frequency resolution and the requirement of obtaining flat magnitude and linear phase response in the composite filter bank. The sequence $d(nT)$ is independent of the prototype lowpass filter and is a function of the frequency spacing, the relative gains and phases, and the number of bandpass filters. Thus, for *a given choice of prototype lowpass filter*, the parameters of $d(nT)$ can be adjusted to achieve the best approximation to the ideal spectrum analyzer. To see how this occurs, we shall first examine in detail the characteristics of the sequence $d(nT)$.

As will be shown in the remainder of this section, a particularly useful choice of the complex coefficients D_k in (11) is

$$\begin{aligned} D_k &= e^{j\omega_k n_0 T} \qquad k = \pm 1, \pm 2, \cdots, \pm M. \\ D_0 &= 1, \end{aligned} \tag{13}$$

where n_0 is an integer. That is,

$$\Phi_k = \omega_k n_0 T, \quad \text{and} \quad |D_k| = 1.$$

(The condition $D_0 = 1$ implies that the band around $\omega = 0$ is included in the filter bank; $D_0 = 0$ implies that it is not.) It can be shown that for this choice of D_k, (11) becomes

$$d(nT) = D_0 + \frac{2 \sin [M\, \Delta\omega(nT + n_0T)/2]}{\sin [\Delta\omega(nT + n_0T)/2]} \cos [\omega_a(nT + n_0T)] \tag{14}$$

where $\omega_a = \omega_1 + (M - 1)\, \Delta\omega/2$, and D_0 is 1 or 0 depending on whether or not the lowpass channel is included.

The properties of the sequence $d(nT)$ determine the character of

the impulse response of the filter bank. Some of these properties are summarized below:

(*i*) The parameter n_0 shifts the sequence $d(nT)$ by n_0 samples with respect to $h(nT)$.

(*ii*) The sequence $d(nT)$ is even about the sample $n = -n_0$; i.e., $d(nT - n_0T) = d(-nT - n_0T)$.

(*iii*) The maximum value of $d(nT)$ occurs at $n = -n_0$. This value is $d(-n_0T) = 2M + D_0$.

(*iv*) If $2\pi/(\Delta\omega T)$ and $\omega_1/\Delta\omega$ are both integers, then the sequence $d(nT)$ is periodic with period $2\pi/\Delta\omega$. Otherwise, $d(nT)$ will be an almost periodic sequence which will peak up at time intervals of $2\pi/\Delta\omega$.

Insight into the properties of $d(nT)$ can be gained by considering a simple example. Assume that $\omega_1 = \Delta\omega$ and $2\pi/(\Delta\omega T) = N$ where N is an odd integer. That is, the entire frequency range $-\pi/T \leqq \omega \leqq \pi/T$, is divided into N equal bands. If $M = (N - 1)/2$, the entire frequency range is covered. Under these conditions (14) can be written

$$d(nT) = \frac{\sin\left[\left(\frac{2M+1}{2}\right)\Delta\omega(nT + n_0T)\right]}{\sin\left[\Delta\omega(nT + n_0T)/2\right]} - 1 \quad \text{if} \quad D_0 = 0 \tag{15a}$$

$$= \frac{\sin\left[\left(\frac{2M+1}{2}\right)\Delta\omega(nT + n_0T)\right]}{\sin\left[\Delta\omega(nT + n_0T)/2\right]} \quad \text{if} \quad D_0 = 1. \tag{15b}$$

(If N is even, the $k = 0$ filter is not used, i.e., $D_0 = 0$, and ω_1 is $\Delta\omega/2$.) It is clear from (15b) that for these conditions $d(nT)$ is a periodic sequence with period $NT = 2\pi/\Delta\omega$. In fact, $d(nT)$ may be thought of as samples of a continuous-time periodic Dirichlet kernel as shown in Fig. 2a. If $M = (N - 1)/2$ and $D_0 = 1$, $d(nT)$ is a periodic discrete-time impulse train with impulses occurring at multiples of NT. This is because the sample points on the periodic Dirichlet kernel occur at the maxima and the zero crossings, as indicated by the small circles in Fig. 2a.

The conditions for $d(nT)$ to be periodic are that both $\omega_1/\Delta\omega$ and $2\pi/(\Delta\omega T)$ be equal to integers. To see this, we must examine (14) in detail. If $2\pi/(\Delta\omega T)$ is an integer, and M is an odd integer, the sequence $2 \sin [M\ \Delta\omega(nT + n_0T)/2]/\sin [\Delta\omega(nT + n_0T)/2]$ is periodic with period $NT = 2\pi/\Delta\omega$. If $\omega_1/\Delta\omega$ is an integer and M is odd, the sequence $\cos [(\omega_1 + (M - 1)\ \Delta\omega/2)(nT + n_0T)]$ is also periodic with a period

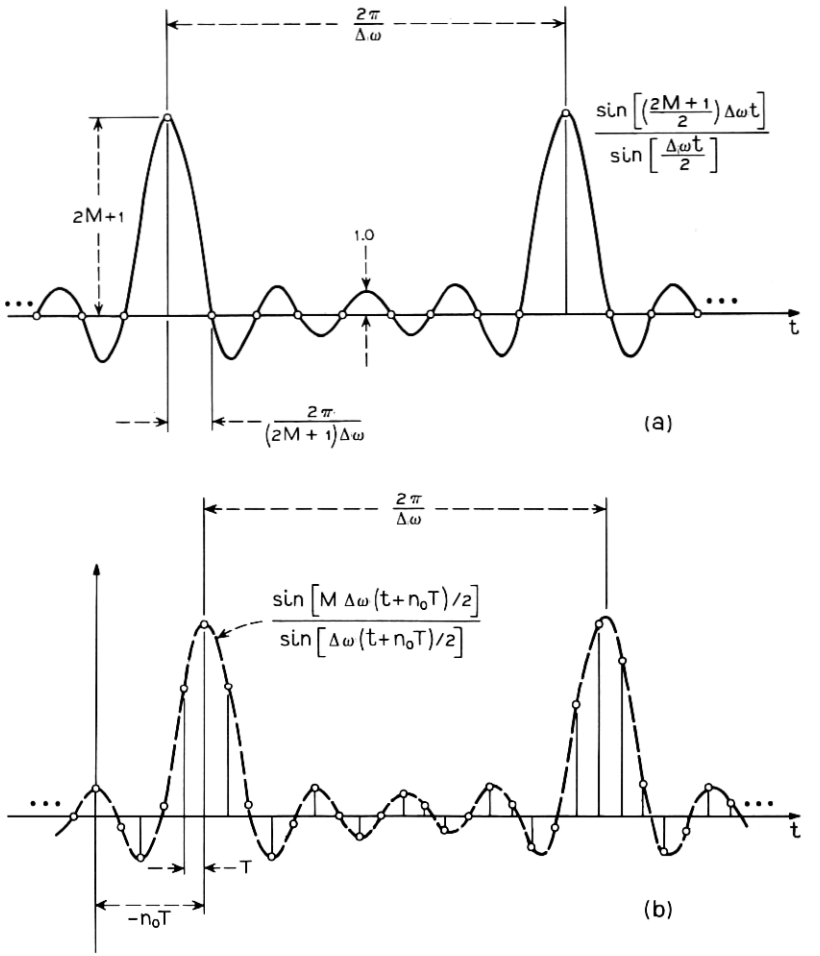

Fig. 2—(a) Periodic continuous-time Dirichlet kernel, (b) continuous-time envelope and sequence $d(nT)$ when either $\omega_1/\Delta\omega$ or $2\pi/(\Delta\omega T)$ are not integers.

that is an integer multiple of $2\pi/\Delta\omega$. Thus the product of these two sequences is periodic with period $2\pi/\Delta\omega$. The identical result holds for M an even integer and $2\pi/\Delta\omega$ and $\omega_1/\Delta\omega$ integers although the interaction between the component sequences is slightly different.

If either $2\pi/(\Delta\omega T)$ or $\omega_1/\Delta\omega$ are not integers, $d(nT)$ will not be periodic, but will still peak up at time intervals of $2\pi/\Delta\omega$. Such a case is depicted in Fig. 2b where the samples $d(nT)$ are marked by the

small circles and the dotted curve shows the factor

$$\sin [M\ \Delta\omega(nT + n_0T)/2]/\sin [\Delta\omega(nT + n_0T)/2]$$

when M is odd. As shown in Fig. 2b, $d(nT)$ will always have even symmetry about sample $n = -n_0$.

III. DESIGN OF UNIFORM FILTER BANKS USING BESSEL FILTERS

In the preceding section we presented a detailed analysis of a filter bank composed of equally spaced equal-bandwidth filters. In this section we will show how the results of that analysis can be employed in filter bank design.

The objective of flat amplitude response and linear phase is most easily achieved with bandpass filters having these same properties. For this reason, Bessel (maximally flat delay) filters are often used in filter banks.[5] In the examples shown in this paper, we have used digital filters obtained from Bessel prototype designs using impulse invariance.[6] It should be noted that the digital filters obtained this way do not have the maximally flat delay property. J. P. Thiran[7] has shown that the denominator of the system function of maximally flat delay digital filters is a Gauss hypergeometric function. It is reasonable to expect however, that for the narrow-band filters of interest here, the differences should be negligible.

As an example a digital filter derived from a sixth-order Bessel lowpass filter with asymptotic cutoff frequency of 60 Hz is shown in Fig. 3. The impulse response is shown in Fig. 3a, and the amplitude and phase responses are shown in Fig. 3b and Fig. 3c. The filter shown in Fig. 3 was used in a filter bank* with the following choice of parameters: $D_0 = 0$, $T = 10^{-4}$ sec, $\Delta\omega = 2\pi(100)$, $\omega_1 = 2\pi(100)$, $n_0 = 0$, and $M = 30$. The resulting filter bank characteristics are shown in Fig. 4. The filter bank impulse response, $\tilde{h}(nT)$, is shown in Fig. 4a along with the prototype lowpass impulse response $h(nT)$. For the above choice of parameters, $d(nT)$ is obtained from (15a) as

$$d(nT) = \frac{\sin [0.61\pi n]}{\sin [0.01\pi n]} - 1, \tag{16}$$

which is periodic with period 100 samples (10 msec), with peaks occurring at $nT = 0, \pm 10, \pm 20, \cdots$ msec. From Fig. 4a, it can be seen that in the product $h(nT) \cdot d(nT)$, the peak of $d(nT)$ at $nT = 0$ will be attenuated since $h(nT)$ is small around $nT = 0$. On the other hand,

* Note that the resulting bandpass filters are twelfth order.

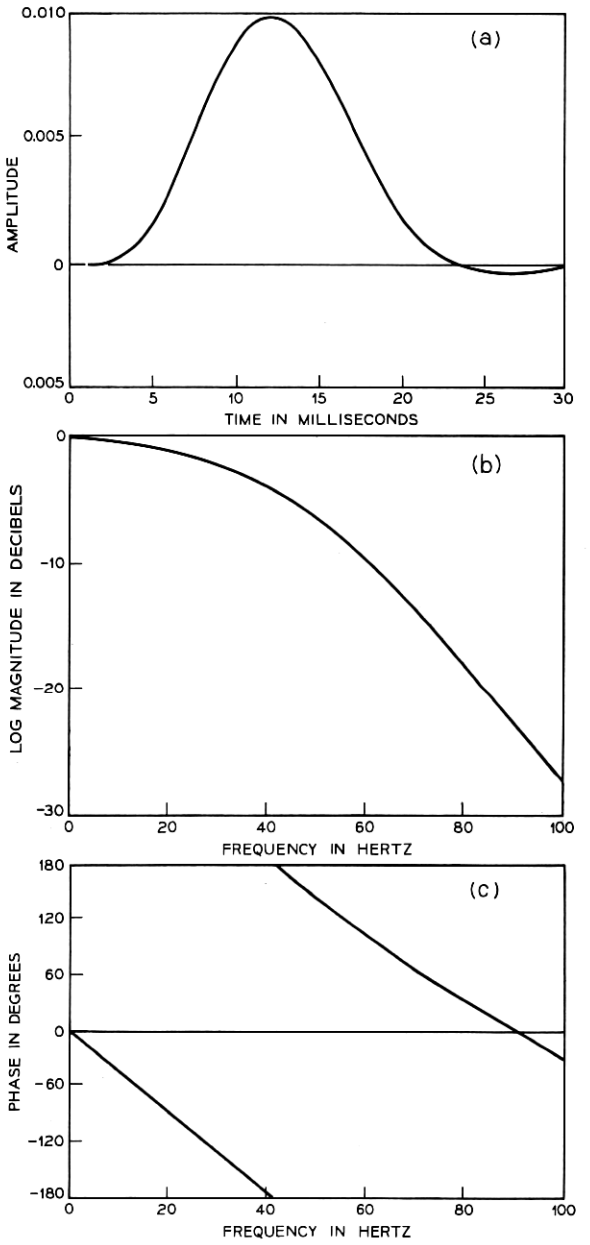

Fig. 3—Sixth-order Bessel filter characteristics. (a) Impulse response, (b) magnitude response, (c) phase response.

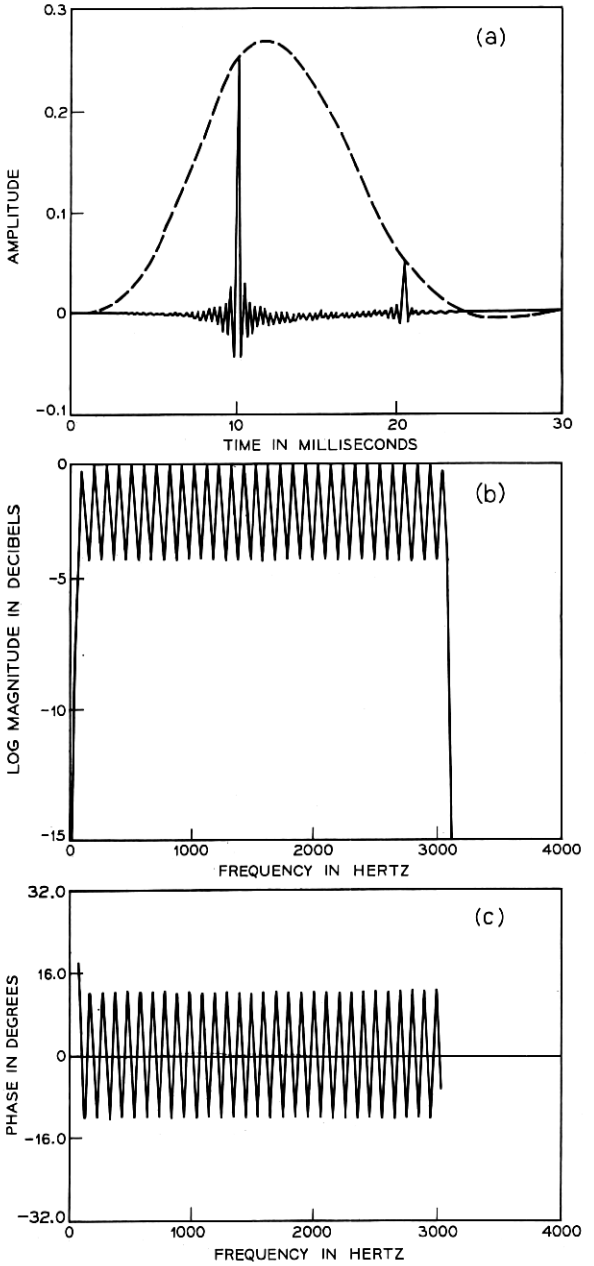

Fig. 4—Characteristics of 30-channel filter bank. (a) Impulse response (dotted curve is the impulse response of the prototype lowpass filter in Fig. 3), (b) composite magnitude response, (c) composite phase response after subtracting 10-msec delay.

the peak of $d(nT)$ at $nT = 10$ msec occurs at approximately the peak of $h(nT)$, and at $nT = 20$ msec, $h(nT)$ is large enough to produce a significant echo in the impulse response of the filter bank. As is shown in Fig. 4b and 4c, this corresponds to a 3.9-dB ripple in the amplitude response and a 25.5-degree peak-to-peak ripple in the phase response (after removing a linear phase component corresponding to a 10-msec or 100-sample delay). To decrease this amplitude and phase ripple, we should attempt to eliminate the echo in the impulse response. Furthermore, the phase ripple will be eliminated if the impulse response $h(nT)$ has even symmetry about some delay time $n_D T$. One approach is to broaden the filter bandwidths, or equivalently reduce the spacing $\Delta\omega$, so that $h(nT)$ is contracted relative to the spacing of pulses in $d(nT)$. This is generally not an acceptable solution since $h(nT)$ and $\Delta\omega$ are usually fixed by some frequency resolution criterion. However, if we refer to the properties of $d(nT)$ which were previously summarized, we note that a negative value of n_0 will shift $d(nT)$ to the right relative to $h(nT)$ so that $d(nT)$ will have even symmetry about time $n_D T = -n_0 T$. If n_0 can be chosen so than $\tilde{h}(nT) = h(nT) \cdot d(nT)$ has approximately even symmetry and consists of only one significant pulse, then the amplitude and phase ripple will be small. The manner in which this is achieved is shown in Fig. 5 where it is assumed for simplicity that $d(nT)$ is a train of digital impulses as would be the case for $M = (N - 1)/2$. Figure 5a depicts the case where $n_0 = 0$. Figure 5b shows the situation where n_0 was chosen to shift the impulse which was at $nT = 0$ in Fig. 5a to the right and into the vicinity of the peak of $h(nT)$. If it is assumed that only three impulses have nonzero amplitudes (α_1, α_2, α_3) such that $4 \mid \alpha_1 \mid \cdot \mid \alpha_3 \mid < \mid \alpha_1 + \alpha_3 \mid \cdot \mid \alpha_2 \mid$, then it can be shown (see Appendix) that the peak-to-peak amplitude ripple of the filter bank is

$$R_A = 20 \log_{10} \left[\frac{\mid \alpha_2 + \alpha_1 + \alpha_3 \mid}{\mid \alpha_2 - \alpha_1 - \alpha_3 \mid} \right]. \tag{17}$$

Similarly, if $\mid \alpha_1 + \alpha_3 \mid < \mid \alpha_2 \mid$, the peak-to-peak phase ripple about a linear phase corresponding to a delay of $-n_0 T$ is given by

$$R_P = 2 \tan^{-1} \left[\frac{\alpha_1 - \alpha_3}{(\alpha_2^2 - (\alpha_1 + \alpha_3)^2)^{1/2}} \right]. \tag{18}$$

The conditions for (17) and (18) to hold are satisfied when α_1 and α_3 are small relative to α_2, which is the normal situation. It can be seen from (18) and (17) that the phase ripple will be zero if $\alpha_1 = \alpha_3$, and the amplitude ripple will be small if $(\alpha_1 + \alpha_3)/\alpha_2$ is small.

Although these results were derived for the idealized case when $d(nT)$

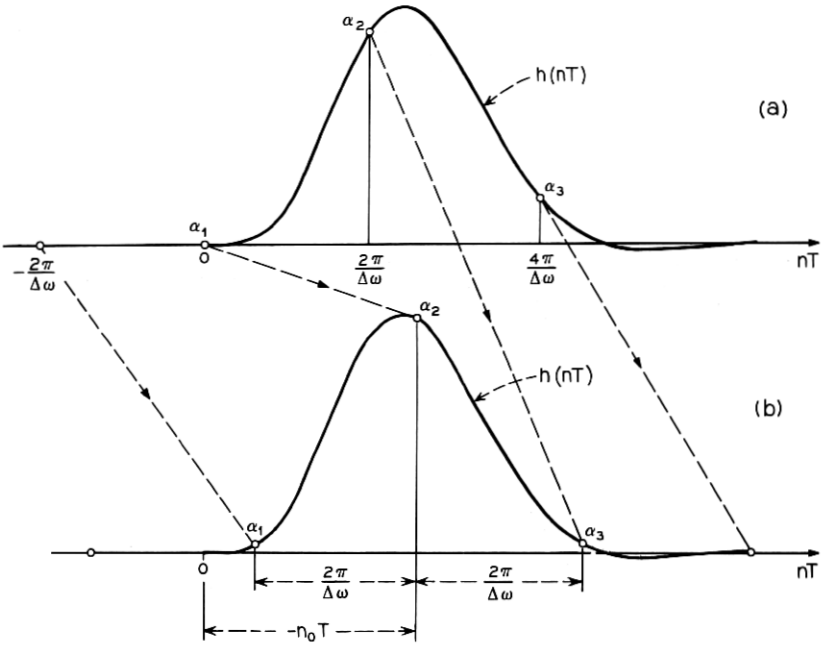

Fig. 5—Illustration of how to adjust the parameter n_0. (a) Composite impulse response for $n_0 = 0$, (b) n_0 chosen to minimize magnitude and phase ripple (dotted lines indicate movement of individual pulses in $d(nT)$).

is an impulse train, we have found that amplitude and phase ripple can be determined quite accurately using (17) and (18) in more general situations. With the foregoing principles in mind we have written an interactive computer program for filter bank design. Using this program we can design a filter bank with low amplitude and phase ripple by the following process:

(*i*) Choose ω_1, $\Delta\omega$, and M to cover the desired analysis band and choose $h(nT)$ to provide desired frequency resolution. This results in an $h(nT)$ that has a duration of approximately $4\pi/\Delta\omega$ as shown in Fig. 5.

(*ii*) Evaluate $h(nT)$ and determine n_0 such that $\alpha_1 \approx \alpha_3$ as in Fig. 5b.

(*iii*) If the resulting filter bank is not satisfactory, steps *i* and *ii* are repeated.

In cases where $\omega_1/\Delta\omega$ is not an integer, it is important to choose n_0

so that the point of even symmetry in $d(nT)$ is shifted into the vicinity of the peak of $h(nT)$. Otherwise, it may be impossible to achieve a very good approximation to linear phase. An example of the improvement gained by proper choice of n_0 is shown in Fig. 6. In this example all the parameters were the same as in the example of Fig. 4 except a value of $n_0 = -129$ was chosen by the above process. In this case the in-band amplitude ripple is 0.8 dB and the phase ripple is 0.6 degree, as compared to 3.8 dB and 25.5 degrees when $n_0 = 0$.

R. M. Golden[5] has shown that inverting the sign of alternating channels often significantly improves the characteristics of a filter bank. This technique has a simple interpretation in terms of our results. It can be shown that inverting the sign of alternating channels is equivalent to delaying the sequence $d(nT)$ by $n_0 = -\pi/(\Delta\omega T)$ samples. This amount of delay may be nearly correct if the duration of $h(nT)$ is approximately $3\pi/\Delta\omega$; however for the situation shown in Fig. 4a, such a delay would produce a worse filter bank than no delay at all ($n_0 = 0$). Also, to achieve linear phase when $\omega_1/\Delta\omega$ is not an integer, the point of even symmetry in $d(nT)$ should be delayed to the vicinity of the peak of $h(nT)$. This does not occur when the signs of alternate channels are inverted.

IV. DESIGN OF NONUNIFORM BANDWIDTH FILTER BANKS

In speech applications it is common to take advantage of the frequency resolution characteristics of the ear[4,5] by using increasing bandwidth filters at higher frequencies. The previously discussed techniques can be applied to this situation if the filter bank consists of several sub-banks, each with different resolution. Each sub-bank can be designed as discussed above, with care being taken to ensure that the entire frequency band of interest is covered by the combination of the sub-banks. It may be necessary to equalize the delay between sub-banks by providing additional delay for all but one of the sub-banks.* This is depicted in Fig. 7 for three sub-banks with increasing-bandwidth sixth-order Bessel filters. Figure 7a shows the lowpass prototype impulse response and shifted $d(nT)$ sequence† for the first sub-bank. The lowpass asymptotic cutoff used was 78 Hz, the spacing of filters was $\Delta\omega_1 = 2\pi(125)$, the first filter was centered at $\omega_{11} = 2\pi(250)$, and a value of $n_{01} = -100$ (10-msec delay) was required to

* Golden[5] has shown that the delays can be approximately equalized by increasing the order of the lowpass prototype in direct proportion to the increase in bandwidth.

† The sequence $d(nT)$ is shown as an impulse train for convenience in plotting. The actual sequences would look like those in Fig. 4 and Fig. 6.

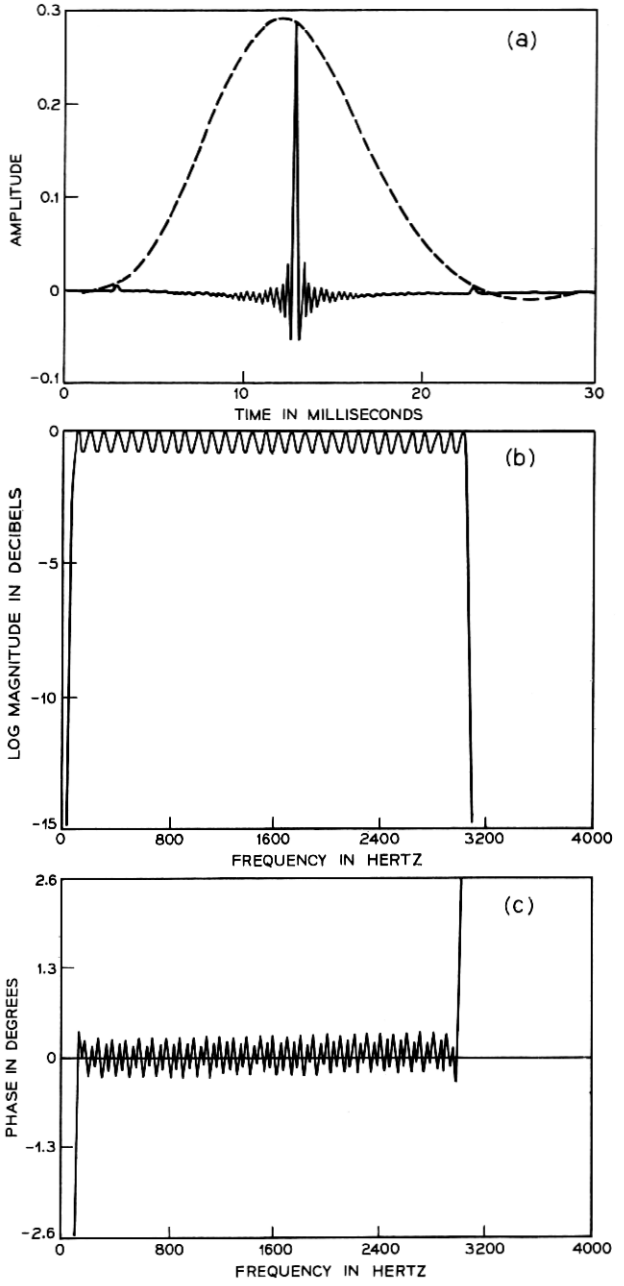

Fig. 6—Characteristics of 30-channel filter bank. (a) Impulse response for $n_0 = -129$ (dotted curve is the impulse response in Fig. 3), (b) composite magnitude response, (c) composite phase response after subtracting 12.9-msec delay.

minimize the amplitude and phase ripple. Figure 7b shows the second sub-bank in which the basic parameters were: lowpass asymptotic cutoff 136 Hz, $\Delta\omega_2 = 2\pi(218)$, $\omega_{12} = 2\pi(1296.5)$, and $n_{02} = -57$ (5.7-msec delay). To line up the central peaks, an additional delay of $n_2 = 43$ samples (4.3 msec) was required. Figure 7c shows the third sub-bank where the lowpass cutoff was 192 Hz, $\Delta\omega_3 = 2\pi(307)$, $\omega_{13} = 2\pi(2213)$, and $n_{03} = -40$ (4-msec delay). A value of $n_3 = 60$ samples (6.0 msec) is required to line up the central peak with those in Fig. 7a and 7b. The response of the combination of these three sub-banks is shown in Fig. 8. Figure 8a shows the impulse response, Fig. 8b shows the amplitude response, and Fig. 8c shows the phase after a linear phase corresponding to 10-msec delay has been subtracted. It can be

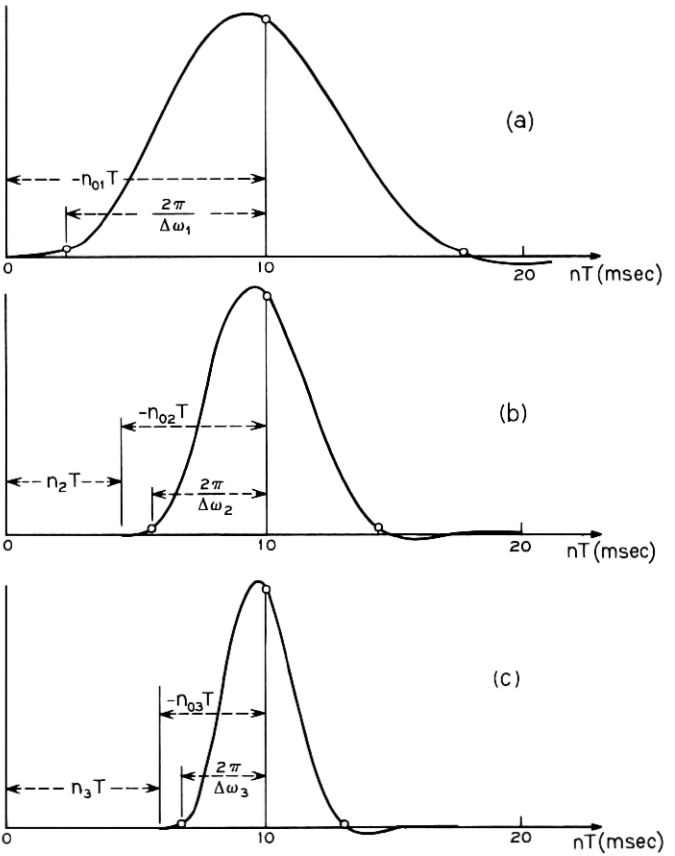

Fig. 7—Illustration of the design of nonuniform filter banks: (a) impulse response for narrow bandwidth filters, (b) impulse response for intermediate bandwidth filters, (c) impulse response for wide bandwidth filters.

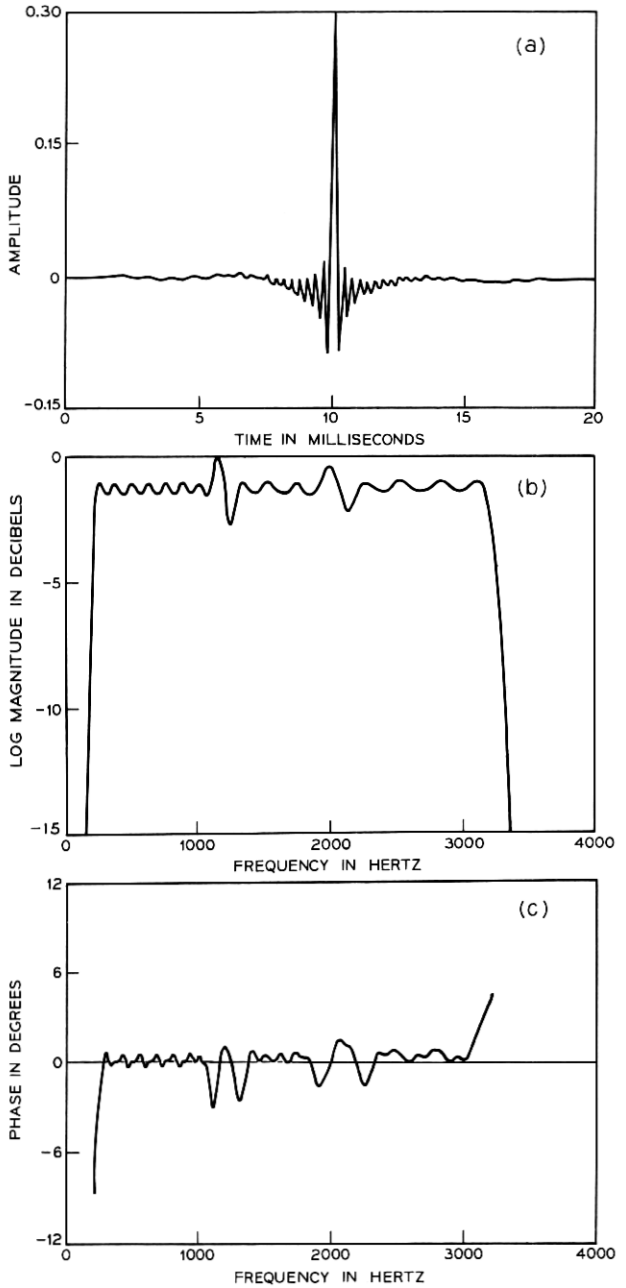

Fig. 8—Composite filter bank characteristics with three different sub-banks: (a) impulse response, (b) magnitude response, (c) phase response after subtracting 10-msec delay.

seen in Fig. 8b and 8c that the ripple in the sub-banks is quite low as would be expected from Fig. 7. At the boundary between sub-banks, however, the ripple increases significantly due to the fact that the last filter in the lower sub-bank drops off more rapidly than the first filter in the next sub-bank. This excessive variation at the boundary between sub-banks can be eliminated to some extent by using increasingly higher-order filters in the sub-banks. Alternatively, nonuniform resolution can be obtained by using equal-bandwidth filters and adding together groups of two or more of their outputs to achieve the desired bandwidth. Such an approach would require increased computation but would produce filter bank characteristics comparable to those in Fig. 6.

V. CONCLUSION

We have discussed the analysis and design of digital filter banks and have shown how the incorporation of a linearly increasing phase shift in each bandpass filter can significantly improve the overall filter bank characteristics. We also showed how the techniques can be used in nonuniform bandwidth filter banks.

The examples which we gave were based on Bessel lowpass prototypes which have impulse responses of desirable shape but rather poor amplitude response. Recent results in the design of finite duration impulse response filters[8] offer attractive possibilities for filter bank design. Such filters can have precisely linear phase and can be designed using iterative techniques with constraints on both the impulse response shape and the amplitude response. The use of such filters, together with the basic principles discussed in this paper, should yield filter banks with excellent properties.

APPENDIX

Derivation of Magnitude and Phase Ripple Formulas

Assume an impulse response sequence

$$\begin{aligned} h(n) &= \alpha_1 \qquad n = 0 \\ &= \alpha_2 \qquad n = n_p \\ &= \alpha_3 \qquad n = 2n_p \\ &= 0 \qquad \text{elsewhere.} \end{aligned} \tag{19}$$

The system function of this system is

$$H(e^{j\omega T}) = \alpha_1 + \alpha_2 e^{-j\omega n_p T} + \alpha_3 e^{-j\omega 2n_p T}. \tag{20}$$

The squared magnitude response is

$$| H(e^{j\omega T}) |^2 = [\alpha_2 + (\alpha_1 + \alpha_3) \cos (\omega n_p T)]^2 + (\alpha_1 - \alpha_3)^2 \sin^2 (\omega n_p T), \tag{21}$$

and the phase response is

$$\arg [H(e^{j\omega T})] = \tan^{-1} \left[\frac{(\alpha_1 - \alpha_3) \sin (\omega n_p T)}{\alpha_2 + (\alpha_1 + \alpha_3) \cos (\omega n_p T)} \right] \tag{22}$$

where a linear phase component $-\omega n_p T$ has been removed. Clearly, both (21) and (22) are periodic functions of ω with period $2\pi/n_p T$. To determine the amplitude and phase ripple, we must locate the maxima and minima of (21) and (22).

If we differentiate (21) with respect to ω, we find that the maxima and minima occur for values of ω satisfying

$$\sin (\omega n_p T) = 0 \tag{23a}$$

$$\cos (\omega n_p T) = -\alpha_2 \frac{(\alpha_1 + \alpha_3)}{4\alpha_1 \alpha_3}. \tag{23b}$$

The second equation is satisfied by a real value of ω if and only if

$$4 | \alpha_1 | \cdot | \alpha_3 | > | \alpha_1 + \alpha_2 | \cdot | \alpha_2 |. \tag{24}$$

In a good filter bank design, α_1 and α_3 will be positive and much smaller than α_2, and (24) will not be satisfied. Evaluating the second derivative shows that in this case the maxima and minima of $| H(e^{j\omega T}) |$ will alternate and occur at values of ω satisfying (23a); i.e., $\omega = 0, \pm\pi/n_p T, \pm 2\pi/n_p T, \cdots$. In this case the amplitude ripple in dB is given by

$$R_A = 20 \log_{10} \left[\frac{| \alpha_2 + \alpha_1 + \alpha_3 |}{| \alpha_2 - \alpha_1 - \alpha_3 |} \right]. \tag{25}$$

If (22) is differentiated with respect to ω, we find that the maxima and minima occur at values of ω satisfying

$$\cos \omega n_p T = -\left(\frac{\alpha_1 + \alpha_3}{\alpha_2} \right). \tag{26}$$

Equation (26) is satisfied by real values of ω if $| \alpha_1 + \alpha_3 | < | \alpha_2 |$. In this case the maxima and minima again alternate, and the peak-to-peak phase ripple is

$$R_p = 2 \tan^{-1} \left[\frac{\alpha_1 - \alpha_3}{(\alpha_2^2 - (\alpha_1 + \alpha_3)^2)^{1/2}} \right]. \tag{27}$$

If $|\alpha_1 + \alpha_3| > |\alpha_2|$, the phase curve will be discontinuous with a jump of 2π radians occurring at $\omega = \pm\pi/n_pT, \pm 3\pi/n_pT, \cdots$.

REFERENCES

1. Flanagan, J. L., and Golden, R. M., "Phase Vocoder," B.S.T.J., *45*, No. 9 (November 1966), pp. 1493–1509.
2. Flanagan, J. L., and Lummis, R. C., "Signal Processing to Reduce Multipath Distortion in Small Rooms," J. Acoust. Soc. Am., *47*, No. 1 (June 1970), pp. 1475–1481.
3. Jackson, L. B., Kaiser, J. F., and McDonald, H. S., "An Approach to the Implementation of Digital Filters," IEEE Trans. Audio and Electroacoust., *AU-16*, No. 3 (September 1968), pp. 413–421.
4. Flanagan, J. L., *Speech Analysis, Synthesis and Perception*, New York: Academic Press, 1965.
5. Golden, R. M., "Vocoder Filter Design: Practical Considerations," J. Acoust. Soc. Am., *43*, (December 1967) pp. 803–810.
6. Gold, B., and Rader, C. M., *Digital Processing of Signals*, New York: McGraw-Hill Book Co., 1969.
7. Thiran, J. P., "Recursive Digital Filters with Maximally Flat Group Delay," IEEE Trans. Ckt. Theory, *CT-18*, No. 4 (November 1971).
8. Rabiner, L. R., "Techniques for Designing Finite Duration Impulse Response Digital Filters," IEEE Trans. Com. Tech., *COM-19*, No. 2 (April 1971), pp. 188–195.