ISCA Archive
http://www.isca-speech.org/archive

First European Conference on
Speech Communication and Technology
EUROSPEECH '89
Paris, France, September 27-29, 1989

Non-Uniform RFT Filterbank Design For Speech Processing

F J Owens * & M S Murphy **

* University of Ulster ** Marconi Underwater Systems

ABSTRACT

This paper describes a technique for designing uniform and non-uniform filterbanks based on the Running Fourier Transform (RFT). The RFT is implemented by convolving the input signal with one of a family of windows, $h(nT) = (nT)^k e^{-\alpha nT}$, where k and α may be chosen to specify the order and bandwidth, respectively, of each analysing filter. An expression for the equivalent composite impulse response has been derived which can be used to optimise its composite amplitude and phase responses. Non-uniform filterbanks can be designed by splitting the frequency band of interest into a number of uniform sections and then optimising the equivalent composite impulse response of each section. Finally, a modified cepstral smoothing technique for non-uniform spectra is presented and shown to be superior to conventional bi-pass filtering.

1. Introduction

The Running Fourier Transform [Flanagan, 1972] is a method for deriving the spectral content, $F(\omega,nT)$, of a quasi-stationary signal at any instant in time. For a discrete-time signal, it is expressed as

$$F(\omega,nT) = \sum_{r=-\infty}^{n} f(rT).h(nT-rT)e^{-j\omega rT} \quad (1)$$
$$= a(\omega,nT) - jb(\omega,nT),$$

where

$$a(\omega,nT) = [f(nT).\cos(\omega nT)]*h(nT), \quad (2)$$
$$b(\omega,nT) = [f(nT).\sin(\omega nT)]*h(nT),$$

h(nT) is the impulse response of a lowpass filter and * denotes discrete convolution. By definition, the magnitude spectrum, $|F(\omega,nT)|$, is given by

$$|F(\omega,nT)| = [\bar{F}(\omega,nT).F(\omega,nT)]^{1/2} \quad (3)$$
$$= [a^2(\omega,nT) + b^2(\omega,nT)]^{1/2}$$

where $\bar{F}(\omega,nT)$ denotes the complex conjugate of $F(\omega,nT)$. Fig 1 shows a system, using real operations only, for carrying out the spectral measurement indicated by equation (1).

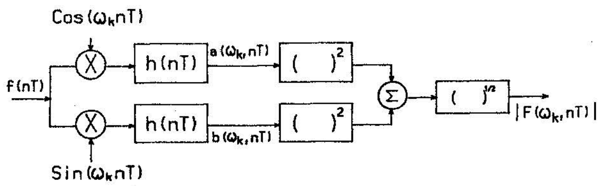

Fig 1 : Implementation of Running Fourier Transform

From a frequency-domain perspective, Fig 1 shows that the input signal is heterodyned to base-band and then filtered by a lowpass filter with impulse response h(nT). The result is effectively a bandpass filter with centre frequency ω and a passband extending from $(\omega - \omega_c)$ to $(\omega + \omega_c)$, where ω_c is the radian bandwidth of the lowpass filter. By carrying out the measurement indicated by Fig 1 at a number of different analysing frequencies, ω_k, a running spectrum can be obtained. The advantage of this method of spectral derivation over, for example, the Discrete Fourier Transform is that analysing bandwidth and spectral resolution are independently variable and therefore spectral leakage across channels can be controlled.

An appropriate family of lowpass filter impulse responses, h(nT), for use with the Running Fourier Transform are defined by [Owens, 1988]

$$h(nT) = (nT)^k.e^{-\alpha nT}, \text{ k integer.} \quad (4)$$

The above family of lowpass functions have order k+1 and a 3dB radian cut-off frequency ω_c given by

$$\omega_c = \alpha.(2^{1/(k+1)} - 1)^{1/2}. \quad (5)$$

The z-domain transfer function, H(z), of this family of filters is

$$H(z) = T^k[\sum_{n=1}^{k} a_n(e^{-\alpha T}z^{-1})^n]/[1 + \sum_{m=1}^{k+1} b_m(e^{-\alpha T}z^{-1})^m]. \quad (6)$$

A wide variety of filterbank configurations can therefore be realised by appropriate choice of ω, α and k.

2. Non-Uniform Filterbank Design For Spectral Analysis

A desirable feature in filterbank design is that the composite frequency response of the bank should be flat with linear phase. This is not too difficult to achieve if uniform channel spacing and constant bandwidth suffice. However if non-uniform spectral resolution and definition are required then the problem of obtaining a flat composite response becomes extremely difficult. The composite amplitude response of the filterbank can be measured by summing the response of each channel to a swept-frequency input signal. This can be efficiently computed from an expression for the frequency response of each channel. Replacing the dummy summation variable r in equation (1) by n-r gives

$$F(\omega,nT) = \sum_{r=0}^{\infty} f(nT-rT).h(rT).e^{-j\omega(n-r)T} \quad (7)$$
$$= e^{-j\omega nT}[a_1(\omega,nT) + jb_1(\omega,nT)]$$

10.21437/Eurospeech.1989-157

where

$$a_1(\omega,nT) = f(nT)*h(nT)\cos(\omega nT)$$

and (8)

$$b_1(\omega,nT) = f(nT)*h(nT)\sin(\omega nT).$$

The magnitude spectrum, $|F(\omega,nT)|$, is now given by

$$|F(\omega,nT)| = [a_1^2 + b_1^2]^{1/2} \quad (9)$$

The above equations show that the RFT may be alternatively viewed as filtering the input signal with bandpass filters with impulse responses $h(nT)\cos(\omega nT)$ and $h(nT)\sin(\omega nT)$, squaring and adding the magnitude of each filter output and then taking the square-root. By deriving the z-domain transfer function of each bandpass filter, squaring and adding the magnitude responses and finally taking the square-root, an expression for the overall amplitude response of each channel can be derived.

Consider first the case of uniform channel spacing. As previously indicated, the filtering action of the system in Fig 1 can be viewed in terms of bandpass filtering, though its impulse response is the window function $h(nT)$. A true bandpass filter would have an impulse response $h(nT)\cos(\omega_k nT)$. Assuming no channel centred on dc, the composite impulse response, $p(nT)$, of an M-channel bank of bandpass filters is

$$p(nT) = h(nT)d(nT), \qquad d(nT) = \sum_{k=1}^{M} \cos(\omega_k nT) \quad (10)$$

If the channel spacing is fixed, then the function $d(nT)$ depends only on the number of channels M and the channel spacing $\delta\omega$ and can be expressed in closed form as

$$d(nT) = [\sin(N\delta\omega nT/2)]/[\sin(\delta\omega nT/2)] \quad (11)$$

For an odd number of channels covering the double-sided frequency band of interest, $N = 2M + 1$ and for an even number, $N = 2M$. In general, $d(nT)$ is a sequence of pulses, of maximum amplitude N, occurring at intervals of $2\pi/\delta\omega$. When the double-sided band of interest covers the range $-\pi/T$ to $+\pi/T$, then $d(nT)$ becomes an impulse train of amplitude N.

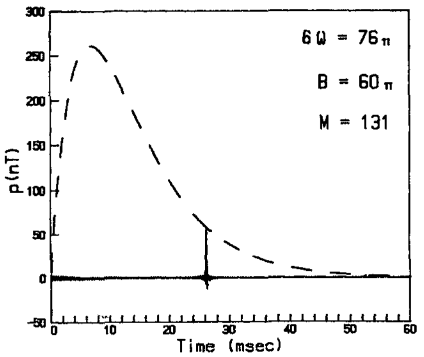

Fig 2 : Conceptual Composite Impulse Response of RFT Filterbank.

In the composite impulse response of the filterbank (Eq 10), the function $d(nT)$ is weighted by the impulse response of the lowpass filter, $h(nT)$. By iteratively adjusting the channel spacing and bandwidth, a composite impulse response can be obtained which approaches the ideal, that is a single delayed impulse. This is illustrated in Fig 2. Note that an RFT filterbank measures the spectral content of a signal by repeatedly heterodyning and lowpass filtering, whereas the composite impulse response shown is for an equivalent bank of bandpass filters.

It is possible to extend the above method of design to obtain a degree of non-uniformity in the channel spacing and bandwidth of the filterbank. This is achieved by dividing the frequency band of interest into a number of sections. Each section must have uniform resolution and definition but the values of each can vary from section to section. The equivalent composite impulse response can be computed for each section and the design procedure applied to each in order to optimise the overall composite frequency response.

To illustrate the design procedure, consider a filterbank, covering the range dc to 4500Hz, which is split into three 1500Hz sections with the channel spacing and bandwidth increasing by a factor of 2 from section to section. The complete initial specification for this filterbank (FB1) is given in Table 1. To minimise distortion in the composite amplitude response due to step increases in bandwidth at section boundaries, the filter orders also increase from section to section. Fig 3 gives its composite amplitude response. It is clear that the response deviates significantly from the ideal.

By iteratively adjusting the channel bandwidth, the number of channels, the channel spacing and filter order, in conjunction with viewing the resulting composite impulse response, $p(nT)$, for each section, a composite amplitude response which approaches the ideal flat characteristic can be achieved. By further iteratively adjusting the bandwidth in each section and observing the resulting composite amplitude response a certain degree of fine-tuning of the filterbank design can be achieved. Fig 4 shows the composite amplitude response for a filterbank (FB2) with a specification given in Table 1, which was designed by optimising the specification of FB1 in the above way.

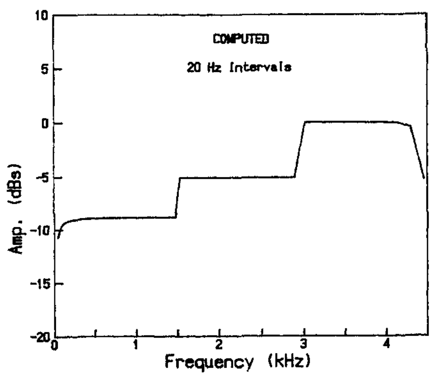

Fig 3 : Composite Amplitude Response of RFT Filterbank FB1.

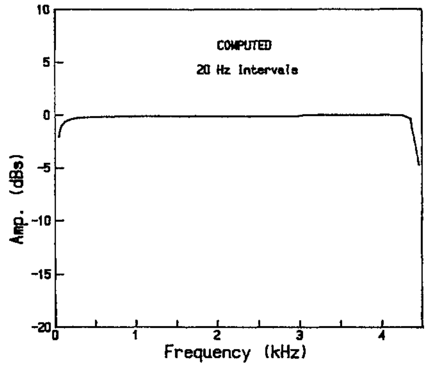

Fig 4 : Composite Amplitude Response of Optimised Filterbank FB2

Parameter	Section 1		Section 2		Section 3	
	FB1	FB2	FB1	FB2	FB1	FB2
Channels	37	37	18	25	10	15
1^{st} Chan.(Hz)	40	40	1540	1540	3020	3060
Spacing (Hz)	40	40	80	60	160	100
Filter Order	2	2	3	3	5	5
Bandwidth(Hz)	32	35	64	44	128	72

Table 1 Filterbank Specifications.

3. Non Uniform Filterbank Design For Analysis/Synthesis

In the design of filterbanks for speech analysis/synthesis applications, the main goal is to obtain a flat amplitude response and a linear phase response. Reconstruction of the input signal f(nT) can be achieved by adding a synthesis stage to the analysis procedure as shown in Fig 5, where a delay variable n_o has now been incorporated into the front-end heterodyning vector. With the inclusion of the synthesis stage, the overall impulse response of a single channel is $h(nT)\cos[\,\omega_k(n-n_o)T\,]$. Including n_o in Eq 10 gives,

$$p(nT) = h(nT)d(nT - n_oT), \qquad (12)$$

where n_o (fixed) has the effect of introducing a

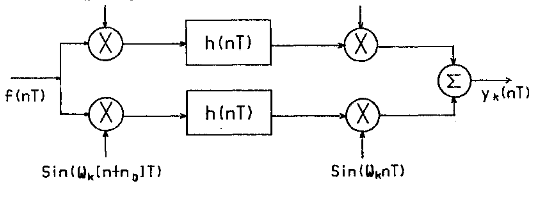

Fig 5 : Schematic of RFT Analysis/Synthesis Channel

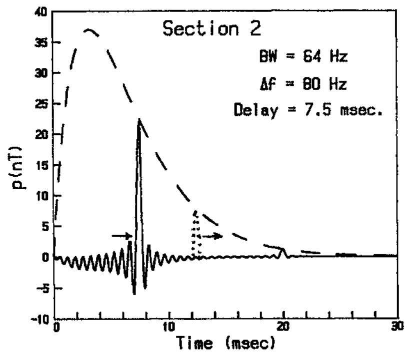

Fig 6 : Typical Composite Impulse Response of Filterbank Section

linear phase component to each channel and can be used as an additional parameter for improving composite responses. Specifically, n_o is used to alter the position of d(nT) relative to h(nT). From a plot of the composite impulse response function, p(nT), it is possible to determine : (i) the phase offset of each frequency component due to the delay of the main pulse; (ii) the additional delay n_o required to minimise amplitude and phase ripple; and (iii) the likely influences of secondary impulses (reverberation).

In order to minimise the amplitude and phase ripple for the family of windows used here, the optimum filterbank configuration should initially (ie $n_o = 0$) produce a single low-amplitude pulse in the composite impulse response. Since the function d(nT) is periodic and has a narrow pulse occurring at zero time, it can be time-shifted to give a high-amplitude pulse in p(nT). The objective is then to choose a time delay which shifts the existing pulse to a location where its amplitude is greater than that of the original pulse. This is illustrated in Fig 6 where the original pulse is shown dotted . It may also be necessary to align the composite impulse responses [p(nT)] in order to equalise the delay between sections. The results of applying this design procedure to filterbank FB1 in Table 1 is illustrated in Figs 7 and 8. Fig 7 shows the original composite amplitude response with no added delays while Fig 8 shows the composite amplitude response after adding delays n_o of 15ms,

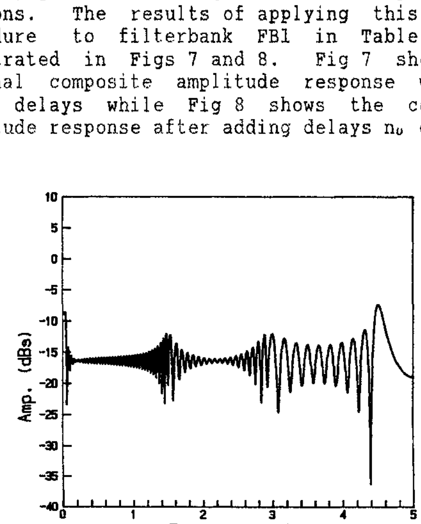

Fig 7 : Original FB1 Composite Amplitude Response

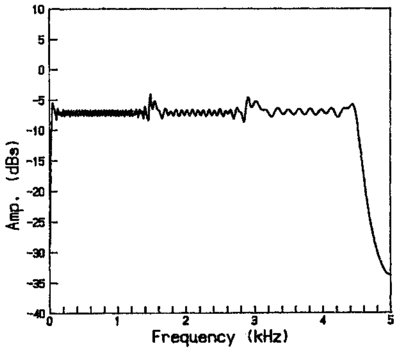

Fig 8 : Optimised FB1 Composite Amplitude Response

7.5ms and 3.7ms to sections 1, 2 and 3 respectively. The composite impulse responses [p(nT)] of each section were time-aligned by adding delays of 7.5ms and 11.3ms to sections 2 and 3 respectively.

4. Spectral Smoothing.

One way of computing the envelope of a discrete spectrum is to treat the spectral components as a time series and smooth them using a bipass (forward- and reverse-time) filter [Kormylo, 1974]. However, the non-uniform sampling interval that results from using non-uniform channel spacing must be removed by interpolation. The result is a spectrum with non-uniform resolution but with linear definition. Such a spectrum can be smoothed not only by lowpass filtering but also by homomorphic filtering.

Fig 9 shows the smoothed short-time spectrum of a synthetic /a/ vowel, computed by linearly interpolating the output of the optimised filterbank, FB2, in Table 1, and then lowpass filtering with a variable, second-order, Butterworth lowpass filter. Different cut-off frequencies were used for each filterbank section. For comparison, the response of the model used to synthesise the signal is shown overplotted (dashed line). Only over the first two sections is there a reasonable spectral match. To remove all of the harmonic detail from section 3 requires an excessive level of smoothing.

In contrast to bipass filtering, homomorphic filtering results in the same level of smoothing being applied to all sections. However, multiple peaks of unequal amplitude occur in the cepstrum due to the different channel spacing and bandwidth in each section and, therefore, the spectrum of the high-time part of the cepstrum will not be flat. In cepstral smoothing, this spectrum is effectively subtracted from the original spectrum and so the smoothed spectrum will contain the spectral trend of the high-time part of the cepstrum. The amount of spectral trend can be determined by carrying out a linear regression on the peak amplitudes of the high-time harmonics and this can then be removed from the smoothed spectrum. Fig 10 illustrates the result of applying this process to the same synthetic /a/ vowel. A comparison of Fig 10 with Fig 9 highlights the superiority of using cepstral smoothing with trend removal over bipass filtering for deriving the spectral envelope of a non-uniform spectrum.

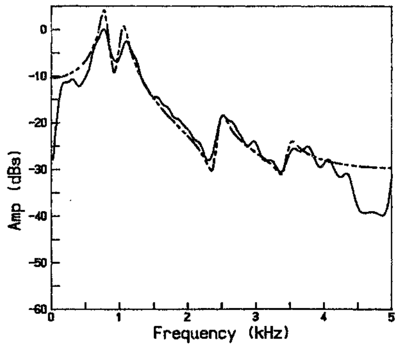

Fig 9 : Bipass Filtered Short-Time Amplitude Spectrum of Vowel /a/.

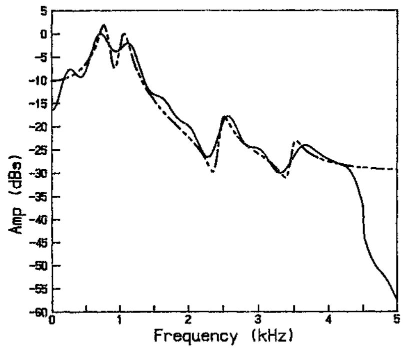

Fig 10 : Cepstrally Smoothed Filterbank (FB2) Spectrum of Synthetic /a/ with Spectral Trend Removed.

5. Conclusions

This paper has described a technique for designing uniform and non-uniform filterbanks based on the Running Fourier Transform (RFT). The RFT is implemented by convolving the input signal with one of a family of windows, $h(nT) = (nT)^k e^{-\alpha nT}$, where k and α may be chosen to specify the order and bandwidth, respectively, of each analysing filter. It has been shown how computation of the composite impulse response can be used to optimise the composite amplitude responses of any RFT filterbank. Finally, a modified cepstral smoothing technique for non-uniform spectra has been presented and has been shown to be superior to conventional bi-pass filtering.

References

Flanagan (1972), J L, Speech Analysis, Synthesis and Perception, Springer-Verlag, Berlin, 2nd edition, 1972, p 142.

Kormylo (1974), J J & Jain, V K, Two-Pass Recursive Digital Filter with Zero Phase Shift, IEEE Trans, ASSP-22, No 5, 1974, pp 384-387.

Owens [1988], F J & Murphy, M S, A Short-Time Fourier Transform, Signal Processing, North-Holland, Vol 14, No 1, 1988, pp 3-10.