‘PERFECT RECONSTRUCTION’ TIME-SCALING FILTERBANKS*

T.F. Quatieri[1] *and T.E. Hanna*[2]

[1]MIT Lincoln Laboratory, Lexington, MA, USA

[2]NSMRL, Box 900, Naval Submarine Base New London, Groton, CT, USA

Email: quatieri@ll.mit.edu

ABSTRACT

A filterbank-based method of time-scale modification is analyzed for elemental signals including clicks, sines, and AM-FM sines. It is shown that with the use of some basic properties of linear systems, as well as FM-to-AM filter transduction, “perfect reconstruction” time-scaling filterbanks can be constructed for these elemental signal classes under certain conditions on the filterbank. Conditions for perfect reconstruction time-scaling are shown analytically for the uniform filterbank case, while empirically for the nonuniform constant-Q (gammatone) case. Extension of perfect reconstruction to multi-components signals is shown to require both filterbank and signal-dependent conditions and indicates the need for a more complete theory of “perfect reconstruction” time-scaling filterbanks.

1 FILTERBANK FRAMEWORK

Consider a discrete-time signal $x(n)$ passed through a bank of filters $h_k(n)$ where each filter is given by a modulated version of a baseband prototype filter $h(n)$, i.e., $h_k(n) = h(n)\exp[j(2\pi/R)kn]$ where $h(n)$ is assumed to lie over a duration $-N/2 \leq n < N/2$ (N even without loss of generality), and R, the frequency sampling factor, is the number of filters. In the context of this paper, the filters are zero phase in time and frequency, e.g., a Gaussian or Hamming function. The filters are designed to satisfy a perfect reconstruction constraint in frequency, i.e, $\sum_k h_k(n) = \delta(n)$, where $\delta(n)$ is the unit sample sequence. One condition for perfect reconstruction is that the length of $h(n)$ be less than twice the frequency sampling factor, i.e., $N < 2R$ [4].

Each filter output $y_k(n) = x(n) * h_k(n)$ is complex [each filter response $h_k(n)$ is complex] so that the temporal envelope of the output of the kth channel is $a_k(n) = |y_k(n)|$ and the phase of each bandpass output is $\theta_k(n) = \tan^{-1}(Im[y_k(n)]/Re[y_k(n)])$. Thus the output of each filter can be viewed as an amplitude and phase modulated (complex) sine wave

$$y_k(n) = a_k(n)\exp[j\theta_k(n)]$$

and reconstruction of the signal can be viewed as a sum of complex exponentials

$$x(n) = \sum_k a_k(n)\exp[j\theta_k(n)] \quad (1)$$

*THIS WORK WAS SUPPORTED BY THE NAVAL SUBMARINE MEDICAL RESEARCH LABORATORY. OPINIONS, INTERPRETATIONS, CONCLUSIONS, AND RECOMMENDATIONS ARE THOSE OF THE AUTHORS AND ARE NOT NECESSARILY ENDORSED BY THE UNITED STATES AIR FORCE OR THE UNITED STATES NAVY.

2 TIME-SCALE MODIFICATION

The approach to time-scale modification relies on the subband signal representation in (1). The output of each filter is viewed as an amplitude- and phase-modulated sine wave, the amplitude and unwrapped phases of which are interpolated to perform time-scale modification, as in the phase vocoder [2]. With time-scale modification by a factor ρ, the modified filter output is given by

$$\tilde{y}_k(n) = \tilde{a}_k(n)\exp[j\rho\tilde{\theta}_k(n)] \quad (2)$$

where $\tilde{a}_k(n)$ is the channel envelope and $\tilde{\theta}_k(n)$ is the unwrapped phase, both interpolated by the factor ρ. The interpolated phase function in (2) is scaled by ρ to maintain the original frequency trajectory, i.e., phase derivative, of each filter output. We would like be be able to state conditions under which a “perfect reconstruction” objective holds for time-scale modification, in particular, for time-scale expansion for the purpose of improving the audibility of closely-spaced signals and/or signals in noise. For an arbitrary signal class, it is not straightforward to obtain such constraints due to the nonlinear nature of the transformation represented by (2) and due to ambiguity in defining time-scale modification for different signal classes. However, there do exist a number of simplifying cases where a specific time-scale modification can be defined and achieved under certain constraints.

2.1 Click

Consider an input “click” of the form $x(n) = \delta(n - n_o)$. The perfect reconstruction time-scale modification of the click is defined to be a displacement in time ρn_o samples, i.e., the modified signal is given by $\tilde{x}(n) = \delta(n - \rho n_o)$ where we have assumed only integer rate changes. There is no change in the signal character, only a signal shift. The output of each filter in the bank is given by $h_k(n - n_o)$, i.e., the impulse response of each filter is shifted to time n_o. Then the output in polar form is given by

$$\begin{aligned} \tilde{y}_k(n) &= h_k(n) * x(n) \\ &= h(n - n_o)\exp[j(2\pi/R)k(n - n_o)] \end{aligned}$$

and so the envelope of each filter output is simply $a_k(n) = h(n - n_o)$ and the phase is $\theta_k(n) = (2\pi/R)k(n - n_o)$.

In doing time-scale modification, we interpolate the envelope and unwrapped phase. The envelope is transformed to a new function $\tilde{a}_k(n - \rho n_o)$ that has length ρN where N is the length of the prototype filter $h(n)$, and that is centered at time $n = \rho n_o$. The phase is transformed to

$$\begin{aligned} \tilde{\theta}_k(n) &= (2\pi/R)k(n/\rho - n_o)\rho \\ &= (2\pi/R)k(n - n_o\rho) \end{aligned}$$

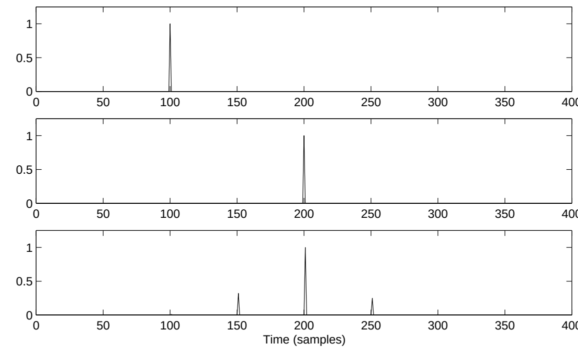

Figure 1: Time-scale modification of click: original (upper); modified with $N = 25$ (middle); modified with $N = 50$ (lower).

The transformed channel response is then given by

$$\tilde{h}_k(n) = \tilde{a}_k(n - \rho n_o)\exp[j(2\pi/R)k(n - n_o\rho)]$$

and, recalling that $a_k(n)$ is simply the prototype filter $h(n)$, for perfect reconstruction we require that

$$\begin{aligned} y(n) &= \sum_k \tilde{h}(n-\rho n_o)\exp[j(2\pi/R)k(n-\rho n_o)] \\ &= \tilde{h}(n-\rho n_o)\sum_k \delta(n-\rho n_o - kR) \\ &= \delta(n-\rho n_o) \end{aligned} \qquad (3)$$

where $\tilde{h}(n)$ is the time-scaled version of $h(n)$. From (3), the new condition for perfect reconstruction is that the duration of the time-expanded prototype filter be less than twice the frequency sampling factor, i.e., $\rho N < 2R$ or $N < 2R/\rho$. An implication of this constraint can be stated as

If we design the filterbank to just meet the perfect reconstruction constraint without modification, i.e., $N = 2R - 1$, we will need ρ times as many filters with modification, i.e., we need a denser sampling (by ρ) of filters to account for the time scaling.

Example 1: A complex filterbank was designed with a Gaussian prototype and with filters spaced by 200 Hz over a 5000 Hz bandwidth. Because the number of filters over the full 5000Hz band is $R = 25$, then the filter length constraint for perfect reconstruction is $N < 50/\rho$. For time expansion by a factor of two, Figure 1 shows examples with a prototype filter length[1] $N = 25$ and of length $N = 50$, the former just meeting our constraint and the later violating the constraint and resulting in unwanted pulses 50 samples to the right and left of the desired pulse.

2.2 Sine

Consider time-expanding a sine input of the form $x(n) = \cos(\omega n + \phi_o)$ by increasing its duration. Because $x(n)$ is an eigenfunction of a linear time-invariant system, and thus of each filter in our filterbank, we have (ignoring the $\frac{1}{2}$ scale factor)

$$\begin{aligned} y_k(n) &= |H_k(\omega)|e^{j\theta_k(\omega)}e^{j\omega n+\phi_o} \\ &= |H_k(\omega)|e^{j[\omega n+\phi_o+\theta_k(\omega)]} \end{aligned}$$

[1]A particular length is achieved with a Gaussian nearly zero at its endpoints.

where $H_k(\omega) = |H_k(\omega)|e^{j\theta_k(\omega)}$ is a channel frequency response. After time-scaling, the output of each channel becomes

$$\begin{aligned} \tilde{y}_k(n) &= |H_k(\omega)|e^{j\rho[\omega n/\rho+\phi_o+\theta_k(\omega)]} \\ &= |H_k(\omega)|e^{j\rho\theta_k(\omega)}e^{j(\omega n+\rho\phi_o)} \end{aligned}$$

Summing all channels, we have

$$\tilde{y}(n) = e^{j(\omega n+\rho\phi_o)}\sum_k |H_k(\omega)|e^{j\rho\theta_k(\omega)} \qquad (4)$$

At this point, we invoke the property that the prototype filter $h(n)$ is assumed zero-phase and therefore that the modulated filters $h_k(n) = h(n)e^{j\frac{2\pi k}{R}n}$ are also zero-phase[2], i.e., $\theta_k(\omega) = 0$,. Under this condition, (4) can be written as

$$\tilde{y}(n) = e^{j(\omega n+\rho\phi_o)}\sum_k H_k(\omega) = e^{j(\omega n+\rho\phi_o)}$$

because the filterbank $h_k(n)$ is assumed perfect reconstruction. Unlike the click input, therefore, for a sine input to be time-expanded without distortion, we do not require a denser frequency sampling on the filterbank that just meets the perfect reconstruction condition.

2.3 AM-FM Sine

Consider now time-expanding an AM-FM sine input of the form $x(n) = a(n)e^{j\phi(n)}$, where $\phi(n) = \int_0^n \omega(\tau)d\tau$, by increasing its duration while slowing the rate of change of its AM and FM. Under certain slow-varying conditions [1] on $a(n)$ and $\omega(n)$, $x(n)$ is a "pseudo-eigenfunction" of a linear time-invariant system, and thus of each filter in our filterbank, i.e.,

$$y_k(n) \approx a(n)|H_k[\omega(n)]|e^{j\theta_k[\omega(n)]}e^{j\phi(n)}$$

where the signal FM has been *transduced* to an AM [5] within each channel. With time-scaling, the output of each channel becomes

$$\tilde{y}_k(n) \approx \tilde{a}(n)|H_k[\tilde{\omega}(n)]|e^{j\rho\theta_k[\tilde{\omega}(n)]}e^{j\rho\tilde{\phi}(n)}$$

where $\tilde{a}(n)$, $\tilde{\omega}(n)$, and $\tilde{\phi}(n)$ are the time-scaled versions of $a(n)$, $\omega(n)$, and $\phi(n)$, respectively. Summing all channels, we have

$$\begin{aligned} \tilde{y}(n) &\approx \sum_k \tilde{a}(n)|H_k[\tilde{\omega}(n)]|e^{j\rho\theta_k[\tilde{\omega}(n)]}e^{j\rho\tilde{\phi}(n)} \\ &\approx \tilde{a}(n)e^{j\rho\tilde{\phi}(n)}\sum_k H_k[\tilde{\omega}(n)] \end{aligned}$$

where we again invoke the property that the prototype filter $h(n)$ is assumed zero-phase, and therefore that the modulated filters $h_k(n) = h(n)e^{j\frac{2\pi k}{R}n}$ are also zero-phase. Each channel filter is swept by $\omega(n)$ and is "synchronized" due to the zero-phase property. Because the original filterbank is an identity, we then have the approximation

$$\tilde{y}(n) \approx \tilde{a}(n)e^{j\rho\tilde{\phi}(n)}$$

[2]More typically, we may constrain the filters to be symmetric in time. Although in this case the sidelobes in frequency are possibly negative, and thus strictly not zero-phase, they are oftened designed to be negligible compared to the mainlobe.

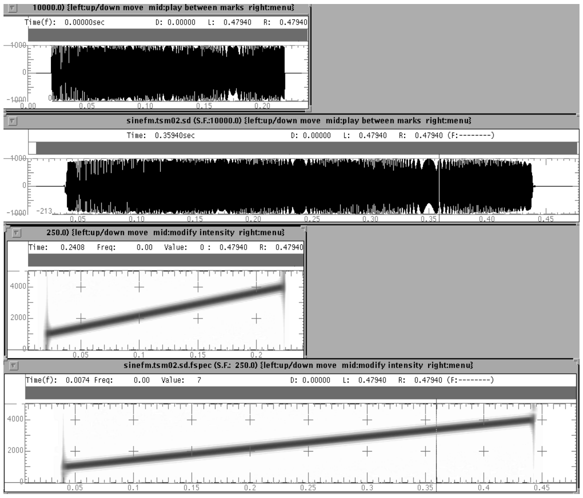

Figure 2: Time-scale modification of FM sine with $N = 25$: original waveform and time-scaled version (upper pair); spectrograms of original waveform and time-scaled version (lower pair).

As with the steady sine input, therefore, we do not require an additional constraint on a filterbank, designed to just meet the perfect reconstruction constraint, for an AM-FM sine input to be time-expanded without distortion. We do require, however, that the filters are "wide enough" in frequency so that the transduction approximation holds.

Example 2: A complex filterbank was again designed with a Gaussian prototype with filters spaced by 200 Hz over a 5000 Hz bandwidth. The prototype filter is selected to just meet the perfect reconstruction constraint when no modification is applied, i.e., $N = 2R - 1$. Because the number of filters over the full 5000Hz band is $R = 25$, the filter length $N = 25$. Figure 2 shows an example of time-scale expansion by two for an AM-FM sine input beginning at 1000 Hz with a 15000 Hz/sec sweep rate.

3 MULTI-COMPONENT SIGNALS

We have seen that single component signals (i.e., a click or AM-FM sine) are time-scalable by perfect reconstruction filterbanks. Multiple closely-spaced components, on the other hand, present in time and frequency new difficulties and stricter conditions for "perfect reconstruction."

3.1 Clicks

Suppose the input signal is of the form $x(n) = \delta(n) + \delta(n - n_o)$. With time scaling by a factor ρ, ideally, we want an output $y(n) = \delta(n) + \delta(n - \rho n_o)$. Without time-scale modification, the output of each filter in the filterbank is given by $y_k(n) = x(n) * h_k(n) = h_k(n) + h_k(n - n_o)$. To develop a sufficient condition for time-scale modification, we first observe that if $n_o > N$, N being the filter support, the two responses $h_k(n)$ and $h_k(n - n_o)$ are nonoverlapping. The time-scale system then "acts linear", i.e., the time-scale modification of the sum is the sum of the time-scale modified signals, and the time-scaled output is given by $\tilde{y}(n) = \sum_k h_k(n) + h_k(n - \rho n_o) = \delta(n) + \delta(n - \rho n_o)$. Therefore, our constraint on the filter length for perfect reconstruction time-scale modification becomes $N < min[2R/\rho, n_o]$.

Example 3: A complex filterbank is designed as in Example 1. As before, because the number of filters over the full 5000 Hz band is $R = 25$, then the filter length constraint for perfect reconstruction is $N < 50/\rho$. For time expansion by

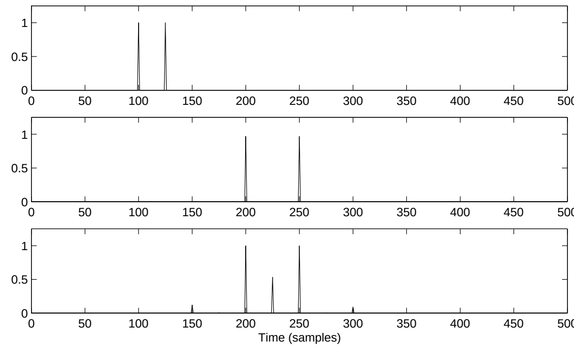

Figure 3: Time-scale modification of two closely-spaced clicks where $n_o = 25$: original (upper); time-scaled click pair with $N = 25$ (middle); time-scaled click pair with $N = 100$ (lower).

two of two clicks spaced n_o samples apart, our constraint is $N < min[25, n_o]$. Figure 3 shows an example of two closely-spaced clicks where $n_o = 25$ and with two different filter lengths $N = 25$ and $N = 100$. is violated. Although not our original goal of spreading the clicks, this later transformation is useful for periodic signals, as with the use of long filters in the phase vocoder [2], where an increase in the number of periods is desired. A continuum between the two time-scaling objectives may therefore be possible through filter length.

3.2 Sines

Suppose the input signal is of the form $x(n) = e^{j\omega_1 n} + e^{\omega_2 n}$. Without time-scale modification, the output of each filter in the filterbank is given by $y_k(n) = x(n) * h_k(n) = H_k(\omega_1)e^{j\omega_1 n} + H_k(\omega_2)e^{\omega_2 n}$. To develop a sufficient condition for time-scale modification, we observe that if $|\omega_1 - \omega_2| > BW$, BW being the filter bandwidth defined, for example, as the distance between the 3 dB attenuation points, the two components are essentially "independent", nonoverlapping in frequency. The time-scale system then "acts linear" and the amplitude and phase interpolation are invoked as though on independent components. Therefore, the time-scaled output is as desired. On the other hand, when the condition $|\omega_1 - \omega_2| > BW$ is not satisfied, then additional unwanted components are introduced. Consider, for example, two closely spaced sines. Then one predicts from the duality of time and frequency, that other spectral components will be introduced. The following example illustrates this property.

Example 4: A complex filterbank was designed as in Example 2. Figure 4 shows an example of time-scale expansion by two of two closely-spaced sines where our sufficient constraint, i.e., $|\omega_1 - \omega_2| > BW$, is violated.

4 CONSTANT-Q FILTERBANKS

4.1 Gammatone Filters

Thusfar we have investigated uniform filterbanks, i.e., uniformly spaced and with equal bandwidth. Our current focus is the front-end auditory stage, i.e., the basilar membrane, whose filters are approximated by gammatone functions, i.e., a damped cosine weighted by a gamma function, that in discrete time are of the form $h(n) = n^N \alpha^n \cos(\omega n)$. Along the basilar membrane, these functions have logarithmically increasing bandwidth (decreasing α) and frequency spacing. Enforcing zero phase on each filter, a set of noncausal filters that are time-aligned result. We have found empirically that the sum of these filter responses approximates

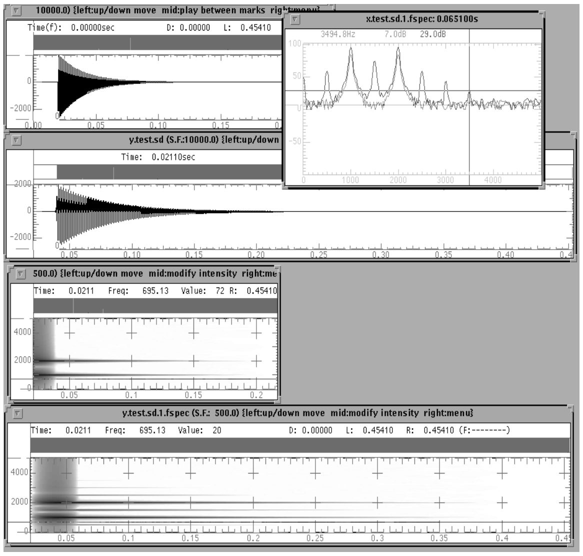

Figure 4: Time-scale expansion of two closely-spaced sines with $|\omega_1 - \omega_2| > BW$: original and modified waveforms (upper pair); spectrograms of original and modified (lower pair); superimposed spectral slices of original and modified (upper right).

an impulse, which is consistent with other designs of analysis/synthesis systems based on a Gammatone filterbank and its time-reversed version [3]. Because we ultimately desire auditory-like filters and because low-frequency constant-Q gammatone filters are very long, we have invoked spacing and bandwidth linearization in the low-end below about 1000 Hz, while maintaining an approximate perfect reconstruction property.

4.2 Time-scale modification

The constraints on the gammatone filterbank are similar in style to those of the uniform case for perfect reconstruction time-scale modification. It has been found empirically that if the frequency spacing-to-bandwidth ratio is "small enough" relative to the time-scale factor, then an impulse and sine are nearly perfectly reconstructed. Based on this observation, we venture the following condition.

Suppose that the frequency spacing-to-bandwidth (denoted by Δ/BW) factor is the minimum possible for perfect reconstruction. Then perfect reconstruction with time-scale modification by a factor of ρ requires that Δ/BW be decreased by a factor of ρ.

With multi-component signals, however, there is a difference from the uniform filterbank case in that different frequency regions dominate for different signal classes, as illustrated in the following example.

Example 5: Figure 5 shows time-scale expansion by two of two click pairs using a linearized gammatone filterbank based on filters from the Matlab Auditory Toolbox developed by Slaney [6]. Spacing between the two clicks of the first and second pair are 20 ms and 10ms, respectively. We see that the 20ms-spaced clicks are time-scaled without distortion. Between the 10ms-spaced clicks however there appears a small unwanted pulse; this pulse resides in the low-frequency region where the filterbank impulse responses are the longest. In fact, the impulse response length in the linearized region is greater than 10ms, implying interaction across signal components. We have also considered the dual problem of two sine-wave pairs. In one example, one sine-wave pair sits at 100Hz and 600Hz, while the second sine-wave pair sits at 3000Hz and 3700Hz. In this case, there is less distortion to the low-frequency pair because the filters in this region treat the sines more "independently."

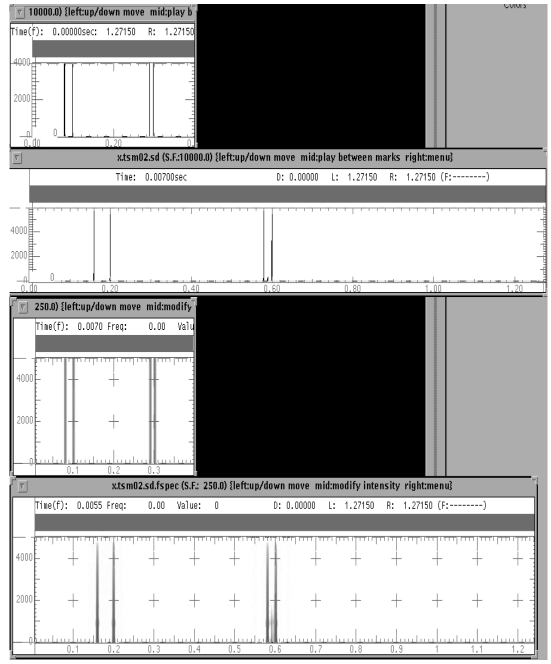

Figure 5: Time-scale modification of two click pairs with linearized gammatone filterbank: original and modified waveforms (upper pair); spectrograms of original and modified (lower pair).

5 CONCLUSIONS

This paper has introduced the concept of "perfect reconstruction" time-scaling filterbanks, using as examples uniform and constant-Q gammatone filterbanks, for a class of elemental signals as well as multi-component signals derived from this set. A more complete filterbank theory is required to address the improvement of time-scaling with multi-components, the analytical treatment of constant-Q and other nonuniform filter types, and the generalization to more complex signal models and alternate definitions of time-scale modification.

REFERENCES

[1] A.C. Bovik, J.P. Havlicek, M.D. Desai, and D.S. Harding, "Limits on discrete modulated signals", *IEEE Trans. on Signal Processing*, Vol. 45, No. 4, April 1997.

[2] M. Dolson, "The phase vocoder: A tutorial", *Comput. Music J.*, vol. 10, no. 4, Winter 1996.

[3] T. Irino and M. Unoki, "A time-varying, analysis/synthesis auditory filterbank using the gammachirp", ICASSP98.

[4] H. Nawab and T.F. Quatieri, "Short-Time Fourier Transform", Chapter 7 in *Advanced Topics in Signal Processing*, Edited by J.S. Lim and A.V. Oppenheim, Prentice Hall, Englewood Cliffs, NJ, 1988.

[5] T.F. Quatieri, T.E. Hanna and O'Leary, "AM-FM separation using auditory-motivated filters", *IEEE Trans. on Speech and Audio. Proc.*, Vol. 5,N0. 5, Sept. 1997.

[6] M. Slaney, *Matlab Auditory Toolbox*, Apple, 1995.