On timbre stamps and other frequency-domain filters

Miller Puckette *

Abstract

FFT-based filters find wide use in both live and 'tape' electronic music. Here an attempt is made to develop straightforward guidelines for choosing parameters such as window size and overlap in order to obtain desired time and frequency resolution and minimize artifacts. As an application, filters derived from other sound sources ("FFT vocoders" or "timbre stamps") are discussed in detail.

1 Introduction

On the frequent occasions when one reaches for one or another sort of filter, one can choose either a time-domain "classical" filter or an FFT-based one. Typically, time-domain ones can achieve very sharp frequency definition and low time latency, and are often cheaper to implement than FFT-based ones. But FFT-based ones have other graces such as explicit phase control and greater ease of varying the filter characteristics in time. Furthermore, certain applications lend themselves naturally to FFT filtering: for example, frequency-band-variable spatialization (Torchia and Lippe 2003) or delay (Kim-Boyle 2004), or "vocoders" or "timbre stamps" in which the spectrum of one sound is used to derive a filter for another (Settel and Lippe 1998) (Puckette 2007).

Here we will consider some issues that arise in FFT-based filtering, particularly for timbre stamping. The following section sets a framework and defines parameters used. Next, we consider whether, and when, FFT filtering really works correctly with arbitrary time-varying filter gains.

In Sections 5 and 6 we turn our attention to the the timbre stamping algorithm. Several possible variations are developed. All of them boil down to computations of various time-varying FFT channel gains, thus fitting into the framework developed and analyzed in Sections 1 through 4.

2 Setup

Using variable names and conventions as in (Puckette 2007), the filters under discussion take an input signal $X[n]$, possibly complex-valued, and compute short-time spectra

$$S[m,k] = \mathcal{FT}\{w_a[n]X[n+mH]\} \equiv \sum_{n=0}^{N-1} e^{-2\pi ink/N} w_a[n]X[n+mH] \quad (1)$$

where N is the window size, H is a hop size, $i^2 = -1$, k is the frequency in bins, m is the frame number, and $w_a[n]$ is the analysis windowing function. In general the spectra $S[m,k]$ are complex valued.

We will use "linear-phase" filters in which we multiply the spectra by real-valued gains $g[m,k]$, which may depend both on frequency k and frame number m. (For a non-time-varying filter there is no m dependence so that we may write the gain as $g[k]$.)

The output is then computed by windowing and overlap-adding the inverse Fourier transform:

$$Y[n] = \sum_m w_s[n-mH]\left(\mathcal{FT}^{-1}\{g[m,k]S[m,k]\}\right)[n-mH] \quad (2)$$

Here $\mathcal{FT}^{-1}$ denotes the inverse of the discrete Fourier transform $\mathcal{FT}$ and $w_s[n]$ is a resynthesis windowing function.

It is reasonable to ask that the signal $X[n]$ be correctly reconstructed when the filter gains are all 1, which implies:

$$\sum_m w_a[n-mH]w_s[n-mH] = 1 \quad (3)$$

for all n (tacitly putting $w_a = w_s = 0$ outside the window).

A possible choice for the analysis and resynthesis window function is the Hann window:

$$h[n] = \frac{1}{2}(1-\cos(2\pi n/N)) \quad (4)$$

We will also consider the "squeezed" version:

$$h_p[n] = \begin{cases} h(\frac{(1-p)N}{2} + pn) & 0 \le \frac{(1-p)N}{2} + pn < N \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

where $0 < p < 1$. This is just the Hann window function rescaled to occupy a segment of length pN in the middle of the window $0, .., N-1$, and zero padded elsewhere. The Fourier transform is

$$\widehat{H}_p = \mathcal{FT}\{h_p[n]\} \approx \quad (6)$$

*CRCA, $\mathrm{Cal(it)}^2$, UCSD. To appear in *Proceedings*, ICMC 2007.

$$e^{-\pi ik}\left\{\frac{N}{2}\mathrm{sinc}(\frac{k}{p})+\frac{N}{4}\mathrm{sinc}(\frac{k+1}{p})+\frac{N}{4}\mathrm{sinc}(\frac{k-1}{p})\right\}$$

where $\mathrm{sinc}(k) = \sin(k)/k$ and $\mathrm{sinc}(0) = 1$. (The phase term comes in because the window function is centered at $n = N/2$ but the Fourier analysis phase is zero at $n = 0$). The "main lobe" of $\widehat{H_p}$ extends over $-2/p \le k \le 2/p$ for a total bandwidth of $4/p$.

The filter is completely specified by the gains g, the window size N, the analysis and resynthesis window functions (including "squeeze factors" p), and the *overlap*, defined as N/H.

3 Simplest low-pass filter

Our analysis will loosely follow that given in (Allen 1977) and appendix B of (Laroche and Dolson 1999). Assuming the filter is time-invariant (i.e., the gain g does not depend on the frame number m), we can predict its behavior from that of the low-pass filter with $g[0] = 1$ and $g[k] = 0$ for $k \neq 0$. This is possible because, first, the filter's output is a linear function of g, and second, passing a signal through a filter admitting only the k^{th} bin gives the same results as for the zeroth bin, except with a frequency shift.

Suppose we introduce the sinusoid $X[n] = Z^n$ with angular frequency ω_0 and whose frequency in bins is $k_0 = 2\pi\omega_0/N$. The Fourier transform at DC is

$$S[m, 0] = Z^{N/2+mH} e^{\pi i k_0} \widehat{W}(k_0) \tag{7}$$

Here the first phase term is that of the incoming signal at the middle of the m^{th} analysis window; the rest is the Fourier-transformed window function evaluated at the point $-k_0$.

We now take the inverse FT and overlap-add using the resynthesis window function. Viewed in the time domain, this convolves the resynthesis window function with the signal:

$$Z^{N/2},\ \underbrace{0,\ldots,0}_{H-1 \text{ times}},\ Z^{N/2+H},\ \underbrace{0,\ldots,0}_{H-1 \text{ times}},\ Z^{N/2+2H},\ \ldots \tag{8}$$

This pulse train contains the bin frequencies:

$$\ldots,\ k_0 - N/H, k_0, k_0 + N/H,\ \ldots \tag{9}$$

and the result of convolving it with the resynthesis windowing function is to apply a low-pass filter with transfer function $\widehat{W_s}$. For the "correct" result we should filter out all but the k_0 term; this works provided $N/H > |k_0| + c_s$ where c_s is the cutoff frequency of the windowing function. Given that we do not wish to have to control the range of frequencies in the incoming signal, the only factor that controls the magnitude of k_0 is the bandwidth of the analysis window function (call it c_a). Then the condition for not aliasing is:

$$c_a + c_s \le \frac{N}{H} \tag{10}$$

If we are using Hann windows with squeeze factors p_a and p_s, we get:

$$\frac{2}{p_a} + \frac{2}{p_s} \le \frac{N}{H} \tag{11}$$

For an overlap of four, we barely get away with it at $p_a = p_s = 1$; no squeezing is allowed.

The signal is attenuated at the analysis stage by $\widehat{W_a}(k_0)$, and again at the resynthesis stage by $\widehat{W_s}(k_0)$, so the frequency response is the product of the magnitudes of the two. If we use Hann windows with no squeezing, we get 12 dB reduction at $k_0 = 1$ and 2.84 at $k_0 = 0.5$; so the bandwidth can reasonably be stated as one bin. But if, for example, we wish to place a filter at a center frequency of $k = 0.5$, we have to superpose filters at $k = 0$ and $k = 1$. The gain then only falls off 1.9 dB one half bin off peak and 5.7 dB one bin off (at $k = 1.5$, e.g.). If desired, the uniformity of bandwidth can be improved by zero-padding the Fourier transforms, effectively doubling N and using squeeze factors of 0.5 so that the filter may be expressed at a resolution of $1/2$ bin instead of 1.

4 Time-varying filters

The low-pass filter (from which we may understand the behavior of any other filter) may be made time-varying by specifying that the gain $g[m, k]$ be zero except when $k = 0$, but varying with m. Since the filter output is a linear function of the gain g, it suffices to know the behavior of a sinusoidally varying filter:

$$g[m, k] = \begin{cases} e^{2\pi i m H k_f/N} & k = 0 \\ 0 & \text{otherwise} \end{cases} \tag{12}$$

This oscillates at the frequency k_f in bins. (The factor mH appears because the m^{th} frame starts at sample mH.)

Everything goes as before and out come the frequencies:

$$\ldots,\ k_0 + k_f - N/H, k_0 + k_f, k_0 + k_f + N/H,\ \ldots \tag{13}$$

For the result not to alias, we must limit the frequency k_f so that

$$|k_f| + c_a + c_s \le \frac{N}{H} \tag{14}$$

Using Hann windows with an overlap of four does not allow any time variation bandwidth at all, the Convolution Brothers' well-known practices notwithstanding.

As before, the transfer function is the product of the two window functions, but the resynthesis window function acts

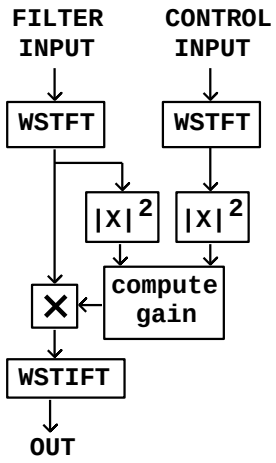

Figure 1: FFT vocoder (timbre stamp) block diagram.

at the aliased frequency; the frequency response is equal to $|W_a(k_0)| \cdot |W_s(k_0 + k_f)|$. If we wish, therefore, for the frequency response to behave "properly", that is, as a function of k_0 alone, we should squeeze the resynthesis window so that its larger bandwidth makes the frequency response less dependent on k_f. This can be done only at the expense of raising the minimum attainable bandwidth.

5 The timbre stamp

Figure 1 shows an overall block diagram for the timbre stamp. The three operations at left are the analysis/resynthesis chain of Section 1, with the input now renamed "FILTER INPUT" to distinguish it from a new, second input that alters the filter. The filter input passes first through a windowed short-time Fourier transform (WSTFT), whose outputs are complex-valued. These are multiplied by a real-valued gain (i.e., their magnitudes are changed but their phases maintained). Then the output is computed using a windowed short-time inverse Fourier transform (WSTIFT).

The gain is a function of the magnitudes of two spectra: that of the original input and that of a second, "control" input. In the simplest procedure we would simply compute the ratio of the control amplitude to the original amplitude (individually for each bin) so that the gain multiplication replaces the original amplitude with the new one; but there are many possible refinements as discussed below.

In light of the previous discussion of allowable bandwidth of the filter coefficients, we can now make preliminary bounds on overlap and squeeze factors. We'll continue to assume squeezed Hann windows so that the windowing bandwidth is $2/p$ bins. If we consider the gain computation as being approximated by a polynomial function of the two spectra (the complex amplitudes and their conjugates, say, so that the square magnitude is of degree two), then terms of degree n will yield at most frequencies of $2n/p$ where p is the minimum (i.e., worse case) of the squeeze factors of the two analysis windows. To control terms up to degree d, we must choose an overlap factor N/H of at least

$$\frac{2d}{p} + \frac{2}{p_a} + \frac{2}{p_s} \tag{15}$$

or, for unsqueezed windows, $2d + 4$. An overlap of eight will cover us up to quadratic terms.

6 Computing suitable gain functions

Figure 2 shows a block diagram for computing an appropriate gain for the timbre stamp, including several possible variations that are useful at times. The main idea is simply to divide the two spectra bin by bin, returning the quotient in linear amplitude units. The two inputs are assumed to be in units of power (squared amplitude). The operations labeled "convolve" and "squelch", and the division, may be carried out in those units. The next operation ("depth" control) is best carried out in so-called Sones (Rossing, Moore, and Wheeler 2002, p. 108), which we here approximate as square root of amplitude (fourth root of power). Finally, if needed, a low-pass filter may be added to control foldover; it should be applied to the gain expressed in linear amplitude units.

The first, "convolve" operation in effect averages neighboring power measurements in order to prevent peaks arising from the filter input from falling between neighboring, relevant peaks in the control signal (Penrose 2001). This may also help in averaging out interference patterns between peaks of the incoming signals.

At frequency bands in which the filtering signal has very low level, it might give unfortunate results to divide by its power spectrum. For this reason it is usually wise to put some sort of limit on the gain that will be applied when filtering it. There are two places in the chain where this might be done. The most logical-sounding spot is after computing the gain as a quotient of the two power spectra. This control appears as "max gain" in the block diagram. Gains greater than a fixed threshold are simply limited to that threshold.

An alternative viewpoint is to regard the filter as having two stages, the first in which the filtering signal is "whitened" by dividing by its own amplitude (so that the resulting spectrum has equal energy at all frequencies), and then applying the spectrum of the control signal as a further stage. It often yields good results to limit the gain of the "whitening" stage

instead of limiting the quotient of the two gains. This control appears as "squelch" in the block diagram. Squelching effectively sets a minimum strength below which the filter input is considered silent, by limiting it below before dividing by it. It is often useful to set squelch to decrease as a function of frequency. (All these controls may vary with time and/or frequency as desired).

Another possible control is the "depth" of the effect. If we consider the identity filter (with unit gain) as one extreme, and fully applying the timbre stamp as the other extreme, then a continuum of mixtures is available between the two. Crossfading between the two is best done in units of Sones. One can even choose "depth" values outside the range from zero to one to generate deeper than 100% filtering, or to filter the original input "away" from the timbre of the control input.

It is possible to morph one sound into another using two timbre stamps applied in opposing directions, with one "depth" ramped from 0 to 1 and the other from 1 to 0. One then crossfades from the first timbre stamp to the second one over a suitably chosen sub-interval of the ramping period.

Finally, either to control foldover or as an effect in its own right, one can low-pass filter the filter gains. This can be brought about naturally by increasing the analysis window size (or making the squeeze factor of the control input analysis greater than that of the filtering input), but this also would have the effect of narrowing the analysis bandwidth. If a higher bandwidth is desired one can return to a smaller window size and, in compensation, low-pass filter the filter gains. This is an alternative to the strategy of convolving a suitable kernel into the power spectra at the top of the diagram; each has its own advantages and drawbacks.

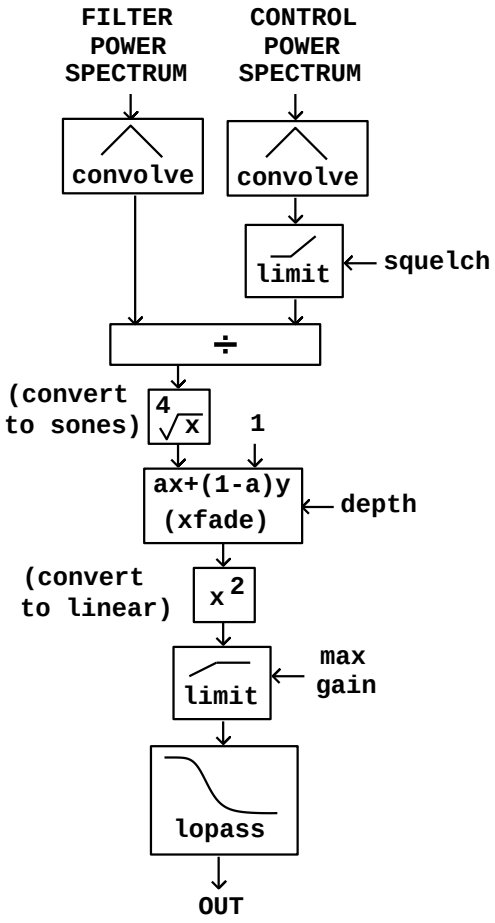

Figure 2: Computing the per-bin filter gain.

References

Allen, J. B. (1977). Short term spectral analysis, synthesis, and modification by discrete fourier transform. *IEEE Transactions on Acoustics, Speech, and Signal Processing 25*(3), 235–238.

Kim-Boyle, D. (2004). Spectral delays with frequency domain processing.

Laroche, J. and M. Dolson (1999). New phase-vocoder techniques for real-time pitch shifting. *Journal of the Audio Engineering Society 47*(11), 928–936.

Penrose, C. (2001). Frequency shaping of audio signals. In *Proceedings of the International Computer Music Conference*, Ann Arbor, pp. 334–337. International Computer Music Association.

Puckette, M. S. (2007). *The Theory and Technique of Electronic Music*. Singapore: World Scientific Press.

Rossing, T. D., F. R. Moore, and P. A. Wheeler (2002). *The Science of Sound* (Third ed.). San Francisco: Addison Wesley.

Settel, J. Z. and A. C. Lippe (1998). Real-time frequency-domain digital signal processing on the desktop. In *Proceedings of the International Computer Music Conference*, Ann Arbor, pp. 142–149. International Computer Music Association.

Torchia, R. H. and A. C. Lippe (2003). Techniques for multichannel real-time spatial distribution using frequency-domain processing. In *Proceedings of the International Computer Music Conference*, Ann Arbor, pp. 41–44. International Computer Music Association.