THE CONSTANT-Q IIR FILTERBANK APPROACH TO SPECTRAL FLUX

Risto Holopainen
Department of Musicology, University of Oslo, Norway
risto.holopainen@imv.uio.no

ABSTRACT

Spectral flux is usually measured with the FFT, but here a constant-Q IIR filterbank implementation is proposed. This leads to a relatively efficient sliding feature extractor with the benefit of keeping the time resolution of the output as high as it is in the input signal. Several applications are considered, such as estimation of sensory dissonance, uses in sound synthesis, adaptive effects processing and visualisation in recurrence plots. A novel feature called second order flux is also introduced.

1. INTRODUCTION

Many feature extractors are conventionally implemented with an FFT, although in several cases there exist alternative time domain implementations. Sometimes well-known time-frequency dualities can be evoked, such as carrying out the spectral centroid estimation with a differentiating filter in the time domain instead of multiplying the spectral bins with a ramp function [1]. Certain feature extractors are most naturally defined in the time domain, as is the case with the zero crossing rate. Yet, the idea of processing a buffer of samples at once, then hopping to the next buffer, seems to be prevailing even in situations where a sliding feature extractor is practical. The sliding DFT [2,3] makes it possible to have sliding versions of feature extractors that usually rely on the FFT. However, sliding feature extractors do not have to rely on the DFT. The principle of any sliding processing is that an entire buffer of arbitrary length may be efficiently processed and updated a single sample at a time, and that the output sample rate is the same as the input sample rate. Here we will consider a sliding filterbank implementation of spectral flux.

The various feature extractors that are commonly in use are more or less correlated. More precisely, for typical signals, there are feature extractors that measure the same thing in slightly different ways, such as the spectral roll-off, centroid and slope which are all sensitive to the presence of high frequency content. Hence, the parallel use of a set of feature extractors may be redundant. As Peeters et al. have recently suggested [4], studies of musical sound should at least use four types of mutually complementary descriptors including a measure of the time-varying central tendency of the spectrum (e.g. the centroid); a measure of the temporal dispersion of time-varying spectral features (presumably including spectral flux and statistical measures of the spread of other features); a descriptor related to the temporal envelope of the energy, and finally a descriptor related to periodicity, such as fundamental frequency or noisiness. In some studies of timbre perception with multidimensional scaling, the level of spectral fluctuation over time (i.e. spectral flux) has been found to correspond to one of the perceptual dimensions [5]. Spectral flux has also been found to be useful for onset detection [6]. When temporally accurate onset detection is the goal, a sliding onset detector may be preferred rather than one with a hop size on the order of several milliseconds. The proposed filterbank flux can be used for onset detection, although we will consider other applications.

In the present implementation, the spectral flux is measured with a small number of second-order IIR filters with constant frequency ratio as well as a constant Q factor. The idea is to measure the difference in amplitude in each band of the filterbank between the current position and a delayed position. In FFT-based flux, the difference in amplitude is taken between the current and the previous non-overlapping windows. In contrast, because the filterbank flux filters have an infinite impulse response, the current and the past windows cannot be completely isolated. Moreover, the filters' variable bandwidth leads to varying decay times of the impulse responses, with decay times decreasing with increasing centre frequency. Therefore it is necessary to insert a delay of suitable length, proportional to the decay time at each band, in order to keep the amount of overlap at a reasonably low level. Otherwise spurious correlations from the tail of the current impulse response into the delayed position will bias the estimated flux towards smaller values.

The single sample hop size leads to a sliding feature extractor, which is particularly useful when the temporal details of the analysed signal are of interest, or when an analysis with high temporal accuracy is preferred. Sliding feature extractors output samples at the audio sample rate; in the case of constant-Q filterbank flux its output is restricted to the range $[0, 1]$, which means that the output can be used as an audio signal, although it comes with a constant DC offset. Hence, if one listens to the filterbank flux as an audio signal, it tends to sound like a distorted, albeit often recognisable version of the input sound. In particular, harmonic and spectrally rich sounds tend to sound like a more buzzing version of the original tone. Anyone who would like to use the filterbank flux for something it was

not intended for may consider passing its output through a highpass filter and treat it as an unconventional distortion effect.

In the following section, the details of the filterbank flux implementation are given. Then, the remaining sections provide several examples of applications of filterbank flux. First, a second order flux will be introduced, then the filterbank flux is used as the core component in a sensory dissonance model. The sliding implementation is particularly suitable for use in certain kinds of feedback systems as well as in adaptive effects processing, some examples of which are mentioned towards the end of this paper.

2. FILTERBANK FLUX IMPLEMENTATION

The input signal $x[n]$ is passed through a constant-Q filter bank with *N* biquad bandpass filters b_k with octave spacing. When all channels of the filterbank are combined, a nearly flat frequency response can be achieved by setting $Q = 1/\sqrt{2}$. The lower and upper edges of the spectrum are not within the passband of the filterbank. Let y_k be the output signals of the filterbank. Then, the amplitude a_k of each band is measured with a sliding RMS amplitude feature extractor, which is efficiently implemented using a moving average filter structure [7], hence

$$a_k[n] = \sqrt{\frac{1}{N_k} \sum_{j=0}^{N_k-1} (y_k[n-j])^2} \qquad (1)$$

where each band has its own length N_k which is set inversely proportional to the centre frequency.

The amplitude signals a_k are then treated analogously to the amplitude bins of an FFT in the computation of flux; in either case, the point is to measure the amount of change between the current frame and a past frame. In the IIR filter bank implementation, however, each filter b_k has its own effective impulse response length, and therefore it is preferable to make the delay time d_k between the current frame and the past frame a function of this duration. Since IIR filters are used, the effective duration may be defined in relation to the amplitude of the impulse response $h_k[n]$ of the filter b_k. Let us define this effective duration $\tau_k(\epsilon)$ as the *last* time when $|h_k[n]| \geq \epsilon$.

For fixed ϵ, it turns out that $\tau_k(\epsilon)$ follows a power law $\log \tau_k \sim \log f_k$, where f_k is the filter's centre frequency. Hence, the decay time of the filter with centre frequency f_k is $\tau_k = C f_k^{\alpha}$ for some constants *C* and α. As $\epsilon \to 0$, α approaches -1 from above. In the present implementation, the constants C and α have been estimated from the measured effective durations of impulse responses using the time when the impulse response has decayed by 60 dB, or $\epsilon = 10^{-3}$, which yields $\alpha \approx -0.8$.

There are several alternative formulas in the literature for the calculation of flux, and furthermore, it may be computed from different signal representations [4]. Perhaps most commonly, the amplitude spectrum is computed from an FFT, but other representations may be used, including the power spectrum, sinusoidal signal models, or even auditory models such as gamma-tone filterbanks. Often, the spectral flux is calculated as 1 minus the normalised correlation of the current and previous amplitude bins [4]. Onset detection using spectral flux depends on identifying rising but not decreasing energy profiles, so the Heaviside step function (a.k.a. half-wave rectifier) may be applied to the difference between the current and the previous frames [6]. For the filterbank version, the average magnitude difference

$$\Phi_n = \frac{\sum_{k=1}^{N} |a_k[n] - a_k[n-d_k]|}{\sum_{k=1}^{N} a_k[n] + a_k[n-d_k]} \qquad (2)$$

is used. As we shall see in the next section, this formulation allows for a generalisation to second order flux.

It is important to set the time constants to suitable lengths in order to avoid spurious correlations between the current and past frames. In the current implementation, the length of the RMS window is doubled at each lower octave. Then, the combined decay time of the filter b_k followed by the RMS processing gives the delay time $d_k = \tau_k + N_k$.

This implementation becomes computationally intensive if there are many channels in the filterbank, but with octave spacing, as little as 8 filters covers most of the audio range. In all applications described next, the filter centre frequencies have been set to $f_k = 100 \cdot 2^k, \quad k = 0, 1, \ldots, 7$, covering the range from 100 Hz to 12.8 kHz.

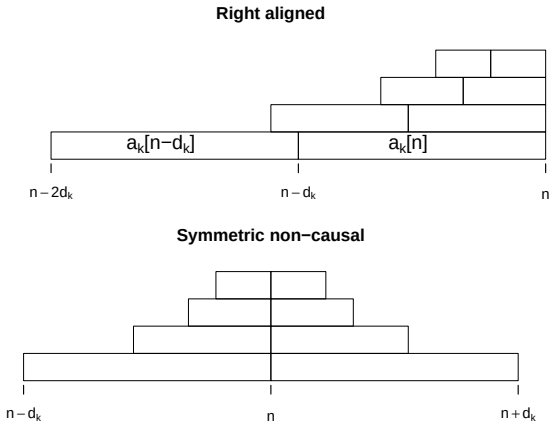

Figure 1. Different alignments of the current and delayed frames.

There is scope for variation in the implementation details of filterbank flux. Apart from the alternative formulas for the flux measure itself, the amplitude may be measured in different ways including the peak amplitude or the energy a_k^2. The filterbank may have denser spaced filters, and the delays d_k may be increased by selecting a smaller ϵ in the formula for the effective durations $\tau_k(\epsilon)$. In the proposed version, the total delay $\sum_k d_k$ is minimised under the condition that the processing is causal. Then, the past frame becomes more delayed towards bass frequencies (as in the top of Figure 1). It would also be conceivable to center the frames symmetrically on top of each other (see bottom of Figure 1), but this would require the additional complication of more delays. In fact, the sliding DFT might be used instead of the IIR filterbank, not least since a constant-Q

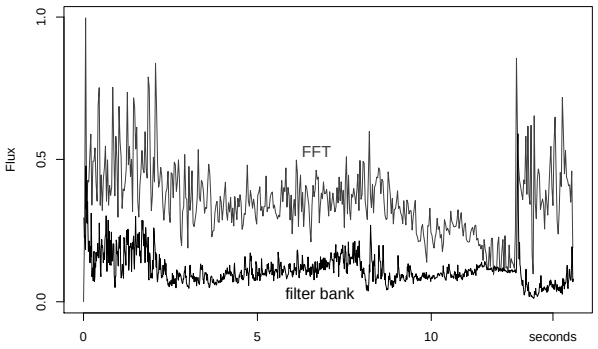

Figure 2. Comparison of flux as measured with the filter bank method and with an FFT with a hop size of 0.03 seconds. The analysed signal comes from the beginning of S.709 by Xenakis.

version of the sliding DFT has been proposed [8]. However, that would lead to yet another different flux implementation.

Due to the constant Q of the present implementation, the bandwidth and the energy of h_k are both proportional to the centre frequency. Consequently, the frequencies have a constant weighting across the spectrum, similarly to the FFT implementation. Alternatively, one could normalise the energy of each filter channel instead, which might qualify as a perceptually motivated weighting. All these possible variations of the filterbank flux implementation would deserve a closer study, but here we shall be contented with a brief comparison with the common FFT version of flux.

The opening of Xenakis' S.709 has been analysed with both the filterbank and the FFT method using the average magnitude difference (see Figure 2). The flux of the filterbank version has been averaged over 512 samples for the sake of visual clarity. Note that the FFT flux usually takes higher values, although both versions of flux respond to onsets. A direct comparison between the filterbank and the FFT flux is complicated for several reasons. First, the output sample rates need to be equal, preferably by setting the hop size to one sample in the FFT version. Second, a suitable window length has to be chosen for the FFT which is in some sense representative for the time constants of the filterbank. Third, the lag at which the cross correlation between the FFT and filterbank flux signals reaches its maximum depends on the analysed sound. All of these factors need to be considered before a sensible comparison can be made between the FFT and filterbank implementations of flux. However, the lack of correlation at some fixed time lag indicates that these are two independent and complementary feature extractors although they are both related to short time spectral fluctuations.

The filterbank flux has been implemented as a C++ class with a member function that processes one sample at a time, although buffer processing can also be used. By default, all filters and delays are initialised with zero input. Therefore, as soon as the analysed input signal contains its first non-zero sample, the flux signal saturates at its highest value. This can be seen in Figure 2 where there is a high peak at the start. In order to minimise these initial transients, one may initialise the filters with low amplitude white noise or other suitable signals. Next, we consider some further elaborations based on the basic flux feature.

3. SECOND ORDER FLUX

Apart from responding to onsets, spectral flux also indicates the amount of short time spectral fluctuation. Any sound that perceptually appears static or smooth may be expected to have a low flux level. However, the relationship between spectral flux and the perception of a sound's smoothness or granularity may be complicated. For instance, one may experience a sound to be granular at one level but static at a higher level, as is the case with coloured noise. This distinction may be explored with the aid of the second order flux, which we will introduce now.

The spectral flux Φ, as defined in (2), may be compared to a differentiating operator. Then, a second order flux Φ^2 can be introduced in analogy with the second derivative. When spectral flux is computed from the FFT amplitude spectrum, a second order flux using the average magnitude difference $|\Delta X[m]| = |X[m] - X[m-1]|$ may be defined in terms of the second difference $\left|\Delta^2 X[m]\right|$. Doing so and applying the result to the filterbank flux, one obtains a formula with delays of d_k and $2d_k$ which means that the time support is highly asymmetric (this is equivalent to adding another block on the left side of each of those in the top of Figure 1). However, the total delay is minimised for this right-aligned filter structure. If instead a non-causal filter is used, the symmetric second order flux

$$\Phi^2[n] = \frac{\sum_k |a_k[n+d_k] - 2a_k[n] + a_k[n-d_k]|}{\sum_k a_k[n+d_k] + 2a_k[n] + a_k[n-d_k]} \quad (3)$$

becomes possible. In practice, further delays would be inserted to make (3) causal. Therefore, the right-aligned structure is simpler to implement and is also faster in its response to events in the input signal than the causal version of (3).

In general, the second order flux is positively correlated to first order flux. Locally, however, there can be deviations from a strict correlation. Since the first and second order flux have already been likened to the first and second derivatives, one may consider a scatter plot of these variables as similar to the phase space portraits commonly used in dynamic systems theory. Two such flux phase portraits are given in Figure 3 using the right-aligned versions of both first and second order flux. The fine details of the phase space portraits are only visible if the flux variables come from sliding feature extractors; the details get blurred if the flux variables are smoothed and downsampled by any significant amount. As a plausible conjecture, the inequality $\Phi^2 < \Phi$ should be expected to hold for most sounds that have a static perceptual appearance, regardless of the level of first order flux.

Since the output of the flux feature extractor $\Phi[n]$ is itself an audio rate signal, it may be submitted to any further analysis that is commonly applied to audio signals. For instance, the spectrum of $\Phi[n]$ will reveal something about

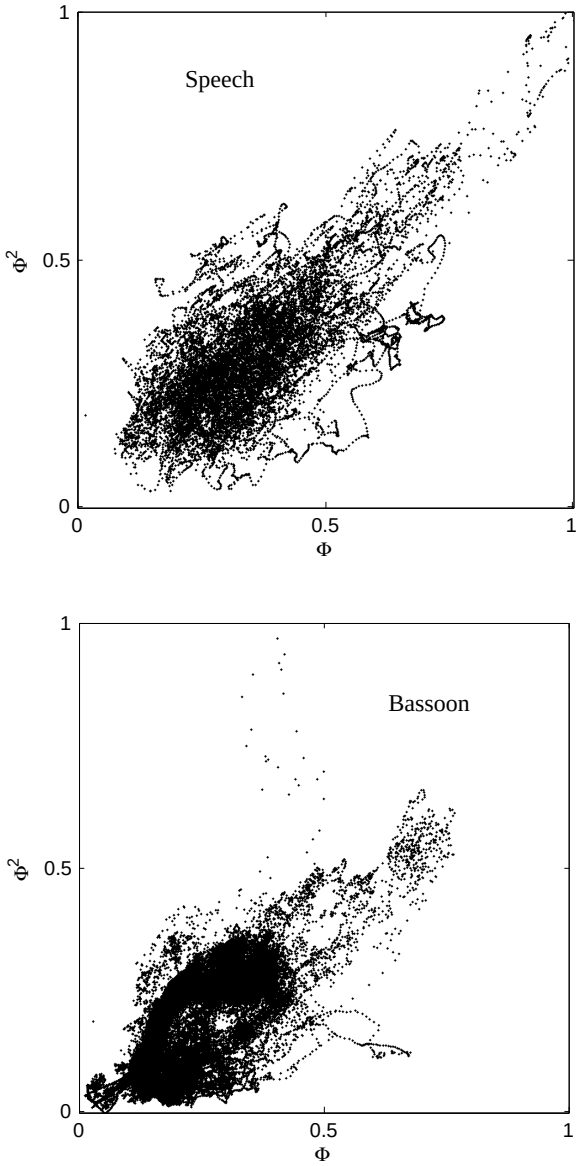

Figure 3. Two scatter plots of the first vs. second order flux. Top: a speech phrase; bottom: a low bassoon note.

the balance of fast and slow variations in the signal $x[n]$. The same procedures can of course be applied to Φ^2.

4. SENSORY DISSONANCE

Roughness or sensory dissonance arises from the occurrence of closely spaced partials, in particular from partials that fall within the same critical band. Partials that are close in frequency will cause beats, which contributes to a higher flux value. Therefore, it should be possible to use the filterbank flux as a sensory dissonance model.

Sethares has proposed a time domain sensory dissonance model [9] that bears a certain similarity to the filterbank flux. The sensory dissonance model of Sethares first splits the input signal $x[n]$ into several bands $y_k[n]$ with critical band filters. Then the amplitude envelope of each channel is extracted with a half-wave rectifier followed by a lowpass filter. Each channel's amplitude envelope is then

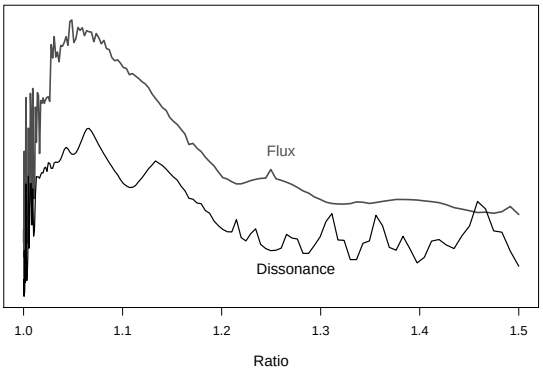

Figure 4. Flux and sensory dissonance score of two square-wave tones detuned by an increasing ratio.

passed through a bandpass filter with pass band in the range 15 – 35 Hz, the purpose of which is to capture the fast modulation or beats that are the cause of perceived roughness. Finally, the energy of the sum of all these signals yields the dissonance level.

In filterbank flux, the RMS amplitude replaces the lowpass filtered half-wave rectifier of the dissonance model. There are other differences as well, such as the delays in the amplitude variables in (2). However, if the flux signal $\Phi[n]$ is submitted to the same processing as the last stages of Sethares' dissonance model, then it may be used as a dissonance model as well. Thus, the flux signal is bandpass filtered and the filter's output RMS amplitude is used as the dissonance value. Specifically, we use a biquad filter again and set the bandpass filter's centre frequency to 25 Hz and use a Q factor 2. Although the flux-based dissonance obviously cannot be claimed to model the auditory system, it seems to give a decent estimate of the degree of sensory dissonance.

In Figure 4, two detuned bandlimited square-wave tones are measured with respect to flux and sensory dissonance according to the proposed method. The dissonance curve shows a similar characteristic to other dissonance curves obtained by the method described by Sethares. Most notably, the highest dissonance peak occurs around the ratio 1.066, or 110 cent, and several troughs can be found around consonant ratios. The dissonance curve actually shows the logarithm of the dissonance measure, because the minima at consonances are more readily seen that way.

5. A FEEDBACK SYSTEM USING FLUX

There are also more creative applications for the filterbank flux, some of which will be described next. Feature extractor feedback systems, or FEFS for short, are synthesis models consisting of the following parts: a signal generator synthesises the output signal which is analysed with a feature extractor, and the time varying feature is mapped to the synthesis parameters of the signal generator [10]. In such feedback systems, block based processing using an FFT and some fixed hop size will tend to introduce discontinuous changes every time the analysis window is advanced. In contrast, a sliding feature extractor leads to smoother

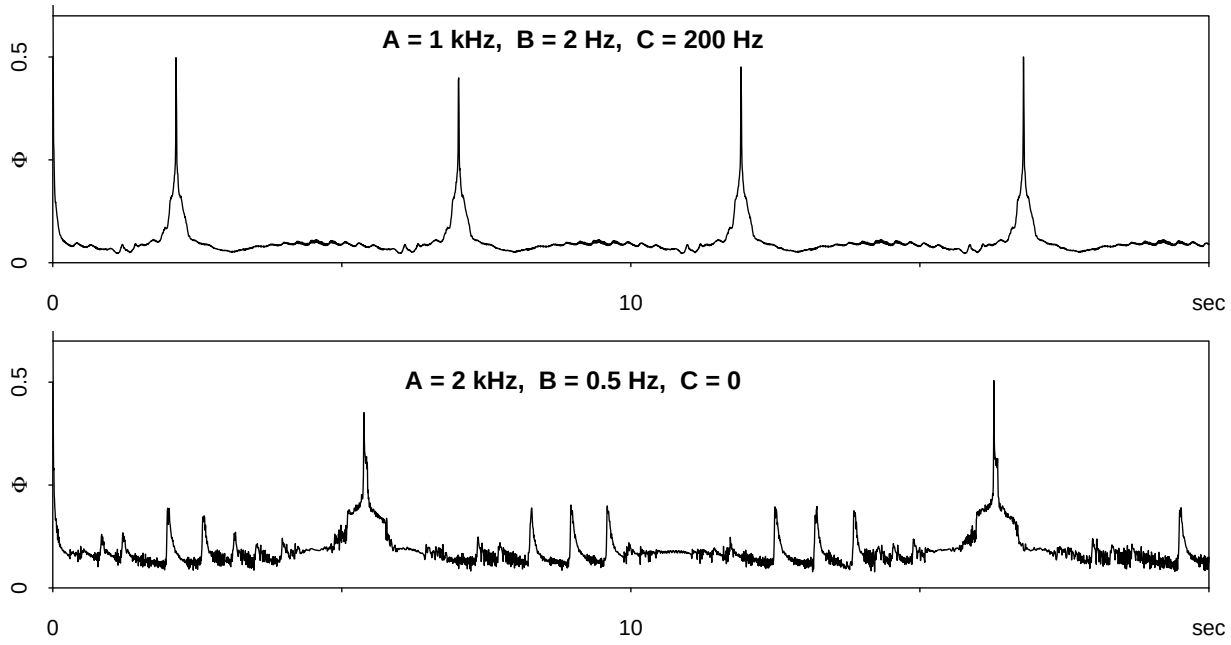

Figure 5. Flux time series of the FEFS for two parameter values. Top: slow periodic behaviour; bottom: more irregular with intermittent spikes.

dynamics. Hence, the filterbank flux is suitable for use in a FEFS.

For the signal generator, let us consider something that is capable of producing both smooth sounds having low values of flux as well as less smooth sounds where closely tuned partials cause beats. Two wavetable oscillators that can be more or less detuned will be used for the generator. The waveform is arbitrary but assumed to be fixed over time. Here, the two oscillators will use the bandlimited square-wave that was investigated with respect to flux and sensory dissonance level in the previous section. We write the output of the signal generator symbolically as

$$x[n] = \mathrm{osc}(f[n] + \delta[n]) + \mathrm{osc}(f[n] - \delta[n]) \qquad (4)$$

where $\mathrm{osc}(f[n] \pm \delta[n])$ denotes the wavetable oscillator running at the instantaneous frequency $f[n] \pm \delta[n]$ Hz, and the oscillators are assumed to have equal amplitude. The flux

$$\Phi[n] = \mathrm{flux}(x[n]) \qquad (5)$$

of the signal is measured using the filterbank approach (2). Then, the flux is mapped to the parameters of the signal generator. The mapping from flux to the two synthesis parameters may be any function $g : \mathbb{R} \rightarrow \mathbb{R}^2$, but let us consider an affine map

$$\begin{array}{rcl} f[n+1] & = & A\Phi[n] + C \\ \delta[n+1] & = & B\Phi[n] \end{array} \qquad (6)$$

with the three parameters A, B, C in units of Hz. The complete system (4, 5, 6) may be considered as a deterministic autonomous dynamic system whose temporal evolution depends entirely upon its control parameters and its initial condition. When $B = 0$, this model may be compared to feedback FM with *C* playing the role of a carrier frequency, and *A* being the modulation index. Typically, cycles of slow periodic pitch contours may arise, but setting $C = 0$ and $A > 0$ results in irregular behaviour.

The flux time series at two distinct parameter configurations can be seen in Figure 5. Note that the affine map (6) guarantees that the frequency f and deviation δ have the same shape as the flux signal, only scaled and shifted according to the parameters. A slow but periodic pitch contour can be seen in the upper plot, whereas a more irregular contour results when $C = 0$. One might think that higher values of B and increased amounts of detuning would lead to more irregularities; however, the opposite is the case. If C is close to zero, very slow developments will occur, including periods when the two oscillators gradually phase out and cancel each other, leading to moments when the sound fades out and then fades in again.

It should be emphasised that these relatively complicated pitch contours result even though the control parameters are constant over time. FEFS in general are high dimensional autonomous dynamic systems; specifically, they are iterated maps whose dimension is given by the number of unit delays in the feature extractor in addition to all the other variables [10]. As with other, better known dynamic systems such as the Lorenz attractor or the Hénon map [11], these systems may reach fixed points, periodic, quasi-periodic or chaotic states. It is hard to tell whether the system (4, 5, 6) is chaotic or merely quasi-periodic, although the detuned oscillators by themselves cause quasi-periodicity.

The mapping (6) does not contain any nonlinear terms, which would usually be needed in the feedback path for a system to be capable of chaos. However, the filterbank flux is already a nonlinear system that may be the cause of

chaos. The extended time support of the flux extractor also affects the dynamics of this system by smoothing it out over its duration. Although the resulting pitch contours are quite complicated and have a relatively long period with the affine mapping (6), even more complicated and irregular pitch contours can easily be generated by using nonlinear mappings.

6. RECURRENCE PLOTS

Systems such as the above example of a FEFS can benefit from being visualised with recurrence plots. Originally conceived as a tool for dynamic systems [12], recurrence plots have found important applications in other areas including music analysis [13]. The recurrence plot is a matrix constructed from the distances $d(X[n], X[m])$ between all pairs of observations in a signal, where the observations in audio applications are often FFT frames or MFCC frames. However, any feature extractor may be used for the observations. The filterbank flux is well suited, provided care is taken to make the successive observations non-overlapping.

When the temporal patterns of a FEFS are displayed with a recurrence plot, it is particularly interesting to use the same feature or set of features that are used internally in the system as the basis for the distance measure. In visualisations of dynamic systems, the recurrence plot would normally show the distances between each iteration of the map if the system is in discrete time. Such a fine-grained resolution is unpractical for audio data, which typically has much slower time scales. Therefore, it is reasonable to use some kind of averaging over a longer time segment for the data points.

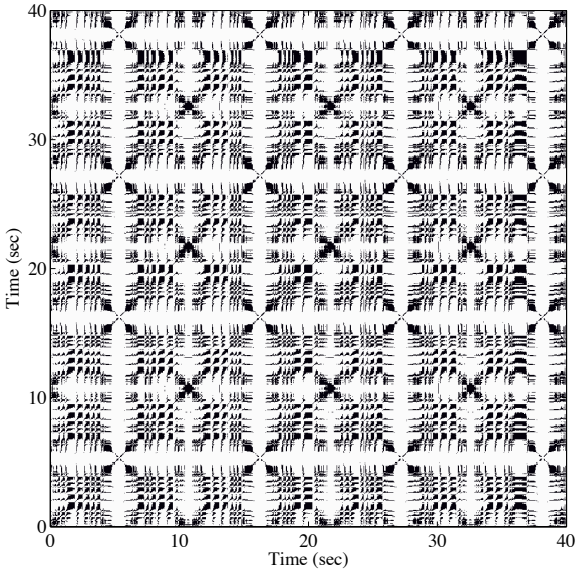

Figure 6. Recurrence plot using flux as distance measure.

The recurrence plot in Figure 6 uses the Euclidian distance over the local mean and standard deviation of the filterbank flux over windows of 2048 samples. The audio signal comes from the system described above and with the same parameter settings as in the bottom of Figure 5. As can be gleaned from the regular block structure, there are periodically repeating patterns, although these repetitions are never exact. It is remarkable that such a simple FEFS with comparatively short temporal support in its feature extractor can exhibit such long-term variability. This is however not unique to this particular system, but has been observed in other FEFS as well [10].

7. ADAPTIVE EFFECTS

Signal-adaptive effects processing uses feature extractors to control effect parameters [1]. By analysing the features coming from the same signal as is being processed, adaptive processing can follow and exaggerate the variation of a sound. Let us consider some applications of the filterbank flux to adaptive effects. For instance, dynamic processing similar to a de-esser can be achieved by letting the flux level determine the amount of gain. Because the filterbank flux signal tends to have a substantial high frequency component, it may be necessary to lowpass filter it in order to avoid noisy artefacts. Smoothing of the flux signal would be needed even if an FFT version of flux were used, but in that case it is needed for interpolation of the sudden jumps that may occur as the window hops several samples forward. In contrast, the filterbank flux typically has continuous high frequency content that introduces audio rate amplitude modulation if used directly as a gain control.

As is the case with ordinary dynamics processing including compressors, noise gates, expanders etc, there are as many possibilities for the use of flux as a gain control. Either portions of the signal containing high flux may be suppressed, or these portions may be emphasised and the static portions suppressed. A processing scheme that gives interesting results is the following. First, the flux $\Phi[n]$ is obtained from the signal $x[n]$. A smoothing one-pole filter is applied to the flux signal, to obtain the smoothed flux $\bar{\varphi}[n] = (1-b)\Phi[n] + b\bar{\varphi}[n-1]$. The variations in $\bar{\varphi}[n]$ are typically too shallow to be used directly as a gain function. Therefore this signal is passed through a sigmoid function such as

$$s(x) = (\tanh(((x-\beta)G) + 1))/2 \qquad (7)$$

where β controls the cutoff point of the transition region, and G is the sharpness of the transition region. Then, $g[n] = s(\bar{\varphi}[n])$ is used for the gain that multiplies $x[n]$. The parameters may be set by trial and error, unless the entire signal is available for offline processing. If so, the statistics of the flux feature can be collected in a first analysis phase and, based on the probability distribution of $\Phi[n]$, a certain proportion can be specified where the signal should come through and where it should be blocked. The probability distribution of flux in a low bassoon tone is shown in Figure 7. From the inverse of the cumulative distribution (not shown), a cutoff point can be found that gives the parameter β in (7).

If the input is a percussive sound, it is possible to keep only the onsets since they will have a higher flux value than other parts of the signal. This can be used for instance

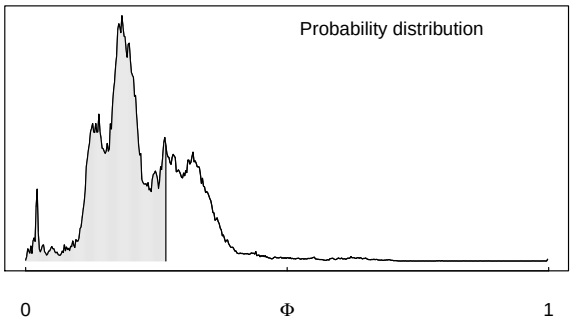

Figure 7. The probability distribution of flux of a low bassoon tone.

on plucked string sounds to impose a staccato articulation by chopping off the gradual decay. Conversely, if $1 - g[n]$ is used as the gain, one has a flux suppressor. When used for attack suppression this can produce smoothly fading in attacks, although the time delays introduced both by the filterbank flux and the smoothing lowpass filter may need to be compensated for by delaying the input signal. Then the output is $(1 - g[n])x[n - D]$ for some suitable delay time D.

Gain processing using flux—as opposed to the amplitude envelope that is used in standard dynamics processing—is not sensitive to the absolute amplitude level of the signal. Thus, the flux level of a note onset does not necessarily depend on whether it is a pianissimo or a forte note.

Adaptive effects can also use the second order flux or the dissonance features introduced above. Furthermore, all these feature extractors can of course be applied to any audio effect one can think of. Adaptive effects processing is only limited by one's imagination.

8. CONCLUSION

Sliding feature extractors can be used in any application where block based processing would ordinarily be used, including feedback systems and signal-adaptive processing. Since the output is an audio rate signal, traditional feature extraction applications that expect highly downsampled signals may need the flux feature to be smoothed and downsampled as well. In other applications such as the FEFS, the temporal resolution of an audio rate feature is more useful.

The current implementation of constant-Q filterbank flux is but one of many possible and, perhaps, equally valid implementations. There is scope for variation as to how the amplitude of each frequency band is measured. In the flux formula itself, the distance measure between the current and the past frames may be formulated in several different ways. More channels may be used in the filterbank, with closer spacing of the centre frequencies. The current implementation has most of all tried to achieve efficiency. If the first order flux is computed, the second order flux and dissonance are obtained at little extra cost.

We have not tried to argue that the filterbank flux is better than the traditional FFT or other implementations of spectral flux, rather it is complementary to them. Which implementation is most advantageous depends on the application at hand, and here we have suggested a number of contexts where a sliding flux feature extractor can be useful. There is reason to believe that other sliding feature extractors of all sorts will likewise find their uses.

9. REFERENCES

[1] V. Verfaille, "Effets audionumériques adaptatifs : théorie, mise en œuvre et usage en création musicale numérique," Ph.D. dissertation, Université Aix-Marseille II, 2003.

[2] E. Jacobsen and R. Lyons, "The sliding DFT," *IEEE Signal Processing Magazine*, vol. 20, no. 2, pp. 74–80, March 2003.

[3] R. Bradford, R. Dobson, and J. ffitch, "Sliding is smoother than jumping," in *Proceedings of the ICMC 2005*, Barcelona, Spain, 2005, pp. 287–290.

[4] G. Peeters, B. Giordano, P. Susini, N. Misdariis, and S. McAdams, "The timbre toolbox: Extracting audio descriptors from musical signals," *Journal of the Acoustic Society of America*, vol. 130, no. 5, pp. 2902–2916, November 2011.

[5] J. Grey, "Multidimensional perceptual scaling of musical timbres," *Journal of the Acoustic Society of America*, vol. 61, no. 5, pp. 1270–1277, May 1977.

[6] S. Dixon, "Onset detection revisited," in *Proc. of the 9th Int. Conf. on Digital Audio Effects (DAfX-06)*, Montreal, Canada, 2006, pp. 133–137.

[7] R. Lyons and A. Bell, "The Swiss army knife of digital networks," *IEEE Signal Processing Magazine*, vol. 21, no. 3, pp. 90–10, May 2004.

[8] R. Bradford, J. ffitch, and R. Dobson, "Sliding with a constant Q," in *Proc. of the 11th Int. Conference on Digital Audio Effects (DAFx-08)*, Espoo, Finland, 2008.

[9] W. Sethares, *Tuning, Timbre, Spectrum, Scale*, 2nd ed. Springer, 2005.

[10] R. Holopainen, "Self-organised sounds with autonomous instruments: Aesthetics and experiments," Ph.D. dissertation, University of Oslo, Norway, 2012.

[11] S. Strogatz, *Nonlinear Dynamics and Chaos. With Applications to Physics, Biology, Chemistry, and Engineering*. Westview Press, 1994.

[12] J.-P. Eckmann, S. Oliffson Kamphorst, and D. Ruelle, "Recurrence plots of dynamical systems," *Europhysics Letters*, vol. 4, no. 9, pp. 973–977, 1987.

[13] J. Foote, "Automatic audio segmentation using a measure of audio novelty," in *Proceedings of IEEE International Conference on Multimedia and Expo*, vol. I, 2000, pp. 452–455.