An efficient algorithm for the calculation of a constant Q transform

Judith C. Brown
Physics Department, Wellesley College, Wellesley, Massachusetts 01281 and Media Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

Miller S. Puckette
IRCAM, 31 rue St Merri, Paris 75004, France

(Received 5 February 1992; revised 10 April 1992; accepted 16 June 1992)

An efficient method of transforming a discrete Fourier transform (DFT) into a constant Q transform, where Q is the ratio of center frequency to bandwidth, has been devised. This method involves the calculation of kernels that are then applied to each subsequent DFT. Only a few multiples are involved in the calculation of each component of the constant Q transform, so this transformation adds a small amount to the computation. In effect, this method makes it possible to take full advantage of the computational efficiency of the fast Fourier transform (FFT). Graphical examples of the application of this calculation to musical signals are given for sounds produced by a clarinet and a violin.

PACS numbers: 43.60.Gk, 43.75.Yy, 43.75.De, 43.75.Ef

I. THEORY

In many cases, such as that of musical signals, a constant Q transform gives a better representation of spectral data than the computationally efficient fast Fourier transform. Various solutions to this problem using constant Q filterbanks or a "bounded Q" Fourier transform have been proposed (Harris, 1976; Schwede, 1983; Mont-Reynaud, 1985). The music group at Marseilles has proposed a "wavelet transform" (Kronland-Martinet, 1988). Brown (1991) describes results applied to musical signals based on a direct evaluation of the DFT for the desired components.

We have calculated a constant Q transform based on transforming a fast Fourier transform into the log frequency domain. The FFT is calculated using a standard FFT program, and the entire calculation takes only slightly longer to run than the FFT since there are few operations involved in the computation of the transformation. The transformation is based upon the following. A constant Q transform can be calculated directly (Brown, 1991) by evaluating:

$$X^{cq}[k_{cq}] = \sum_{n=0}^{N[k_{cq}]-1} w[n,k_{cq}]x[n]e^{-j\omega_{k_{cq}}n}, \qquad (1)$$

where $X^{cq}[k_{cq}]$ is the k_{cq} component of the constant Q transform. Here $x[n]$ is a sampled function of time, and, for each value of k_{cq}, $w[n,k_{cq}]$ is a window function of length $N[k_{cq}]$. The exponential has the effect of a filter for center frequency $\omega_{k_{cq}}$.

In a constant Q filterbank the center frequencies are geometrically spaced; for musical applications, the calculation is often based on the frequencies of the equal tempered scale with

$$\omega_{k_{cq}} = (2^{(1/12)})^{k_{cq}}\omega_{\min} \qquad (2)$$

for semitone spacing.

The Q of a filter is defined as $f/\Delta f$, where Δf denotes bandwidth and f the center frequency. In the case of the filter defined in Eq. (1), this bandwidth depends upon the choice of the windowing function $w[n,k_{cq}]$, but it is inversely proportional to $N[k_{cq}]$. We may therefore keep Q constant by choosing values of $N[k_{cq}]$ inversely proportional to those of $\omega_{k_{cq}}$. Often the bandwidth is chosen as $\omega_{k_{cq}+1} - \omega_{k_{cq}}$, which is proportional to the center frequencies $\omega_{k_{cq}}$ because of their geometric spacing. In the case of the equal tempered scale, this leads to

$$Q = 1/(2^{(1/12)} - 1) \cong 17.$$

The direct evaluation of Eq. (1) is computationally inefficient. However, it can be shown that for any two discrete functions of time $x[n]$ and $y[n]$:

$$\sum_{n=0}^{N-1} x[n]y^*[n] = \frac{1}{N}\sum_{k=0}^{N-1} X[k]Y^*[k], \qquad (3)$$

where $X[k]$ and $Y[k]$ are the discrete Fourier transforms of $x[n]$ and $y[n]$, and $Y^*[k]$ is the complex conjugate of $Y[k]$. Equation (3) is a form of Parseval's equation (Oppenheim, 1975).

We can use Eq. (3) to evaluate Eq. (1) as follows. Letting

$$w[n,k_{cq}]e^{-j\omega_{k_{cq}}n} = \mathcal{K}^*[n,k_{cq}]. \qquad (4)$$

Equation (3) gives

$$X^{cq}[k_{cq}] = \sum_{n=0}^{N-1} x[n]\mathcal{K}^*[n,k_{cq}]$$

$$= \frac{1}{N}\sum_{k=0}^{N-1} X[k]K^*[k,k_{cq}], \qquad (5)$$

where $K[k,k_{cq}]$ is the discrete Fourier transform of $\mathcal{K}[n,k_{cq}]$; that is

$$K[k,k_{cq}] = \sum_{n=0}^{N-1} w[n,k_{cq}]e^{j\omega_{k_{cq}}n}e^{-j2\pi kn/N}. \qquad (6)$$

We will refer to the $K[k,k_{cq}]$ in the frequency domain as the spectral kernels of the transformation and to the $\mathcal{K}[n,k_{cq}]$ as the temporal kernels. We have used a Hamming window:

$$w[n,k_{cq}] = \alpha - (1-\alpha)\cos(2\pi n/N[k_{cq}]),$$

 0001-4966/92/112698-04$00.80

where $\alpha = 25/46$. In practice, we have chosen the window and the exponential to be symmetric about the center of the interval and thus evaluated

$$K[k,k_{cq}] = \sum_{n=0}^{N-1} w\left[n - \left(\frac{N}{2} - \frac{N(k_{cq})}{2}\right), k_{cq}\right] \times e^{j\omega_{k_{cq}}(n-N/2)} e^{-j2\pi kn/N}. \quad (7)$$

Here the window is zero outside the interval $(N/2 - N(k_{cq})/2,\ N/2 + N(k_{cq})/2)$. With the choice of frequencies of Eq. (2), we are calculating 12 constant Q components per octave corresponding to the 12 tones of the equal tempered scale. The choice of $f_{\min} = 174.6$ Hz corresponding to F_3, the F below middle C, leads to 60 components of the constant Q transform from F_3 to the Nyquist frequency. In some cases quarter tone spacing is preferable, and for this

$$f_{k_{cq}} = (2^{(1/24)})^{k_{cq}} f_{\min}$$

leading to 24 components per octave for a total of 120 components. In the following section, we will give examples of these two choices.

The spectral kernels are real since $\mathcal{K}[n,k_{cq}] = \mathcal{K}^*[-n,k_{cq}]$; that is, the temporal kernels are conjugate symmetric, which is the condition for a real discrete Fourier transform. We have evaluated the spectral kernels using only the real part of the temporal kernels, since we are looking at positive frequency components only.

Figure 1 is a display of the real part of the temporal kernels as defined in Eq. (4) for the first 30 kernels $\mathcal{K}[n,k_{cq}]$, that is, k_{cq} goes from 0 to 29 in Eq. (4). The vertical axis is labeled with the kernel number k_{cq} corresponding to the frequency of the kernel as defined in Eq. (2), and the horizontal axis is labeled with the sample number n as it appears in Eq. (4). These kernels are normalized with the window length so that sinusoidal components in the signal having the same amplitude will occur with the same amplitude for their constant Q transform. Thus the amplitude in the figure increases with the number of the component.

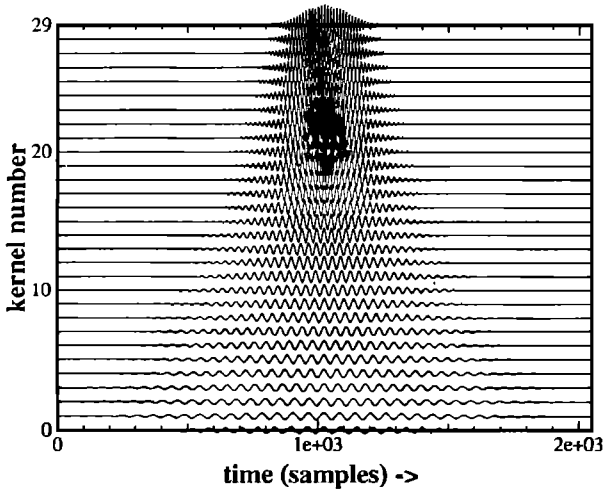


FIG. 1. Real part of temporal kernels plotted against sample number for the first 30 temporal kernels.

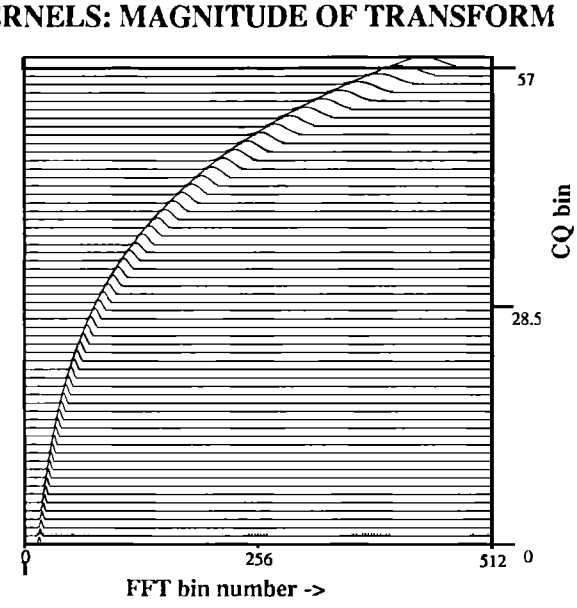


FIG. 2. Magnitude of kernels against number of Fourier frequency component k. Each frequency bin represents 10.8 Hz.

Figure 2 is a graph of the magnitude of the spectral kernels as defined in Eq. (6). The lower dotted and upper solid lines are artifacts of the plotting program and should be ignored. These kernels were obtained by taking a standard 1024 point FFT of the temporal kernels graphed in Fig. 1. The vertical axis is labeled with the kernel number k_{cq} corresponding to the frequency of the kernel as defined in Eq. (2), and the horizontal axis is labeled with the FFT frequency component number k as used in Eq. (7) or Eq. (6). These kernels need only be calculated once and can then be called for use in Eq. (5). It is clear from this figure that the kernels are near zero over most of the spectrum so there few multiplies involved in the evaluation of Eq. (5).

We have estimated the error in keeping only the values

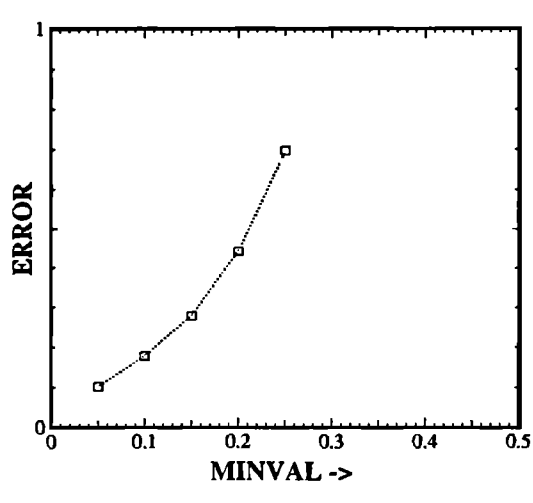


FIG. 3. Error in dropping small values of kernels (defined in text) plotted against cutoff value.

greater than an adjustable parameter called MINVAL by summing the absolute values of the numbers which are dropped and dividing by the sum of the absolute value of all values of the kernel. The results are given in Fig. 3 where we have plotted the error in the approximation as defined above against MINVAL.

Choosing MINVAL equal to 0.15, then there are only two multiplications (or one in a few cases) for components up to $X^{cq}[41]$ and two to six multiplications for the remaining components. In all there are roughly 280 multiplications for quarter tone spacing. These are negligible compared to the multiplications involved in the evaluation of the FFT. For a Q of 17 at a frequency of 175 Hz (the lowest frequency for which the constant Q transform is desired) and a sample rate of 11 025, a 1024 point FFT is needed. This gives rise to $1024 \log_2(1024) = 10\,240$ complex multiplies.

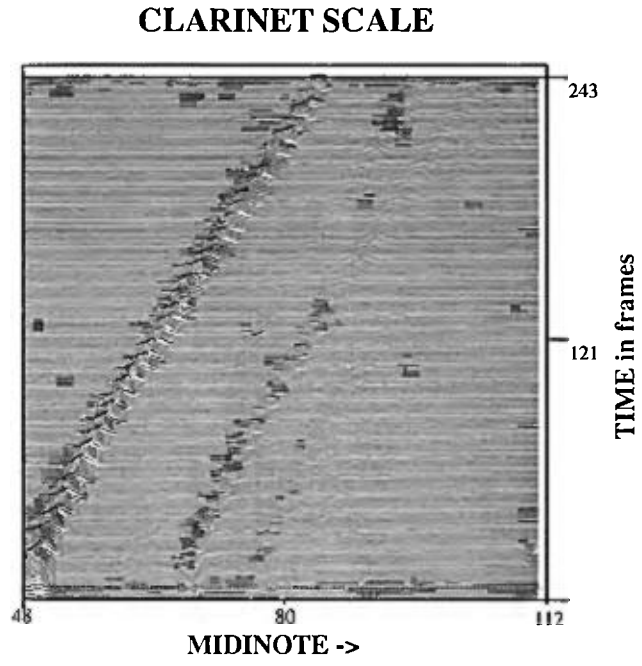
The direct evaluation of Eq. (1) involves sums over windows ranging from 1074 samples at the low-frequency end to 34 samples at the upper end. For quarter tone spacing this leads to roughly 35 000 complex terms. Thus our method leads to an increase of a factor of roughly $3\frac{1}{2}$ in computational efficiency.

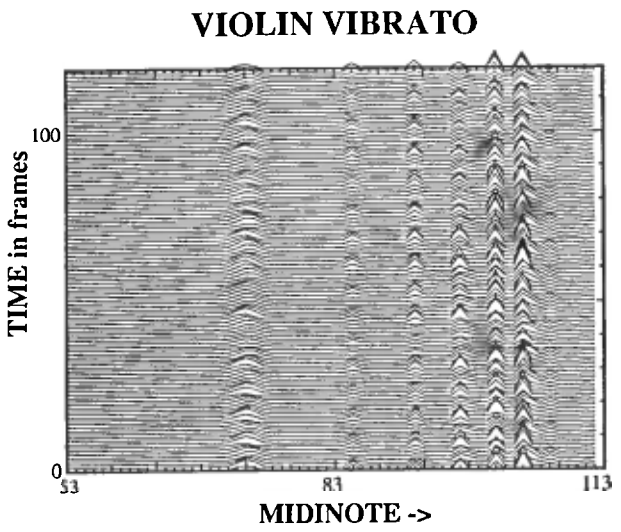
We have estimated the run time for our algorithm with calculations carried out on a 40-MHz Intel i860 using a hand-coded routine. With a 512-point FFT and quarter tone spacing over three octaves, the FFT takes 343 μs and the transform 166 ± 2 μs (measured on an oscilloscope). This can be compared to a time advance of 25 ms between frames, so the calculation can easily be carried out in real time. In a subsequent article, we shall describe approximations which increase the computational efficiency and are appropriate for the application of fundamental frequency tracking.

Although we only discuss the case of constant Q and centered windows here, it is sometimes useful to make other choices. For example, for low center frequencies one often wishes to decrease Q in order that the temporal kernels do not exceed a given length of time. This requirement usually arises from a need for temporal precision and is not related to our method of calculating the transform. For real-time applications, in which delay must be kept to a minimum, the temporal kernels may be aligned to the end of the sample window instead of centering them.

II. EXAMPLES

Graphical examples of the output are given in Figs. 4 to 8 where we have plotted Fourier amplitude, as calculated using Eq. (5) versus midi note for time frames with a 25-ms separation. The use of midi note for labeling is equivalent to that of frequency with middle C (261.6 Hz) assigned the value of 60, and each integer added corresponds to a step up in the equal tempered scale (or to multiplication by $2^{1/12}$). The sample rate was 11 025 for all examples. The sound source of Fig. 4 is a clarinet playing a chromatic scale. Particularly prominent in this graph is the absence of even harmonics, which is a feature of the spectra of tubes with one open end and one closed end. Figure 5 is the transform of the same clarinet in a portion of a performance of the piece "Dialogue de l'Ombre Double," by Pierre Boulez.





Figures 6 and 7 represent the constant Q spectrum of a violin with examples of vibrato in Fig. 6 and glissando in Fig. 7. The vibrato is not resolved in Fig. 6, so we have redone the calculation with quarter tone spacing and an appropriately larger window size in Fig. 8. We have chosen these particular sounds as examples since they will be used for fundamental frequency tracking in a subsequent article.

ACKNOWLEDGMENT

JCB is very grateful to Daniel P. W. Ellis of the MIT Media Lab for invaluable discussions. She is also grateful to Wellesley College for its generous Sabbatical leave policy and to IRCAM for the use of its facilities during her stay there.

Brown, J. C. (1991). "Calculation of a Constant Q Spectral Transform," J. Acoust. Soc. Am. **89**, 425–434.

Harris, F. J. (1976). "High-Resolution Spectral Analysis with Arbitrary Spectral Centers and Arbitrary Spectral Resolutions," Compt. Elect. Eng. **3**, 171–191.

Kronland-Martinet, R. (1988). "The Wavelet Transform for Analysis, Synthesis, and Processing of Speech and Music Sounds," Comput. Music J. **12**, 11–20.

Mont-Reynaud, B. (1985). "The Bounded-Q Approach to Time-Varying Spectral Analysis," Department of Music Technical Report STAN-M-28.

Oppenheim, A.V., and Schafer, R. W. (1975). *Digital Signal Processing* (Prentice-Hall, Englewood Cliffs, NJ).

Schwede, G. W. (1983). "An Algorithm and Architecture for Constant-Q Spectrum Analysis," Proc. ICASSP **29.2**, 1384–1387.