5

Transforms

Transforms model a signal as a collection of waveforms of a particular form: sinusoids for the Fourier transform, mother wavelets for the wavelet transforms, periodic basis functions for the periodicity transforms. All of these methods are united in their use of inner products as a basic measure of the similarity and dissimilarity between signals, and all may be applied (with suitable care) to problems of rhythmic identification.

Suppose there are two signals or sequences. Are they the same or are they different? Do they have the same orientation or do they point in different directions? Are they periodic? Do they have the same periods? The inner product is one way of quantifying the similarity of (and the dissimilarity between) two signals. It can be used to find properties of an unknown signal by comparing it to one or more known signals, a technique that lies at the heart of many common transform methods. The inner product is closely related to (cross) correlation, which is a simple form of pattern matching useful for aligning signals in time. A special case is autocorrelation which is a standard way of searching for repetitions or periodicities. The inner product provides the basic definitions of a variety of transform techniques such as the Fourier and wavelet transforms as well as the nonorthogonal periodicity transforms.

The first section reviews the basic ideas of the angle between two signals or sequences in terms of the inner product, and sets the mathematical notations that will be used throughout the chapter. Sect. 5.2 defines the cross correlation between two signals or sequences in terms of the inner product and interprets the correlation as a measure of the fit or alignment between the signals. Sect. 5.3 shows how the Fourier transform of a signal is a collection of inner products between the signal and various sinusoids. Some cautionary remarks are made regarding the applicability of transforms to the rhythm-finding problem. Two signal processing technologies, the short time Fourier transform and the phase vocoder are then described in Sects. 5.3.3 and 5.3.4. Wavelet transforms are discussed in Sect. 5.4 in terms of their operation as an inner product between a "mother wavelet" and the signal of interest. The final section describes the Periodicity transforms, which are again introduced in terms of an inner product, and some advantages are noted in terms of rhythmic processing.

5.1 Inner Product: The Angle Between Two Signals

The angle between two vectors gives a good indication of how closely aligned they are: if the angle is small then they point in nearly the same direction; if the angle is near 90 degrees, then they point in completely different directions (they are at right angles). The generalization of these ideas to sequences and signals uses the *inner product* to define the "angle." When the inner product is large, the sequences are approximately the same ("point in the same direction") while if the inner product is zero (if the two are *orthogonal*) then they are like two vectors at right angles.

The most common definition of the inner product between two vectors x and y is

$$\langle x\,,y\rangle = \sum_k x[k]y[k]. \tag{5.1}$$

The length (or *norm*) of a vector is the square root of the sum of the squares of its elements, and can also be written in terms of the inner product as

$$||x|| = \left(\sum_k x^2[k]\right)^{\frac{1}{2}} = \sqrt{\langle x\,,x\rangle}. \tag{5.2}$$

For example, consider the two vectors $x = (2,1)$ and $y = (1,0)$ shown in Fig. 5.1. The lengths of these vectors are $||x|| = \sqrt{2^2+1^2} = \sqrt{5}$ and $||y|| = \sqrt{1^2+0^2} = 1$ and the inner product is $\langle x\,,y\rangle = 2\cdot 1 + 1\cdot 0 = 2$. The angle between x and y is the θ such that

$$\cos(\theta) = \frac{\langle x\,,y\rangle}{||x||\,||y||}. \tag{5.3}$$

For the vectors in Fig. 5.1, $\cos(\theta) = \frac{2}{\sqrt{5}}$ and so $\theta = 0.46$ radians or about 26 degrees.

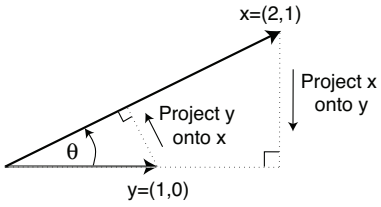

Fig. 5.1. The angle θ between two vectors x and y can be calculated from the inner product using (5.3). The projection of x in the direction y is $\cos(\theta)||x||$ (which is the same as $\frac{\langle x\,,y\rangle}{||y||}$). This is the dotted line forming a right angle with y. The projection of y onto x, given by $\frac{\langle x\,,y\rangle}{||x||}$, is also shown. If a projection is zero, then x and y are already at right angles (orthogonal).

The inner product is important because it extends the idea of angle (and especially the notion of a right angle) to a wide variety of signals. The definition (5.1) applies directly to sequences (where the sum is over all possible k) while

$$\langle x\,,y\rangle = \int_{-\infty}^{\infty} x(t)y(t)dt \tag{5.4}$$

defines the inner product between two functions $x(t)$ and $y(t)$ by replacing the sum with an integral. As before, if the inner product is zero, the two signals are said to be orthogonal. For instance, the two sequences

$$x = \{\ldots 1, 1, -1, -1, 1, 1, -1, -1, \ldots\}$$
$$y = \{\ldots 1, -1, 1, -1, 1, -1, 1, -1, \ldots\}$$

are orthogonal, and

$$z = \{\ldots 1, 1, 1, 1, 1, 1, 1, 1, \ldots\}$$

is orthogonal to both x and y. Taking all linear combinations of x, y, and z (i.e., the set of all $a_1 x + a_2 y + a_3 z$ for all real numbers a_i) defines a subspace with three dimensions. Similarly, the two functions

$$x(t) = \sin(2\pi f_1 t) \text{ and } y(t) = \sin(2\pi f_2 t)$$

are orthogonal whenever $f_1 \neq f_2$. The set of all linear combinations of sinusoids for all possible frequencies f_i is at the heart of the Fourier transform of Sect. 5.3 and orthogonality plays an important role because it simplifies many of the calculations. If the signals are complex-valued, then $y(t)$ in (5.4) (and $y[k]$ in (5.1)) should be replaced with their complex conjugates.

Suppose there is a set of signals x_i that all have the same norm, so that $||x_i||^2 = ||x_k||^2$ for all i and k. Given any signal y, the inner product can be used to determine which of the x_i's is closest to y where "closeness" is defined by the norm of the difference. Since

$$||y - x_i||^2 = ||y||^2 - 2\langle y, x_i \rangle + ||x_i||^2 \tag{5.5}$$

and since $||y||$ and $||x_i||$ are fixed, the i that minimizes the norm on the left hand side is the same as the i that maximizes the inner product $\langle y, x_i \rangle$.

5.2 Correlation and Autocorrelation

The (cross) correlation between two signals $x(t)$ and $y(t)$ with shift τ can be defined directly or in terms of the inner product:

$$\begin{aligned} R_{xy}(\tau) &= \int_{-\infty}^{\infty} x(t) y(t+\tau) dt \\ &= \langle x(t), y(t+\tau) \rangle . \end{aligned} \tag{5.6}$$

When the correlation $R_{xy}(\tau)$ is large, x and y point in (nearly) the same direction. If $R_{xy}(\tau)$ is small (near zero), $x(t)$ and $y(t+\tau)$ are nearly orthogonal. The correlation can also be interpreted in terms of similarity or closeness: large $R_{xy}(\tau)$ mean that $x(t)$ and $y(t+\tau)$ are similar (close to each other) while small $R_{xy}(\tau)$ mean they are different (far from each other). These follow directly from (5.5).

In discrete time, the (cross) correlation between two sequences $x[k]$ and $y[k+j]$ with time shift j is

$$R_{xy}(j) = \sum_{k=-\infty}^{\infty} x[k]y[k+j] = \langle x[k], y[k+j] \rangle . \quad (5.7)$$

Correlation shifts one of the sequences in time and calculates how well they match (by multiplying point by point and summing) at each shift. When the sum is small then they are not much alike; when the sum is large, many terms are similar. Equations (5.6) and (5.7) are recipes for the calculation of the correlation. First, choose a τ (or a j). Shift the function $y(t)$ by τ (or $y[k]$ by j) and then multiply point by point times $x(t)$ (or times $x[k]$). The area under the resulting product (or the sum of the elements) is the cross correlation. Repeat for all possible τ (or j).

Seven pairs of functions $x(t)$ and $y(t)$ are shown in Fig. 5.2 along with their correlations. In (a), a train of spikes is correlated with a Gaussian pulse. The correlation reproduces the pulse, once for each spike. In (b), the spike train is replaced by a sinusoid. The correlation smears the pulse and inverts it with each undulation of the sine. In (f), two random signals are generated: their correlation is small (and random) because the two random sequences are independent.

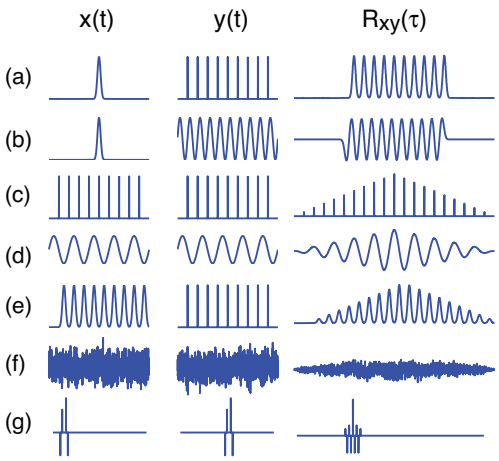

Fig. 5.2. Seven examples of the crosscorrrelation between two signals x and y. The examples consider spike trains, Gaussian pulses, sinusoids, pulse trains, and random signals. When $x = y$ (as in (c) and (d)), the largest value of the correlation occurs at a shift $\tau = 0$. The distance between successive peaks of $R_{xx}(\tau)$ is directly related to the periodicity in the input.

One useful situation is when x and y are two copies of the same signal but displaced in time. The variable τ shifts y and at some shift τ^* they become aligned. At this τ^*, $x(t)$ is the same as $y(t + \tau^*)$ and the product is positive everywhere: hence, when integrated, $R_{xy}(\tau^*)$ achieves its largest value. This situation is depicted in Fig. 5.2(g) which shows a y that is a shifted version of x. The maximum value of the correlation occurs at at the τ^* where $x(t) = y(t + \tau)$, where the signals are closest. Correlation is an ideal tool for aligning signals in time.

A special case that can be very useful is when the two signals x and y happen to be the same. In this case, $R_{xx}(\tau) = \langle x(t), x(t + \tau) \rangle$ is called the *autocorrelation* of x. For any x, the largest value of the autocorrelation always occurs at $\tau = 0$, that

is, when there is no shift. This is particularly useful when x is periodic since then $R_{xx}(\tau)$ has peaks at values of τ that correspond precisely to the period. For example, Fig. 5.2(c) shows a periodic spike train with one second between spikes. The autocorrelation has a series of peaks that are precisely one second apart. Similarly, in (d) the input is a sinusoid with frequency 0.5 Hz. The peaks of the autocorrelation occur 2 seconds apart, exactly the periodicity of the sine wave.

5.3 The Fourier Transform

Computer techniques allow us to look inside a sound; to dissect it into its constituent elements. But what are the fundamental elements of a sound? Are they sine waves, sound grains, wavelets, notes, beats, or something else? Each of these kinds of elements requires a different kind of processing to detect the regularities, the frequencies, scales, or periods.

As sound (in the physical sense) is a wave, it has many properties that are analogous to the wave properties of light. Think of a prism, which bends each color through a different angle and so decomposes sunlight into a family of colored beams. Each beam contains a "pure color," a wave of a single frequency, amplitude, and phase.[1] Similarly, complex sound waves can be decomposed into a family of simple sine waves, each of which is characterized by its frequency, amplitude, and phase. These are called the *partials*, or the *overtones* of the sound, and the collection of all the partials is called the *spectrum*. Fig. 5.3 depicts the *Fourier transform* in its role as a "sound prism."

This prism effect for sound waves is achieved using the Fourier transform. Mathematically, the Fourier transform of a function $x(t)$ is defined as

$$\begin{aligned} X(f) &= \int_{-\infty}^{\infty} x(t) e^{-j2\pi ft} dt \\ &= \left\langle x(t), e^{j2\pi ft} \right\rangle \end{aligned} \tag{5.8}$$

which is the inner product of the signal $x(t)$ and the complex-valued sinusoid[2] $e^{-j2\pi ft}$.

Consider the meaning of the Fourier transform (5.8). First, $X(f)$ is a function of frequency: for each f the integral defined by the inner product is evaluated to give a complex-valued number with magnitude m and angle θ. Since $X(f)$ is the correlation (inner product) between the signal $x(t)$ and a sinusoid of frequency f, m is the magnitude (and θ the phase) of the sine wave that is closest[3] to $x(t)$. Since sine waves of different frequencies are orthogonal[4] there is no interaction between

[1] For light, frequency corresponds to color, and amplitude to intensity. Like the ear, the eye is predominantly blind to the phase of a single sinusoid.

[2] Euler's formula specifies the relationship between real and complex sinusoids: $e^{\pm j\theta} = \cos(\theta) \pm j\sin(\theta)$.

[3] Recall (5.5).

[4] That is, the inner product of two sinusoids is $\left\langle e^{-j2\pi f_1 t}, e^{j2\pi f_2 t} \right\rangle = \delta(f_1 - f_2)$ where $\delta(z)$ is the "delta function" that has unit area and is zero except when $z = 0$.

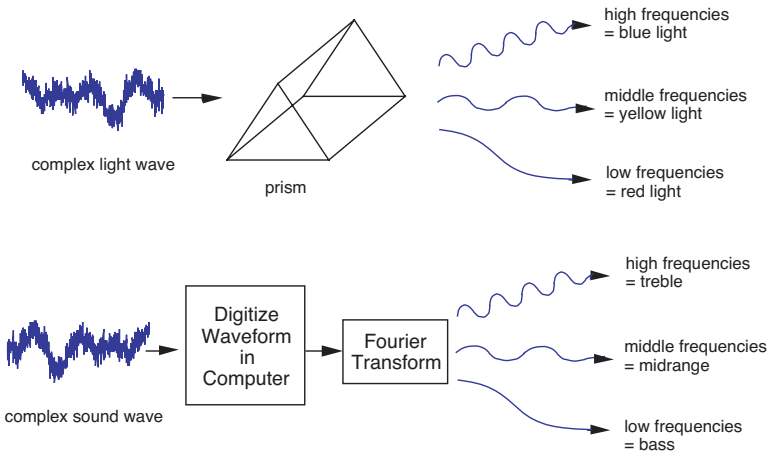

Fig. 5.3. Just as a prism separates light into its simple constituent elements (the colors of the rainbow), the Fourier Transform separates sound waves into simpler sine waves in the low (bass), middle (midrange), and high (treble) frequencies. Similarly, the auditory system transforms a pressure wave into a spatial array that corresponds to the various frequencies contained in the wave, as shown in Fig. 4.2 on p. 75.

different frequencies and m is the amount of the frequency f present in the signal $x(t)$. The Fourier transform shows how $x(t)$ can be uniquely decomposed into (and rebuilt from) sums of sinusoids.

Second, the Fourier transform is invertible. The inversion formula $x(t) = \int_{-\infty}^{\infty} X(f) e^{j2\pi ft} df = \langle X(f), e^{-j2\pi ft} \rangle$ reverses the role of the time and frequency variables and ensures that the transform neither creates nor destroys information.

5.3.1 Frequency via the DFT/FFT

The spectrum gives important information about the makeup of a sound and is most commonly implemented in a computer by running a program called the Discrete Fourier Transform (DFT) or the more efficient Fast Fourier Transform (FFT). Standard versions of the DFT and/or the FFT are available in audio processing software and in numerical packages (such as MATLAB and Mathematica) that can manipulate sound data files.

Like the Fourier transform, the DFT decomposes a signal into its constituent sinusoidal elements. Like the Fourier transform, the DFT is an invertible, information preserving transformation. But the DFT differs from the Fourier transform in three useful ways. First, it applies to discrete-time sequences which can be stored and manipulated directly in computers (rather than to functions or analog waveforms). Second, it is a sum rather than an integral, and so is easy to implement in either hardware or software. Third, it operates on a finite data record (rather than operating on a function that must be defined over all time). Given a sequence $x[k]$ of length N,

the DFT is defined by

$$X[n] = \sum_{k=0}^{N-1} x[k] e^{-j2\pi nk/N} \quad n = 0, 1, 2, ..., N-1$$
$$= \left\langle x[k], e^{j2\pi nk/N} \right\rangle . \tag{5.9}$$

For each value n, (5.9) multiplies each term of the data by a complex exponential and then sums. Compare this to the Fourier transform; for each frequency f, (5.8) multiplies each point of the waveform by a complex exponential and then integrates. Thus $X[n]$ is a function of frequency in the same way that $X(f)$ is a function of frequency. Indeed, the term $e^{-j2\pi nk/N}$ is a discrete-time sinusoid with frequency proportional to n.

A good example of the use of the DFT/FFT for spectral analysis appears in Fig. 2.19 on p. 43 which shows the waveform and corresponding spectrum of the pluck of a guitar string. While the time evolution of the signal is clear from the waveform, the underlying nature of the sound as a sum of a number of harmonically related sinusoids is clear from the spectrum. The two plots are complementary and display different aspects of the same sound.

One source of confusion is that the frequency f in the Fourier transform can take on any value while the frequencies present in (5.9) are all integer multiples n of $2\pi/N$. This "fundamental frequency" is precisely the sine wave with period equal to the length N of the window over which the DFT is taken. Thus the frequencies in (5.9) are constrained to a discrete set and the frequencies are separated by a constant difference. This resolution is equal to the sampling rate divided by the window size (the number of samples used in the calculation), that is,

$$\text{resolution in Hz} = \frac{\text{sampling rate}}{\text{window size}}. \tag{5.10}$$

For example, the window used for the guitar pluck in Fig. 2.19 contains $32,000$ samples and the sampling rate is 44.1 KHz. Thus the resolution is 1.38 Hz. The peak of the spectrum occurs at entry $n = 142$ in the output of the FFT, which corresponds to a frequency of $142 \cdot 1.38$ which is approximately (but not exactly) 196 Hz, as annotated in the figure. Observe that the units of (5.10) are inverse seconds (the units of the numerator are samples per second while the denominator has units of samples). Thus an accuracy of 10 Hz requires a duration of only 0.01 sec and an accuracy of 1.38 Hz requires a time window of 0.72 sec, as used with the guitar pluck. To achieve an accuracy of $\frac{1}{10}$ Hz would require a time window of at least 10 sec, irrespective of the sampling rate.

Why not simply use long windows for increased resolution? Because long windows do not show when (in time) events occur. For example, Fig. 5.4 shows a signal that consists of two sinusoids: a sine wave with frequency 150 Hz is followed by a somewhat larger wave with frequency 100 Hz. The magnitude spectrum shows peaks near the expected values of 100 and 150 Hz. But it does not show the order of the sine waves. Indeed, the magnitude spectrum is the same if the sine waves are reversed in

order, and even if they both sound for the entire time interval.[5] Thus, use of the FFT requires a compromise: long windows are desired in order to have good frequency resolution while short windows are desired in order to locate events accurately in time.

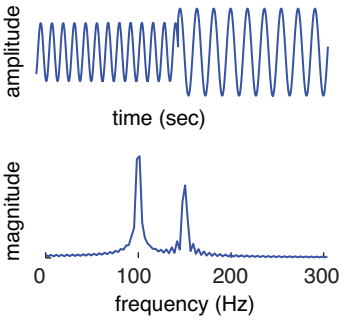

Fig. 5.4. A signal consists of two sine waves. The first, at 150 Hz, lasts for 0.5 sec and the second, at 100 Hz, begins when the first ends. The spectrum shows peaks corresponding to both sine waves but does not show their temporal relationship. The spectrum would look the same if the order of the sine waves were reversed or if they occurred simultaneously (rather than successively).

Windowing also influences the accuracy of frequency estimation through the effect called "smearing." Fig. 5.5 shows two different analyses of the same 200 Hz sine wave. In the top case, the window size is 0.5 seconds and so exactly 100 repetitions of the wave fit into the window. Accordingly, all of the inner products in (5.9) are zero except for one that has frequency exactly equal to 200 Hz. The algorithms in MATLAB report these values as less than 10^{-14}, which is numerically indistinguishable from zero. In contrast, the bottom analysis uses a window of 0.503 sec and so an integer number of waves does not fit exactly within the window. This implies that none of the terms in the inner product have frequency exactly 200 Hz. A large number of terms become nonzero in order to compensate, to represent a frequency that falls between the cracks of its resolution.[6]

5.3.2 Three Mistakes

Over the years, the Fourier transform has found many uses throughout science and engineering and it is easy to develop a naive overconfidence in its use. In terms of the rhythm finding goals of *Rhythm and Transforms*, the naive argument goes something like this:

> The Fourier transform is an ideal tool for finding frequencies and/or periodicities in complex data sets. The beat of a piece of music and the larger

[5] The phase spectrum of the three cases differs, but the relationship between the phase and the temporal order is notoriously difficult to decipher.

[6] The effect of smearing can be studied by observing that the windowed signal is equal to the product of the signal and the window. Consequently, the spectrum of the windowed signal is equal to the convolution of the spectrum of the signal (the spike as in the top part of Fig. 5.5) with the spectrum of the window (in this case, the rectangular window has a spectrum that is a sinc function). Thus the smearing can be controlled, but never eliminated, by careful choice of window function. See [B: 169] for details.

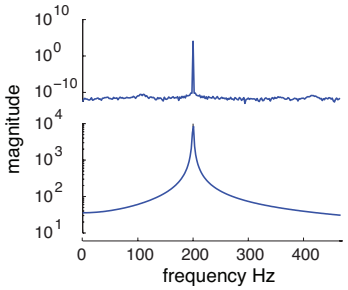

Fig. 5.5. A sine wave of frequency 200 Hz is analyzed twice, resulting in two spectra. The window used in top spectrum is 0.5 sec, and so an integer number of periods of the signal fits exactly. This means that one of the terms in the inner product (5.9) has frequency exactly equal to 200 Hz: this one is large and all others are (numerically) zero. In the bottom spectrum, the window width 0.503 does not support an integer number of periods. No single term in the inner product has frequency 200 Hz and the representation is "smeared."

> rhythmic structures are, at heart, different frequencies and/or periodicities that exist in the sound. Accordingly, it should be straightforward to apply the Fourier transform to find the beat and higher metrical structures within a musical passage.

This section shows that this argument is fundamentally flawed in three separate ways.

The first flaw is the easiest to see since it has been repeatedly emphasized throughout the earlier chapters: rhythmic phenomenon are only partially defined by the sound itself, they are heavily influenced by the perceptual apparatus of the listener. Accordingly, it is only sensible to expect to be able to locate the part of the rhythm that is in the sound itself using a technique such as the Fourier transform.

The second flaw arises from a misunderstanding of the nature of rhythmic phenomena. Consider naively applying the FFT to the first 100 sec of an audio CD in the hopes of finding "the beat" of a performance that occurs at (say) two times per second. As shown in Fig. 1.4 on p. 7, the phenomenon of musical beats occur at rates between about 0.2 Hz and 2 Hz. Formula (5.10) shows that 100 sec corresponds to a frequency resolution of 1/100 Hz which should allow detection within the needed range with a fair degree of accuracy. But surprise! The output of this FFT contains no measurable energy below 20 Hz. How can this be? We clearly *hear* the beat at 2 Hz, how can the FFT show nothing near 2 Hz?

The FFT says that there is no match between the signal (in this case the sound) and sinusoids with frequencies near 2 Hz. This should come as no surprise, since human hearing extends from a high of 20 KHz down to a low of about 20 Hz and we cannot directly perceive a 2 Hz sinusoid.[7] Yet we clearly perceive *something* occurring two times each second. In other words, the perception of rhythm is not a perception of sinusoids at very low frequencies. Rather, it is a perception of changes in energy at the specified rate. Thus the "hearing" of a pitch at 200 Hz is a very different phenomenon from the "hearing" of a rhythm at 2 Hz. While the Fourier transform is adept at displaying the physical characteristics of the sine waves associated with the perception of pitch, it does not straightforwardly display the physical characteristics of patterns of energy associated with rhythmic perception.

[7] It is common practice to filter out all frequencies below about 20 Hz on recordings. Even in live situations, music contains no purposeful energy at these frequencies.

Using this insight, it is easy to modify the sound wave so that the transform does reveal something. The simplest approach is to take the FFT of the energy of the sound wave (rather than of the sound wave itself). This is a primitive kind of perceptually-motivated data preprocessing that might lead to better replication of the ear's abilities. But it is a slippery slope: what kind of criteria will specify the best kind of preprocessing to use? Maybe it would be better to take the absolute value of the sound wave? Or to take the percent change in the absolute value of the energy? There are many possibilities, and it is hard to know what criteria for success look like.

The third flaw in the argument arises from the nature of the FFT itself. Consider the simplest situation where a drum beats at a regular rhythm. Some kind of simple preprocessing (such as taking the energy of the sound wave) is applied. The input to the transform looks like a train of somewhat noisy pulses. The output of the FFT is: a train of somewhat noisy-looking pulses. Fig. 5.6 shows three cases. In each case the signal is a set of regularly spaced noises with period T_i seconds. The transform is a set of regularly spaced pulses separated by $\frac{1}{T_i}$. As the time-pulses grow further apart, the frequency-pulses grow closer together.

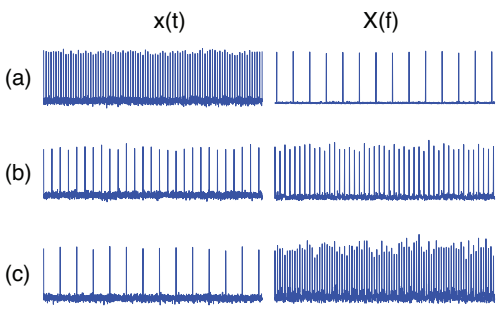

Fig. 5.6. The FFT is applied to a train of noisy pulses. The spectrum is again a train of noisy pulses. Close pulses in time imply widely separated pulses in frequency and distant pulses in time imply small separation in frequency. Three cases are shown with progressively longer period.

Students of the Fourier transform will recognize Fig. 5.6 as somewhat noisy versions of a fundamental result from Fourier series. Let $\delta(t)$ represent a single spike at time t. Then a train of such spikes with T sec between spikes is the sum $s(t) = \sum_k \delta(t - kT)$. The Fourier transform of $s(t)$ is[8]

$$S(f) = \frac{1}{T} \sum_{n=-\infty}^{\infty} \delta(f - \frac{n}{T})$$

which is itself a spike train in frequency. Thus the behavior in Fig. 5.6 is not a pathology of deviously chosen numerical parameters: it is simply how the transform works.

The goal of the analysis is to locate a regular succession in the input. In the case of a pulse train this requires locating the distance between successive pulses. As Fig. 5.6 suggests, it is no easier to locate the distance between pulses in the

[8] This result can be found in most tables of Fourier transforms since it is the key to the sampling theorem. See [B: 103].

transformed data than in the original data itself. Thus, at least in the situation of simple regular inputs like the pulse train, there is no compelling reason to believe that the transform provides any insight: it simply returns another problem with the same character as the original.

To summarize: application of the Fourier transform to the problem of describing rhythmic phenomena is neither straightforward nor obvious. First is the problem that only part of the perception of rhythm is located in the sound wave (this critique applies to all such signal-based approaches). Second is the problem that some kind of preprocessing of the audio signal is required in order for the transform to show anything. Finally, even in the idealized case where a rhythm consists of exact repetitions of a brief sound, the Fourier transform provides little insight.

These critiques do not, however, mean that the Fourier transform is incapable of playing a role in the interpretation of rhythmic phenomenon. Rather, they show that it is necessary to carefully consider proper uses of the FFT within a larger system. For example, it can be used as part of a method of pre-processing the audio so as to emphasize key features of a sound and to locate auditory boundaries where the character of a sound changes.

5.3.3 Short-time Fourier Transform

The short-time Fourier transform (STFT) is often used when a signal is too long to be analyzed with a single transform or when it is desirable to have better time-localization. The idea is to use a window function $w(t)$ that zeroes all but a short time interval. All events in the FFT are then localized to that interval. The windows are shaped so that when they are overlapped (shifted by S samples and summed) their sum $\sum_n w(t - nS)$ is constant for all t. This is shown schematically in Fig. 5.7 where the windows are overlapped by half their support.[9]

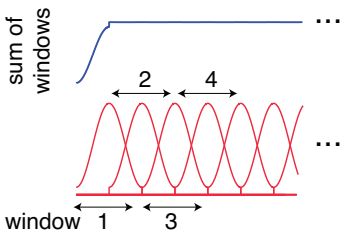

Fig. 5.7. A set of overlapping windows is used to zero all but a short segment of the signal. The FFT can then be applied to that segment in order to localize events in time. An overlap factor of 2 is shown; 4 might be more common in applications.

Using the window functions, the STFT can be described mathematically in much the same way as the Fourier transform itself

[9] Let $W(f)$ be the Fourier transform of the window $w(t)$ and $X(f)$ be the transform of the data within the time span of the window. Then the convolution of $W(f)$ and $X(f)$ describes the effect of the windowing on the data analysis. See [B: 169] or [B: 217] for a detailed comparison of various window functions.

$$X_{STFT}(f,\tau) = \left\langle x(t), w(t-\tau)e^{j2\pi ft} \right\rangle .$$

Observe that X_{STFT} is a function of both time (τ specifies where the window is nonzero) and frequency (f has the same meaning as in (5.8)). Similarly, the discrete-time version parallels the definition of the DFT in (5.9)

$$X_{STFT}[n,i] = \left\langle x[k], w[k-i]e^{j2\pi nk/N} \right\rangle$$

where n is the frequency variable and i specifies (via the window $w[k-i]$) where in time the FFT is taken. Thus the STFT provides a series of spectral snapshots that move through time. Plotting the snapshots sequentially is like looking at a multi-banded graphic equalizer. A common plotting technique is to change the magnitude of the spectra into colors (or into grayscale). Placing the frequency on the vertical axis and time on the horizontal axis leads to a spectrogram such as that depicting the *Maple Leaf Rag* in Fig. 2.20 on p. 44.

The operation of an STFT-based signal processor is diagrammed in Fig. 5.8. The signal is partitioned into segments by the windows. The FFT is applied to each segment separately (only one processing path is shown). The resulting spectrum may then be manipulated in any desired way; later chapters demonstrate some of the possibilities. In the figure, no changes are made to the spectrum and so the inverse transform (the IFFT block) rebuilds each segment without modification. When summed together, the segments reconstruct the original signal. Thus the STFT is invertible: it is possible to break the signal into spectral snapshots and then reconstruct the original signal from the snapshots.

In typical use, the support of the window (the region over which it is nonzero) is between 512 and 4096 samples. Using a medium window of size 2048 and a sampling rate of 44.1 KHz, the resolution in frequency is, from (5.10), about 21.5 Hz. The resolution in time is $\frac{2048}{44100} \approx 46$ ms (about 1/20 of a sec). This may be adequate to specify high frequencies (where 21.5 Hz is a small percentage of the frequency in question) but it is far too coarse at the low end. A low note on the piano may have a fundamental near 80 Hz. The resolution of this FFT is only good to within 25%! For comparison, the distance between consecutive notes on the piano is a constant 6%. Musical keyboards and scales are designed so that all equidistant musical intervals are a constant percentage apart in frequency, mimicking the constant percentage pitch perception of the auditory system.

This discussion raises two questions. First, is there a way to improve the frequency resolution of the STFT without overly harming the time resolution? The phase vocoder makes improved frequency estimates by using phase information that the STFT ignores; this is explored in Sect. 5.3.4. Second, is there a way to create a transform that operates at constant percentages (like the ear) rather than at constant differences? Brown's "constant-Q spectral transform" [B: 19] uses variable length windows (small ones to analyze high frequencies and large ones to capture low frequencies) that are tuned logarithmically like the steps of the 12-tone equal tempered scale (the chromatic scale). But it has not become popular, probably due to its non-invertibility (hence it cannot be used in a signal processing system like the STFT of

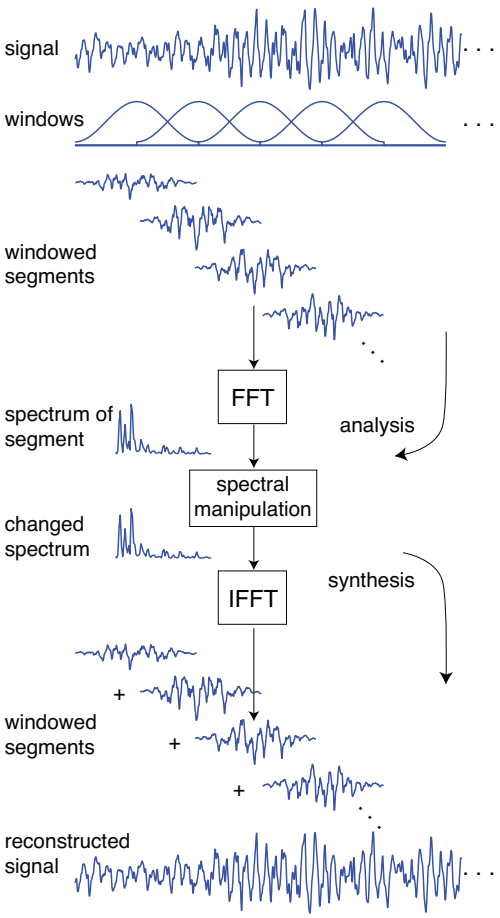

Fig. 5.8. A short-time Fourier transform (STFT) signal processor is an analysis/synthesis method that begins by windowing a signal into short segments. The FFT is applied to each segment separately and the resulting spectral snapshot can be manipulated in a variety of ways. After the desired spectral changes, the resynthesis is handled by the inverse FFT to return each segment to the time domain. The modified segments are then summed. For the special case where no spectral manipulations are made (as shown), the output of the STFT is identical to the input.

Fig. 5.8). Perhaps the most successful method that can operate with constant percentages is the wavelet transform, which is discussed in Sect. 5.4.

5.3.4 The Phase Vocoder

Like the short-time Fourier transform, the phase vocoder (PV) is an analysis-resynthesis technique based on the FFT. The analysis portion of the PV begins by slicing the signal into windowed segments that are analyzed using the FFT. If the PV used only the magnitude spectrum, the frequency resolution of each segment would be dictated by (5.10). Instead, the PV compares the phases of corresponding partials in successive segments and uses the comparison to improve the frequency estimates. The gains can be considerable. The resynthesis of the PV calculates a vector that can be inverted using the IFFT. The resulting signal has the same frequency content as the original but it is stretched or compressed in time.

Phase vocoders based on banks of (analog) filters were introduced by Flanagan [B: 62] for the compression of speech signals. Portnoff [B: 170] showed how the same idea can be implemented digitally using the FFT, and Dolson's tutorial [B: 49] helped bring the method to the attention of the computer music community. Recent work such as Laroche [B: 125] focuses on fine-tuning the resynthesis portion of the algorithm for various applications such as pitch shifting and time-stretching. Two MATLAB implementations are currently available on the internet: see Brandorff and Møller-Nielsen's pVoc [W: 6] and Ellis' pvoc.m [W: 13]. Also notable is Klingbeil's graphical interface called SPEAR (Sinusoidal Partial Editing Analysis and Resynthesis) [B: 114] [W: 21]. There is also a version on the CD in the software folder.

Analysis Using the Phase Vocoder

To see how the analysis portion of the PV can use phase information to make improved frequency estimates, suppose there is a sinusoid of unknown frequency but with known phases: at time t_1 the sinusoid has phase θ_1 and at time t_2 it has phase θ_2. The situation is depicted in Fig. 5.9. The sinusoid may have a frequency that moves it directly from θ_1 to θ_2 in time $\Delta t = t_2 - t_1$. Or it may begin at θ_1, move completely around the circle, and end at θ_2 after one full revolution. Or it may revolve twice, or n times.[10] Thus the frequency must be

$$f_n = \frac{(\theta_2 - \theta_1) + 2\pi n}{2\pi \Delta t} \tag{5.11}$$

for some integer n. Without more information, it is not possible to uniquely determine f, though it is constrained to one of the above values.

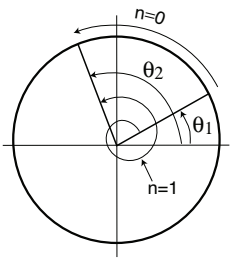

Fig. 5.9. The phases θ_1 and θ_2 of a sinusoid are known at two different times t_1 and t_2. The frequency must then fulfill f_n of (5.11), where the integer n specifies the number of revolutions around the circle (the number of periods of the sinusoid that occur within the specified time Δt). The two cases $n = 0$ (the lowest possible frequency) and $n = 1$ (one complete revolution) are shown.

The phase vocoder exploits (5.11) by locating a common peak in the magnitude spectrum of two different segments. It then chooses the f_n that is closest to the frequency of that peak. This is shown diagrammatically in Fig. 5.10 where the signal is assumed to be a single sinusoid that spans the time interval over which the calculations are made. The output of the windowing is a collection of short sinusoidal bursts.

[10] In other words, the frequency multiplied by the change in time must equal the change in angle, that is, $2\pi f(t_2 - t_1) = \theta_2 - \theta_1$ or some 2π multiple. Solving for f gives (5.11).

The FFT is applied to each burst, resulting in magnitude and phase spectra. For the case of a pure sinusoidal input, the magnitude spectra of the successive spectra are the same (as shown). But the phase spectra differ, and these provide the needed values of θ_1 (the phase corresponding to the peak of the first magnitude spectrum) and θ_2 (the phase corresponding to the peak of the second magnitude spectrum). The time difference Δt can be determined directly from the window length, the overlap factor, and the sampling rate. These values are then substituted into (5.11) and the f_n that is closest in frequency to the peak is the PV's frequency estimate.

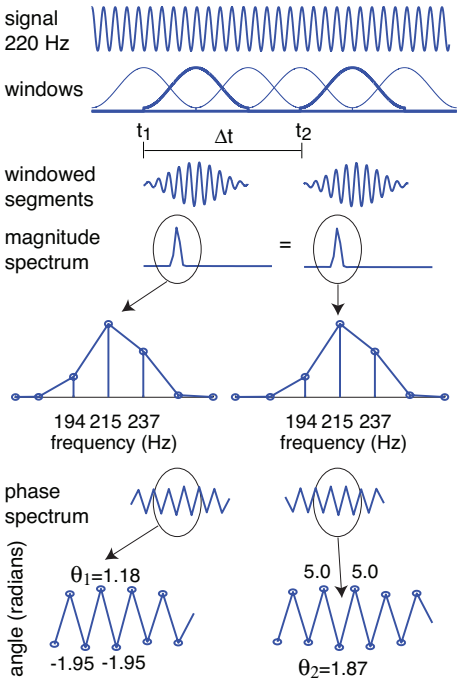

Fig. 5.10. The analysis portion of the phase vocoder rests on the assumption that a sinusoid remains fixed in frequency for the duration of the calculation. The input (shown here as a 220 Hz sinusoid) is windowed and the FFT is taken of the resulting bursts. The common peaks in the magnitude spectra are located (in this case at 215 Hz) and their phases recorded (in this case, the phase corresponding to the first and second bursts are $\theta_1 = 1.18$ and $\theta_2 = 1.87$). Information about sampling rate, window size, and overlap factor specify the time interval between the bursts (in this case, $\Delta t = 0.07$). These parameters are entered into (5.11) and the f_n closest to the frequency of the peak is chosen as the frequency estimate. In more interesting signals, when there are many sinusoids, the method is repeated for each peak in the magnitude spectrum.

To see the PV in action, and to give an idea of its accuracy, consider the problem of estimating the frequency of a 220 Hz sinusoid using a $2K$ FFT (assuming a sampling rate of 44.1 KHz). According to (5.10), the resolution of the FFT is 21.5 Hz, that is, it is only possible to find the frequency of the sinusoid to within about 10 Hz. Indeed, the nearby frequencies that are exactly representable are 193.8, 215.3, and 236.9, as shown in the enlargement of the magnitude spectrum in Fig. 5.10. Since the peak at 215.3 is the largest, an actual error of 4.7 Hz occurs when using only the FFT magnitude. The PV improves this by exploiting phase information. The phases corresponding to the peaks at 215.3 are $\theta_1 = 1.18$ and $\theta_2 = 1.87$ and so

$$f_n = \frac{(\theta_2 - \theta_1) + 2\pi n}{2\pi \Delta t} = \frac{(1.18 - 1.87) + 2\pi n}{2\pi \cdot 0.07}$$

since[11] $\Delta t = \frac{2048}{2} \cdot \frac{1}{44100} = 0.023$. With these values, the first six f_n are

$$47.7472,\ 90.8136,\ 133.8800,\ 176.94649,\ 220.01290,\ \text{and}\ 263.0793.$$

Clearly, the fifth term is closest to 215.3, and the error in the frequency estimate is 0.0129, a vast improvement over 4.7 Hz. This kind of accuracy is typical and is not just a numerical fluke. In fact, [B: 176] shows that, under certain conditions (for a signal consisting of a single sinusoid and with a Δt corresponding to a single sample) the phase vocoder estimate of the frequency is closely related to the maximum likelihood estimate.

In more complex situations, when the input signal consists of many sine waves, the phase manipulations are repeated for each peak individually, which is justified as long as the peaks are adequately separated in frequency. Once the analysis portion is complete, it is possible to change the signal in a variety of ways: by modifying the rate at which time passes (spacing the output bursts differently from the input bursts), by changing the frequencies in the signal (so that the output will contain different frequencies than the input), by adding or by subtracting partials.

Resynthesis Using the Phase Vocoder

Once the modifications are complete, it is necessary to synthesize the output waveform. One possibility is to use a straightforward additive-synthesis where the partials (each with its desired frequency and amplitude) are generated individually and then summed together. This is computationally intensive[12] when there are a large number of partials. Fortunately, there is a better way: the PV creates a complex-valued (frequency) vector. This is inverted using the IFFT and the resulting output bursts are time-shifted and summed as in the STFT.

Specification of the magnitude spectrum of the output is straightforward since it can be inherited directly from the input. The phase values are chosen to ensure continuity of the most prominent partials through successive bursts, as shown in Fig. 5.11 for a single sinusoid. For each peak j in the magnitude spectrum, the required phase

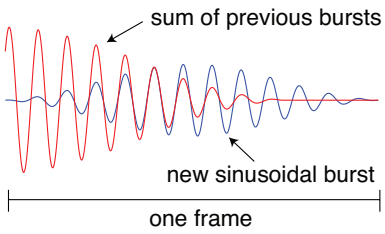

Fig. 5.11. Each new windowed sinusoidal element (burst) is added in phase with the already existing signal. The result is a continuous sinusoid of the specified frequency.

[11] The window width is 2048 with an overlap of 2, the sampling rate is 44.1 KHz, and the second burst is one step ahead of the first.

[12] There is still the problem of assigning appropriate phase values to the generated sine waves, a problem that the phase vocoder handles elegantly.

can be calculated directly from the frequency f_j and the time interval Δt between frame k and frame $k-1$. This is

$$\theta_k^j = \theta_{k-1}^j + 2\pi f^j \Delta t.$$

It is also necessary to choose the nearby phases (those under the same peak in the magnitude spectrum). If these are chosen to be $\theta_k^j + \text{mod}(n, 2)\pi$ (where n is the number of bins away from the peak value), the burst generated by the IFFT will be windowed with tapered ends, as in Fig. 5.11. For example, in the phase spectrum plots of Fig. 5.10, the values to the left and right of θ_1 and θ_2 are (approximately) either π or 0 away.[13] The MATLAB savvy reader will find an implementation of the phase vocoder called `PV.m` in the software folder on the CD.

5.4 Wavelet Transforms

In the STFT and the phase vocoder, sinusoids are used as basis functions and windows are used to localize the signal to a particular time interval. In the wavelet transforms, the windows are incorporated directly into the basis functions and a variety of nonsinusoidal shapes are common. Several different "mother wavelets" are shown in Fig. 5.12.

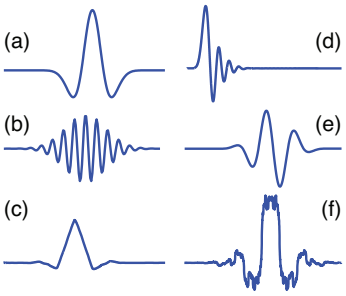

Fig. 5.12. There are many kinds of wavelet basis functions, including the (a) Mexican Hat wavelet, (b) complex Morlet wavelet, (c) Coiflets wavelet, (d) Daubechies wavelet, (e) complex Gaussian wavelet, and the (f) biorthogonal spline wavelet. The wavelet transform operates by correlating the signal with scaled and shifted versions of a basis function.

The wavelet transform uses the mother wavelet much as the STFT uses a windowed sinusoid: one parameter specifies where in time the wavelet is centered (analogous to the windowing) and another parameter stretches or compresses the wavelet. This latter is called the *scale* of the wavelet and is analogous to frequency in the STFT.

Let $\psi(t)$ be the mother wavelet (for example, any of the signals in Fig. 5.12[14]), and define

[13] A formal justification of this choice requires observing the phase values of a Bartlett (and a Parzen window) have exactly this pattern of values. Other patterns of 0 and π, such as that in [W: 32], correspond to different choices of output windowing functions.

[14] To be considered a wavelet function, $\psi(t)$ must be orthogonal to the function $x(t) = 1$ and must have finite energy. Thus $\langle 1, \psi^*(t) \rangle = 0$ and $\langle \psi(t), \psi^*(t) \rangle < \infty$.

$$\psi_{a,b}(t) = \frac{1}{\sqrt{|a|}}\,\psi\left(\frac{t-b}{a}\right).$$

The parameter b shifts the wavelet in time while the parameter a scales the wavelet by stretching or compressing it (and also by adjusting the amplitude). Fig. 5.13 illustrates the effects of the two parameters for several different values.

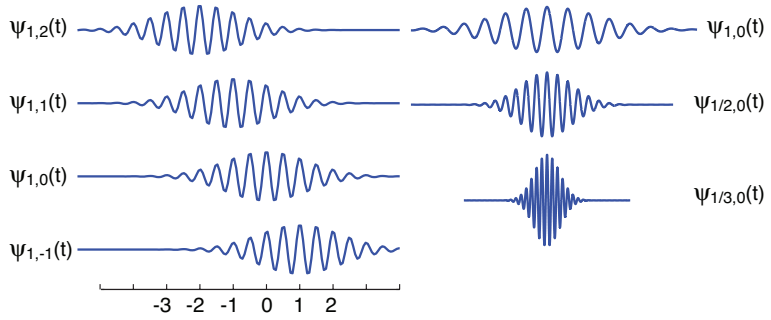

Fig. 5.13. The complex Morlet wavelet is a complex-valued sinusoid windowed by a Gaussian envelope $\psi_{a,b}(t) = \frac{1}{\sqrt{3\pi|a|}} e^{j2\pi 2t} e^{-\frac{(t-b)^2}{3a}}$. These plots show the real part of the Morlet wavelet for a variety of shifts b and scales a. As b decreases, the wavelet moves to the right; as a decreases, the wavelet compresses and grows.

The continuous wavelet transform uses the shifted and scaled functions as a basis for representing a signal $x(t)$ via the inner product

$$W(a,b) = \left\langle x(t), \psi^*_{a,b}(t) \right\rangle. \tag{5.12}$$

For every (a,b) pair, the coefficient $W(a,b)$ is the inner product of the signal with the appropriately scaled and shifted basis function $\psi_{a,b}(t)$. Where the signal is aligned with the basis function (when $x(t)$ locally looks like the basis function), the coefficient is large. Where the signal is very different from the basis function (the extreme being orthogonal) then the coefficient is small. As a and b change, the inner product scans through the signal looking for places (in time) and values (in scale) where the signal correlates well with the wavelet. This suggests that prior information about the shape or general character of the signal can be usefully exploited by the wavelet transform by tailoring the wavelet to the signal. When the information is correct, the set of parameters $W(a,b)$ can provide a concise and informative representation of the signal. When the information is incorrect, the wavelet representation may be less useful.

When the wavelet is real-valued, $W(a,b)$ is real; when the wavelet is complex-valued (like the Morlet wavelet used in Fig. 5.13) then $W(a,b)$ is complex. It is common to plot the magnitude of $W(a,b)$ (using a color or grayscale mapping) with

axes defined by the scale a and time b, much as the STFT and the spectrogram are plotted with axes defined by frequency and time. When $W(a, b)$ is complex, it is also common to plot a similar contour with the phase, though sometimes plots of the real and/or imaginary parts are useful. For example, Fig. 5.14 shows separate plots of the magnitude and phase when the complex Gaussian wavelet (from Fig. 5.12(e)) is applied to an input that is a train of spikes separated by one second. The temporal locations of the spikes are readily visible in both the magnitude and the phase plots (the vertical stripes).

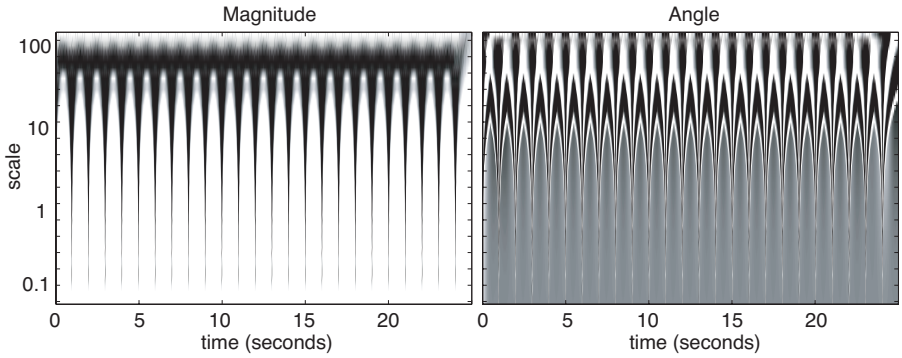

Fig. 5.14. The complex Gaussian wavelet is applied to an input spike train with period one second. The left plot shows values of the magnitude of the wavelet coefficients $W(a, b)$ of (5.12) while the right hand plot shows the phase. The location in time of the spikes is easy to see.

There is an interesting parallel between the wavelet transform of the spike train in Fig. 5.14 and the corresponding Fourier transform of a spike train in Fig. 5.6. In both cases, the transform returns a display (plot) that contains data of the same general character as the input. The output of the FT is a spike train; the output of the wavelet transform is a collection of regularly spaced ridges in a two-dimensional field. This suggests that the wavelet transform is not going to be able to magically solve the rhythm finding problem. In many cases (such as the spike train) it is no simpler to determine regularity from the output of the wavelet transform than it is to determine regularity directly from the input itself. Again, as with the FT, this does imply that wavelet transforms cannot play a role in rhythm analysis. Rather, it means that they must be used thoughtfully and in proper contexts.

This discussion has stressed the similarities between the STFT and the wavelet transforms. There are also important differences. In the windowed FFT and the granular techniques (such as the "Gabor grains" of Fig. 2.23 on p. 46), the frequency of the waveform is independent of the grain duration. In wavelets, there is an inverse relation maintained between the frequency of the waveforms and the duration of the wavelet. Unlike a typical grain, most wavelets contain the same number of cycles irrespective of the scale (roughly, frequency) of the wavelet. Thus the duration of the

wavelet window grows or shrinks as a function of the scale; wavelets that capture low frequency information are dilated (wide in time) while those that represent high frequencies are contracted. This allows more precise localization of the high frequency components. This can be seen in Fig. 5.13 where the Morlet wavelet maintains the same number of cycles at all scale values.

5.5 Periodicity Transform

The Periodicity Transform (PT) decomposes data into a sum of periodic sequences by projecting onto a set of "periodic subspaces" $\mathcal{P}_p$, leaving residuals whose periodicities have been removed. As the name suggests, this decomposition is accomplished directly in terms of periodic sequences and not in terms of frequency or scale, as do the Fourier and Wavelet Transforms. In consequence, the representation is linear-in-period, rather than linear-in-frequency or linear-in-scale. Unlike most transforms, the set of basis vectors is not specified a priori, rather, the Periodicity Transform finds its own "best" set of basis elements. In this way, it is analogous to the approach of Karhunen-Loeve [B: 23], which transforms a signal by projecting onto an orthogonal basis that is determined by the eigenvectors of the covariance matrix. In contrast, the periodic subspaces $\mathcal{P}_p$ lack orthogonality, which underlies much of the power of (and difficulties with) the Periodicity Transform. Technically, the collection of all periodic subspaces forms a *frame* [B: 24], a more-than-complete spanning set. The PT specifies ways of sensibly handling the redundancy by exploiting some of the general properties of the periodic subspaces.

This section describes the PT and compares its output to other transforms in a number of examples. Later chapters will detail how the PT may be applied to the problem of detecting rhythmic patterns in a musical setting. Much of this is based on the discussion in [B: 206] and [B: 207], which may be found on the accompanying CD.

5.5.1 Periodic Subspaces

A sequence of real numbers $x(k)$ is called *p-periodic* if there is an integer p with $x(k+p) = x(k)$ for all integers k. Let

$\mathcal{P}_p$ be the set of all p-periodic sequences, and
$\mathcal{P}$ be the set of all periodic sequences.

In practice, a data vector x contains N elements. This can be considered to be a single period of an element $x_N \in \mathcal{P}_N \subset \mathcal{P}$, and the goal is to locate smaller periodicities within x_N, should they exist. The strategy is to "project" x_N onto the subspaces $\mathcal{P}_p$ for $p < N$. When x_N is "close to" some periodic subspace $\mathcal{P}_p$, then there is a p-periodic element $x_p \in \mathcal{P}_p$ that is close to the original x. This x_p is an ideal choice to use when decomposing x. To make these ideas concrete, it is necessary to understand the structure of the various spaces, and to investigate how the needed calculations can be realized.

Observe that $\mathcal{P}_p$ is closed under addition since the sum of two sequences with period p is itself p-periodic. Similarly, $\mathcal{P}$ is closed under addition since the sum of x_1 with period p_1 and x_2 with period p_2 has period (at most) $p_1 p_2$. Thus, with scalar multiplication defined in the usual way, both $\mathcal{P}_p$ and $\mathcal{P}$ form linear vector spaces, and $\mathcal{P}$ is equal to the union of the $\mathcal{P}_p$.

For every period p and every "time shift" s, define the sequence $\delta_p^s(j)$ for all integers j by

$$\delta_p^s(j) = \begin{cases} 1, \text{ if } (j-s) \bmod p = 0 \\ 0, \qquad \text{otherwise} \end{cases} . \tag{5.13}$$

The sequences δ_p^s for $s = 0, 1, 2, ..., p-1$ are called the p-periodic basis vectors since they form a basis for $\mathcal{P}_p$.

Example 5.1. For $p = 4$, the 4-periodic basis vectors

j	$\cdots$	-4	-3	-2	-1	0	1	2	3	4	5	6	7	$\cdots$
$\delta_4^0(j)$	$\cdots$	1	0	0	0	1	0	0	0	1	0	0	0	$\cdots$
$\delta_4^1(j)$	$\cdots$	0	1	0	0	0	1	0	0	0	1	0	0	$\cdots$
$\delta_4^2(j)$	$\cdots$	0	0	1	0	0	0	1	0	0	0	1	0	$\cdots$
$\delta_4^3(j)$	$\cdots$	0	0	0	1	0	0	0	1	0	0	0	1	$\cdots$

span the 4-periodic subspace $\mathcal{P}_4$.

An inner product can be imposed on the periodic subspaces by considering the function from $\mathcal{P} \times \mathcal{P}$ into $\Re$ defined by

$$< x, y > = \lim_{k \to \infty} \frac{1}{2k+1} \sum_{i=-k}^{k} x(i) y(i), \tag{5.14}$$

for arbitrary elements x and y in $\mathcal{P}$. For the purposes of calculation, observe that if $x \in \mathcal{P}_{p_1}$ and $y \in \mathcal{P}_{p_2}$, the product sequence $x(i)y(i) \in \mathcal{P}_{p_1 p_2}$ is $p_1 p_2$-periodic, and (5.14) is equal to the average over a single period, that is,

$$< x, y > = \frac{1}{p_1 p_2} \sum_{i=0}^{p_1 p_2 - 1} x(i) y(i). \tag{5.15}$$

The corresponding norm on $\mathcal{P}$ is called the *Periodicity Norm*

$$||x|| = \sqrt{< x, x >}. \tag{5.16}$$

These definitions of inner product and norm are slightly different from (5.1) and (5.2). The extra term ($\frac{1}{p_1 p_2}$ in the example above) ensures that the norm gives the same value whether x is considered to be an element of $\mathcal{P}_p$, of $\mathcal{P}_{kp}$ (for positive integers k), or of $\mathcal{P}$.

Example 5.2. Let $x \in \mathcal{P}_3$ be the 3-periodic sequence $\{\cdots, 1, 2, 3, \cdots\}$ and let $y \in \mathcal{P}_6$ be the 6-periodic sequence $\{\cdots, 1, 2, 3, 1, 2, 3, \cdots\}$. Using (5.16), $||x|| = ||y||$.

As usual, the signals x and y in $\mathcal{P}$ are said to be orthogonal if $< x, y >= 0$.

Example 5.3. The periodic basis elements δ_p^s for $s = 0, 1, ..., p-1$ are orthogonal, and $||\delta_p^s|| = \sqrt{1/p}$.

The idea of orthogonality can also be applied to subspaces. A signal x is orthogonal to the subspace $\mathcal{P}_p$ if $< x, x_p >= 0$ for all $x_p \in \mathcal{P}_p$, and two subspaces are orthogonal if every vector in one is orthogonal to every vector in the other. Unfortunately, the periodic subspaces $\mathcal{P}_p$ are not orthogonal to each other.

Example 5.4. If p_1 and p_2 are mutually prime, then

$$< \delta_{p_1}^s, \delta_{p_2}^s >= < \delta_{p_1 p_2}^s, \delta_{p_1 p_2}^s >= \frac{1}{p_1 p_2} \neq 0.$$

Suppose that $p_1 p_2 = p_3$. Then $\mathcal{P}_{p_1} \subset \mathcal{P}_{p_3}$ and $\mathcal{P}_{p_2} \subset \mathcal{P}_{p_3}$, which restates the fact that any sequence that is p-periodic is also np-periodic for any integer n. But $\mathcal{P}_{p_3}$ can be strictly larger than $\mathcal{P}_{p_1} \cup \mathcal{P}_{p_2}$.

Example 5.5. Let $x = \{\cdots, 1, 2, 1, -1, -2, -1, \cdots\} \in \mathcal{P}_6$. Then x is orthogonal to both $\mathcal{P}_2$ and $\mathcal{P}_3$, since direct calculation shows that x is orthogonal to δ_2^s and to δ_3^s for all s.

In fact, no two subspaces $\mathcal{P}_p$ are linearly independent, since $\mathcal{P}_1 \subset \mathcal{P}_p$ for every p. This is because the vector $\mathbf{1}$, (the 1-periodic vector of all ones) can be expressed as the sum of the p periodic basis vectors

$$\mathbf{1} = \sum_{s=0}^{p-1} \delta_p^s$$

for every p. In fact, $\mathcal{P}_1$ is the only commonality between $\mathcal{P}_{p_1}$ and $\mathcal{P}_{p_2}$ when p_1 and p_2 are mutually prime. More generally, $\mathcal{P}_{np} \cap \mathcal{P}_{mp} = \mathcal{P}_p$ when n and m are mutually prime. The structure of the periodic subspaces reflects the structure of the integers.

5.5.2 Projection onto Periodic Subspaces

The primary reason for formulating this problem in an inner product space is to exploit the projection theorem. Let $x \in \mathcal{P}$ be arbitrary. Then a minimizing vector in $\mathcal{P}_p$ is an $x_p^* \in \mathcal{P}_p$ such that

$$||x - x_p^*|| \leq ||x - x_p||, \quad \text{for all } x_p \in \mathcal{P}_p.$$

Thus x_p^* is the p-periodic vector "closest to" the original x. The projection theorem, from Luenberger [B: 135]], is stated here in slightly modified form, shows how x_p^* can be characterized as an orthogonal projection of x onto $\mathcal{P}_p$.

Theorem 5.6 (The Projection Theorem). *Let $x \in \mathcal{P}$ be arbitrary. A necessary and sufficient condition that x_p^* be a minimizing vector in $\mathcal{P}_p$ is that the error $x - x_p^*$ be orthogonal to $\mathcal{P}_p$.*

Since $\mathcal{P}_p$ is a finite (p-dimensional) subspace, x_p^* will in fact exist, and the projection theorem provides, after some simplification, a simple way to calculate it. The optimal $x_p^* \in \mathcal{P}_p$ can be expressed as a linear combination of the periodic basis elements δ_p^s as

$$x_p^* = \alpha_0 \delta_p^0 + \alpha_1 \delta_p^1 + \cdots + \alpha_{p-1} \delta_p^{p-1}.$$

According to the projection theorem, the unique minimizing vector is the orthogonal projection of x on $\mathcal{P}_p$, that is, $x - x_p^*$ is orthogonal to each of the δ_p^s for $s = 0, 1, ..., p-1$. Thus

$$0 = < x - x_p^*, \delta_p^s > = < x - \alpha_0 \delta_p^0 - \alpha_1 \delta_p^1 - ... - \alpha_{p-1} \delta_p^{p-1}, \delta_p^s > .$$

Since the δ_p^s are orthogonal to each other, this can be rewritten using the additivity of the inner product as

$$\begin{aligned} &= < x - \alpha_s \delta_p^s, \delta_p^s > \\ &= < x, \delta_p^s > - \alpha_s < \delta_p^s, \delta_p^s > \\ &= < x, \delta_p^s > - \frac{\alpha_s}{p}. \end{aligned}$$

Hence α_s can be written as

$$\alpha_s = p < x, \delta_p^s > .$$

Since $x \in \mathcal{P}$, it is periodic with some period N. From (5.15), the above inner product can be calculated

$$\alpha_s = p \frac{1}{pN} \sum_{i=0}^{pN-1} x(i) \delta_p^s(i).$$

But δ_p^s is zero except when $(s - i) \bmod p = 0$, and this simplifies to

$$\alpha_s = \frac{1}{N} \sum_{n=0}^{N-1} x(s + np). \tag{5.17}$$

If, in addition, N/p is an integer, then this reduces to

$$\alpha_s = \frac{1}{N/p} \sum_{n=0}^{N/p-1} x(s + np). \tag{5.18}$$

Example 5.7. N=14, p=2. Let $x \in \mathcal{P}_{14}$ be the 14-periodic sequence

$$x = \{\cdots, 2, -1.1, -1.1, 2, -1.2, -1.2, 2, -1.1, -1.1, 2, -1.2, -1.1, 2, -1.1, \cdots\}.$$

Then the projection of x onto $\mathcal{P}_2$ is $x_2 = \{\cdots, 0.2, -0.228, \cdots\}$.

This sequence x_2 is the 2-periodic sequence that best "fits" this 14-periodic x. But looking at this x closely suggests that it has more of the character of a 3-periodic sequence, albeit somewhat truncated in the final "repeat" of the $2, -1, -1$. Accordingly, it is reasonable to project x onto $\mathcal{P}_3$.

Example 5.8. N=14, p=3. Let $x \in \mathcal{P}_{14}$ be as defined in example 5.7. Then the projection of x onto $\mathcal{P}_3$ (using (5.17)) is $x_3 = -0.2\{\cdots, 1, 1, 1, \cdots\}$.

Clearly, this does not accord with the intuition that this x is "almost" 3-periodic. In fact, this is an example of a rather generic effect. Whenever N and p are mutually prime, the sum in (5.17) cycles through all the elements of x, and so $\alpha_s = \frac{1}{N}\sum_{i=0}^{N-1} x(i)$ for all s. Hence the projection onto $\mathcal{P}_p$ is the vector of all ones (times the mean value of the x). The problem here is the incommensurability of the N and p.

What does it mean to say that x (with length N) is p-periodic when N/p is not an integer? Intuitively, it should mean that there are $\lfloor N/p \rfloor$ complete repeats of the p-periodic sequence (where $\lfloor z \rfloor$ is the largest integer less than or equal to z) plus a "partial repeat" within the remaining $\tilde{N} = N - p\lfloor N/p \rfloor$ elements. For instance, the $N = 14$ sequence

$$x_1,\ x_2,\ x_3,\ x_1,\ x_2,\ x_3,\ x_1,\ x_2,\ x_3,\ x_1,\ x_2,\ x_3,\ x_1,\ x_2$$

can be considered a (truncated) 3-periodic sequence.

There are two ways to formalize this notion: to "shorten" x so that it is compatible with p, or to "lengthen" δ_p^s so that it is compatible with N. Though roughly equivalent (they differ only in the final $\tilde{N}$ elements), the first approach is simpler since it is possible to replace x with $x_{\tilde{N}}$ (the $\tilde{N}$-periodic sequence constructed from the first $\tilde{N} = p\lfloor N/p \rfloor$ elements of x) whenever the projection operator is involved. With this understanding, (5.17) becomes

$$\alpha_s = \frac{1}{\lfloor N/p \rfloor} \sum_{n=0}^{\lfloor N/p \rfloor - 1} x_{\tilde{N}}(s + np). \tag{5.19}$$

Example 5.9. N=14, p=3. Let $x \in \mathcal{P}_{14}$ be as defined in example 5.7. Then the projection of x onto $\mathcal{P}_3$ (using (5.19)) is $x_3 = \{\cdots, 2, -1.14, -1.125, \cdots\}$.

Clearly, this captures the intuitive notion of periodicity far better than example 5.8, and the sum (5.19) forms the foundation of the Periodicity Transform. The calculation of each α_s thus requires $\lfloor N/p \rfloor$ operations (additions). Since there are p different values of s, the calculation of the complete projection x_p requires $\tilde{N} \approx N$ additions. A MATLAB subroutine that carries out the needed calculations is available at the web site [W: 46].

Let $\pi(x, \mathcal{P}_p)$ represent the projection of x onto $\mathcal{P}_p$. Then

$$\pi(x, \mathcal{P}_p) = \sum_{s=0}^{p-1} \alpha_s\, \delta_p^s \tag{5.20}$$

where the δ_p^s are the (orthogonal) p-periodic basis elements of $\mathcal{P}_p$. Clearly, when $x \in \mathcal{P}_p$, $x = \pi(x, \mathcal{P}_p)$. By construction, when x is projected onto $\mathcal{P}_{np}$ it finds the best np-periodic components within x, and hence the residual $r = x - \mathcal{P}_{np}$ has no

np-periodic component. The content of the next result is that this residual also has no p-periodic component. In essence, the projection onto $\mathcal{P}_{np}$ "grabs" all the p-periodic information.

Theorem 5.10. *For any integer n, let $r = x - \pi(x, \mathcal{P}_{np})$ be the residual after projecting x onto $\mathcal{P}_{np}$. Then $\pi(r, \mathcal{P}_p) = 0$.*

All proofs are found in our paper [B: 206] which can also be found on the CD. The next result relates the residual after projecting onto $\mathcal{P}_p$ to the residual after projection onto $\mathcal{P}_{np}$.

Theorem 5.11. *Let $r_p = x - \pi(x, \mathcal{P}_p)$ be the residual after projecting x onto $\mathcal{P}_p$. Similarly, let $r_{np} = x - \pi(x, \mathcal{P}_{np})$ denote the residual after projecting x onto $\mathcal{P}_{np}$. Then*

$$r_{np} = r_p - \pi(r_p, \mathcal{P}_{np}).$$

Combining the two previous results shows that the order of projections doesn't matter in some special cases, that is

$$\pi(x, \mathcal{P}_p) = \pi(\pi(x, \mathcal{P}_p), \mathcal{P}_{np}) = \pi(\pi(x, \mathcal{P}_{np}), \mathcal{P}_p),$$

which is used in the next section to help sensibly order the projections.

5.5.3 Algorithms for Periodic Decomposition

The Periodicity Transform searches for the best periodic characterization of the length N signal x. The underlying technique is to project x onto some periodic subspace giving $x_p = \pi(x, \mathcal{P}_p)$, the closest p-periodic vector to x. This periodicity is then removed from x leaving the residual $r_p = x - x_p$ stripped of its p-periodicities. Both the projection x_p and the residual r_p may contain other periodicities, and so may be decomposed into other q-periodic components by projection onto $\mathcal{P}_q$. The trick in designing a useful algorithm is to provide a sensible criterion for choosing the order in which the successive p's and q's are chosen. The intended goal of the decomposition, the amount of computational resources available, and the measure of "goodness-of-fit" all influence the algorithm. The analysis of the previous sections can be used to guide the decomposition by exploiting the relationship between the structure of the various $\mathcal{P}_p$. For instance, it makes no sense to project x_p onto $\mathcal{P}_{np}$ because $x_p \in \mathcal{P}_{np}$ and no new information is obtained. This section presents several different algorithms, discusses their properties, and then compares these algorithms with some methods available in the literature.

One subtlety in the search for periodicities is related to the question of appropriate boundary (end) conditions. Given the signal x of length N, it is not particularly meaningful to look for periodicities longer than $p = N/2$, even though nothing in the mathematics forbids it. Indeed, a "periodic" signal with length $N - 1$ has $N - 1$ degrees of freedom, and surely can match x very closely, yet provides neither a convincing explanation nor a compact representation of x. Consequently, we restrict further attention to periods smaller than $N/2$.

Probably the simplest useful algorithm operates from small periods to large, as shown in Table 5.1. The Small-To-Large algorithm is simple because there is no need to further decompose the basis elements x_p; if there were significant q-periodicities within x_p (where "significant" is determined by the threshold T), they would already have been removed by x_q at an earlier iteration. The algorithm works well because it tends to favor small periodicities, to concentrate the power in $\mathcal{P}_p$ for small p, and hence to provide a compact representation.

Table 5.1. Small-To-Large Algorithm

```
pick threshold T ∈ (0, 1)
let r = x
for p = 2, 3, ..., N/2
        x_p = π(r, P_p)
        if ||r − x_p|| / ||x|| > T
                r = r − x_p
                save x_p as basis element
        end
end
```

Thinking of the norm as a measure of power, the threshold is used to insure that each chosen basis element removes at least a factor T of the power from the signal. Of course, choosing different thresholds leads to different decompositions. If T is chosen too small (say zero) then the decomposition will simply pick the first linear independent set from among the p-periodic basis vectors

$$\overbrace{\delta_2^1,\ \delta_2^2,}^{\mathcal{P}_2}\ \overbrace{\delta_3^1,\ \delta_3^2,\ \delta_3^3,}^{\mathcal{P}_3}\ \overbrace{\delta_4^1,\ \delta_4^2,\ \delta_4^3,\ \delta_4^4,}^{\mathcal{P}_4}\ \delta_5^1,\ \delta_5^2, \ldots,$$

which defeats the purpose of searching for periodicities. If T is chosen too large, then too few basis elements may be chosen (none as $T \to 1$). In between "too small" and "too large" is where the algorithm provides interesting descriptions. For many problems, $0.01 < T < 0.1$ is appropriate, since this allows detection of periodicities containing only a few percent of the power, yet ignores those p which only incidentally contribute to x.

An equally simple "Large-To-Small" algorithm is not feasible, because projections onto x_p for composite p may mask periodicities of the factors of p. For instance, if $x_{100} = \pi(x, \mathcal{P}_{100})$ removes a large fraction of the power, this may in fact be due to a periodicity at $p = 20$, yet further projection of the residual onto $\mathcal{P}_{20}$ is futile since $\pi(x - x_{100}, \mathcal{P}_{20}) = 0$ by Theorem 5.10. Thus an algorithm that decomposes from large p to smaller p must further decompose both the candidate basis element x_p as well as the residual r_p, since either might contain smaller q-periodicities.

The M-Best Algorithm deals with these issues by maintaining lists of the M best periodicities and the corresponding basis elements. The first step is to build

the initial list. This is described in Table 5.2. At this stage, the algorithm has compiled a list of the M periodicities q_i that remove the most "energy" (in the sense of the norm measure) from the sequence. But typically, the q_i will be large (since by Theorem 5.10, the projections onto larger subspaces np contain the projections onto smaller subspaces p). Thus the projections x_{q_i} can be further decomposed into their constituent periodic elements to determine whether these smaller (sub)periodicities remove more energy from the signal than another currently on the list. If so, the new one replaces the old.

Table 5.2. M-Best Algorithm (step 1)

pick size M
let $r_0 = x$
for $i = 1, 2, ..., M$
find q_i with $\|\|\pi(r_{i-1}, \mathcal{P}_{q_i})\|\| \geq \|\|\pi(r_{i-1}, \mathcal{P}_q)\|\| \, \forall q \in [1, 2, ..., N/2]$
$r_i = r_{i-1} - \pi(r_{i-1}, \mathcal{P}_{q_i})$
concatenate q_i and $x_{q_{i-1}} = \pi(r_i, \mathcal{P}_{q_i})$ onto respective lists
end

Fortunately, it is not necessary to search all possible periods $p < q_i$ when decomposing, but only the factors. Let $\rho_i = \{n; q_i/n \text{ is an integer}\}$ be the set of factors of q_i. The second step in the algorithm, shown in Table 5.3, begins by projecting the x_{q_i} onto each of its factors $Q \in \rho_i$. If the norm of the new projection x_Q is larger than the smallest norm in the list, and if the sum of all the norms will increase by replacing x_{q_i}, then the new Q is added to the list and the last element x_{q_M} is deleted. These steps rely heavily on Theorem 5.11. For example, suppose that the algorithm has found a strong periodicity in (say) $\mathcal{P}_{140}$, giving the projection $x_{140} = \pi(x, \mathcal{P}_{140})$. Since $140 = 2^2 \cdot 5 \cdot 7$, the factors are $\rho = \{2, 4, 5, 7, 10, 14, 20, 28, 35, 70\}$. Then the inner loop in step 2 searches over each of the $\pi(x_{140}, \mathcal{P}_Q) \; \forall Q \in \rho$. If x_{140} is "really" composed of a significant periodicity at (say) 20, then this new periodicity is inserted in the list and will later be searched for yet smaller periodicities. The M-Best Algorithm is relatively complex, but it removes the need for a threshold parameter by maintaining the list. This is a sensible approach and it often succeeds in building a good decomposition of the signal. A variation called the M-Best algorithm with γ-modification (or M-Best$_\gamma$) is described in Appendix A, where the measure of energy removed is normalized by the (square root of) the length p.

Another approach is to project x onto all the periodic basis elements δ_p^s for all p and s, essentially measuring the correlation between x and the individual periodic basis elements. The p with the largest (in absolute value) correlation is then used for the projection. This idea leads to the Best-Correlation Algorithm of Table 5.4, which presumes that good p will tend to have good correlation with at least one of the p-periodic basis vectors. This method tends to pick out periodicities with large regular spikes over those that are more uniform.

Table 5.3. M-Best Algorithm (step 2)

repeat until no change in list
  for $i = 1, 2, ..., M$
    find Q^* with $||\pi(x_{q_i}, \mathcal{P}_{Q^*})|| \geq ||\pi(x_{q_i}, \mathcal{P}_Q)|| \; \forall \; Q \in \rho_i$
    let $x_{Q^*} = \pi(x_{q_i}, \mathcal{P}_{Q^*})$ be the projection onto $\mathcal{P}_{Q^*}$
    let $x_{q^*} = x_{q_i} - x_{Q^*}$ be the residual
    if ($||x_{q^*}|| + ||x_{Q^*}|| > ||x_{q_M}|| + ||x_{q_i}||$)
      & ($||x_{q^*}|| > \min_k ||x_{q_k}||$ & $||x_{Q^*}|| > \min_k ||x_{q_k}||$)
        replace q_i with q^* and x_{q_i} with x_{q^*}
        insert Q^* and x_{Q^*} into lists at position $i - 1$
        remove q_M and x_{q_M} from end of lists
    end if
  end for
end repeat

Table 5.4. Best-Correlation Algorithm

M = number of desired basis elements
let $r = x$
for $i = 1, 2, ..., M$
    $\rho = \underset{p}{\text{argmax}} \; | < r, \delta_p^s > |$
    save $x_\rho = \pi(r, \mathcal{P}_\rho)$ as basis element
    $r = r - x_\rho$
end

A fourth approach is to determine the best periodicity p by Fourier methods, and then to project onto $\mathcal{P}_p$. Using frequency to find periodicity is certainly not always the best idea, but it can work well, and has the advantage that it is a well understood process. The interaction between the frequency and periodicity domains can be a powerful tool, especially since the Fourier methods have good resolution at high frequencies (small periodicities) while the periodicity transform has better resolution at large periodicities (low frequencies).

Table 5.5. Best-Frequency Algorithm

M = number of desired basis elements
let $r = x$
for $i = 1, 2, ..., M$
    $y = ||DFT\{r\}||$
    $p =$ Round$(1/f)$, where $f =$ frequency at which y is max
    save $x_p = \pi(r, \mathcal{P}_p)$ as basis element
    $r = r - x_p$
end

At present, there is no simple way to guarantee that an optimal decomposition has been obtained. One foolproof method for finding the best M subspaces would be to search all of the possible $\binom{N}{M}$ different orderings of projections to find the one with the smallest residual. This is computationally prohibitive in all but the simplest settings, although an interesting special case is when $M = 1$, that is, when only the largest periodicity is of importance.

5.5.4 Signal Separation

When signals are added together, information is often lost. But if there is some characteristic that distinguishes the signals, then they may be recoverable from their sum. Perhaps the best known example is when the spectrum of x and the spectrum of y do not overlap. Then both signals can be recovered from $x + y$ with a linear filter. But if the spectra overlap significantly, the situation is more complicated. This example shows how, if the underlying signals are periodic in nature, then the Periodicity Transform can be used to recover signals from their sum. This process can be thought of as a way to extract a "harmonic template" from a complicated spectrum.

Consider the signal z in Fig. 5.15, which is the sum of two zero mean sequences, x with period 13 and y with period 19. The spectrum of z is quite complex, and it is not obvious just by looking at the spectrum which parts of the spectrum arise from x and which from y. To help the eye, the two lattices marked A and B point to the spectral lines corresponding to the two periodic sequences. These are inextricably interwoven and there is no way to separate the two parts of the spectrum with linear filtering.

When the Periodicity Transform is applied to z, two periodicities are found, with periods of 13 and 19, with basis elements that are exactly $x_{13} = x + c_1$ and $y_{19} = y + c_2$, that is, both signals x and y are recovered, up to a constant. Thus the PT is able to locate the periodicities (which were assumed a priori unknown) and to reconstruct (up to a constant offset) both x and y given only their sum. Even when z is contaminated with 50% random noise, the PT still locates the two periodicities, though the reconstructions of x and y are noisy. To see the mechanism, let η be the noise signal, and let $\eta_{13} = \pi(\eta, \mathcal{P}_{13})$ be the projection of η onto the 13-periodic subspace. The algorithm then finds $x_{13} = x + c_1 + \eta_{13}$ as its 13-periodic basis element. If the x and y were not zero mean, there would also be a component with period one.

For this particular example, all four of the PT variants behave essentially the same, but in general they do not give identical outputs. The Small-To-Large algorithm regularly finds such periodic sequences. The Best-Correlation algorithm works best when the periodic data is spiky. The M-Best algorithm is sometimes fooled into returning multiples of the basic periodicities (say 26 or 39 instead of 13) while the M-Best$_\gamma$ is overall the most reliable and noise resistant. The Best-Frequency algorithm often becomes 'stuck' when the frequency with the largest magnitude does not closely correspond to an integer periodicity. The behaviors of the algorithms are

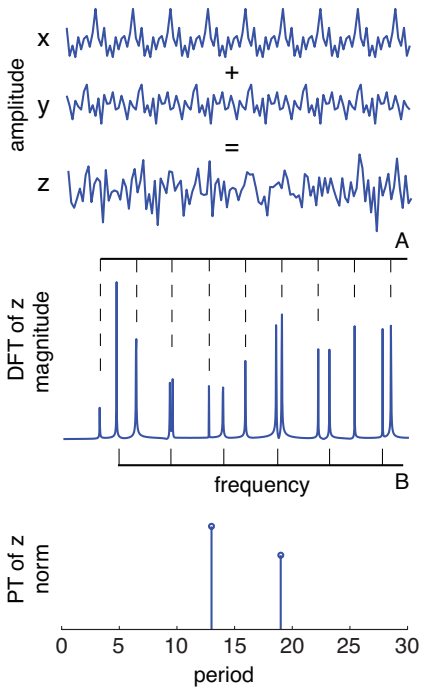

Fig. 5.15. The signal z is the sum of the 13-periodic x and the 19-periodic y. The DFT spectrum shows the overlapping of the two spectra (emphasized by the two lattices labeled A and B), which cannot be separated by linear filtering. The output of the M-Best$_\gamma$ Periodicity Transform, shown in the bottom plot, locates the two periodicities (which were a priori unknown) and reconstructs (up to a constant offset) both x and y given only z.

explored in detail in four demonstration files that accompany the periodicity software.[15]

Two aspects of this example deserve comment. First, the determination of a periodicity and its corresponding basis element is tantamount to locating a "harmonic template" in the frequency domain. For example, the 13-periodic component has a spectrum consisting of a fundamental (at a frequency f_1 proportional to $1/13$), and harmonics at $2f_1,\ 3f_1,\ 4f_1, \ldots$. Similarly, the 19-periodic component has a spectrum consisting of a fundamental (at a frequency f_2 proportional to $1/19$), and harmonics at $2f_2,\ 3f_2,\ 4f_2, \ldots$. These are indicated in Fig. 5.15 by the lattices A and B above and below the spectrum of z. Thus the PT provides a way of finding simple harmonic templates that may be obscured by the inherent complexity of the spectrum. The process of subtracting the projection from the original signal can be interpreted as a multi-notched filter that removes the relevant fundamental and its harmonics. For a single p, this is a kind of "gapped weight" filter familiar to those who work in time series analysis [B: 117].

The offsets c_1 and c_2 occur because $\mathcal{P}_1$ is contained in both $\mathcal{P}_{13}$ and in $\mathcal{P}_{19}$. In essence, both of these subspaces are capable of removing the constant offset (which is an element of $\mathcal{P}_1$) from z. When x and y are zero mean, both c_1 and c_2 are zero. If

[15] `MATLAB` versions of the periodicity software can be found on the CD in `files/matlab/periodicity/` and online at [W: 46]. The demos are called `PTdemoS2L`, `PTdemoBC`, `PTdemoMB`, and `PTdemoBF`.

they have nonzero mean, the projection onto (say) $\mathcal{P}_{13}$ grabs all of the signal in $\mathcal{P}_1$ for itself (Thus c_1 = mean(x) + mean(y), and further projection onto $\mathcal{P}_{19}$ gives c_2 =-mean(y)). This illustrates a general property of projections onto periodic subspaces. Suppose that the periodic signals to be separated were $x_{np} \in \mathcal{P}_{np}$ and $x_{mp} \in \mathcal{P}_{mp}$ for some mutually prime n and m. Since $\mathcal{P}_{np} \cap \mathcal{P}_{mp}$ is $\mathcal{P}_p$, both $\mathcal{P}_{np}$ and $\mathcal{P}_{mp}$ are capable of representing the common part of the signal, and x_{np} and x_{mp} can only be recovered up to their common component in $\mathcal{P}_p$. In terms of the harmonic templates, there is overlap between the set of harmonics of x_{np} and the harmonics of x_{mp}, and the algorithm does not know whether to assign the overlapping harmonics to x_{np} or to x_{mp}. The four different periodicity algorithms make different choices in this assignment.

It is also possible to separate a deterministic periodic sequence $z \in \mathcal{P}_p$ from a random sequence y when only their sum $x = y + z$ can be observed. Suppose that y is a stationary (independent, identically distributed) process with mean m_y. Then $E\{\pi(y, \mathcal{P}_p)\} = m_y \cdot \mathbf{1}$ (where $\mathbf{1}$ is the vector of all ones), and so

$$E\{\pi(x, \mathcal{P}_p)\} = E\{\pi(y + z, \mathcal{P}_p)\} = E\{\pi(y, \mathcal{P}_p)\} + E\{\pi(z, \mathcal{P}_p)\} = m_y \cdot \mathbf{1} + z$$

since $E\{\pi(z, \mathcal{P}_p)\} = E\{z\} = z$. Hence the deterministic periodicity z can be identified (up to a constant) and removed from x. Such decomposition will likely be most valuable when there is a strong periodic "explanation" for z, and hence for x. In some situations such as economic and geophysical data sets, regular daily, monthly, or yearly cycles may obscure the underlying signal of interest. Projecting onto the subspaces $\mathcal{P}_p$ where p corresponds to these known periodicities is very sensible. But appropriate values for p need not be known a priori. By searching through an appropriate range of p (exploiting the various algorithms of Sect. 5.5.3), both the value of p and the best p-periodic basis element can be recovered from the data itself.

5.5.5 Patterns in Astronomical Data

To examine the performance of the Periodicity Transform in the detection of more complex patterns, a three minute segment of astronomical data gathered by the Voyager spacecraft (published on audio CD in [D: 42]) was analyzed. When listening to this CD, there is an apparent pulse rate with approximately 18 (not necessarily equal length) pulses in each 32 second segment. Because of the length of the data, significant downsampling was required. This was accomplished by filtering the digital audio data in overlapping sections and calculating the energy value in each section. The resulting sequence approximates the amplitude of the Voyager signal with an effective sampling rate of $\frac{44100}{2520} = 17.5$ Hz.

The downsampled data was first analyzed with a Fourier Transform. The most significant sinusoidal components occur at 0.078, 0.137, 0.157, 0.167, 0.177 and 0.216 Hz, which correspond to periodicities at 12.8, 7.3, 6.3, 6.0, 5.7 and 4.6 seconds. Because the Fourier Transform is linear-in-frequency, the values are less accurate at long periods (low frequencies). For example, while the time interval between

adjacent Fourier bins is only ± 0.2 sec for the shortest of the significant periodicities (4.6 sec), the time between bins at the longest detected periodicity (12.8 sec) is approximately ± 1.6 sec.

Applying the PT to the downsampled data using the M-Best$_\gamma$ algorithm (with $M = 10$) gives the output shown in Table 5.6. This is plotted graphically in Fig. 5.16.

Table 5.6. M-Best$_\gamma$ Analysis of Voyager Data

Period p	31	217	434	124	868	328	656	799	880	525
Time (seconds)	1.77	12.4	24.8	7.09	49.6	18.74	37.49	45.66	50.29	30
Norm (in percent)	29.6	6.2	4.3	2.8	8.6	3.7	5.9	4.9	3.6	2.7

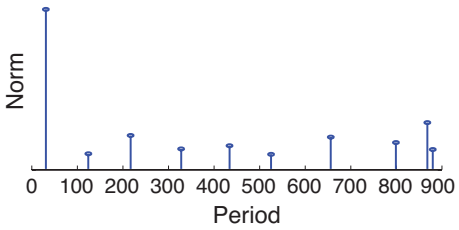

Fig. 5.16. Applying the PT to the downsampled data using the M-Best$_\gamma$ algorithm locates many of the most significant periodic features of the data.

The shortest periodicity at 1.77 second (period 31) corresponds well with the pulse rate that is apparent when listening to the data directly. The order in which the results appear in Table 5.6 mimics the operation of the algorithm. For example, there are three sets of related periodicities. The first three $31 : 217 : 434$ are in the ratio $1 : 7 : 14$, the next two $124 : 868$ are in the ratio $1 : 7$ (and are also in the ratio 4:28 relative to the periodicity 31). The third set $328 : 656$ are in the ratio $1 : 2$. These are indicative of the decomposition of large periods by Step 2 of the M-Best$_\gamma$ algorithm. While these longer periodicities are not immediately obvious when "listening to" or "looking at" at the raw data, their importance is reinforced by a comparison with the Fourier results. The PT periodicities at 7.09 and 12.4 correspond to the Fourier periodicities at 7.3 and 12.8 (recall the inherent margin of error in the bin width of the FFT). Also, the periodicity at 12.4 contains the $2 : 1$ subperiod at 6.2 detected by the FFT. The 1.77 second pulse is approximately $1/3$ of the 5.7 second Fourier result, reinforcing the interpretation that this was an underlying "fundamental."

On the other hand, the appearance of several periodicities without mutual factors clustered at one time scale (i.e., the periodicities at 45.66, 50.29 and 49.6 seconds), suggests that one long periodicity in the data may have been inexactly decomposed into several related components. This relates to what may be the most severe limitation to the general applicability of the PT: when the sample interval does not correspond to a factor of some periodicity in the data, the decomposition of small periods from larger ones is difficult. Qualitatively, this can be thought of as the converse of

the limitation of the Fourier method; while the linear-in-frequency behavior of the FFT increases the error at long periods, the linear-in-period behavior of the PT causes inefficiencies at short periods. Just as an increase in the amount of data used in the FFT can provide better precision, a decrease in the sample interval can improve the performance of the PT. Nonetheless, there is no general method of ensuring a priori that a particular periodicity in unknown data will correspond to a multiple of the sample rate. To see this effect clearly, we resampled the data at an effective sampling interval of 0.065 second, and then reapplied the M-Best$_\gamma$ algorithm. In this case, the longer periods were not as successfully decomposed. Similar sensitivities to the choice of sampling rates were observed in an early critique of the Buys-Ballot method [B: 21].

This "integer periodicity" limitation of the PT can be mitigated by the proper choice of a highly factorable integer as the sample interval. In general, the identification of small periodicities within an arbitrary data set will be most efficient when the sample interval itself contains many factors (many exact periodicities). These intervals, each composed of many periodic "building blocks", are easily combined to identify larger multiples. In fact, this was the reason we chose the effective sampling rate based on a subsampling interval of 2520, which factors as $2^3 \cdot 3^2 \cdot 5 \cdot 7$. This interval has a large number of factors within the desired range of effective sampling rates (0.05 seconds to 0.1 seconds) consistent with a downsampled data set of reasonable length.

5.5.6 Discussion of PT

The Periodicity Transform is designed to locate periodicities within a data set by projecting onto the (nonorthogonal) periodic subspaces. The method decomposes signals into their basic periodic components, creating its own "basis elements" as linear combinations of delta-like p-periodic basis vectors.

In some cases, the PT can provide a clearer explanation of the underlying nature of the signals than standard techniques. For instance, the signal z of Fig. 5.15 is decomposed into (roughly) 14 complex sinusoids by the DFT, or into two periodic sequences by the PT. In a strict mathematical sense, they are equivalent, since the residuals are equal in norm. But the PT "explanation" is simpler and allows the recovery of the individual elements from their sum. When periodicity provides a better explanation of a signal or an event than does frequency, then the PT is likely to outperform the DFT. Conversely, when the signal incorporates clear frequency relationships, the DFT will likely provide a clearer result. In general, an analysis of truly unknown signals will benefit from the application of all available techniques.

Like the Hadamard transform [B: 245], the PT can be calculated using only additions (no multiplications are required). As shown in Sect. 5.5.2, each projection requires approximately N operations. But the calculations required to project onto (say) $\mathcal{P}_p$ overlap the calculations required to project onto $\mathcal{P}_{np}$ in a nontrivial way, and these redundancies can undoubtedly be exploited in a more efficient implementation.

Several methods for finding the "best" basis functions from among some (possibly large) set of potential basis elements have been explored in the literature [B:

24], many of which are related to variants of general "projection pursuit" algorithms [B: 99]. Usually these are set in the context of choosing a representation for a given signal from among a family of prespecified frame elements. For instance, a Fourier basis, a collection of Gabor functions, a wavelet basis, and a wavelet packet basis may form the elements of an over-complete "dictionary." Coifman [B: 33] proposes an algorithm that chooses a basis to represent a given signal based on a measure of entropy. In [B: 139], a greedy algorithm called "matching-pursuit" is presented that successively decomposes a signal by picking the element that best correlates with the signal, subtracts off the residual, and decomposes again. This is analogous to (though somewhat more elaborate than) the Best-Correlation algorithm of Sect. 5.5.3. Nafie [B: 152] proposes an approach that maintains "active" and "inactive" dictionaries. Elements are swapped into the active dictionary when they better represent the signal than those currently active. This is analogous to the M-Best algorithm. The "best basis" approach of [B: 119] uses a thresholding method aimed at signal enhancement, and is somewhat analogous to the Small-To-Large algorithm. Using an l^1 norm, [B: 29] proposes a method that exploits Karmarkar's interior point linear programming method. The "method of frames" [B: 39] essentially calculates the pseudo-inverse of a (large rectangular) matrix composed of all the vectors in the dictionary.

While these provide analogous approaches to the problems of dealing with a redundant spanning set, there are two distinguishing features of the Periodicity Transform. The first is that the p-periodic basis elements are inherently coupled together. For instance, it does not make any particular sense to chose (say) δ_3^1, δ_4^3, δ_7^3, and δ_9^2 as a basis for the representation of a periodic signal. The p-periodic basis elements are fundamentally coupled together, and none of the methods were designed to deal with such a coupling. More generally, none of the methods is able (at least directly) to exploit the kind of structure (for instance, the containment of certain subspaces and the equality of certain residuals) that is inherent when dealing with the periodic subspaces of the PT.

5.6 Summary

A transform must ultimately be judged by the insight it provides and not solely by the elegance of its mathematics. Transforms and the various algorithms encountered in this chapter are mathematical operations that have no understanding of psychoacoustics or of the human perceptual apparatus. Thus a triangle wave may be decomposed into its appropriate harmonics by the Fourier transform irrespective of the time axis. It makes no difference whether the time scale is milliseconds (in which case we would hear pitch) or on the order of seconds (in which case we would hear rhythm). It is, therefore, up to us to include such extra information in the interpretation of the transformed data.

B: Bibliography

References in the body of the text to the bibliography are coded with [B:] *to distinguish them from references to the discography, sound examples, and websites.*

[1] K. Agawu, *African Rhythm*, Cambridge University Press, 1995.

[2] M. M. Al-Ibrahim, "A multifrequency range digital sinusoidal oscillator with high resolution and uniform frequency spacing," *IEEE Trans. on Circuits and Systems-II*, Vol. 48, No. 9, Sept. 2001.

[3] Safî al-Din al-Urmawî, *Kitâb al-Adwâr* 1252, trans. R. Erlanger in *La Musique arabe*, Paul Geuthner, Paris, 1938.

[4] W. Anku, "Circles and time: a theory of structural organization of rhythm in African music," *Music Theory Online*, Vol. 6, No. 1, Jan. 2000.

[5] Abramowitz and Stegun, *Handbook of Mathematical Functions*, Dover Pubs. Inc. NY 1972.

[6] American Standards Association, *USA Standard Acoustical Terminology*, New York, 1960.

[7] D. Bañuelos, *Beyond the Spectrum of Music* Ph.D. Thesis, University of Wisconsin, 2006. [Thesis appears (with permission) on the CD in `files/papers/BeyondtheSpectrum.pdf`].

[8] I. Barrow, *Lectiones Geometricae*, (trans. E. Stone), London, 1735.

[9] Beever, B. *Guide to Juggling Patterns*, `http://www.jugglingdb.com/articles/index.php`, 2000.

[10] W. E. Benjamin, "A theory of musical meter," *Music Perception*, Vol. 1, No. 4, 355-413, Summer 1984.

[11] H. Bergson, *Time and Free Will: An Essay on the Immediate Data of Consciousness*, translated by F.L. Pogson, M.A. London: George Allen and Unwin 1910, reprinted Dover, 2001.

[12] J. M. Bernardo and A. F. M. Smith, *Bayesian Theory*, Wiley, 2001.

[13] W. Berry, "Metric and rhythmic articulations in music," *Music Theory Spectrum*, 7: 7-33, 1985.

[14] J. A. Bilmes "Techniques to foster drum machine expressivity," *Proc. 1993 Int. Comp. Music Conf.*, San Francisco, 276-283, 1993.

[15] J. Blacking, *How Musical is Man?* Univ. Washington Press, Seattle, 1973.

[16] T. L. Bolton, "Rhythm," *American Journal of Psychology* 6: 145-238, 1894.

[17] A. Bregman, *Auditory Scene Analysis: The Perceptual Organization of Sound*, MIT Press, Cambridge, MA 1990.
[18] G. Brelet, *Le Temps Musical*, Paris, 1949.
[19] J. Brown, "Calculation of a constant Q spectral transform," *J. Acoustical Society of America*, 89, 425-434, 1991.
[20] J. Brown, "Determination of the meter of musical scores by autocorrelation," *J. Acoustical Society of America*, 94 (4), 1953-1957, Oct. 1993.
[21] H. Burkhardt, *Trigonometrische Interpolation*, Enzykl. Math. Wiss., Bd II, Teil 1, 1904.
[22] D. Butler, *The Musician's Guide to Perception and Cognition*, Schirmer Books, Macmillan Inc., 1992.
[23] J. B. Burl, "Estimating the basis functions of the Karhunen-Loeve transform," *IEEE Trans. Acoustics, Speech, and Signal Processing*, Vol. 37, No. 1, 99-105, Jan. 1989.
[24] C. S. Burrus, R. A. Gopinath, and H. Guo, *Wavelets and Wavelet Transforms*, Prentice Hall, NJ 1998.
[25] P. Cariani, "Temporal coding of periodicity pitch in the auditory system: an overview," *Neural Plasticity* 6, No. 4, 147-172, 1999.
[26] P. Cariani, "Neural timing nets for auditory computation," in *Computation Models of Auditory Function*, Ed. S. Greenberg and M. Slaney, IOS Press, 1999.
[27] A. T. Cemgil, B. Kappen, P. Desain and H. Honing "On tempo tracking: tempogram representation and Kalman filtering," *J. New Music Research*, Vol. 28, No. 4, 2001.
[28] A. T. Cemgil and B. Kappen, "Monte Carlo methods for tempo tracking and rhythm quantization," *J. Artificial Intelligence Research*, Vol 18., 45-81, 2003.
[29] S. Chen and D. L. Donoho, "Basis pursuit," *Proc. 28th Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, 41-44 Nov. 1994.
[30] J.M. Chernoff, *African Rhythm and African Sensibility*," Univ. of Chicago Press,,Chicago, IL, 1973.
[31] E. F. Clark and C. L. Krumhansl, "Perceiving musical time," *Music Perception*, Vol. 7, No. 3, 213-252, Spring 1990.
[32] T. Clifton, *Music as Heard: A Study in Applied Phenomenology*, Yale University Press, New Haven 1983.
[33] R. R. Coifman and M. V. Wickerhauser, "Entropy-based algorithms for best-basis selection," *IEEE Trans. Information Theory*, Vol. 38, No. 2, March 1992.
[34] E.Condon and T. Sugrue *We Called It Music: A Generation of Jazz* Da Capo Press, 1988.
[35] G. Cooper and L. B. Meyer, *The Rhythmic Structure of Music*, University of Chicago Press, 1960.
[36] D. R. Courtney, *Fundamentals of Tabla*, 3rd Edition, Sur Sangeet Services, Houston, 1998.
[37] H. Cowell, *New Musical Resources* Cambridge University Press 1996. (Original publication Alfred A. Knopf, 1930.
[38] P. E. Allen and R. B. Dannenberg, "Tracking Musical Beats in Real Time," *Int. Computer Music Conf.* 1990.

[39] I. Daubechies, "Time-frequency localization operators: A geometric phase space approach," *IEEE Trans. Information Theory*, Vol. 34, No. 4, 605-612, July 1988.
[40] S. De Furia, *Secrets of Analog and Digital Synthesis*, Third Earth Productions Inc., NJ, 1986.
[41] G. De Poli, A. Piccialli, and C. Roads Ed., *Representations of Musical Signals*, The MIT Press, Cambridge, MA 1991.
[42] F. Densmore, *Yuman and Yaqui Music*, Washington, 1932.
[43] P. Desain, "A (de)composable theory of rhythm perception," *Music Perception*, Vol. 9, No. 4, 439-454, Summer 1992.
[44] P. Desain and H. Honing, "Quantization of musical time: a connectionist approach," *Computer Music Journal*, 13(3): 56-66, 1989.
[45] D. Deutsch, ed., *The Psychology of Music*, Academic Press Inc., San Diego, CA (1982).
[46] D. Deutsch, "The processing of structured and unstructured tonal sequences," *Perception and Psychophysics*, 28, 381-389, 1980.
[47] J. Diamond, "Gamelan programs for children from the cross-cultural to the creative," *Ear Magazine*, Vol. VIII, No. 4, Sept. 1983.
[48] S. Dixon, "Automatic extraction of tempo and beat from expressive performances," *J. New Music Research*, Vol. 30, No. 1, 2001.
[49] M Dolson, "The Phase Vocoder: A Tutorial," *Computer Music Journal*, Spring, Vol. 10, No. 4, 14-27, 1986.
[50] A. Doucet, N de Freitas and N. Gordon, eds., *Sequential Monte Carlo Methods in Practice*, Springer-Verlag 2001.
[51] A. Doucet, N. de Freitas and N. Gordon. "An introduction to sequential monte carlo methods." In *Sequential Monte Carlo Methods in Practice.* Arnaud Doucet, Nando de Freitas and Neil Gordon (editors). Springer-Verlag. 2001.
[52] A. Doucet, N. de Freitas, K. Murphy and S. Russell. "Rao-Blackwellised particle filtering for dynamic Bayesian networks." *Proceedings of Uncertainty in Artificial Intelligence*, 2000.
[53] W. J. Dowling, "Rhythmic groups and subjective chunks in memory for melodies," *Perception and Psychophysics*, 14, 37-40, 1973.
[54] J. Duesenberry, "The world in a grain of sound," *Electronic Musician*, Nov. 1999.
[55] D. Eck, "A network of relaxation oscillators that finds downbeats in rhythms," in G. Dorffner, ed. *Artificial Neural Networks - ICANN 2001*, 1239-1247, Berlin, Springer, 2001.
[56] D. Eck, "Finding downbeats with a relaxation oscillator," *Psychological Research*, 66(1), 18-25, 2002.
[57] D. Ehresman and D. Wessel, "Perception of timbral analogies," Rapports IRCAM 13/78, 1978.
[58] W. A. Gardner and L. E. Franks, "Characterization of cyclostationary random signal processes," *IEEE Trans. Inform Theory*, Vol. IT-21, 414, 1975.

[59] L. J. Eriksson, M. C. Allie, and R. A. Griener, "The selection and application of an IIR adaptive filter for use in active sound attenuation," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, ASSP-35, No. 4, Apr. 1987.
[60] M. C. Escher, *The Graphic Work of M. C. Escher* Harry N Abrams, Pubs., 1984.
[61] R. Fitzhugh, "Impulses and physiological states in theoretical models of nerve induction," *Biophysical Journal*, 1, 455-466, 1961.
[62] J. L. Flanagan and R. M. Golden, "Phase vocoder," *Bell System Technical Journal*, 1493-1509, 1966.
[63] J. E. Flannick, R. W. Hall, and R. Kelly, "Detecting meter in recorded music," Bridges Proceedings, 2005.
[64] N. J. Fliege and J. Wintermantel, "Complex digital oscillators and FSK modulators," *IEEE Trans. Signal Processing*, Vol. 40, No. 2, Feb. 1992.
[65] P. Fraisse, *The Psychology of Time*, Harper & Row, New York, 1963.
[66] P. Fraisse, "Rhythm and tempo," in [B: 45].
[67] J. T. Fraser and N. Lawrence, *The Study of Time*, Springer-Verlag, NY 1975.
[68] D. Gabor, "Acoustical quanta and the theory of hearing" *Nature*, 159(4044), 591-594, 1947.
[69] S. Ganassi, *Fontegara*, Venice 1535. Ed. H. Peter, trans. D. Swainson, Berlin 1956.
[70] T. H. Garlanda and C. V. Kahn, *Math and Music: Harmonious Connections* Dale Seymour Publications, 1995.
[71] C. Gerard and M. Sheller, *Salsa: the Rhythm of Latin Music*, White Cliffs Media, Tempe, AZ. 1998.
[72] N. Ghosh, *Fundamentals of Raga and Tala*, Popular Prakashan, Bombay, India 1968.
[73] J. J. Gibson, "Events are perceivable but time is not," in [B: 67], 1975.
[74] R. O. Gjerdingen, " 'Smooth' rhythms as probes of entrainment," *Music Perception*, Vol. 10, No., 4, 503-508, Summer 1993.
[75] N. J. Gordon, D .J. Salmond and A. F. M. Smith, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," *IEEE Proceedings-F*, 140(2): 107-113, April 1993.
[76] M. Goto, "An audio-based real-time beat tracking system for music with or without drum-sounds," *J. New Music Research*, Vol. 30, No. 2, 159-171, 2001.
[77] M. Goto and Y. Muraoka, "Real-time beat tracking for drumless audio signals: chord change detection for musical decisions," *Speech Communication*, Vol.27, No. 3-4, 311-335, April 1999.
[78] F. Gouyon and S. Dixon, "A review of automatic rhythm description systems," *Computer Music Journal* 29:1, 3454, Spring 2005.
[79] F. Gouyon and P. Herrera, "A beat induction method for musical audio signals," *Proc. of 4th WIAMIS-Special session on Audio Segmentation and Digital Music* London, UK, 2003.
[80] F.Gouyon and P. Herrera, "Determination of the meter of musical audio signals: seeking recurrences in beat segment descriptors," *Proc. of AES*, 114th Convention, 2003.

[81] F. Gouyon and B. Meudic, "Towards rhythmic content processing of musical signals: fostering complementary approaches," *J. New Music Research*, Vol. 32, No. 1, 159-171, 2003.
[82] N. Gray, *Roots Jam : Collected Rhythms for Hand Drum and Percussion* Cougar WebWorks, 1996.
[83] J. M. Grey, *An Exploration of Musical Timbre* Ph.D. Thesis in Psychology, Stanford, 1975.
[84] J. M. Grey and J. W. Gordon, "Perceptual effects of spectral modifications on musical timbres," *J. Acoustical Society of America*, 65, No. 5, 1493-1500 (1978).
[85] J. M. Grey and J. A. Moorer, "Perceptual evaluation of synthesized musical instrument tones," *J. Acoustical Society of America*, 62, 454-462 (1977).
[86] G. S. Hall and J. Jastrow, "Studies of rhythm," *Mind* 11, 55-62, 1886.
[87] R. Hall and P. Klingsberg, "Asymmetric rhythms, tiling canons, and Burnsides lemma," Bridges Proceedings, 189-194, 2004.
[88] S. Handel, "Temporal segmentation of repeating auditory patterns," *J. Exp. Psych.* 101, 46-54, 1973.
[89] S. Handel and J. S. Oshinsky, "The meter of syncopated auditory polyrhythms," *Perception and Psychophysics*, 30 (1) 1-9, 1981.
[90] S. Handel and G. R. Lawson, "The contextual nature of rhythmic interpretation," *Perception and Psychophysics*,19 34 (2) 103-120, 13.
[91] S. Handel, "Using polyrhythms to study rhythm," *Music Perception*, Vol 1, No. 4, 465-484, Summer 1984.
[92] M. S. Harvey, "Jazz time and our time," in [B: 156].
[93] S. Haykin, *Adaptive Filter Theory*, Prentice-Hall, Englewood Cliffs, NJ, 1991.
[94] R. J. Heifetz, *On the Wires of OUr Nerves*, Bucknell University Press, 1989.
[95] H. Helmholtz, *On the Sensations of Tones*, (1877). Trans. A. J. Ellis, Dover Pubs., New York 1954.
[96] P. Hindemith, *The Craft of Musical Composition*, (1937). Trans. A Mendel, Associated Music Publishers, New York, NY (1945).
[97] D. R. Hofstadter, *Gödel, Escher, Bach: An Eternal Golden Braid* Basic Books, 1979.
[98] E. M. von Hornbostel, "African Negro Music, *Africa*, 1(1), 1928.
[99] P. J. Huber, "Projection pursuit," *Annals of Statistics*, Vol. 13, No. 2, 435-475, 1985.
[100] A. Hutchinson, "History of dance notation," *The Dance Encyclopedia*, Simon and Schuster, 1967.
[101] A. Hutchinson, *Labanotation: The System of Analyzing and Recording Movement*, Theatre Arts Books, 1970.
[102] C. Huygens, *Horologium* 1658.
[103] C. R. Johnson, Jr. and W. A. Sethares, *Telecommunication Breakdown: concepts of communications transmitted via software-defined radio*, Prentice-Hall 2004.
[104] M. R. Jones, "Time, our lost dimension: toward a new theory of perception, attention, and memory," *Psychological Review*, Vol. 83, No. 5, Sept. 1976.

[105] M. R. Jones and M. Boltz, "Dynamic attending and responses to time," *Psychological Review*, 96 (3) 459-491, 1989.
[106] M. R. Jones, G. Kidd, and R. Wetzel, "Evidence for rhythmic attention," *J. Experimental Psychology,* Vol. 7, No. 5, 1059-1073, 1981.
[107] M. R. Jones, "Musical events and models of musical time," in R. Block, Ed., *Cognitive Models of Psychological Time*, Lawrence Erlbaum Associates, Hillsdale NJ, 1990.
[108] S. Joplin, *Maple Leaf Rag*, John Stark & Sons, 1899. [Musical score appears on the CD in `files/mapleafscore.pdf`. Standard MIDI files of this are found in `files/mids/mapleafrag.mid`, `files/mids/mapleleafrag.mid`, `files/mids/mapleaf.mid`, and `files/mids/mapleleafrag.mid`.]
[109] D. Keane, "Some practical aesthetic problems of computer music," in [B: 94]
[110] R. Kelly, "Mathematics of musical rhythm," Honors Thesis, 2002, see [W: 15].
[111] K. Kharat, "The 'tala' system in Indian music," *Second ASEA Composer Forum on Traditional Music*, Singapore, April 1993.
[112] A. King, "Employments of the "standard pattern" in Yoruba music." *African Music* 2(3) 1961.
[113] P. Kivy, *Authenticities*, Cornell University Press, Ithaca NY, 1995.
[114] M. Klingbeil, "Software for spectral analysis, editing, and synthesis," *Proc. of the 2005 International Computer Music Conference,* Barcelona, 2005.
[115] K. Koffka, *Principles of Gestalt Psychology,* Harcourt, Brace, and Co., New York, NY 1935.
[116] W. Köhler, *Gestalt Psychology,* Liveright, New York, NY 1929.
[117] L. H. Koopmans, *The Spectral Analysis of Time Series,* Academic Press, San Diego 1995.
[118] J. D. Kramer, *The Time of Music*, MacMillan, Inc. 1998.
[119] H. Krim, S. Mallat, D. Donoho, and A. Willsky, "Best basis algorithm for signal enhancement," *Proc. IEEE Conf. on Acoustics, Speech and Signal Processing*, ICASSP-95, 1561-1564, Detroit, May 1995.
[120] A. Krims, *Rap Music and the Poetics of Identity*, Cambridge University Press, Cambridge UK, 2000.
[121] A. B. Kristofferson, "A quantal step function in duration discrimination" *Perception & Psychophysics* 27, 300-306, 1980.
[122] J. Kunst, *Music in Java*, Martinus Nijhoff, The Hague, Netherlands (1949).
[123] S. Langer, *Feeling and Form*, Scribners, NY 1953.
[124] E. W. Large and J. F. Kolen, "Resonance and the perception of musical meter," *Connection Science* 6: 177-208, 1994.
[125] J. Laroche and M. Dolson, "Improved phase vocoder time-scale modification of audio," *IEEE Trans. on Audio and Speech Processing,* Vol. 7, No. 3, May 1999.
[126] M. Leman, *Music and Schema Theory: Cognitive Foundations of Systematic Musicology*, Berlin, Heidelberg: Springer-Verlag 1995.
[127] M. Leman, Ed. *Music, Gestalt, and Computing* Springer, 1997.

[128] M. Leman and A. Schneider, "Origina and nature of cognitive and systematic musicology: an introduction," in [B: 127].

[129] F. Lerdahl and R. Jackendoff, *A Generative Theory of Tonal Music,* MIT Press, Cambridge, 1983.

[130] D. Locke, *Drum Gahu*, White Cliffs Media, Tempe AZ, 1998.

[131] J.London, "Rhythm," *Grove Music Online* Ed. L. Macy (Accessed 19-05-2005), `http://www.grovemusic.com`

[132] H. C. Longuet-Higgins and C. S. Lee, "The perception of musical rhythms," *Perception*, Vol. 11, 115-128, 1982.

[133] H. C. Longuet-Higgins and C. S. Lee, "The rhythmic interpretation of monophonic music," *Music Perception*, Vol. 1, No. 4, 424-441, Summer 1984.

[134] D. G. Luenberger, *Introduction to Dynamic Systems: Theory, Models, and Applications*, John Wiley and Sons, Inc. NY, 1979.

[135] D. G. Luenberger, *Optimization by Vector Space Methods*, John Wiley and Sons, Inc. NY, 1968.

[136] E. Mach, *Die Analyse der Empfindungen*, Jena: Fischer, 1886.

[137] D. J. C. MacKay, *Information Theory, Inference, and Learning Algorithms*, Cambridge University Press, 2003.

[138] R. C. Maher and J. W. Beauchamp, "Fundamental frequency estimation of musical signals using a two-way mismatch procedure," *J. Acoustical Society of America* 95 (4), April 1994.

[139] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Processing*, Vol. 41, No. 12, Dec. 1993.

[140] W. P. Malm, *Music Cultures of the Pacific, the Near East, and Asia*, Prentice-Hall, NJ, 1996.

[141] M. V. Mathews and J. R. Pierce, "Harmony and inharmonic partials," *J. Acoustical Society of America*, 68, 1252-1257 (1980).

[142] M. V. Mathews, J. R. Pierce, A. Reeves, and L. A. Roberts, "Theoretical and experimental explorations of the Bohlen-Pierce scale," *J. Acoustical Society of America* 84, 1214-1222 (1988).

[143] J. D. McAuley, *Perception of time as phase: toward an adaptive-oscillator model of rhythmic pattern processing*, Ph.D Thesis, Indiana University, 1995.

[144] J. D. McAuley, Time as phase: a dynamic model of time perception," *Proc. Sixteenth Conf. of Cognitive Science Society*, Lawrence Erlbaum, 607-612, 1994.

[145] M. K. McClintock, "Menstrual synchrony and suppression," *Nature*, 229:244-245, 1971.

[146] R. Meddis and L. O'Mard, "A unitary model of pitch perception," *J. Acoustical Society of America*, 102(3): 1811-1820, Sept. 1997.

[147] E. Meumann, "Untersuchungen zur psychologie und ästhetik des rhythmus," *Phiosophische Studien* 10, 1894.

[148] G. A. Miller, "The magical number seven, plus or minus two: some limits on our capacity for processing information," *The Psychological Review* vol. 63, 81-97, 1956.

[149] J. D. Miller, L. P. Morin, W. J. Schwartz, and R. Y. Moore, "New insights into the mammalian circadian clock," *Sleep*, Vol. 19., No. 8., 641-667, 1996.

[150] B. C. J. Moore, *An Introduction to the Psychology of Hearing*, Academic Press, Inc., 1982.

[151] R. D. Morris and W. A. Sethares, "Beat Tracking," *Int. Society for Bayesian Analysis*, Valencia Spain, June 2002. [On CD in `files/papers/valenciaposter.pdf`.]

[152] M. Nafie, M. Ali, and A. Tewfik, "Optimal subset selection for adaptive signal representation," *Proc. IEEE Conf. on Acoustics, Speech and Signal Processing*, ICASSP-96, 2511-2514, Atlanta, May 1996.

[153] U. Neisser, *Cognitive Psychology*, Appleton Century-Crofts, New York, 1967.

[154] J. H. K. Nketia, *Drumming in Akan Communities of Ghana*, Edinburgh: Thomas Nelson and Sons, Ltd., 1963.

[155] L. van Noorden and D. Moelants, "Resonance in the perception of musical pulse," *J. New Music Research* Vol. 28, No. 1, March 1999.

[156] A. M. S. Nelson, *This is How We Flow*, Univ. S. Carolina Press, 1999.

[157] R. Parncutt, "A perceptual model of pulse salience and metrical accent in musical rhythms," *Music Perception*, Vol. 11, No. 4, Summer 1994.

[158] R. D. Patterson, M. H. Allerhand, and C. Giguere, "Time domain modelling of peripheral auditory processing: a modular architecture and a software platform," *J. Acoustical Society of America*, 98: 1890-1894, 1995.

[159] R. D. Patterson and B. C. J. Moore, "Auditory filters and excitation patterns as representations of frequency resolution," *Frequency Selectivity in Hearing*, Ed. B. C. J. Moore, Academic Press, London 1986.

[160] L. S. Penrose and R. Penrose, "Impossible objects: a special type of visual illusion," *British J. of Psychology*, 1958.

[161] D. N. Perkins, "Coding position in a sequence by rhythmic grouping," *Memory and Cognition*, 2, 219-223, 1974.

[162] A. Pikovsky, M. Rosenblum, and J. Kurths, *Synchronization: a universal concept in nonlinear science*, Cambridge University Press, Cambridge, UK, 2001.

[163] J. R. Pierce, "Periodicity and pitch perception," *J. Acoustical Society of America*, 90,No. 4, 1889-1893, 1991.

[164] H. Platel, C. Price, J-C Baron, R. Wise, J. Lambert, R. S. J. Frackowiak, B. Lechevalier, and F. Eustache, "The structural components of music perception: a functional anatomical study," *Brain*, 120, 229-243, 1997.

[165] R. Plomp and W. J. M. Levelt, "Tonal consonance and critical bandwidth," *J. Acoustical Society of America*, 38, 548-560 (1965).

[166] R. Plomp, "Timbre as a multidimensional attribute of complex tones," in *Frequency Analysis and Periodicity Detection in Hearing*, ed. R. Plomp and G. F. Smoorenburg, A. W. Sijthoff, Lieden, 1970.

[167] R. Plomp, *Aspects of Tone Sensation*, Academic Press, London, 1976.

[168] I. Pollack, "The information of elementary auditory displays," *J. Acoustical Society of America*, 24, 745-749, 1952.

[169] B. Porat, *Digital Signal Processing*, Wiley 1997.

[170] M. R. Portnoff, "Implementation of the digital phase vocoder using the fast fourier transform," *IEEE Trans. Acoustics, Speech, and Signal Processing,* Vol. ASSP-24, No. 3, June 1976.
[171] D. J. Povel, *Internal representation of simple temporal patterns, J. of Experimental Psychology* Vol. 7, No. 1, 3-18, 1981.
[172] D. J. Povel, "A theoretical framework for rhythmic perception," *Psychological Research* 45: 315-337, 1984.
[173] D. J. Povel and P. Essens, "Perception of temporal patterns," *Music Perception,* Vol. 2, No. 4, 411-440, Summer 1985.
[174] D. J. Povel and H. Okkerman, "Accents in equitone sequences," *Perception and Psychophysics,* Vol. 30, 565-572, 1981.
[175] D. Preusser, "The effect of structure and rate on the recognition and description of auditory temporal patterns," *Perception and Psychophysics,* Vol. 11(3) 1972.
[176] M. S. Puckette and J. C. Browm, "Accuracy of frequency estimates using the phase vocoder," *IEEE Trans. Speech and Audio Processing,* Vol. 6, No. 2, March 1998.
[177] J. P. Rameau, *Treatise on Harmony,* Dover Pubs., New York 1971, original edition, 1722.
[178] C. Raphael. "Automated rhythm transcription" *Proc. Int. Symposium on Music Inform. Retriev.* IN, USA, Oct. 2001.
[179] L. G. Ratner, *The Musical Experience,* W, H, Freeman and Co. New York, 1983.
[180] C. Roads, *Microsound,* MIT Press, 2002.
[181] D. A. Robin, P. J. Abbas, and L. N. Hug, "Neural responses to auditory temporal patterns," *J. Acoustical Society of America,* Vol. 87, 1673-1682, 1990.
[182] J. G. Roederer, *The Physics and Psychophysics of Music*, Springer-Verlag, New York, 1994.
[183] D. Rosenthal, "Emulation of rhythm perception," *Computer Music Journal,* Vol. 16, No. 1, Spring 1992.
[184] T. D. Rossing, *The Science of Sound*, Addison Wesley Pub., Reading, MA, 1990.
[185] F. L. Royer and W. R. Garner, "Perceptual organization of nine-element temporal patterns," *Perception and Psychophysics,* Vol. 7(2) 1970.
[186] C. Sachs, *Rhythm and Tempo*, W. W. Norton and Co., Inc. NY 1953.
[187] E. D. Scheirer, "Tempo and beat analysis of acoustic musical signals," *J. Acoustical Society of America,* 103 (1), 588-601, Jan. 1998.
[188] J. Schillinger, *The Schillinger System of Musical Composition,* Carl Fischer, Inc. NY, 1946.
[189] J. Schillinger, *Encyclopedia of Rhythms,* Da Capo Press, NY, 1976.
[190] A. Schoenberg,*Fundamentals of Music Composition,* Faber and Faber, London, 1967.
[191] C. H. Sears, "A contribution to the psychology of rhythm," *American Journal of Psychology,* Vol 13(1) 28-61, 1913.

[192] A. K. Sen, *Indian Concept of Tala*, Kanishka Pubs. 9/2325 Kailash Nagar, Delhi, India 1984.

[193] J. Seppänen, "Computational models of musical meter recognition," MS Thesis, Tempere Univ. Tech. 2001.

[194] X. Serra, "Sound hybridization based on a deterministic plus stochastic decomposition model," in *Proc. of the 1994 International Computer Music Conference*, Aarhus, Denmark, 348351, 1994.

[195] W. A. Sethares, *Tuning, Timbre, Spectrum, Scale*, Springer-Verlag, 1997. Second edition 2004.

[196] W. A. Sethares, "Repetition and Pseudo-Periodicity," *Tatra Mt. Mathematics Publications*, Dec. 2001. [This paper appears on the CD in `files/papers/pnorm.pdf`].

[197] W. A. Sethares, "Specifying spectra for musical scales," *J. Acoustical Society of America*, 102, No. 4, Oct. 1997.

[198] W. A. Sethares, "Consonance-based spectral mappings," *Computer Music Journal* 22, No. 1, 56-72, Spring (1998).

[199] W. A. Sethares, "Local consonance and the relationship between timbre and scale," *J. Acoustical Society of America*, 94, No. 3, 1218-1228, Sept. (1993). [On CD in `files/papers/consonance.pdf`.]

[200] W. A. Sethares, "Automatic detection of meter and periodicity in musical performance," *Proc. of the Research Society for the Foundations of Music*, Ghent, Belgium, Oct. 1999.

[201] W. A. Sethares, "Clock Induction," See `files/papers/clockinduction.pdf`.

[202] W. A. Sethares, "Some Statistical Models of Periodic Phenomenon," See `files/papers/statmodels.pdf`.

[203] W. A. Sethares and R. D. Morris., "Performance measures for beat tracking," *Int. Workshop on Bayesian Data Analysis, Santa Cruz, Aug. 2003.*

[204] W. A. Sethares and R. D. Morris, "Performance measures for beat tracking," (Technical Report, available on the CD in `files/papers/beatqual.pdf`).

[205] W. A. Sethares, R. D. Morris and J. C. Sethares, "Beat tracking of audio signals," *IEEE Trans. on Speech and Audio Processing*, Vol. 13, No. 2, March, 2005. [This paper appears on the CD in `files/papers/beatrack.pdf`].

[206] W. A. Sethares and T. Staley, "The periodicity transform," *IEEE Trans. Signal Processing*, Vol. 47, No. 11, Nov. 1999. [This paper appears on the CD in `files/papers/pertrans.pdf`].

[207] W. A. Sethares and T. Staley, "Meter and periodicity in musical performance," *J. New Music Research*, Vol. 30, No. 2, June 2001. [This paper appears on the CD in `files/papers/jnmr2001.pdf`].

[208] W. A. Sethares, `TapTimes:` A Max [W: 29] patch designed to allow listeners to tap along with a piece of music and to record the times of the tapping. Available by request from the author.

[209] P. Schaeffer, *A la Recherche d'une Musique Concrète*, Paris, Seuil, 1952.

[210] R. N. Shepard, "Circularity in the judgments of relative pitch," *J. Acoustical Society of America* 36, 2346-2353, 1964.
[211] D. S. Sivia, *Data Analysis: A Baysian Tutorial*
[212] A. Slutsky and C. Silverman, *James Brown Rhythm Sections*, Manhattan Music, Inc., 1997.
[213] P. Singh, "The role of timbre, pitch, and loudness changes in determining perceived metrical structure," *J. Acoustical Society of America*, 101 (5) Pt 2, 3167, May 1997.
[214] A. M. Small, "An objective analysis of artistic violin performances,"
[215] A. M. Small and M. E. McClellan, "Pitch associated with time delay between two pulse trains," *J. Acoustical Society of America*, 35 (8), 1246-1255, Aug. 1963.
[216] L. M. Smith, "Listening to musical rhythms with progressive wavelets," *IEEE TENCON Digital Signal Processing Applications*, 1996.
[217] S. W. Smith, *The Scientist and Engineer's Guide to Digital Signal Processing*, California Technical Publishing, Second Ed., 1999. [See website at www.DSPguide.com.]
[218] B. Snyder, *Music and Memory*, MIT Press, Cambridge, MA 2000.
[219] N. Sorrell, *A Guide to the Gamelan* Faber and Faber Ltd., London, 1990.
[220] M. J. Steedman, "The perception of musical rhythm and metre," *Perception* Vol. 6, 555-569, 1977.
[221] T. Stilson and J. Smith, "Alias-free digital synthesis of classic analog waveforms," `http://www-ccrma.stanford.edu/~stilti/papers`
[222] K. Stockhausen, "... How Time Passes ...," *die Reihe* 3:10-43, English edition trans. by C. Cardew, 1959.
[223] E. Terhardt, "Pitch, consonance, and harmony," *J. Acoustical Society of America*, 55, No. 5, 1061-1069, May 1974.
[224] E. Terhardt, G. Stoll, and M. Seewann "Algorithm for extraction of pitch and pitch salience from complex tone signals," *J. Acoustical Society of America*, 71, No. 3, 679-688, March 1982.
[225] B. Tiemann and B. Magnusson, "A notation for juggling tricks," *Jugglers World*, 1991.
[226] N. P. M. Todd, D. J. O'Boyle and C. S. Lee, "A sensory-motor theory of rhythm, time perception and beat induction," *J. of New Music Research*, 28:1:5-28, 1999.
[227] P. Toiviainen "Modelling the perception of metre with competing subharmonic oscillators," *Proc. 3rd Triennial ESCOM Conf.* Uppsala, Sweden, 1997.
[228] P. Toiviainen, "An interactive MIDI accompanist," *Comp. Music J.*, 22(4):6375, 1998.
[229] B. Truax, "Real-time granular synthesis with a digital signal processor," *Computer Music Journal*, 12(2), 14-26 1988.
[230] G. Tzanetakis and P. Cook, "Musical genre classification of audio signals", *IEEE Trans. Speech and Audio Processing*, 10(5), July 2002.
[231] E. Uribe, *The Essence of Afro-Cuban Percussion and Drum Set*, Warner Bros. Pubs., Miami FL, 1996.

[232] B. van der Pol and J. van der Mark, "The heartbeat considered as a relaxation oscillation and an electrical model of the heart," *Phil. Mag. Suppl.* No. 6, 763-775, 1928.

[233] J. D. Vantomme, "Score following by temporal pattern," *Computer Music Journal*, 19:3, 50-59, Fall 1995.

[234] J. Vermaak, C. Andrieu, A. Doucet, and S. J. Godsill, "Particle methods for Bayesian modelling and enhancement of speech signals," *IEEE Trans. Speech Audio Processing*, 10(3):173-185, 2002.

[235] R. A. Waterman, " 'Hot' rhythm in Negro music," *J. of the American Musicological Society*, 1, Spring 1948.

[236] M. Wertheimer, "Principles of perceptual organization," in D. Beardslee and M. Wertheimer (Eds.) *Readings in Perception*, Van Nostrand, Princeton, NJ 1958.

[237] C. G. Wier, W. Jesteadt, and D. M. Green, "Frequency discrimination as a function of frequency and sensation level," *J. Acoustical Society of America*, 61, 178-184, 1977.

[238] L Wilcken, *The Drums of Vodou*, White Cliffs Media, Tempe, AZ, 1992.

[239] T. Wishart, *Audible Design*, Orpheus the Pantomine, York, 1994.

[240] H. Wong and W. A. Sethares, "Estimation of pseudo-periodic signals," IEEE International Conference on Acoustics, Speech and Signal Processing, Montreal 2004. [This paper appears on the CD in `files/papers/wong2004.pdf`].

[241] H. Woodrow, "A quantitative study of rhythm," *Archives of Psychology* Vol. 14, 1-66, 1909.

[242] H. Woodrow, "The role of pitch in rhythm," *Psychological Review* Vol. 11, 54-76, 1911.

[243] E Wold, T. Blum, D. Keislar, and J. Wheaton, "Content-based classification, search and retrieval of audio," *IEEE Trans. Multimedia*, 3(3):2736, 1996.

[244] I. Xenakis, *Formalized Music*, Indiana University Press, Bloomington, IN, 1971.

[245] R. K. R. Yarlagadda and J. E. Hershey, *Hadamard Matrix Analysis and Synthesis* Kluwer Academic, 1997.

[246] L. D. Zimba and D. A. Robin, "Perceptual organization of auditory patterns," *J. Acoustical Society of America*, Vol. 104, No. 4, Oct. 1998.

[247] E. Zwicker and H. Fastl, *Psychoacoustics*, Springer-Verlag, Berlin, 1990.

[248] E. Zwicker, G. Flottorp, and S. S. Stevens, "Critical bandwidth in loudness summation," *J. Acoustical Society of America*, 29, 548, 1957.