2

Visualizing and Conceptualizing Rhythm

There are many different ways to think about and notate rhythmic patterns. Visualizing and Conceptualizing Rhythm *introduces the notations, tablatures, conventions, and illustrations that will be used throughout* Rhythm and Transforms. *The distinction between symbolic and literal notations is emphasized.*

Rhythmic notations represent time via a spatial metaphor. There are two approaches to the notation of rhythmic activities: symbolic and literal. Symbolic approaches accentuate high level information about a sound while literal representations allow the sound to be recreated. A good analogy is with written (symbolic) vs. spoken (literal) language. Text presents a concise representation of speech but cannot specify every nuance and vocal gesture of a native speaker. Similarly, symbolic representations of music present a concise description of the sound, but cannot specify every nuance and musical gesture that a musician would naturally include in a performance. Standard musical notation, drum tablatures, and MIDI transcriptions are all examples of symbolic notations. Literal notations allow a (near) perfect reproduction of a performance. Somewhat paradoxically, by preserving all the information about a performance, literal representations make it difficult to focus on particular aspects of the sound that may be aurally significant. For example, while a symbolic notation may show the fundamental timepoints on which the music rests, these may not be discernible in a literal representation. Similarly, symbolic representations of the pitches of musical events are easy to comprehend, yet the pitch may not be easy to extract from a literal representation such as a `.wav` file, a spectrogram, or a granular representation.

A third class might be called abstract notations, where the score itself is intended to be a work of (visual) art. Several modern composers have created idiosyncratic notations for particular purposes in individual pieces, including Cage's indeterminate notations, Crumb's emulation of Medieval symbolic designs, Penderecki's ideograms, and Lutoslawski's mobiles. These are certainly interesting from an artistic perspective and may be quite instructive in terms of the composer's intentions for a particular piece. They are not, however, general notations that can be applied to represent a large class of sound; rather, they are designed for a unique purpose. A variety of visual metaphors are described in Sect. 2.3.

This chapter surveys the varieties of notations that have been used over the centuries to represent rhythmic phenomena, beginning with symbolic methods and then exploring various kinds of literal notations.

2.1 Symbolic Notations

From a near infinite number of possible timbres, rhythms, pitches, and sonic gestures, symbolic notations extract a small number of features to emphasize in pictorial, numeric, or geometric form. In many cases, time is viewed as passing at regular intervals and this passage is depicted by equal divisions of space. In other cases, time (and/or duration) is itself represented symbolically.

2.1.1 Lyrical Notation

One of the earliest forms of rhythmic notation were markings used to annotate chants, though metrical symbols for syllables were well known to the ancient Greeks. As codified in the anonymous *Discantus Positio Vulgaris* in the early 13th century, these were built on the distinction between long, strong, accented syllables notated with a dash $-$, and short, weak, unaccented syllables labeled $\smile$. Many of these terms, such as those in the prosodic notation of Table 2.1 are still used in the analysis of stress patterns in poetry.

Table 2.1. The five common elements (feet) of the metrical structure of English verse

name	stress pattern	symbol	examples
trochee	long–short	$- \smile$	singing, pizza, rigid
iamb	short–long	$\smile -$	appear, remark, event
dactyl	long–short–short	$- \smile\smile$	tenderly, bitterly, specimen
anapest	short–short–long	$\smile\smile -$	in the night, on the road, up a tree
amphibrach	short–long–short	$\smile - \smile$	acoustic, familiar, Sethares

There are two parts to *scansion*, the rhythmic analysis of poetic meter. First, partition the phrase into syllables and identify the most common metric units from Table 2.1. Second, name the lines according to the number of feet. The words monometer, dimeter, trimeter, tetrameter, pentameter, hexameter, heptameter, and octameter, describe the names for one through eight feet per line respectively. For example, Shakespeare's witches sing

$$- \smile \quad - \smile \quad - \smile \quad - \smile$$

Double Double toil and trouble

in trochee with four feet per line, hence in trochaic tetrameter. The lines

◡ ◡ – ◡◡ – ◡ ◡ – ◡ ◡

For the moon never beams without bringing me

– ◡ ◡ – ◡ ◡ – ◡ ◡ –

Dreams of the Beautiful Annabel Lee

from Edgar Allan Poe's *Annabel Lee* are in regular anapestic heptameter, while Christopher Marlowe's line about Helen of Troy (from *Dr. Faustus*)

◡ – ◡ – ◡ – ◡ – ◡ –

Was this the face that launched a thousand ships

is iambic pentameter.

This kind of meter is an arrangement of primitive elements (such as those in Table 2.1) into groups based on stress patterns, accented syllables, and the number of syllables per line and defines the rhythmic structure of poetic stanzas. Though such notation can be useful in following the metrical flow of a poem, it is imprecise because it does not describe actual temporal relationships such as the relative durations of the – and ◡s, nor how long pauses should last.

In principle, the rhythm of chant is the rhythm of ordinary spoken Latin since the intent of the recital is to make the meaning as clear as possible. In practice, it may be desirable to standardize the chants so that important words are emphasized. Guido D'Arezzo, a Benedictine monk working in the eleventh century, invented the first form of chant notation that was able to notate both pitch and rhythm using a four lined staff. This early musical notation (using symbols such as the longa, the brevis, the maxima and the semibrevis as shown in Fig. 2.1) allowed a notation where certain vowels could be elongated and others abbreviated. Especially when wishing to notate nonlyrical music, greater precision is needed. D'Arezzo's system eventually evolved into the modern system of musical notation.

Fig. 2.1. This small segment of a chant from an illuminated manuscript looks remarkably modern with lyrics positioned beneath symbols that indicate both pitch and duration. But there is no clef and no time signature, and there are only four lines per staff.

Lyrically inspired notations also play a role in many oral traditions. For example, Indian *tabla* players traditionally use a collection of *bols* (syllables for "mouth drumming," see Sect. 3.8) to help learn and communicate complex drumming patterns. During the Renaissance, articulation techniques such as the tonguing of recorders and cornets and the bowing of violin were conceptualized syllabically. Ganassi's *Fontegara* [B: 69] shows three major kinds of articulations: *taka*, *tara*, and *lara*, which embody hard-hard, hard-soft, and soft-soft syllables. Such (sub)vocalizations are particularly useful when a recorder attempts to mimic the voice. Lyrical notations have also had impact on modern theories of rhythmic perception such as that of Cooper and Myers [B: 35], which use the five prosodic rhythmic groups of Table 2.1 as a basis for metrical hierarchy. This is discussed further in Sect. 3.2.

2.1.2 Musical Notation

Modern musical notation is, at heart, a set of instructions given by a composer (or arranger) to a performer for the purpose of describing how a piece should be performed. Since performers typically create sound by playing notes on an instrument (or using their voice), it is natural that the notation should be written in "notes." Standard notation contains two parts, a staff that represents the pitches and a method of specifying the duration of each note. The discussion here focuses on the rhythmic portion of the notation.

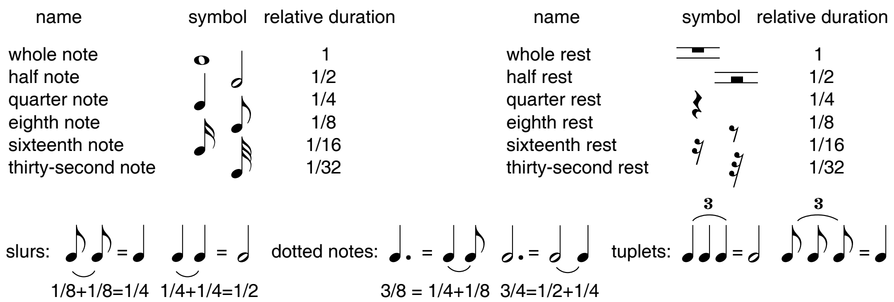

Fig. 2.2. There are six common kinds of notes and six kinds of rests (silences), each with a different duration. The whole tone has a duration of one time unit and all others are specified as fractions of that time. Thus if the whole tone represents one second, the quarter note represents 250 ms. In addition, if a note is followed by a small dot or period, its duration is increased by half. Thus, if the above quarter note were dotted, it would represent a duration of 375 ms. Two notes connected by a slur ⌣ are played as a single note with duration equal to the sum of the two notes. Tuplets allow the specification of durations that are not factors of 2; in one example a half note is divided into three equal parts while in the other a quarter note is divided into three equal parts.

Figure 2.2 shows the various kinds of note symbols and their relative durations. By definition, the whole tone represents a time duration of one time unit, and the others are scaled accordingly. Besides the notes themselves, standard notation also specifies a time signature that looks something like a fraction, for example $\frac{4}{4}$ and $\frac{6}{8}$. The bottom number specifies the note value that defines the beat (4 specifies a quarter note beat while 8 specifies an eighth note beat). The top number states how many beats are in each measure.[1] Thus in $\frac{4}{4}$ time, the length of a measure is equal to the time span of a whole note (since this has the same duration as four quarter notes). In practice, this can be divided in any possible way: as four quarter notes, as four eighth notes plus a single half note, as eight eighth notes, etc. Similarly, in $\frac{6}{8}$ time, there are six eighth note beats in each measure. Again, any combination is possible: six eighth notes, two quarter notes plus two eighth notes, one quarter note plus four eighth notes, etc.

To be concrete, Fig. 2.3 shows the first four measures of Scott Joplin's *Maple Leaf Rag*. The time signature is $\frac{2}{4}$. A quarter note receives one beat and there are two beats per measure. The bass in the lower staff moves along with complete regularity, each measure containing four eighth notes (notes are played simultaneously when they are stacked vertically). In the treble (top) staff, the division is more complex. The first measure contains five sixteenth notes, one eighth note, and one sixteenth rest (totalling two beats). The second measure, in contrast, is divided into a quarter note and four sixteenth notes.[2] Overall, the piece moves at the rate of the sixteenth note, and this forms the tatum of the piece. The complete musical score for the *Maple Leaf Rag* is available on the CD in the files folder [B: 107]. The piece itself can be heard in [S: 5].

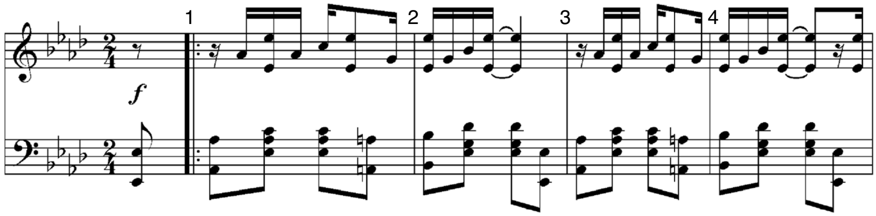

Fig. 2.3. The first four measures of the *Maple Leaf Rag* by Scott Joplin. The full score, drawn by J. Paterson, may be found on the CD [B: 107].

In principle, the same musical passage can be written with different time signatures. The *Maple Leaf Rag*, for instance, can be notated in $\frac{4}{8}$ since two quarter notes per measure is the same as four eighth notes per measure. It could also be notated in $\frac{4}{4}$ by changing all the sixteenth notes to eighth notes,

[1] Measures are typically separated by vertical lines.

[2] One of the sixteenth notes is tied to the quarter note (by the slur) so that there are actually only four notes played in the measure. The three sixteenth notes are followed by a single note with duration $\frac{1}{4} + \frac{1}{16}$. The complete measure is again two beats.

all the eighth notes to quarter notes, etc., and requiring four beats per measure instead of two. Indeed, some MIDI transcriptions represent the meter this way.

Another example is the rhythm represented in Fig. 2.4(a). This rhythm is popular throughout much of Africa; it is typically notated in $\frac{6}{8}$ as shown. Though it could logically be written in $\frac{3}{4}$, it is not. The time signature $\frac{3}{4}$ is reserved for dances that have a “three” feel to them, like a waltz. The difference between such equal-fraction time signatures is stylistic and infused with history.

2.1.3 Necklace Notation

Representing temporal cycles as spatial circles is an old idea: Safî al-Din al-Urmawî, the 13th century theoretician from Baghdad, represents both musical and natural rhythms in a circular notation in the *Book of Cycles* [B: 3]. Time moves around the circle (usually in a clockwise direction) and events are depicted along the periphery. Since the “end” of the circle is also the “beginning,” this emphasizes the repetition inherent in rhythmic patterns.

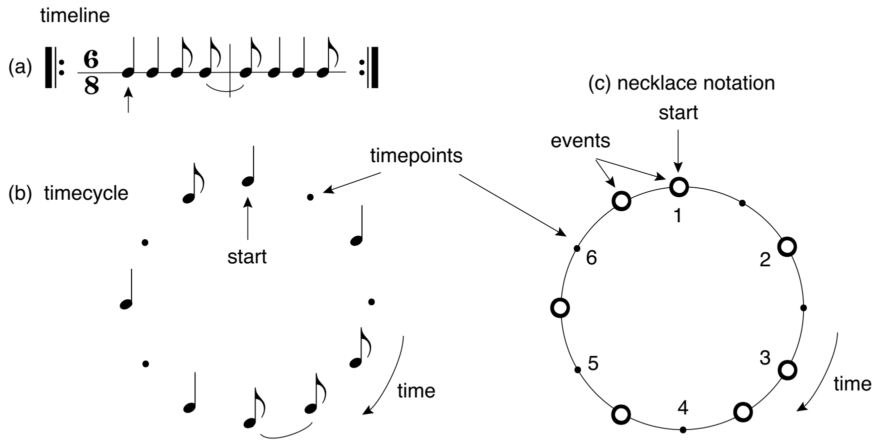

Fig. 2.4. The rhythmic pattern (see also Fig. 1.3 and sound example [S: 2]) is represented in musical notation (a) and then translated into the timecycle notation (b) where the repetition is implicit in the circular structure. The necklace notation in (c) replaces the redundant note symbols with simpler donut-shaped events, and the beats are labeled inside the circle. In both (b) and (c), the twelve timepoints define the tatum, a regular grid of time on which all events lie. The tatum must be inferred from the $\frac{6}{8}$ time signature in the musical notation.

Anku [B: 4] argues that African music is perceived in a circular (rather than linear) fashion that makes the necklace notation particularly appropriate. Performances consist of a steady ostinato against which the master drum performs a series of rhythmic manipulations. The ostinato, visualized as a background of concentric circular rhythms each with its own orientation, reveal staggered entries that sound against (or along with) the regular beat. The master drummer "improvises" by choosing patterns from a collection of rhythmic phrases commonly associated with the specific musical style. "It is in these complex structural manipulations (against a background of a steady ostinato referent) that Africa finds its finest rhythmic qualities" [B: 4].

The necklace notation is useful in showing how seemingly "different" rhythms are related. Part (a) of Fig. 2.5 shows traditional rhythms of the Ewe (from Ghana), the Yoruba (from Nigeria) and the Bemba (from Central Africa). All three are variants of the "standard rhythm pattern" described by King [B: 111]. In (b), the Yoruba pattern (called the *Kànàngó*) is shown along with the accompanying *Aguda*, which can be rotated against the primary rhythm for variety.

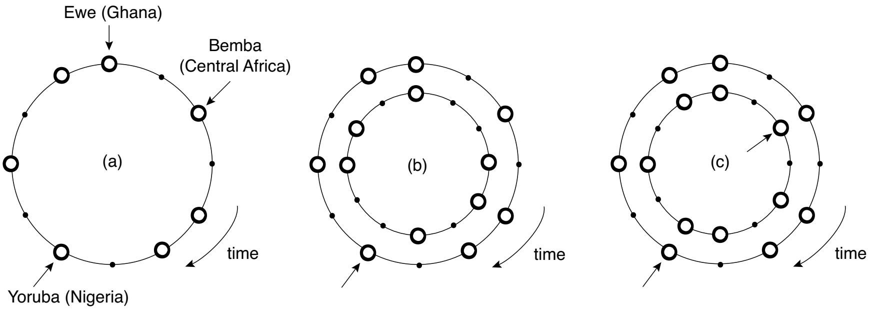

Fig. 2.5. Three traditional rhythms of Africa are variants of King's "standard pattern." They are the same as the rhythmic motif of Figs. 1.3 and 2.4 but interpreted (and perceived) as having different starting points. (b) demonstrates how two drums may play with each other, and in (c) the same pattern appears in both the inner and outer necklace, though rotated in time. These are demonstrated in [S: 9].

In (c), the same *Kànàngó* rhythm is played on the drum and sung, but the two are out of phase with each other. This technique of playing a rhythmic pattern along with a time delayed version of itself can provide "a strong sense of driving forward through the time continuum" [B: 140]. This technique is called a "gap" (*zure*) in Japanese folk music. Similarly, Sachs [B: 187] comments on the use of a rhythmical shift between a singer's melody and the accompaniment that may last for long stretches as the instrument keeps slightly ahead of the singer "by an eighth note or less."

2.1.4 Numerical Notations

Perhaps the simplest form of rhythmic notation begins with a time grid in which each location represents a possible note event. If a note is present at that point in time, the location is labeled "1" while if no event occurs it is labeled “0.” The points of the underlying time grid are assumed to be fixed and known. For example, one cycle of the Ewe rhythm of Fig. 2.5(a) is represented as

$$1\ 0\ 1\ 0\ 1\ 1\ 0\ 1\ 0\ 1\ 0\ 1$$

in the binary notation. One cycle of the two simultaneous rhythms in Fig. 2.5(b) (starting at the arrow) is

$$\begin{array}{c} 1\ 0\ 1\ 0\ 1\ 1\ 0\ 1\ 0\ 1\ 1\ 0 \\ 0\ 0\ 1\ 1\ 0\ 1\ 0\ 0\ 1\ 1\ 0\ 1 \end{array}.$$

Some authors use a variant in which different numbers may appear in the timeslots, depending on some attribute of the instrument or of the event: its volume, the type of stroke, or its duration. For example, Brown [B: 21] weights the amplitude of the note events by the duration, which may be useful when attempting automatic meter detection. In this scheme, the Ewe rhythm is

$$2\ 0\ 2\ 0\ 1\ 2\ 0\ 2\ 0\ 2\ 0\ 1.$$

More sophisticated weighting schemes exploit the results of psychoacoustic experiments on accents. Povel and Okkerman [B: 175] study sequences composed of identical tones. Individual tones tend to be perceptually marked (or accented) if they are

(i) relatively isolated
(ii) the second tone of a cluster of two
(iii) at the start or end of a run (containing three or more elements)

For example, consider the rhythm shown in Fig. 2.6, which compares Brown's duration-weighted scheme with Povel's accent-weighted scheme. Along with the binary notation, these are examples of what Jones [B: 106] calls patterns *in* time, sequences that are defined element by element within the flow of time.

Another kind of numerical notation translates each musical note into a number specifying the duration as an integer multiple of the underlying tatum. For the rhythm in Fig. 2.6 this is the eighth note, and the “duration notation” appears just below the musical notation. Jones [B: 106] calls these patterns *of* time, because each element represents the temporal extent of an event.

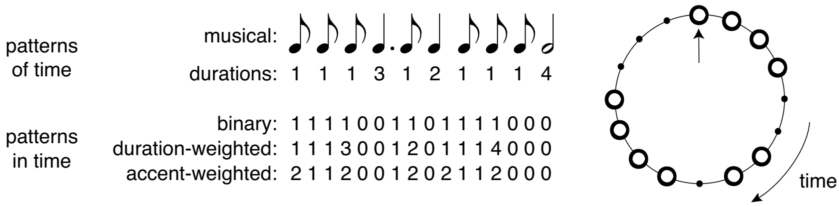

Fig. 2.6. Several different numerical notations for the same rhythm, which is shown in musical and necklace notations and is performed in [S: 10]

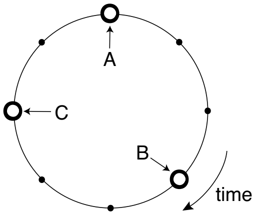

Fig. 2.7. The three functions representing the three rhythms A, B, and C are identified as elements of the same equivalence class under the shift operator

2.1.5 Functional Notation

The binary notation can also be translated into the form of mathematical functions. This is probably the most abstract of the notations, yet it allows the mathematician to demonstrate certain properties of repeating (i.e., rhythmic) patterns. Suppose first that there is an underlying beat (or tatum) on which the rhythm is based and let the integers $\mathcal{Z}$ label the timepoints in the tatum. Following Hall [B: 87], define the function $f : \mathcal{Z} \to \{0, 1\}$ by

$$f(k) = \begin{cases} 1 \text{ if is there is a note onset at } k \\ 0 \text{ otherwise} \end{cases} .$$

The function f represents a periodic rhythm if there is a p such that $f(k) = f(k + p)$ for all $k \in \mathcal{Z}$. The smallest such p is called the *period* of the rhythm. A *rhythm cycle* is defined to be an equivalence class of all p-periodic functions f modulo the shift operator s, which is defined as $(s \cdot f)(k) = f(k - 1)$.

This shift property captures the idea that the rhythm is not just a vector of timepoints, but one that repeats. The equivalence property implies that the 8-periodic functions corresponding to sequences such as $A = \{10010010\}$, $B = \{01001001\}$, and $C = \{10100100\}$ all represent the "same" underlying cyclical pattern, as is demonstrated in Fig. 2.7.

Using such functional notations, it is possible to investigate questions about the number of possible different rhythm patterns, to talk concretely about both symmetric and asymmetric rhythm patterns, and to explore tiling canons (those that "fill" all possible timepoints using a single rhythm cycle with offset starting points, see [B: 87]).

2.1.6 Drum/Percussion Tablature

Drum and percussion tablature is a graphical form of the binary representation in which each possible location on a two-dimensional grid is filled to represent the presence of an event or left empty to indicate silence. Time is presented on the horizontal axis along with verbal directions for counting the beat. The instruments used are defined by the various rows of the grid. For example, Fig. 2.8 shows a two measure phrase played on a full drum kit.

The drum and percussion tablature shows when and how to strike the instruments and there are archives of such tablature available on the web [W: 13]. Percussion grids such as in Fig. 2.8 are also a popular interface for programming drum machines such as Roland's TR-707.

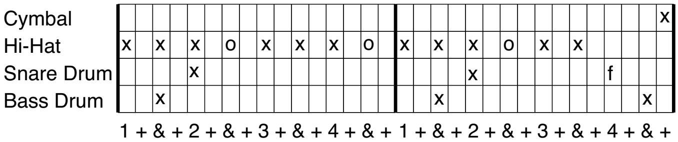

Fig. 2.8. Drum tablature lists the various percussion instruments on the left. The time grid shows when each drum should be hit and the numbering below shows how to count the time. In this case, each of the two measures is counted one-ah-and-ah two-ah-and-ah three-ah-and-ah four-ah-and-ah. The type of stroke is also indicated: ‘x’ means a normal hit, ‘o’ means an open hi-hat, and ‘f’ means a flam on the snare. Typically, there is a legend describing the symbols used on any given tablature. This example is performed in [S: 11].

2.1.7 Schillinger’s Notation

Joseph Schillinger [B: 189] (1895–1943) proposed a graphical technique for picturing musical composition that he hoped would one day replace musical notation. In this notation, a grid of squares represents time moving in the horizontal direction; one square for each timepoint. The vertical dimension is pitch, typically labeled in semitones. Figure 2.9, for instance, shows the start of Bach’s *Two-Part Invention No. 8* [S: 12]. The contour of the melody line is immediately apparent from the graphical representation, and Schillinger considered this an important step in helping to create a scientific approach to melody. Schillinger showed how to modify the musical work graphically through variation of its geometrical properties, and how to compose directly in the graphical domain. Other factors than pitch may also be recorded analogously, for example, a curve might show the loudness at each grid point, the amount of vibrato, or various aspects of tone quality.

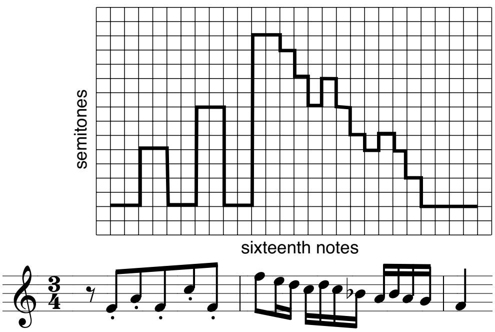

Fig. 2.9. The first two measures of Bach’s *Two-Part Invention No. 8* are shown in musical notation and in Schillinger’s graphical notation. In the grid, each horizontal square represents one time unit (in this case, a sixteenth note) and each vertical square represents one semitone. The contour of the melody is immediately apparent from the contour of the curve.

Schillinger’s theory of rhythm merges this graphical notation with the idea of interference patterns. When two sine waves of different frequencies are added together, they alternate between constructive and destructive interference. This alternation is perceived as beats if the differences between the frequencies is in the range of about 0.5 to 10 Hz. Schillinger schematizes this idea using square waves in which each change in level represents a new note. When two (or more) waves are combined, new notes occur at every change in level. This allows creation of a large variety of rhythmic patterns from simple source material, and the *Encyclopedia of Rhythms* [B: 190] provides a "massive collection of rhythm patterns."

Two examples are shown in Fig. 2.10. In part (a), a square wave with two changes per unit time is combined with a square wave with three changes per unit time. Overall, changes (new notes) occur at four of the six possible times. The resulting rhythm is also given in standard musical notation, and in the necklace notation. Similarly, part (b) shows a four-against-three pattern, which results in a more complex polyrhythm. Schillinger's system is a method of generating polyrhythms (see also Sect. 3.9), that is, a method of combining multiple steady pulse trains, each with its own period. When played simultaneously, the polyrhythm is sounded.

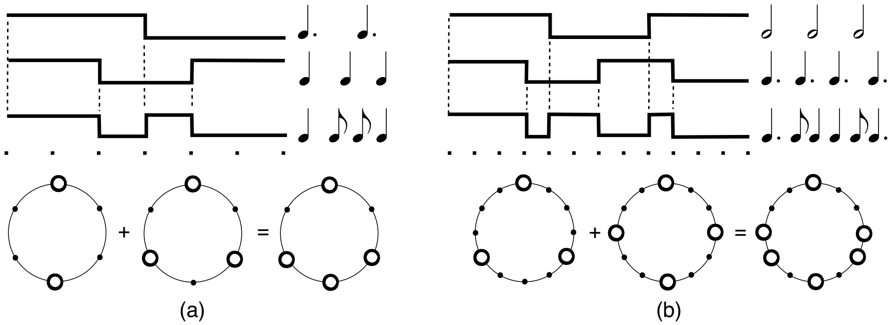

Fig. 2.10. (a) One wave repeats every three time units and the other repeats every two time units. Combining these leads to a three-against-two polyrhythm. (b) One wave repeats every four time units and the other repeats every three time units. Combining these leads to a four-against-three polyrhythm. Both are shown in Schillinger's graphical notation, in musical notation, and in the necklace notation. Dots represent the steady pulse. These and other polyrhythms are demonstrated in [S: 27].

2.1.8 MIDI Notation

In 1982, a number of synthesizer manufacturers agreed to a common specification for the transfer of digital information between electronic musical instruments. While originally conceived as a way to synchronize the performance of multiple synthesizers and keyboards, the MIDI (Musical Instrument Digital Interface) specification is flexible enough to allow computers to interact with electronic keyboards in many ways. Since then, MIDI has been adopted throughout the electronic music industry.

An outgrowth of the MIDI protocol is the Standard MIDI File (SMF) format [W: 49], commonly indicated by the file extension `.mid`, which allows computer programs to store and recall MIDI performances. While not originally intended as a method of notation, the SMF has become a common way of transferring and storing musical information and hence a standard way of representing musical scores and keyboard performances.

Raw MIDI data is difficult to parse directly. For example, the list of numbers in Table 2.2 represents a single note (number $53 = F2$, the F just below middle C) played on channel 1 (MIDI allows 16 simultaneous channels of information). The note is struck with a "velocity" of 100 (out of a maximum of 127). Time passes (represented by the variable length "delta time") and then the note is turned off. Fortunately, the computer can be used to aid in the organization and visualization of the MIDI data.

Table 2.2. The right table shows the MIDI event list for the five note pattern in Fig. 2.11. The left table shows the raw MIDI data for the first note alone.

MIDI Data	Comment
144	note on channel 1
53	note number 53 (F2) – bell
100	note velocity
129	delta time
112	delta time continued
128	note off channel 1
53	note number 53 (F2)
64	note off velocity

Time	Note	Velocity	Duration
1|1|000	F2	↓100 ↑64	1|000
1|2|000	F2	↓64 ↑64	1|000
1|3|000	F2	↓64 ↑64	0|240
1|3|240	F2	↓64 ↑64	1|000
1|4|240	F2	↓64 ↑64	0|240

For example, the right hand side of Table 2.2 shows an "event list," a way of organizing the MIDI data. The first row in this table represents the same information as the previous column of numbers. Each of the other rows represents a subsequent note event, and the five notes are shown in timeline notation in Fig. 2.11. They can be heard in [S: 13]. Observe that the SMF data incorporates information about the metric structure (the time signature) of the piece. For example, the first note occurs at measure 1, beat 1, tick 0. By default, there are four beats in each measure and each beat has 480 ticks. These defaults can be changed in the file header. The duration of the first note is given on the right as 1|000, which specifies the duration as one beat). Similarly, the third and fifth notes have a duration of 0 beats and 240 ticks (1/2 of a beat). This representation makes it easy for software to change the tempo of a MIDI performance by simply redefining the time span of the underlying tick.

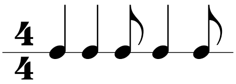

Fig. 2.11. This measure appears as a MIDI event list in Table 2.2 and as the first measure in the piano-roll notation of Fig. 2.12. It is performed using a bell sound in [S: 13].

Even more useful than event lists, however, are visualizations of MIDI data such as the piano-roll of Fig. 2.12. In this representation, the vertical axis represents MIDI note numbers (such as $F2 = 53$ above) and the corresponding notes of the piano keyboard are shown graphically on both sides of the figure. Time moves along the horizontal axis, marked in beats and measures. When representing percussion, each row corresponds to a different instrument (instead of a different note). In the general MIDI drum specification, for instance, the row corresponding to $C1$ is the bass drum with MIDI note number 36, $D1 = 38$ is the snare, and $F\sharp 1 = 42$, $G\sharp 1 = 44$, and $A\sharp 1 = 46$ are various kinds of hi-hat cymbals. These are labeled in the figure along with the corresponding MIDI note number. The drum pattern is played in [S: 14].

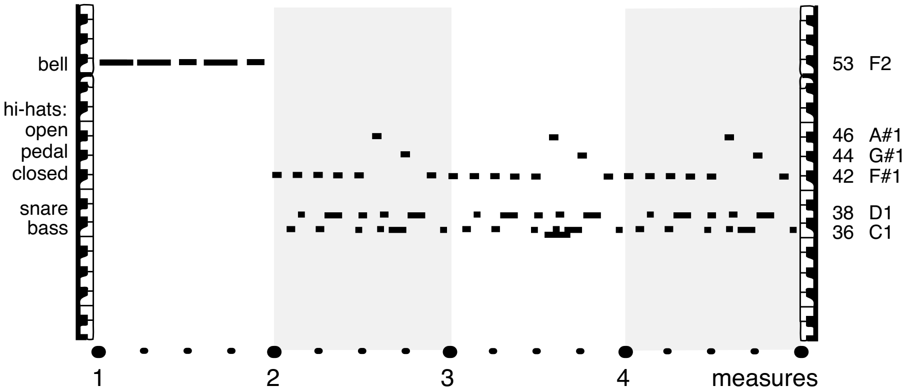

Fig. 2.12. A standard MIDI drum track is shown in piano roll notation. Each of the four measures is divided into four equal beats. The first measure represents the same five notes (performed with a bell sound in [S: 13]) as are given in Table 2.2 and shown in the timeline of Fig. 2.11. The final four measures represent a common "rock" oriented drum pattern, and can be heard in [S: 14].

More common than using the piano-roll for notating percussion is to notate complete musical performances. The first four measures of the *Maple Leaf Rag* by Scott Joplin are displayed in MIDI piano-roll notation in Fig. 2.13. This can be readily compared to the musical score from Fig. 2.3 and is performed in [S: 5]. Thus standard MIDI files, like all the notations discussed so far, are a symbolic representation of the music rather than a direct representation of the sound. When using software to play standard MIDI files, it is easy to change the "sound" of the piece by assigning a different synthesizer "patch," for example, to play the *Maple Leaf Rag* with a violin sound or a xylophone sound instead of the default piano. If you are not familiar with this technique, you may want to experiment with the MIDI file of the *Maple Leaf Rag*, which is available on the CD [B: 107], or download some standard MIDI files from the Classical Music Archives [W: 9].

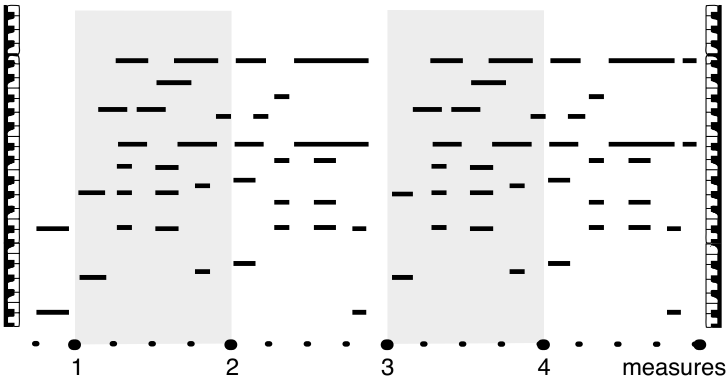

Fig. 2.13. Joplin's *Maple Leaf Rag* is sequenced in MIDI form on the CD (see [B: 107]) and a piano performance can be heard at [S: 5]. Shown here are the first four measures for easy comparison to the standard musical notation in Fig. 2.3.

2.1.9 Harmonic Rhythm

When a rhythmic pattern is played sufficiently rapidly, it becomes a tone. For example, sound example [S: 33] plays a regular sequence of identical clicks separated in time by N milliseconds. When N is large ($N = 500$), it is perceived as a steady rhythm at a rate of two clicks per second; when N is small ($N = 10$) it is perceived as a tone with a pitch that corresponds to a frequency of $\frac{1}{10 \text{ ms per pulse}} = 100$ Hz.

Suppose that the rhythm consists of two pulse trains sounding simultaneously. Cowell [B: 37] draws a parallel between the ratio of the pulse rates and the interval between the pitches of the resulting tones. Suppose that the pulses in the first rhythm are separated by N_1 ms and those in the second rhythm are separated by N_2. Each will have a pitch corresponding to the inverse of the pulse rate, and so the interval (ratio) between the two tones is $\frac{N_2}{N_1}$. For example, with $N_1 = 10$ and $N_2 = 20$ ms, the two pitches are in a 2:1 ratio. Thus they are an octave apart. If N_2 were 15 ms, then the ratio would be 3:2, corresponding to a musical fifth.

This is shown in Fig. 2.14, which lists the first six harmonics of a C note. The ratio formed by successive harmonics and the corresponding musical intervals appear in the next columns. If the period of the fundamental is one time unit, then the time occupied by each of the harmonics is shown graphically. Finally, the ratios are expressed in terms of musical (rhythmic) notation.

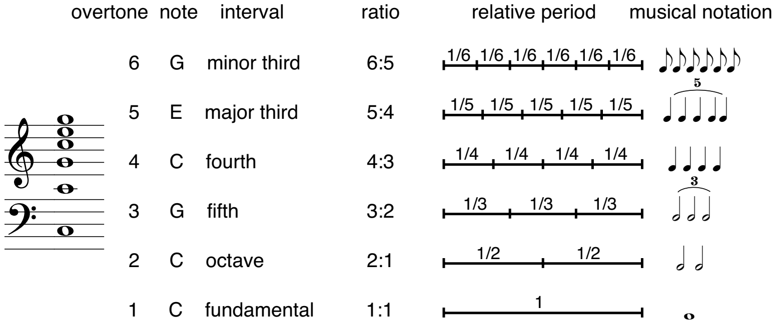

Fig. 2.14. A complex musical tone with fundamental C has six harmonics. The intervals formed by successive overtones and the corresponding ratios are shown. These ratios correspond to the period of vibration of the overtone, which are shown schematically. These periods are translated into musical notation in the final column.

This provides an analogy between the harmonic overtones of musical sounds and the simple-integer relationship found in certain rhythmic patterns. For example, the perfect fifth is represented by a frequency ratio of 3:2, hence it is analogous to a pair of rhythmic pulses with periods in the ratio 3:2. This is the same three-against-two polyrhythm as in Fig. 2.10(a). The polyrhythm in part (b) is in a 4:3 ratio and thus corresponds to a perfect fourth. Similarly, the major and minor thirds correspond to 5:4 and 6:5 polyrhythms. Not shown are the 5:3 (major sixth) and 8:5 (minor sixth) rhythms.

Using this kind of argument, Jay [W: 20] states that harmony and rhythm are different aspects of the same phenomenon occurring at radically different speeds. If a harmony is lowered several octaves, pitches become pulses, and the more consonant harmonic intervals become regularly repeating rhythms. The greater the consonance, the simpler the rhythm. On the other hand, rhythm is converted into harmony by raising it five to ten octaves. Thus harmony is fast rhythm; rhythm is slow harmony.

Stockhausen [B: 224] takes an analogous view in "... *How Time Passes...*" where he argues that pitch and rhythm can be considered to be the same phenomenon differing only in time scale. Stockhausen makes several attempts to create a scale of durations that mirrors the twelve notes of the chromatic scale. These include:

(i) a series of durations that are multiples of a single time unit: T, $2T$, ..., $12T$

(ii) subdividing a unit of time T into twelve fractions T, $\frac{T}{2}$, $\frac{T}{3}$, $\ldots$, $\frac{T}{12}$
(iii) dividing time into twelve durations that are spaced logarithmically between T and $2T$: T, αT, $\alpha^2 T$, $\alpha^3 T$, $\ldots$, $2T$, where $\alpha = \sqrt[12]{2}$

It is not easy to construct a scale of durations that makes sense both logically and perceptually. The idea of a fundamental relationship between harmony and rhythm is enticing and the logical connections are clear. Unfortunately, our perceptual apparatus operates radically differently at very slow (rhythmic) and very fast (pitched) rates. See Sect. 3.9 for sound examples and further discussion.

2.1.10 Dance Notation

> *Labanotation* serves the art of dance much as music notation serves the art of music. Ann Hutchinson [B: 100].

Music is but one human rhythmic activity, and there are a wide variety of notations and conventions that describe dance, juggling, mime, sports (such as gymnastics, ice skating, and karate), and physical movements (such as physical therapy and body language). In [B: 99], Ann Hutchinson reviews a number of dance notations, beginning with the 15th Century Municipal Archives of Cervera, Spain, where various steps are notated by collections of horizontal and vertical lines and a series of abbreviations. Raoul Feuillet's "track drawings" show a floor plan annotated with foot and arm movements. Theleur's method uses stick figures to show leg, body, and arm motions. Saint-Léon's stenochorégraphie places symbolic stick figures (showing motion) on a musical staff (indicating time). Margaret Morris observed that "all human movements take place around an imaginary central axis," and used abstract symbols placed on a pair of three-lined staves to express various movements: gestures of the head and arms on the upper staff, activities of the legs and feet on the lower staff, and movement of the body in between.

Perhaps the most common dance notation today is *Labanotation*, named after the Hungarian Rudolf von Laban (1879–1958). Laban conceived of his "kinetographi," which combines the Greek words *kinētikos* (to move) and *graphos* (writing), as an attempt to analyze and record every aspect of human movement [W: 24]. Like musical notation, Labanotation is based on a series of regular beats, which are typically displayed running vertically from bottom to top. A vertical center line divides two columns that represent the left and right sides of the dancer. Motions occurring to the left (or right) side of the body are thus sensibly displayed on the left (or right) side of the center line. Symbols are placed in the columns: rectangles are altered in shape to indicate direction, color and/or shading are used to show level, and length is used to indicate duration. Special symbols are used to represent the joints (shoulders, knees, fingers), for surfaces of the body (palm, face, and chest), and special signs can be used for a wide variety of actions such as touching, sliding, stamping, clapping, etc. A readily accessible introduction to Labanotation can be found at the Dance Notation Bureau [W: 12].

For example, Fig. 2.15 shows two measures of a tango in Labanotation, as recorded by Andreas Maag [W: 25]. There are two columns because there are two dancers, and the notation is augmented with verbal descriptions of unusual or characteristic motions. Different dance styles have different sets of specialized symbols.

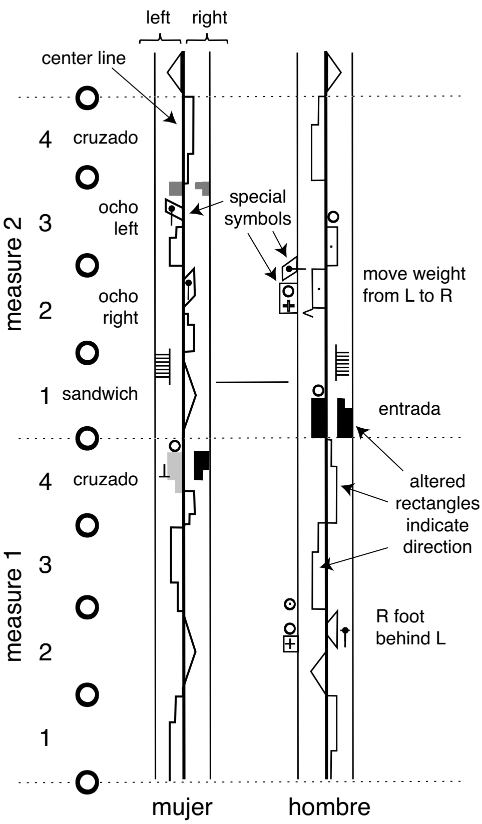

Fig. 2.15. Two measures of a tango notated in Labanotation by Andreas Maag [W: 25]. The two dancers perform somewhat different actions and their motions are notated in separate columns that are synchronized (vertically) in time. The annotation reads, in part: Normal salida (2 steps) then step to change feet (R.F. behind the L.F., do not turn the pelvis); 2 steps. Entrada (footstop) and sandwich (R.F. of the hombre between the feet of the mujer). Bow, hombre R.F. touches mujer feet (but not on top). The hombre leads a turn and interrupts it right away (beat 7) with R.F. Pulls R.F. back and leads ocho to finish. The figure is redrawn and annotated with permission. The circles representing the beat locations are added to emphasize the similarities with beat-based musical notations.

2.1.11 Juggling Notation

Juggling is the action of repeatedly tossing a number of objects (typically balls) into the air and catching them again. It requires skill and agility because no hand holds more than one ball at a time and there are typically more balls than hands. Like music, the actions must be performed rhythmically. Like dance, the actions must respect gravity.

Over the years, jugglers have developed a variety of ways to notate their art. Some notations are more flexible (can describe a large variety of different patterns) and some are more compact. One recent innovation is the "beat-based juggling notation" of Luke Burrage (described at the Internet Juggling Database [W: 19]), which notates what each hand is doing at every beat in time. Beever's comprehensive *Guide to Juggling Patterns* [B: 10] written for "jugglers, mathematicians and other curious people," gives a rich overview of juggling techniques and notations ("like sheet music for jugglers"), and demonstrates a variety of related mathematical results.

Perhaps the most popular notation is the siteswap notation (see [B: 227], [W: 48]), which represents throws by integers that specify the number of beats in the future when the object is thrown again. Because there are many constraints (no hand holds more than one ball, balls must return to a hand after a short time in flight) these integer patterns represent juggling patterns. Beever [B: 10] builds up a logical description of the process that includes the time, site (e.g., left or right hand, elbow, knee), and position (to the left, right or right side of the body) of the throw, the position and site of the catch, and the airtime of the flight. From these, and under nominal assumptions, he derives the siteswap base which provides a shorthand notation for general use.

For example, Fig. 2.16 shows the schematic "ladder" notation of the cascade, a standard three-object juggling pattern. Time moves vertically from the bottom to the top. Beats occurring at roughly equal timepoints are represented by the small circles. The location of each of the three balls is indicated by its horizontal position: held in the left hand, in the air, or held in the right hand. At each beat timepoint, one ball is being thrown, one is being caught, and one is in flight. Since there are three balls and since each ball is thrown every three beats, this pattern is represented succinctly as "3." Except for trivial changes (such as starting with a different hand), this completely specifies the cascade.

By changing the number of balls and the order of arrivals and departures, a large variety of different juggling patterns can be represented. For example, the right hand part of Fig. 2.16 shows a four-ball juggling pattern notated "534" in the siteswap notation. In this pattern, one ball moves regularly back and forth between the two hands every three beats. The other three balls follow a more complex pattern. Observe that odd numbers are always caught by the opposite hand and even numbers are caught by the tossing hand. Hence the five beat throw moves from one hand to the other while every four beat toss is caught by the throwing hand. The throwing pattern of the three balls repeats every 18 beats. The website at [W: 48] has a useful and entertaining visualization tool that automatically simulates any legal siteswap pattern.

The beats (successive timepoints) in juggling are analogous to beats in music since they provide a rhythmic frame on which all activity is based. All of the well known juggling notations are symbolic since they specify only the most important events (the throwing and catching of the balls) and do not specify the exact hand positions, the activities of the hands (other than at

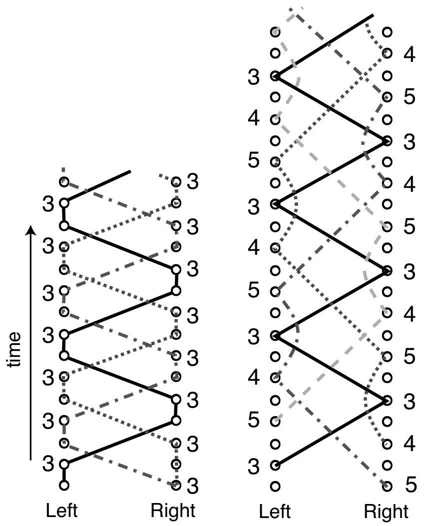

Fig. 2.16. Each rung in the "ladder notation" (described in [B: 10]) represents one (roughly equal) timepoint. The balls are represented by the different lines and the location of each ball is specified (in the left hand, in the right hand, or in the air) at each timepoint. The numbers indicate how many beats must pass before the ball can be thrown again. In the left ladder, the standard three-ball cascade, this is always three beats and this pattern is "3" in the siteswap notation. The "534" on the right is a four ball juggling pattern with one ball that changes hands every three beats (like the cascade) and three balls that follow a more complex pattern. "345" and "453" represent the same pattern with different starting hands.

the instants of catching and throwing) nor the actual trajectories of the balls in flight. Nonetheless, there is enough information in the notation to allow jugglers to learn and invent new patterns.

2.2 Literal Notations

The great strength of symbolic notations for music is that they show the structure of a piece at a high level. The symbols (the notes, rests, MIDI events, tablatures, etc.) represent the underlying structure of a piece by providing instructions that can be readily translated by people or machine into performances of the work. The weakness of symbolic representations is that they do not specify many salient factors such as the timbre of the instruments or the exact timing of the events. In short, they are not a literal record of the sound, but a reminder of what the sound is like.

Literal notations allow full reproduction of a performance even though important aspects of the sound may become lost among a flood of (near) irrelevant data. For example, while pitch is clearly an important aspect of musical performance, waveform representations (such as that in Fig. 2.17) do not display pitch in an obvious way. Similarly, the fundamental timepoints on which a rhythmic passage is built may not be clearly marked. Literal notations are, by their nature, recordings of particular performances and not representations of the underlying composition. This distinction is discussed further in Sect. 12.3.

2.2.1 Waveforms

The most common literal method of musical representation is direct storage of the sound wave. Common analog storage mechanisms are cassette tapes and LP records. Common digital storage techniques sample the waveform and then store a representation of the samples on magnetic tape, in optical form on a CD, or in the memory of a computer. All of these technologies record the variations in a sound pressure wave as it reaches a microphone. Larger numbers (greater deviations) indicate higher pressures, and sound is perceived when the fluctuations in the waves are between (about) 20 and 20,000 cycles each second. When stored digitally, this requires thousands of numbers per second (88,200 numbers for each second of stereo music). The most common way to view the data is to plot the values vs. time, as is done in Fig. 2.17.

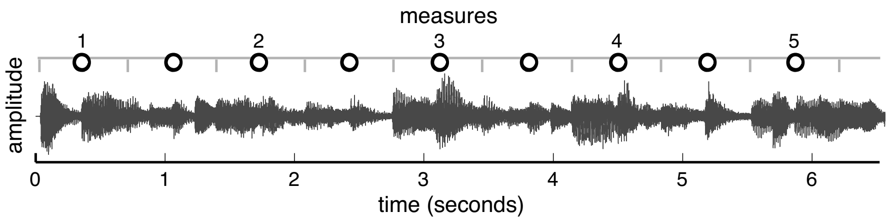

Fig. 2.17. This is a plot of the raw sound wave data for the first six seconds of Joplin's *Maple Leaf Rag.* The larger blobs represent louder passages. The measures and beats are also notated, though these are not directly present in the data, they have been added to help orient the reader and to enable comparison with the musical score Fig. 2.3 and the MIDI notation Fig. 2.13. It is not easy to see individual notes in this representation.

Playing back a sound file such as a `.wav` or a `.mp3` allows a listening experience that is much like that experienced near the microphone at which the recording was made. The timbre of the instruments and the timings are reproduced in (nearly) unaltered form. The performance is fully specified by the sound file. What is not obvious, however, are the higher level abstractions: the notes, the beats, the melodies. Indeed, it is very difficult to determine such high level information directly from a waveform. This problem in musical processing is analogous to the well known problem of creating a computer program that can understand connected speech.

Waveform representations typically represent time linearly, but they may also use other geometrical constructions. Analogous to the necklace notation for symbolic cyclical patterns, the waveform can be plotted around a circle. Figure 2.18, for instance, shows the first two measures of the *Maple Leaf Rag.* The progress of time is specified in beats, which are indicated by the small circles. Such a representation is particularly appropriate when the music repeats regularly at the specified interval.

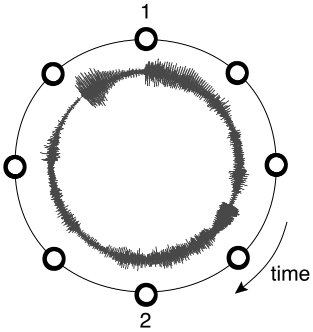

Fig. 2.18. The first two measures of the *Maple Leaf Rag* are displayed in a polar (circular) plot. Time moves in a clockwise direction and the beats are indicated by the small circles.

It is common to use the waveform representation as a method of storing music (this is what CDs do) but it is not common to compose directly with waveforms. In one approach, the composer chooses a collection of individual time points and amplitudes, and then invokes a computer program to algorithmically interpolate the intermediate samples [B: 242]. The algorithms may generate the intervening data using a nonlinear dynamical system, via some kind of random process, or via a process of hierarchical construction. Wavetable synthesis directly generates short snippets of waves that are used as building blocks for larger sound structures. By creative looping and moving the loop points, it is possible to create ever-changing waveforms from a small number of specified tables. This synthesis technique was used in the PPG synthesizers and in the more modern Waldorf wavetable synthesizers.

2.2.2 Spectrograms

The spectrum looks inside a sound and shows how it can be decomposed into (or built up from) a collection of sinusoids. For example, guitar strings are flexible and lightweight, and they are held firmly in place at both ends under considerable tension. When plucked, the string vibrates in a far more complex and interesting way than the simple sine wave oscillations of a tuning fork or an electronic tuner. Figure 2.19 shows the first $\frac{3}{4}$ second of the open G string of my Martin acoustic guitar. Observe that the waveform is initially very complex, bouncing up and down rapidly. As time passes, the oscillations die away and the gyrations simplify. Although it may appear that almost anything could be happening, the string can vibrate freely only at certain frequencies because of its physical constraints.

For sustained oscillations, a complete half cycle of the wave must fit exactly inside the length of the string; otherwise, the string would have to move up and down where it is rigidly attached to the bridge (or nut) of the guitar. This is a tug of war the string inevitably loses, because the bridge and nut are far more massive than the string. Thus, all oscillations except those at certain privileged frequencies are rapidly attenuated. This is why the spectrum shows large peaks at the fundamental and at the integer harmonics: these are the

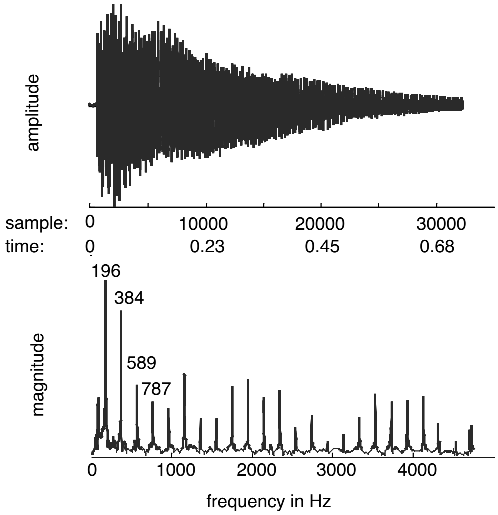

Fig. 2.19. Waveform of a guitar pluck and its spectrum. The top figure shows the first $\frac{3}{4}$ second (32,000 samples) of the pluck of a G string of an acoustic guitar. The spectrum shows the fundamental at 196 Hz and evenly spaced overtones at 384, 589, 787, etc. These are called the harmonics of the sound; they occur near simple integer multiples of the fundamental since $384 \approx 2 \times 196$, $589 \approx 3 \times 196$, and $787 \approx 4 \times 196$. More than twenty harmonics are clearly distinguishable.

frequencies that "fit" inside the length of the string. Spectra such as Fig. 2.19 are typically calculated in a computer using the Discrete Fourier Transform (DFT) and its numerically slicker cousin, the Fast Fourier Transform (FFT).

Spectrum plots are useful when studying the composition of isolated sounds (such as a guitar pluck) and more generally to sounds that do not change over time. But they are not typically useful for complex evolving sounds such as a rendition of the *Maple Leaf Rag*. In this case, it is common to use a sequence of short spectral snapshots called a spectrogram, which shows how the spectrum changes over time. Figure 2.20 shows the waveform and spectrogram of the start of the *Maple Leaf Rag*.

Spectrograms are built by partitioning a sound into small segments of time called frames and then applying the FFT to each frame. All the FFTs are placed side-by-side; large values are represented by dark coloration, small values are shaded lightly. Thus time is represented on the horizontal axis, frequency is represented on the vertical axis, and energy (or loudness) is represented by darkness of the shading. In addition it is possible to create spectrograms dynamically and in other geometrical arrangements. See, for instance, [W: 52].

Since a spectrogram is a visual representation of a sound, it is possible to synthesize sound from an image, and to manipulate the sound graphically. The "image synthesizer" in Eric Wenger's *Metasynth* [W: 31] transforms images into sound by translating each pixel into a short fragment of sound based on its horizontal location (mapped to time), vertical location (mapped to frequency), and color (mapped to stereo placement). Both the vertical and horizontal scales are flexible; frequencies can be specified in almost any scale and time can move at any rate. Using default parameters, the stretched girl

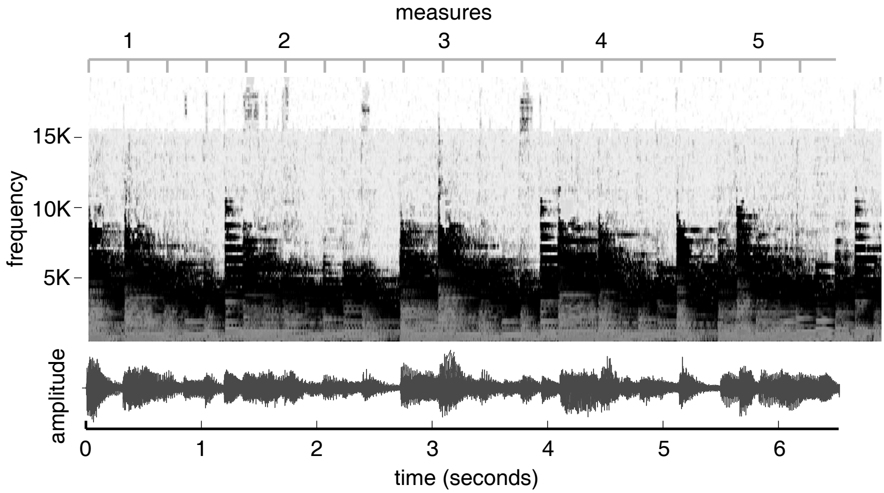

Fig. 2.20. Waveform and spectrogram of the first 6.5 seconds of the *Maple Leaf Rag*

in Fig. 2.21 has a duration of about 24 seconds. She may be heard in sound example [S: 15].

The pictorial representation encourages manipulation of the picture using the kinds of tools familiar from graphics and drawing programs: cutting and pasting, selecting ranges, inverting colors, and drawing with various shaped "pens." *Metasynth* also contains a number of uniquely music-oriented tools: pens that draw the shape of a harmonic series, grid lines that mark off the time axis in a rhythmic pattern, tools that sharpen attacks by emphasizing edges, spray brushes that splatter tiny fragments of sound across the pallette, and tools that add a haze after each sound (reverb) or before each sound (pre-verb).

Another innovation in *Metasynth* is the use of images as filters. The filter palette uses the same mapping (horizontal for time, vertical for frequency) as the image synthesizer, but is not heard directly. Rather, the filter image is applied point-by-point to the sound image. This allows the creation of arbitrarily complex filterings that can be different at each time instant. Figure 2.22, for

Fig. 2.21. Metasynth can transform images into sound: this stretched picture can be heard in sound example [S: 15]

instance, shows two image filters in (a) and (c). These are applied to the first few seconds of the *Maple Leaf Rag* and result in the spectrograms shown in (b) and (d). The vertical stripes of (a) suggest a stuttering effect, and this is borne out the sound examples [S: 16]. Similarly, the graceful sweeping arch of (c) suggests a lowpass effect that slowly changes to highpass and back.

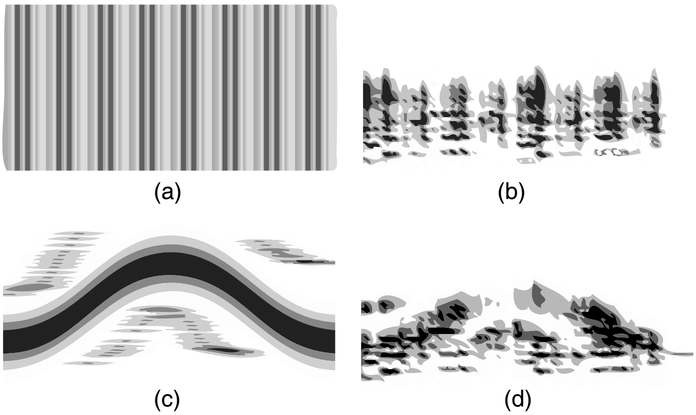

Fig. 2.22. *Metasynth* uses one image to filter another. The two filters in (a) and (c) are applied to the first few seconds of the *Maple Leaf Rag*. The resulting spectrograms are shown in (b) and (d), and may be heard in sound example [S: 16].

Such graphically oriented sound manipulation represents a change in paradigm from conventional methods of creating and modifying sounds. In the standard method, musical instruments are used as the sound source, and these are orchestrated using symbolic notations such as a musical score. In contrast, *Metasynth* generates its sound directly from a picture, directly in its literal notation. The strength of the system is that the composer is supplied with a large variety of perceptually meaningful tools for changing and rearranging the basic sound material. The high level abstractions of the musical score include the pitch of notes and the regular time unit defined by the measure. The high level abstractions in spectrogram “scores” include the frequency content of a signal as displayed in the image and time-grids that can be arbitrarily specified.

2.2.3 Granular Representations

Any signal can be be represented by a collection of (possibly overlapping) acoustic elements called grains (see Fig. 2.23). Though each individual grain sounds like a brief click, aligned masses of grains can represent familiar sounds and make it easy to specify certain kinds of complex sound clouds and auditory textures. This decomposition of sound into quanta of acoustical energy was first proposed by Gabor [B: 68] in 1947, but was popularized in the 1970s by Xenakis [B: 247] and Roads [B: 181], who were among the first to exploit the power of computer-based synthesis using a granular technique.

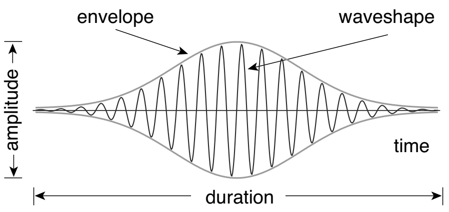

Fig. 2.23. Sound grains are characterized by their duration, amplitude, envelope, and waveshape. If the envelope is a Gaussian curve and the waveshape is a sinusoid (as shown), then it is a "Gabor grain." Other kinds of grains use different waveshapes, envelopes, and durations.

Each grain is characterized by its duration, envelope, amplitude, and waveshape. In Gabor's original work, the envelope had a Gaussian (bell curve) shape (because this provides an optimal trade-off between spread in frequency and spread in time) and the waveshape was sinusoidal. Since the durations are very short, individual grains sound like short clicks, though longer duration grains may give a pitch sensation as well. In the two examples [S: 17], both individual grains and small clouds of grains are clearly audible. [S: 18] is built from a variety of different grain shapes synchronized to a shuffle pattern.

Grains become useful when many are clustered together. Xenakis pictures each grain as a small dot, and places the dots onto a "screen" whose axes specify the frequency and intensity of the grain. A collection of screens (called a "book") is then played back over time (like the individual still frames of a movie) in order to create complex evolving sound textures. Figure 2.24 illustrates.

For example, in Xenakis' piece *Concret PH*, which premiered at the Brussels World's Fair in 1958, the grains were drawn from recordings of a crackling wood fire that were cut into one-second fragments. They were then recombined using tape splicing techniques. In *Analogique B*, Xenakis created grains from the output of an analog sine wave generator by cutting the tape recording into short sections. The grains were then combined by placing them on screens with probabilities drawn from an exponential distribution. The "degree of order" was controlled by random numbers drawn from a first-order Markov chain.

Because so many grains are needed, specifying all the information is a nontrivial task. Roads [B: 181] (and others [B: 41], [B: 231]) have developed sophisticated ways to use stochastic procedures (random numbers) to choose from among the most salient variables. In these schemes, the composer chooses a collection of parameters that specify the probabilities that certain kinds of grains occur at certain times. The computer then generates the bulk of the actual data and realizes the composition. Typical parameters that might be controlled are:

(i) the density of the grains (number of grains per second)

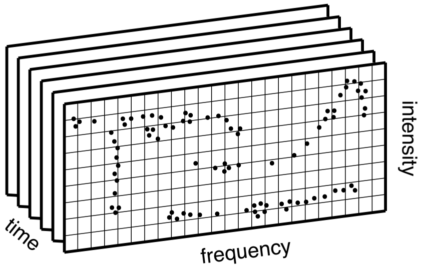

Fig. 2.24. Each grain of sound is represented as a single dot sprayed out onto a grid or screen with differing frequencies and intensities. Collections of screens (a "book of screens") defines the evolution of a complex sound through time.

(ii) the durations of the grains
(iii) the waveshapes within the grains (sinusoid, square wave, noise, impulsive, sampled from a given source, etc.)
(iv) intensity/amplitude of each grain
(v) temporal spacing between grains
(vi) grain envelopes
(vii) frequencies of the grains (e.g., limits on the highest and lowest frequencies)
(viii) frequencies: scattered or aligned
(ix) timing (whether the grains are synchronous or asynchronous)
(x) placement in the stereo field

This represents another conception of musical composition. Rather than working with "notes" and "instruments," the composer specifies a method, or an algorithm for choosing parameters within that method. Rather than conceiving the piece as fixed in a musical score which requires musicians for performance, it is fixed in a computer program which requires appropriate software and hardware for its realization. Granular synthesis is a method of composition in a literal notation (that of sound grains) rather than a method that operates at the symbolic level. Perhaps the most important part of any such method is that the control possibilities provided to the composer must make perceptual sense. They must impose some kind of (relatively) high level order on the literal representation so that it is comprehensible. The list of parameters (i)–(x) above can thus be viewed as an attempt to allow the composer direct control of certain high level variables. These primitives are quite different from the standard idea of a musical "note," with its pitch, volume and timbre, but it is not hard to gain an intuitive feel for parameters such as density, opacity, and transparency.

For example, a cloud containing many 100 ms grains might be perceived as continuous and solid, whereas if it were created from 1 ms grains the cloud would be thin and transparent. If the waveshapes are sinusoids then the sound mass might appear sparkling and clear, while if the waveshapes are jagged and noisy the cloud would occupy more of the spectrum. One of the strengths of sound clouds is that they can evolve over time: in amplitude, density, internal tempo, harmonicity, noise, spectrum, etc. The waveshape might be chosen a priori, generated by an algorithmic process, or perhaps harvested from a sound file.

Computer programs that implement granular synthesis include Road's `Cloud Generator`, Erbe's `Cornbucket` (used in conjunction with `Csound`), Bencina's `Audiomulch`, and McCartney's `SuperCollider`. See [B: 181] and [B: 54]. A variety of experiments with grainlets, pulsars, and glissons are documented in Roads *Microsound*. `Kyma` [W: 23] is a mixed hardware/software package that implements granular synthesis along with a variety of other synthesis techniques.

There are a number similarities between granular techniques and the spectrogram approach of the previous section. Indeed, if the windows used in partitioning the signal are chosen to be the same as the envelope of the grain, and if the waveshape is a sinusoid, then the two representations are logically equivalent. However, there are differences. Grains tend to be of very short duration, while the window length in the FFT-based methods is typically large enough to allow representation of low frequencies (limitations of the spectral methods are discussed further in Chap. 5). The granular technique allows any waveshape within its envelope, while the spectrogram requires the use of sinusoids. Moreover, as we have seen, the spectrogram and the granular communities have developed different kinds of high level abstractions that help to make composition and sound manipulation more transparent for the composer.

As will become apparent in Sect. 5.4, there are also close similarities between granular methods and wavelet transforms. Wavelets also have short "grains," though they are not called this. One difference is that wavelets are not all the same duration; low frequency wavelets are longer and high frequency wavelets are shorter. This helps to make wavelet representations more efficient.

2.3 Visual and Physical Metaphors for Rhythm

Rhythm refers to orderly recurrence in any domain. Visual rhythms may involve alternations of light and dark, of up and down, of colored patterns, or of symbols. Tactile rhythms may involve alternations between strong and weak or may lie in the motion of our bodies. Architectural rhythms may be built from repeated structural or decorative elements such as windows, columns, and arches. Interestingly, we do not appear to be able to sense rhythmic phenomenon directly in either the olfactory or the gustatory senses [B: 65].

For example, Figs. 2.25 and 2.26 show two artistic conceptions of rhythm inspired by Matisse's *Two Dancers* and Mondrian's *Rhythm of Black Lines*. These provide an artistic representation of two fundamental aspects of rhythm: motion and repetition. Like Matisse's famous *Two Dancers*, Fig. 2.25 shows a collection of disembodied shapes that can be seen as representing two people dancing. The darker figure (with disembodied head) is poised to catch his partner after flinging her into the air. Or does this show a flying warrior

Fig. 2.25. Rhythm as motion: like Matisse's *Dancers* paintings, this tries to capture the idea of motion. An instability (the person in the air) drives these figures towards resolution. "Dance," Matisse said, means "life and rhythm."

pouncing on his beheaded victim just as the prey begins to topple? There are several illusions associated with these shapes: why they appear to be people when in reality they are collections of blobs, why they appear to be in motion when in reality the blobs are utterly fixed on the page. Whether dancers or warriors (and other interpretations are possible), the essence of the motionless figure is motion!

Like Mondrian's well known *Rhythm of Black Lines*, Fig. 2.26 shows a collection of crosshatched black lines that repeat at regular intervals. The lines are punctuated by shaded regions and small variations in the lengths and connections of the lines. This is a visual representation of the kinds of *repetition with variation* that is so common in musical phenomenon. A completely regular structure might represent the unchanging beat or the metrical level; the idiosyncratic lengthenings and shortenings correspond to the variations that make the work interesting and exciting.

One of the best known examples of proportion and balance in classical architecture is the Parthenon, the crown of the Acropolis. Built as a temple to Athena around 440 BC, its massive columns, ornamental friezes (showing processions of men and Gods), metopes (representing battles between good and evil, between the Greeks and their enemies) and pediments (with dozens of statues representing stories of various Gods, especially Athena) are elegant, detailed, and exemplify the flowing kinds of elements associated with architectural rhythm. Many of its features are in the proportion of the golden section.

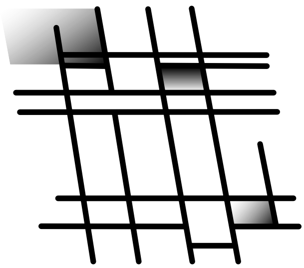

Fig. 2.26. Visual rhythms unfold in space while auditory rhythms evolve over time. Repetition plays a key element in both visual patterns and temporal rhythms.

Figure 2.27 shows the facade, the floor plan, and a small piece of the frieze, all of which demonstrate a mastery of variety within repetition. Goethe refers to such repetitive architectural structures as “frozen music,” though a modern view of architecture as “people moving through light filled spaces”[3] suggests a closer analogy with musical improvisation.

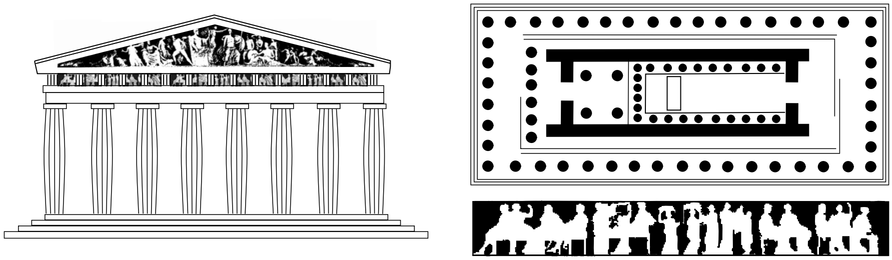

Fig. 2.27. The Parthenon shows the kinds of proportion and balance achieved in classical architecture. The facade, floor plan, and a section of the frieze sometimes called the Elgin marbles (currently in the British museum) all show clear rhythmic features.

Among the oldest metaphors for rhythm are the motions of the bodies of the solar system such as the daily path of the sun, the monthly waxing and waning of the moon and the cyclical behavior of the planets. Many plants have rhythmic behaviors and rhythmic structures: the daily opening and closing of a flower, the yearly cycles of a tree shedding and regrowing its leaves, the bamboo plant with its long stalk punctuated at semi-regular intervals by knots. A necklace of beads provides an analogy in which the beads represent the regular beats while the circularity represents the repetitive metrical structure. Each sand grain dropping through a sand clock is metaphorical of a tiny grain of sound, and the regularity of sand grains passing through the neck represent the regularity of the musical experience. Zuckerkandl [B: 250] suggests that a (water) wave provides a powerful analog of rhythm: the repetitive feeling of relaxation, rising tension, approach to a climax, and then the final ebb.

Rhythm and Transforms deals primarily with the steady recurrence of audible impressions, with rhythmical sounds that represent the organization of time into parts that are perceptible to the ear. Though visual representations and physical metaphors are often useful (especially in a format such as a paperbound book), the subject matter is sound. Accordingly, the sound examples on the CD represent the most important part of the argument. Visualizations provide analogies or metaphors; sound examples are the real thing.

[3] Roger Tucker, private correspondence.

http://www.springer.com/978-1-84628-639-1

Rhythm and Transforms
Sethares, W.A.
2007, XIII, 336 p., Hardcover
ISBN: 978-1-84628-639-1