CHAPTER 4

Rhythm and Timing

4.1 Introduction

Rhythm and timing are inherent in so many aspects of perception, and rhythmic behavior is so common and easy that there does not seem any need to create a framework to understand it. Yet, it is not easy to create an adequate one. We construct our perception of time, and perhaps it is based on the succession of events.

One method is to study the perception of simple auditory rhythms directly. The traditional approach is to consider the perception of auditory rhythms as a cognitive problem and attempt to abstract the rules underlying their organization. At the simplest level, we have the perception of simple isochronous rhythms made up of different sorts of sounds. Alternatively, we have the perception of rhythms based on different timings between the onsets of identical adjacent sounds (for rhythms, it is the onset-to-onset interval that is critical, not the offset-to-onset interval critical for stream segregation). Combining the sound and timing differences creates rhythms that approach those found in music. The understanding of these simple rhythms would be clearly based on the Gestalt principles of organization described in Chap. 2.

An alternate approach to rhythm perception is to enlist the accompanying motor sensations and consider the interplay between the two. This has been termed *embodied rhythm perception* (Maes, Leman, Palmer, & Wanderley, 2014). People tap their feet, clap in time, and sway to rhythms, and these movements are indubitably tied to our perception of those rhythms. By considering those movements timed to the *beat*, it is possible to understand the expressiveness of different rhythms. For physical actions there are the rhythms of hands, arms, bodies, and other individuals that move at different tempos. In addition, there

Electronic Supplementary Material: The online version of this chapter (https://doi.org/10.1007/978-3-319-96337-2_4) contains supplementary material, which is available to authorized users.

S. Handel, *Perceptual Organization*,
https://doi.org/10.1007/978-3-319-96337-2_4

are the multiple timings involved in all kinds of physical activities, ranging from hitting a tennis ball to jumping over a ditch. It is a mistake to argue that rhythm occurs at one level, when each rhythmic level supports the others. The emergent rhythm is due to the interplay of rhythms at different levels.

A second method is to consider the ways that rhythms slice up the continuous ongoing sensations of music, speech, and vision into units of varying duration. For music, there are the rhythms of individual notes, groups of notes, contours of notes, and themes. For speech, there are the rhythms of syllables, words, and sentences. It seems that speech is broken initially into longer, slower units at roughly five syllables/sec, and then the acoustic variations within those units are analyzed at a higher frequency to identify the particular syllable. For vision, timing will affect apparent movement as well as other multistable movements; the timing between the positions of the individual lights in point light displays described in Chap. 2 (Johansson, 1975) determines which motions are perceived. Visually, we perceive smooth motion in movies being projected at 24 frames per second not on account of a sluggish visual system. A sequence of random movements in successive frames is seen as a jumble of unconnected flashes. It is the contour of the movements, rather, that leads to smooth motion. At a slower frame rate, the motion looks jerky once again; it is the combination of the space and time contours that are critical.

Still another method is to stretch concepts like rhythm and timing metaphorically to the visual arts. We can think of rhythms as creating a set of slots into which notes can be placed. In Western music at least, the slots are equally spaced, and as explained later, subsets of more widely separated but still equally spaced slots receive accents. These accents create expectancies that important notes will occur at particular rhythmic intervals and lead listeners to hear the note sequence in synchrony to the accents (Jones, 1976; Large & Jones, 1999). In vision, we can think of space as being filled with a recurring spatial grid, and designs being created by "darkening" some of the grid lines or spaces (DeLong, 2013; Tetlow, 2013). By varying the size and type of grid, it is possible to create a bewildering assortment of designs, whose repetitions in many instances are analogous to the repeating nature of the auditory rhythms. Understanding visual rhythms entail the same principles as the Gestalt approach to auditory rhythms.

At stated above, what is common to all of these aspects of rhythm is the notion that rhythmic organization occurs at several levels simultaneously. Neural codes operating on different time scales might encode complementary stimulus features. We may use the longer intervals of slow cortical network oscillations to create phrase information, namely, the start and end of complex events, and simultaneously use shorter intervals of faster oscillations to analyze the variability within those segmented events. The perceived rhythm comes from the interaction between the timing of the sounds and the internal timing of the listener. Multiple timings allow us to attend to expected events, but it also allows us to react to unexpected changes in the timing of events. It ensures flexibility.

4.2 Auditory Temporal Rhythms

Before starting it is important to distinguish between the physical and phenomenal meanings of the term *rhythm*. We can define the rhythm in terms of the acoustical characteristics of the tonal sequence: the frequency, duration, intensity, and timbre of the individual tones and the timing between the onsets of adjacent notes. We can also define rhythm in terms of the perceptual response. Clapping, finger tapping, and/or swaying to the beat show the perceived rhythm. Finally, we can define rhythm in terms of the notated musical score that indicates the relative timing of the notes. The physical, phenomenal, and notated rhythms may not have a simple match. That is, the beat of a rhythm can occur on an actual tone, but the beat can also fall on a silence.

It seems appropriate to focus on the experience of rhythm. Rhythm is inside us and arises in a specific context. We can specify the physical acoustical characteristics of the sounds that support the experience of rhythm, but at every level we must relate those characteristics to the phenomenal and muscular feelings of listeners to the experienced rhythm. There is no single component of the acoustical signal that can uniquely predict the felt rhythm.

The experience of rhythms involves the feeling of regularity and grouping among the sounds as well as the accentuation and differentiation of those sounds within the groups. Weaker elements are attached to stronger accented ones creating a feeling of regularity among the stronger sounds that influence body motions. The obvious periodic physiological process such as breathing, heartbeats, and walking initially led researchers to believe that such physiological responses underlay the perceptual response. The tempo (i.e., the rate of the events) of heartbeats or breathing rates was thought to set a baseline, so that the tempo of rhythms above that rate would be described as fast and ones below were thought to be sluggish. It soon became quite clear that although physiological changes and movements accompanied the perception of rhythms, they were not the cause of that perception.

Yet, physiological processes do set limits on perceivable rhythms and the ability to synchronize. When synchronizing, listeners anticipate the next sound and respond before it appears; they are not simply reacting to the appearance of that sound and then responding. The critical sense of rhythmic regularity disappears as the onset-to-onset interval approaches 100 msec (10 elements/sec), and this is about the fastest tapping rate that individuals can maintain when synchronizing to a sequence of sounds. Beyond this rate, individuals are even unable to determine if they are synchronized to the sounds. (The 100 msec interval is also the minimum interval between musical notes). At the other extreme, as the onset-to-onset interval approaches 2–2.5 sec, the sense of regularity again disappears and the sounds appear to be discontinuous and isolated. That interval is about the longest that allows for successful synchronization. Synchronization to flashing lights is far more restricted at the faster presentation rates. Individuals are unable to synchronize when the presentation rate is faster than 2/sec, although the limitations at the slower rates are the same as those to audition.

While these limits on rhythmic perceiving and synchronizing are quite broad, the onset-to-onset intervals between 400 and 600 msec (roughly two elements/sec) seem natural and preferred for rhythms. Individuals spontaneously tap at these rates and can more accurately reproduce intervals in this range.

In what follows, we will initially concentrate on the rhythms in Western European music with a single regular beat. These follow a strong-weak beat or strong-weak-weak beat sequence that allows listeners to tap or march along in synchrony. It is a simple rhythmic structure, but it allows listeners to keep in time when the rhythm speeds up or slows down for expressive purposes. The 1:2 and 1:3 timing ratios seem natural, easy to produce, and are preferred across a wide range of musical cultures. The steady beats allow individuals to synchronize in groups; music that is performed solo often does not have a steady beat. Moreover, many movements and motions follow the same strong-weak alternation, so that musical rhythms may be just one example of a more fundamental pattern. Next we will study polyrhythms that contain two or more rhythmic lines so that listeners can select one of them to act as the beat. Finally, we will briefly consider non-regular rhythms that are found in many musical traditions.

4.2.1 Isochronous Pulse Trains

The simplest rhythmic stimulus consists of a series of equally spaced, identical sounds. Early work found that even without physical differences, the sounds grouped into units of two, three, or four depending on their tempo. Listeners are able to vary their tapping rate, at faster rates creating groups of four or eight sounds, or at slower rates creating groups of two or even one element. Even without a physical rhythm, the phenomenal rhythm exists at several levels. In all cases, the first note of the group appears stronger and accented, and the intervals within a group appear shorter than the intervals between the final element of one group and the initial element of the next group as shown in Fig. 4.1A, B & C (Bolton, 1894; Woodrow, 1909). These were termed "subjective rhythms" due to the strong perceptual rhythm in spite of no differences among the elements. (In reality, all rhythms are subjective).

Subsequent research, although keeping the onset-to-onset intervals identical, made the sequences more complex by varying the physical characteristics of some of the elements. If every second, third, or fourth element is made louder, then that louder element is perceived to be the first element of the groups (Fig. 4.1D & E). The intervals within the groups are perceived to be shorter than the intervals between the final softer note of the group and the initial louder sound of the next group. In similar fashion, if every second, third, or fourth sound is increased in duration with the offset-to-onset interval remaining constant, then the longer duration sound is perceived to be the last sound of the groups (Fig. 4.1F & G). Again, the intervals within groups are perceived to be shorter than the intervals between the final longer duration sound of one group and the initial shorter sound of the next group. Finally, if one of the onset-to-onset intervals is physically lengthened, the groups are created by that interval

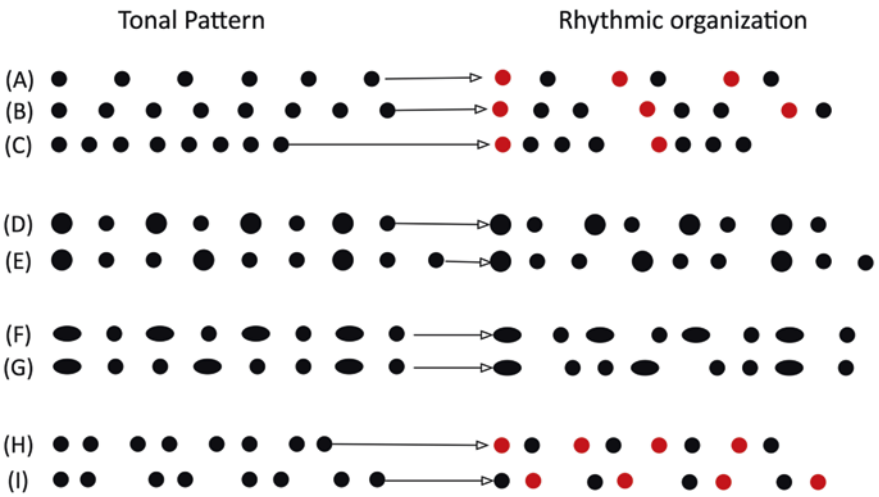


Fig. 4.1 The perception of simple rhythms. Isochronous sequences of identical elements are organized into groups of two, three, or four depending on the rate of the elements (A, B, & C). The initial elements are perceived as accented shown in red. Moreover, the elements seem to be grouped in time so that the interval between the last element in the group and the first element in the next group is perceived to be longer than intervals within the group. If every second or third element is louder, then that element is perceived to start the group (D & E), but if those elements increase in duration, they are perceived as the last element in the group (F & G). As the interval between groups of elements increases, the accent shifts from the first element (red in H) to the last one (red in I)

Sound Files 4.1: Rhythmic grouping created by intensity, duration, and inter-element interval differences

(Fig. 4.1H & I). If the difference between the two onset intervals is small, the first element of the group is perceived to be louder, while if the difference is large, the last element of the group is perceived to be louder.

Although these outcomes seem too simple to explain the complex rhythms of music and speech, I think they form the backbone for understanding those rhythms in much the same ways that Gestalt principles form the backbone for understanding complex visual scenes. In fact, the rhythmic principles are quite analogous to the Gestalt ones. It is so commonplace that it is easy to miss that people spontaneously organize sounds (and visual or tactual) elements into rhythmic groups, which become figures against a homogeneous background of time. The rhythm creates a sound source. In this process, grouping reorganizes the entire sequence. If the intensity of some sounds is increased, it brings about a change in the perception of the intervals between the sounds. Conversely, if the interval itself is changed, it brings about a change in the perception of the intensity of the sounds. Rhythm is relative timing, as the timing and accentuation of any sound is determined relative to the timing and accentuation of all the other sounds; often there is no simple correspondence between the acoustic signal and the perceived rhythm. The perception of an accented sound may be the result of an acoustic change at a distant point in that sequence.

4.2.2 Beats and Meters

As described above, listeners hear stronger accented elements even in isochronous pulse trains. The accents tend to occur at every two, three, or four elements and, at least for Western listeners, occur at equal intervals. Moreover, listeners can alter their tapping rate at will, tapping rapidly every two elements, more slowly every four elements, and slowest every eight elements. (Within limits, people can also vary the limb used to tap the rhythm). This ability to hear a sequence at different time intervals yields a hierarchical representation that results in the perception of beats and meters.

Beats refer to that sense of equally spaced accented elements and usually occur at the onset of tonal elements. However, they can also be felt without an actual element. Meter is the sense of a periodic sequence of subjectively stronger and weaker beats that characterize music; it arises from the beats at different time scales. Meter creates a sense of temporal regularity that tells the listener when to expect the next stronger beat. Dancing, tapping, and marching are all timed to the meter. There are faster meters at lower levels in the hierarchy in which beats occur on every element, or on every two or three elements, and slower meters at higher levels in which beats occur every four, six, or eight elements. Typically, lower pitches are used to play the slower beats, and higher pitches are used to play the faster beats. At each level the beats are presumed to be equal so that it is the addition of the beats across the levels that give rise to the strength of each element. Without the faster beats, the slower beats are simply recurring accents, and without the slower beats, the faster ones also are simply recurring accents. Meter is the mental construct that fuses the beats at different levels together and organizes time.

Any theory of rhythm must explain why some elements become strong beats and why some become weak beats in order to account for the sense of meter. Furthermore the theory must account for the grouping of the beats, which strong beats and weak beats go together. To untangle the two components of meter, beat strength and grouping, we start by considering the grouping of the elements and how that feeds into the perceived location of the beats. The grouping process brings about the sense of strong and weak accents, which are integrated to form the beat. The accents may not occur at the equal intervals of the beat so that there is a constant back and forth between the group and beat organizations.

4.2.3 The Grouping Hierarchy

Many factors determine the partitioning of a sequence of elements into groups. In nearly all instances, each element is placed in one and only one group (as is true for all kinds of perceptual organizations as detailed in the previous chapters), although in many real cases a single element may seem to bridge two groups.

1. First, there is the principle common to all perceptual and cognitive processes to place elements into equally sized groups and to avoid groups of just one element. For the isochronous sequences, the first element of groups of two to four elements is perceived to be accented (Sound Files 4.1A, B, C).
2. Second, for non-isochronous sequences composed of identical elements, the longer onset-to-onset intervals between adjacent elements split the elements into groups. Povel and Essens (1985) and Povel and Okkerman (1981) propose several rules to account for the perceived accents in such sequences: (a) accents would occur on isolated elements; (b) accents would occur on the first element of a two-element group if the temporal interval was short, but on the second element if the interval was longer (as in Fig. 4.1H versus 4.1I and Sound File 4.1H versus 4.1I); and (c) the first and last elements would be accented if the groups consisted of three or more elements. Consider the following 16-unit rhythm (the rhythm would be recycled at the end). The x's represent sound and the dots represent silences equal in duration to the sounds. Applying these rules to the same sequence yields the following accents shown in capitals:

   PHYSICAL RHYTHM x x x . . x . . x x . . . . x x . . . .
   GROUPING RHYTHM X x X . . X . . X x . . . . x X . . . .

3. Third, for isochronous and non-isochronous sequences composed of differing elements, groups tend to occur at points of change. Changes in timbre, frequency range, intensity, and/or tempo may act to slice the sequence into discrete groups in a variety of musical genres, e.g., Turkish makam music investigated by (Mungan, Yazici, & Kaya, 2017). As described in Fig. 4.1, in sequences composed of different intensities, louder elements begin groups and the interval between the prior softer and louder next element is heard as longer than intervals among the softer elements (Fig. 4.1D & E and Sound Files 4.1D & E). For sequences composed of differing frequencies, listeners expect adjacent tones to be close in frequency, so that large frequency shifts lead to the perception of group boundaries at the shift (Fig. 4.2A). In similar fashion, reversals in direction of the frequency changes lead to the perception of a group boundary at the turn-around (Fig. 4.2B). Another general rule is that groups are formed from repeating sequences, regardless of the nature of the repeats (Fig. 4.2C). Every musical culture contains repeating segments and the repetitions allow the listener to look forward in the music. Finally, one more cue to a group boundary is the increase in duration of a tone at the end of phrases (Fig. 4.2D).

Vimeo https://vimeo.com/120517523, Video discussing the importance of melodic repetition.

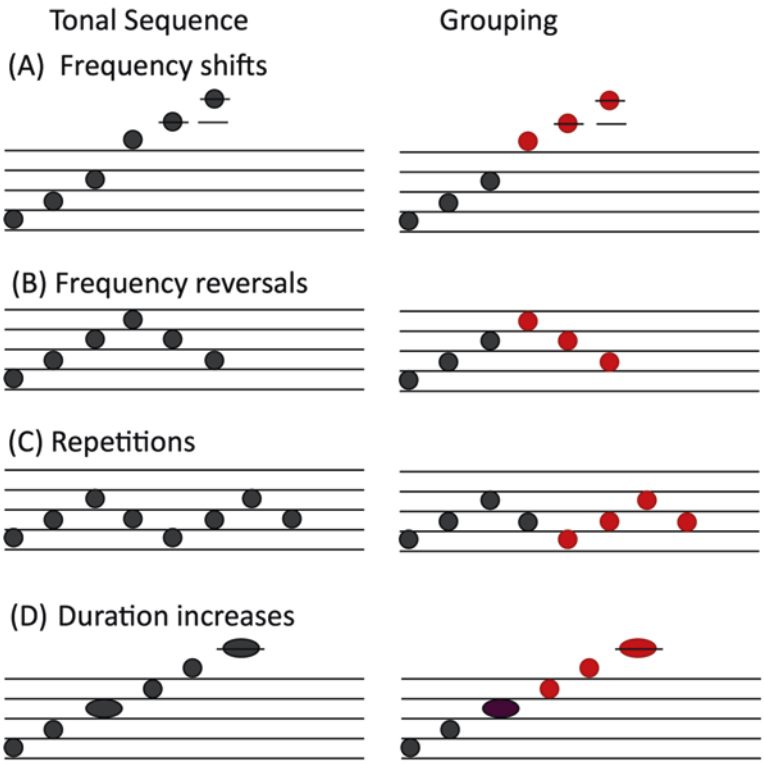


Fig. 4.2 Frequency shifts, frequency reversals, repetitions, and duration lengthening are four properties of tonal sequences that lead to the grouping of the tones (black versus red notes)

Sound Files 4.2: Tonal grouping due to frequency shifts and reversals, repetitions, and tonal duration lengthening as shown in Fig. 4.2A–D

In non-musical settings, repetition can aid the identification and localization of the source. In many species, repeating sequences help maintain contact and cohesion among individuals within a social group and may even serve to identify the social group itself (Zwamborn & Whitehead, 2017). For example, when male humpback whales congregate to mate in warm water, each whale in the region sings pretty much the same song. The song does change each season, but the change is contagious and each whale soon begins to sing the same new song. If you put a hydrophone in the water there is a cacophony of whale sounds, so that if a male is going to successfully attract a female its song must be easy to locate. The repetition reduces the masking of the song due to the background noise (Brumm & Slater, 2006). The first factor that helps identify the same whale and its location is the repetition of a set of sounds, termed a phrase. The second factor is that the sounds in successive phrases are nearly identical, so that the song evolves slowly. An example is shown in Fig. 4.3. The first phrase is the simple alternation RA and that phrase may be repeated 10–20 times. Invariably, the next phrase RRCCC doubles the same R sound and combines it with three

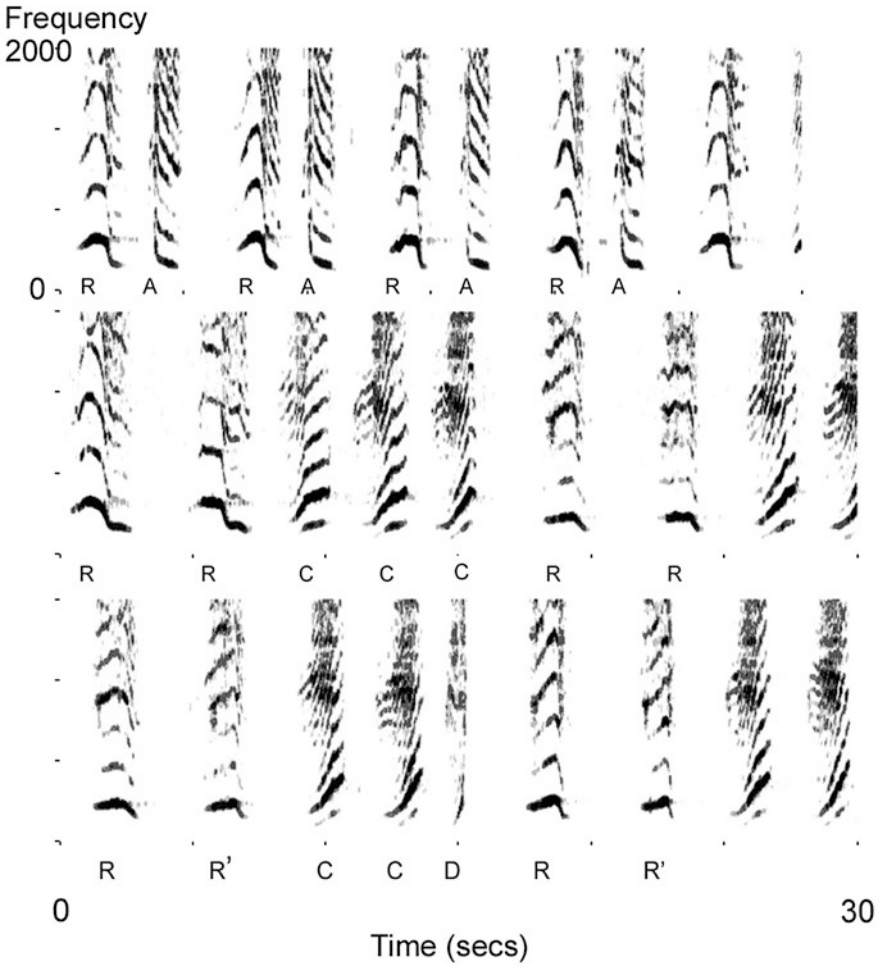


Fig. 4.3 Three successive phrases from a humpback whale recorded in Hawaii during 1994. (Adapted from Handel et al., 2012)

Sound Files 4.3: Three phrases in the humpback whale song (Fig. 4.3)

repetitions of the C sound. This phrase is repeated two or three times and it then evolves into the RR'CCD in which the second R sound is slightly modified and the third C sound is replaced by the D sound. The same pattern occurs in all parts of the song: each phrase is repeated and successive phrases substitute new sounds for some of the old ones (Handel, Todd, & Zoidis, 2012). (I think this minimizes the memory load for the whale, but that is another story.) Sound files 4.3A,B, and C give an example of each phrase.

Given the many factors that affect grouping, it is no wonder that although some sequences are invariably grouped in the same way, others lead to a wide variety of outcomes. Nonetheless, whatever the resulting grouping, it should meet several criteria: (1) only adjacent notes can be placed in a group; (2) no

note can be skipped; and (3) with rare exceptions, any note can only be placed in a single group. In similar fashion, in the hierarchical organization of the groups, any longer group can be split at lower levels, but only adjacent groups can be combined at higher levels.

4.2.4 The Meter Hierarchy

4.2.4.1 Single Rhythm Lines

The splitting of the element sequence into groups creates the pattern of perceived accents and timing. The construction of the meter can be thought as fitting the regular timing of the strong-weak or strong-weak-weak beats to the grouping. The "best" meter would happen when the periodic beats coincide with the onsets of those accented elements. Thus, the beats should occur on the more intense elements, the longer duration elements, on the initial element of large frequency shifts or reversals, and on the initial element of groups of repeating melodic phrases. Needless to say, it is all but impossible to satisfy all of these preferences. Fitting a meter is nearly always a compromise.

A useful way of imagining fitting the metric hierarchy to the grouping pattern of the sequence of elements is by using a hierarchical timing grid based on a series of isochronous dots so that each represents beat strength. At the highest level (the fastest) of the hierarchy, each dot represents the timing of the shortest element, for example, an eighth or sixteenth note in musical terms, so that at this level each note or silence has equal beat strength. At the next level down, the isochronous dots are aligned with every second or third beat of the lowest level. To create a strong-weak duple meter every second beat is represented by one timing dot to represent its added metric importance. To create a strong-weak-weak triple meter, every third beat is represented by a timing dot to represent its added importance. At the third level down, the isochronous dots are aligned to represent every fourth (duple) or sixth beat (triple). The timing dots at this level further show the relative strength of each beat. At lower levels, the isochronous timing dots again and again split the dots at higher levels in half to give the overall strength of each beat. The depth of the stack of dots indicates the strength of the beat.

We can start with nine-element patterns embedded in a 12-element repeating unit depicted in Fig. 4.4. Pattern A is xxx-xxx-xxx- where x represents an element and a dash represents a silent interval equal to an element in duration is shown in (A). (To be more specific, if we present the pattern at a comfortable rate of four elements per sec (i.e., 250 msec between element onsets), an element might be 125 msec long with a silent offset-to-onset interval between adjacent elements of 125 msec, to create the full 250 msec onset-to-onset interval between elements. The silent interval would therefore be 375 msec, composed of the 125 msec silent offset-to-onset interval following an element plus the 250 msec silent onset-to-onset interval). For this pattern, the stronger accents would fall on the first element of each run of three elements, and a weaker accent would fall on the last element of each run of three elements **X**x**X**-**X**x**X**-**X**x**X**-. Given these accents,

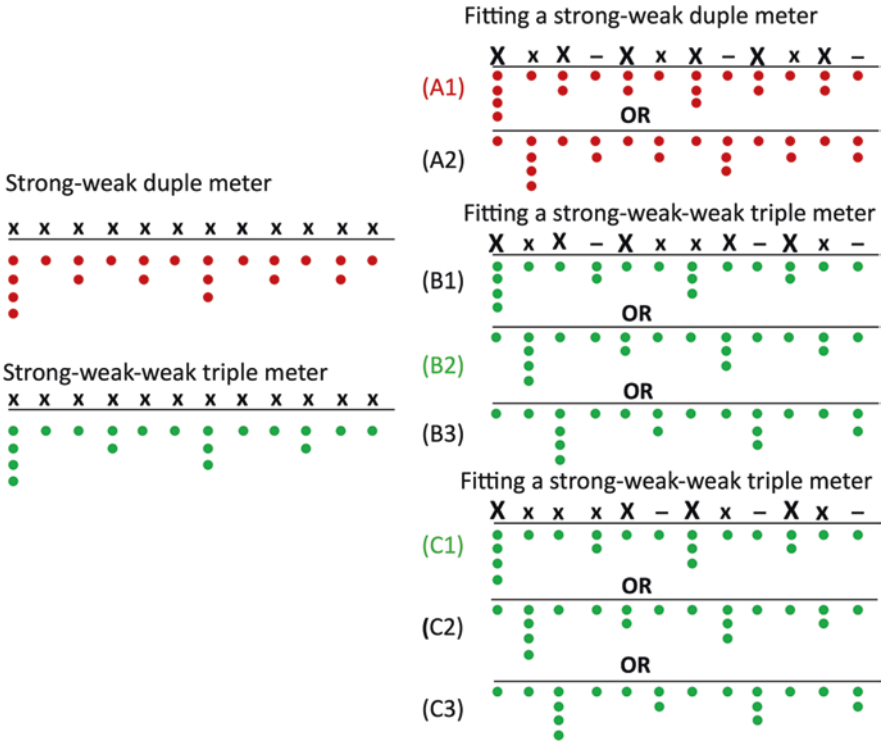


Fig. 4.4 The beat hierarchy of strong-weak (red) and strong-weak-weak meters (green) is illustrated for repeating 12 element patterns. The two alternative strong-weak beat meters are shown for the nine sounds plus three silences in a 12-element pattern xxx-xxx-xxx- in (A1) and (A2). The three alternative strong-weak-weak meters are shown for xxx-xxxx-xx- in (B1, B2, & B3) and xxxxx-xx-xx- in (C1, C2, & C3). In all cases, the goal is to align the strongest beats shown as the depth of the hierarchy with the first element of a run of two elements and the first and last element of a longer run. The best fits are in color: (A1 red), (B2 green), and (C1 green)

the best-fitting meter would be a strong-weak duple meter starting on the initial element of each run (see A1 in Fig. 4.4). This positioning would optimize the next level down in the hierarchy because each strong-weak-strong-weak beat would consist of the run of three elements ending on a silent interval (xxx-). If the duple meter was shifted one element to the right (see A2 in Fig. 4.4), half of the strong-weak duples would begin on a silent interval. A triple strong-weak-weak meter will not work because wherever the meter starts the beats will fall at least once on a silent interval.

Even moving the position of one element can drastically change the meter. Consider Pattern B, xxx-xxxx-xx-, that still contains nine elements embedded in a 12-element repeating unit. The accents would be XxX-XxxX-Xx-. The optimal duple meter found above simply does not work as well here because the fifth strong beat would fall on a silent interval. Actually, there is no possible

shift for the strong-weak meter that does not result in a strong beat falling on a silent interval. A triple strong-weak-weak meter can result in the strong beats falling on actual elements if the meter starts at the second element of the initial run of three elements (see B2 in Fig. 4.4). However, this meter is not optimal for two reasons; first, the beat falls on the middle element not accented in a run of three elements, and second, at the next hierarchical level the six-element meter breaks the runs apart.

Another one element move can shift the best meter again. For Pattern C, xxxxx-xx-xx-, the accents would be XxxxX-Xx-Xx-. The best meter placement is again a triple meter starting on the initial element of the pattern (C1). No other triple or duple meter works as well in matching the accents to the beats. The amplitude × time representations of the rhythms in Fig. 4.4 are pictured in Fig. 4.5, and the sound files are found in Sound File 4.5.

The meter controls the perception of the entire pattern. Povel and Essens (1985) presented single-note patterns and accompanied the pattern either with a drumbeat every three elements (triple meter as in Sound File 4.6D1) or every four elements (duple meter as in Sound File 4.6D2). Listeners were unable to recognize that the patterns were identical when accompanied by the different metric drumbeats. What is important to understand is that the fitting of a meter to a grouping of the elements is usually a compromise between the best grouping and the best meter. It is rare that they perfectly correspond and it may take a while for the meter to emerge (Fig. 4.6).

We can see the fit between the grouping and meter structures for simple songs such as Stephen Foster's "Way Down Upon The Swanee River." In the first phrase, the shortest note is an eighth note, so that the lowest (fastest) level of the metric hierarchy should equal that note. With the exception of the dotted quarter note equal to three eighth notes, the remaining notes are all even multiples of the eighth note. The grouping structure would be based on the differences in duration of the notes, the pattern of frequency changes, and the clear repetitions of the groups. The grouping is perfectly aligned with a duple meter in which there are four strong-weak units.

"Down in the Valley" has an unusual time signature of 9/8, indicating that an eighth note gets one beat and there are nine beats per measure. The lowest level of the meter hierarchy would be based on the individual eighth notes, the next higher on three eighth notes, lining up with the dotted quarter notes, and the highest level would be based on nine eighth notes, lining up with the initial dotted quarter note of the first full measure, the dotted half note in the second and fourth measures, and the first dotted quarter note in the third measure. Another factor that influences the meter is the position of the important tonal elements. In extensive analyses of classical composers, the important tonal elements (i.e., the tonic and fifth) occur at the strongest metric points. This correspondence probably strengthens the musical expectancies that convey musical rhythms and emotions (Prince & Schmuckler, 2014) (Fig. 4.7).

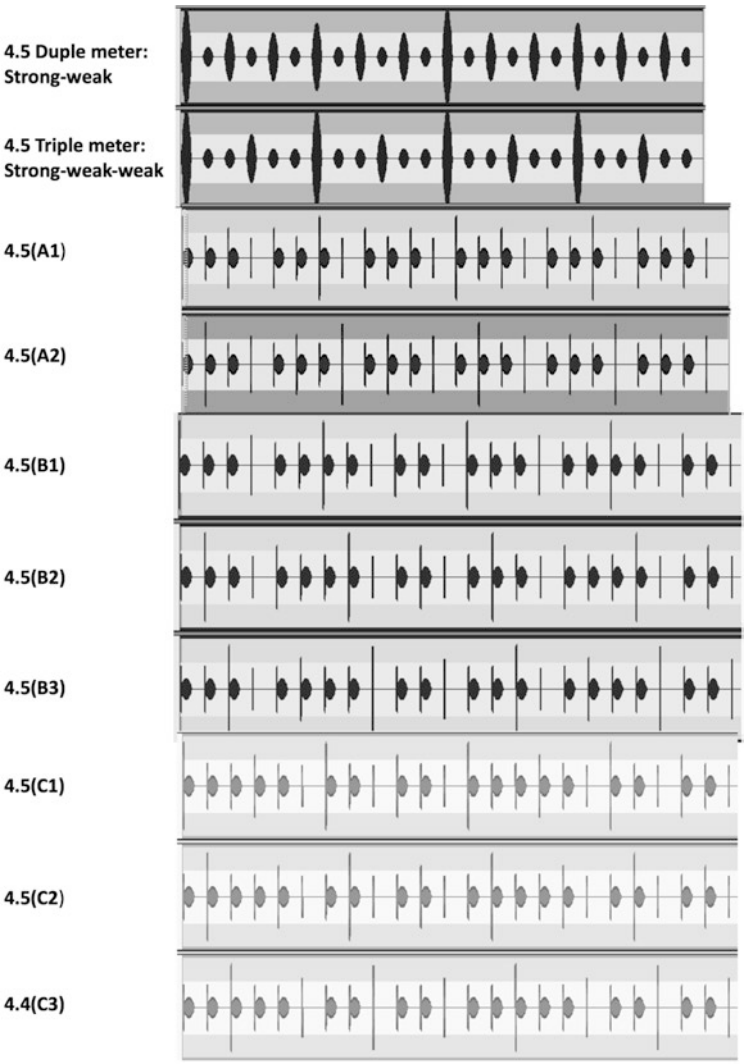


Fig. 4.5 Amplitude by time representation of the sound files for the rhythms displayed in Fig. 4.4. The presentation rate is 4/sec. Each tone is 125 msec and each beat (the thin vertical line) is 10 msec. The strength of the accents for the duple and triple meters is indicated by the amplitude of each sound (its height). The strength of each beat is indicated by the length of the beat line

Sound Files 4.5: Beats in strong-weak duple meters and strong-weak-weak triple meters for the sequences shown in Fig. 4.5

4.2.4.2 Multiple Rhythm Lines

In spite of the obvious strengths, there are clear limitations to this hierarchical analysis of rhythm. First, hierarchical analyses are restricted to one rhythmic line, so they cannot cope with thick polyphonic passages in which instruments are playing at different tempos and differing groupings. Second, these hierarchies

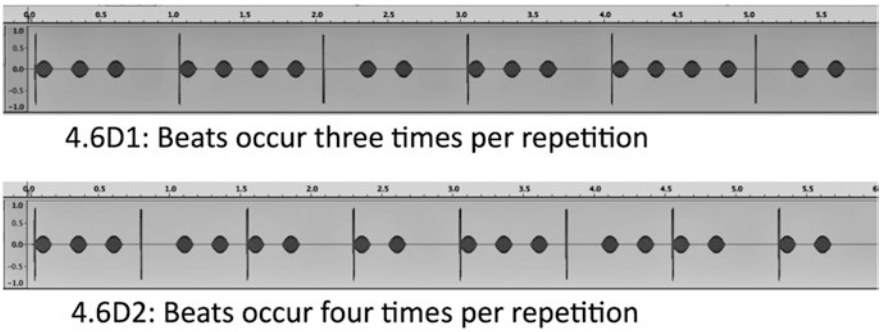


Fig. 4.6 The amplitude × time representation of the alternate beat structures. Sound Files 4.6D1 and 4.6D2 illustrate how incompatible meters can make the same pattern xxx-xxxx-xx- appear to be different rhythms as found by Povel and Essens (1985). It is also possible to hear the effect of shifting the meter by comparing 4.5B1, 4.5B2, and 4.5B3 to one another and 4.5C1, 4.5C2, and 4.5C3 to one another

Sound Files 4.6: Alternative beat structures shown in Fig. 4.6D1 and D2 determine the perception of the pattern

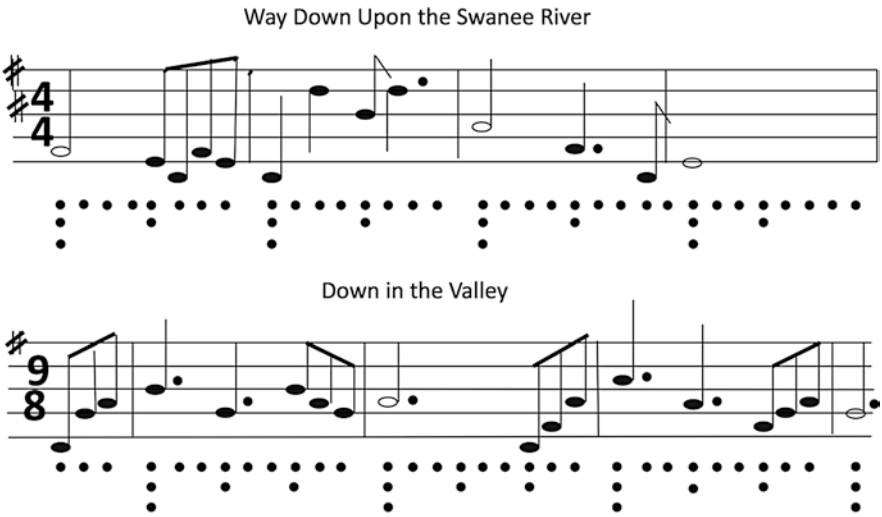


Fig. 4.7 The initial phrases of "Way down on the Swanee River" and "Down in the Valley." The number of dots under each beat indicates the importance of each beat. In these phrases, the important beats always coincide with the onset of a note, but that does not always occur in other songs. (Adapted from Fox & Weissman, 2007)

work for Western music with a strict periodic meter, but will not work for other musical genres with non-metrical rhythms in which the beat is not periodic.

4.2.4.2.1 Polyrhythms

Our argument above is that the meter, beat, and accent for a single rhythmic line arises from the interplay of rhythms at different levels. In essence, the rhythms at slower levels create the accentuation for the rhythm at faster levels.

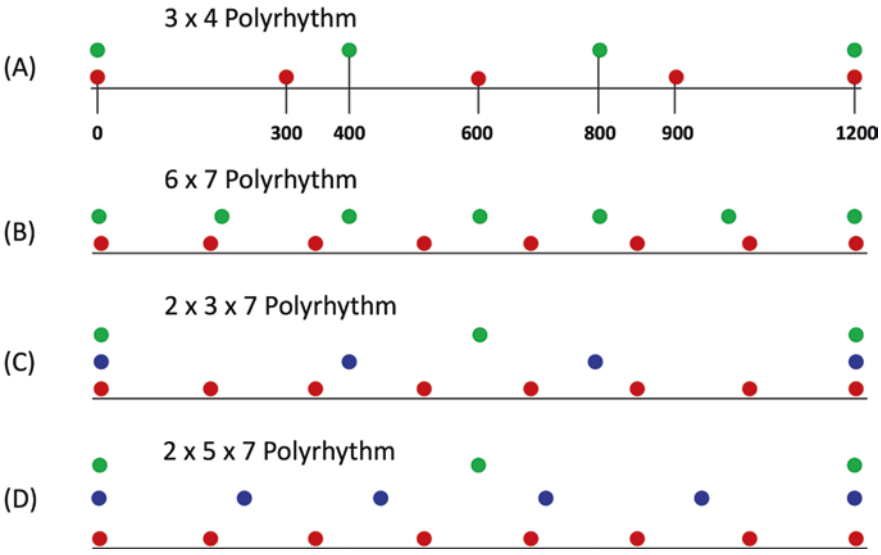


Fig. 4.8 In (A), the timing of three-pulse and four-pulse polyrhythms is shown for each of the pulse trains. In (B), (C), and (D), the two-pulse 6 × 7 polyrhythms, and the three-pulse 2 × 3 × 7 and 2 × 5 × 7 polyrhythms are shown. The pulse trains are synchronous at the first note of each cycle

An alternative approach is to create polyrhythms composed of two or three isochronous pulse trains in order to understand how the interplay of the levels represented by each pulse train determines the overall rhythm (Handel, 1984). Polyrhythms are composed of two or more isochronous pulse trains with identical elements that have conflicting tempos, for example, two per pattern repetition versus three per repetition (i.e., two tones per measure versus three tones per measure) but not two per repetition versus four per repetition.

An example of a 3 × 4 polyrhythm is shown in Fig. 4.8A. Based on a repetition rate of 1200 msec, the onset-to-onset interval for the three-pulse rhythm would be 400 msec, while the onset-to-onset interval for the four-pulse rhythm would be 300 msec. There would be one point at which the pulses from each rhythm would coincide; the order of the pulses alternates and the onset-to-onset intervals between the pulses from each rhythm would be either 200 msec or 100 msec. In polyrhythms with conflicting ratios, the onset timing between elements changes. For more complex polyrhythms like 6 × 7 (Fig. 4.8B), the number of different onset-to-onset intervals between the rhythmic pulses increases and this increase is even greater for three-pulse polyrhythms such as 2 × 3 × 7 (Fig. 4.8C) and 2 × 5 × 7 (Fig. 4.8D). Schematic illustrations of 2- and 3-pulse polyrhythms with different frequency components are shown in Fig. 4.9.

These sorts of polyrhythms can be heard in several ways: (a) one of the two or three pulse trains is heard as the meter; listeners would tap the elements of only one of the pulse trains and the elements of the other pulse train would fall off the meter; (b) the meter of the polyrhythm would occur only on the synchro-

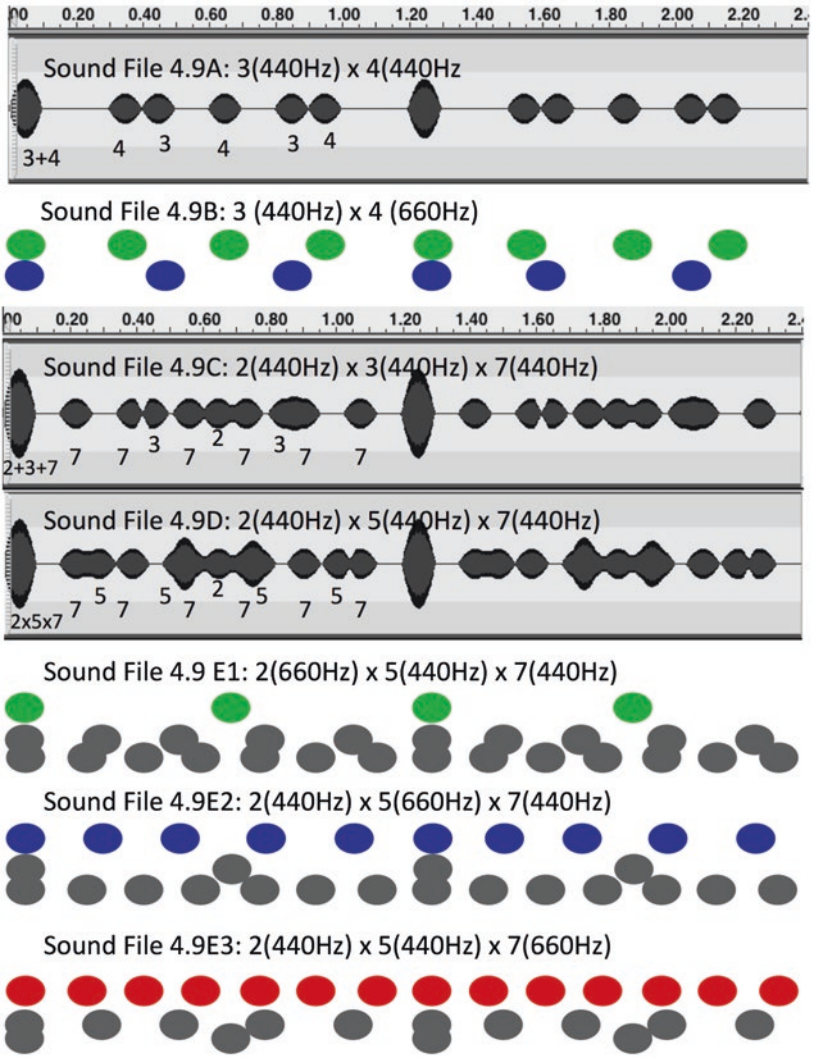


Fig. 4.9 The first panel shows the amplitudes across time of the identical notes in the 3 × 4 polyrhythm. The duration and amplitude of the notes is evident and the higher amplitude of the synchronous first note of each repetition is due to the sum of the two pulses. The next panel is a schematic of the two pulses at different frequencies. The third and fourth panels show the notes of the 2 × 3 × 7 and 2 × 5 × 7 polyrhythms with all notes at the same pitch. The final three panels illustrate schematically the timing of the pulses of the 2 × 5 × 7 polyrhythm when one of the pulses is presented at a different frequency

Sound Files 4.9: Two- and three-pulse polyrhythms pictured in Fig. 4.9

nous elements so that there would be one beat per measure; (c) the two pulse trains could be integrated into a *cross-rhythm* and the listener would tap each element, six for the 3 × 4 polyrhythm (not seven because of the synchronous element). If the polyrhythm were composed of three pulse trains, for example,

2 × 3 × 7, the possible ways of hearing the rhythm would multiply. In addition to the three ways listed above, there would be three ways in which listeners tap along with two of the three pulse trains with the third pulse train heard separately, 2 × 3 versus 7, 2 × 7 versus 3, and 3 × 7 versus 2. One might expect that as the complexity of the polyrhythm increased there would be a tendency to find a simpler meter based on one of the pulse trains or based on the repetition of the polyrhythm at the point where the three elements synchronize.

In experiments investigating the perception of polyrhythms, the task of the listeners was simply to tap along in any way they wished as if they were tapping on the steering wheel at a red light. Although there were differences among listeners, there are some consistent trends:

A) The fundamental factor determining the interpretation of the polyrhythms was the timing among the elements. Remember from the beginning of this chapter the preferred tapping rate occurred when the onset-to-onset interval was between 200 and 800 msec. For polyrhythms composed of two pulse trains, if the interval between the onsets of the elements of only one pulse train falls in the preferred tempo zone of 200–800 msec, there was an extremely strong tendency to tap along with that pulse train. If the intervals between the onsets of the elements for both pulse trains were within that 200–800 msec region, then the low pitch or more staccato pulse train was preferred. If the overall presentation rate of the polyrhythm was so slow that the onsets between the elements of both pulse trains are longer than 800 msec, most listeners shift to tapping every element of both pulse trains. Conversely, if the presentation rate is so fast that the onset intervals of both pulse trains become less than 200 msec, then most listeners switch to tapping once each repetition, on the synchronous elements.

B) The second factor is based on the configuration of the specific pulses within a polyrhythm. Although the outcomes for the 2 × 3 and 2 × 5 polyrhythms are similar, if those polyrhythms are combined with a pulse train in which the elements occur seven times per measure, for example, 2 × 3 × 7 and 2 × 5 × 7, the outcomes become quite different. For the 2 × 3 × 7 polyrhythm, listeners either tapped to the 2 × 3 cross-rhythm or the seven-pulse trains; they rarely tapped separately to the two- or three-pulse train. The seven-pulse train emphasized the similarity between the two- and three-pulse trains so that neither one of them was tapped alone. For the 2 × 5 × 7 polyrhythm, one group of listeners tapped the seven-pulse train at the slower presentation rates and the two-pulse train at the faster rates, and a second group tapped the two-pulse train at all rates. The seven-pulse train pulled the five-pulse train from the two-pulse train so that tapping to the two-pulse train was predominant and 2 × 5 cross-rhythms did not occur. If one of the three-pulse trains was a different frequency, the tendency was to tap to the cross-rhythm of the two-pulse trains with the same frequency at slower tempos, and tap to the different frequency pulse train at the faster tempos.

To summarize, the perception of the meter of a polyrhythm is contextual with respect to the timing of each pulse train, the configuration of the polyrhythm, and the frequency and intensity of the elements making up each pulse train. Each factor influences the rhythmic interpretation; its effect depends on the values of all the other factors. There is no single rhythm in a temporal pattern so that there can be no single stimulus factor that determines the perceived rhythm.

4.2.4.2.2 Non-metric Rhythms

Even though Western music is almost exclusively metric with equally spaced beats, other cultures make use of complex meters in which the bests do not occur at equal intervals (Fig. 4.10). For example, in Turkish music, common complex meters are five or seven beats broken into 2 + 3 or 2 + 2 + 3 beats (Dave Brubeck makes use of these meters in "Take Five" and "Blue Rondo a la Turk"). It is surprising that exclusive listening to the simple Western meters seems to interfere with the perception of more complex non-metric rhythms. For example, six-month-old Western infants can detect meter changes in both simple and complex Balkan meters, but 12-month-old Western infants cannot detect the same changes in the complex Balkan rhythms. At 12 months it is possible to reverse this loss by a short period of exposure to complex Balkan music, but the same period of exposure for adults does not reverse the loss (Hannon & Trehub, 2005). Recently, Kalender, Trehub, and Schellenberg (2013) found that the ability of adults to detect changes in complex Turkish meters did not require experience with specifically Turkish rhythms. As long as the adult had listened even sporadically to music that did

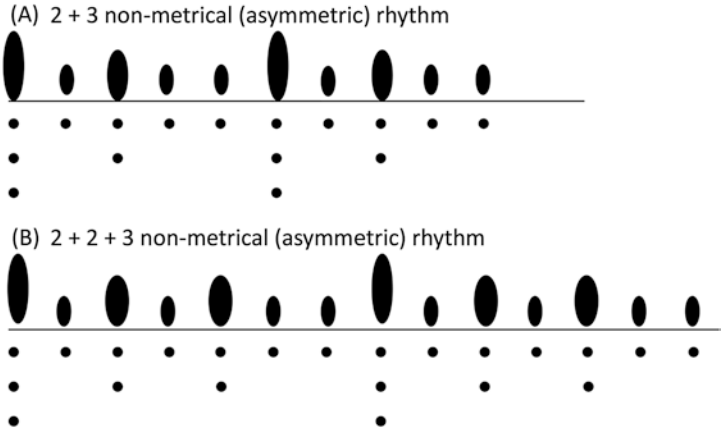


Fig. 4.10 Non-metric rhythms contain equally timed notes, but the beats do not fall at equal intervals. (Non-metric rhythms are sometimes termed asymmetric rhythms). (A) The rhythm 2 + 3 is shown in terms of the actual sounds (heard in Sound File 4.6A). The strong beats occur on the first and third tones. (B) The rhythm 2 + 2 + 3 is shown. The strong beats occur on the first, third, and fifth tones

Sound Files 4.10: The 2+3 and 2+2+3 non-metric rhythms drawn in Fig. 4.10

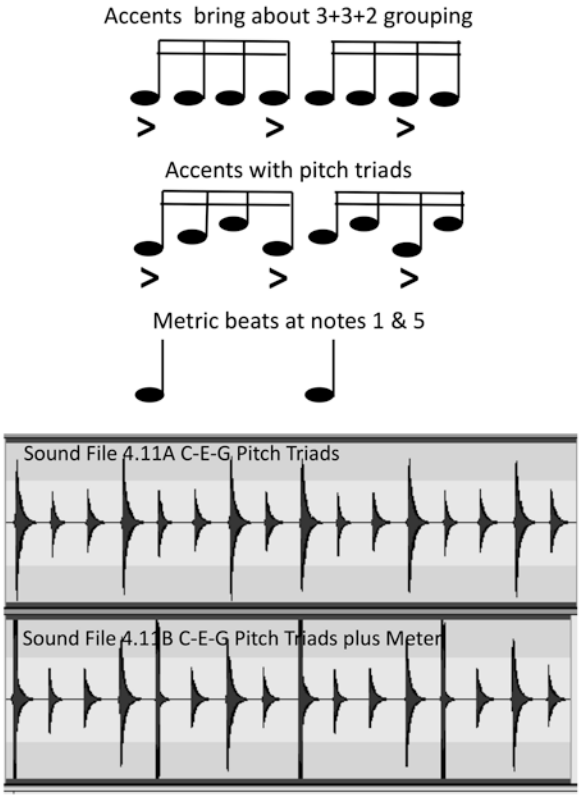


Fig. 4.11 Accents characteristic of bluegrass and ragtime music. If metric beats are added, the rhythm seems to shift over time

Sound Files 4.11: Ragtime rhythms based on the major triad and duple meters shown in Fig. 4.11

make use of non-metric rhythms, for example, Indian music, they were able to detect the changes. It seems that sensitivity to non-metrical rhythms is hard to lose, but then again it is hard to regain.

A second way to create a non-metric rhythm is to accent off the beat. Start with a simple eight-note measure in which the beats would normally fall on the first and fifth note. If the first, fourth, and seventh notes are accented, then a 3 + 3 + 2 grouping is created within a standard eight-note duple meter as illustrated in Fig. 4.11. If the grouping is played as pitch triads it creates a rhythmic phrase that is typical of ragtime and bluegrass music. Furthermore, if this rhythmic phrase is combined with a regular beat on notes one and five, it results in a sense of a shifting rhythmic structure (Fig. 4.11).

4.2.5 Beats, Embodied Rhythms, and Relative Movements

At the beginning of this chapter, two seemingly contrasting conceptualizations of rhythm were presented. The first made use of classic Gestalt principles of grouping and the second, while far more vague, made body movements and

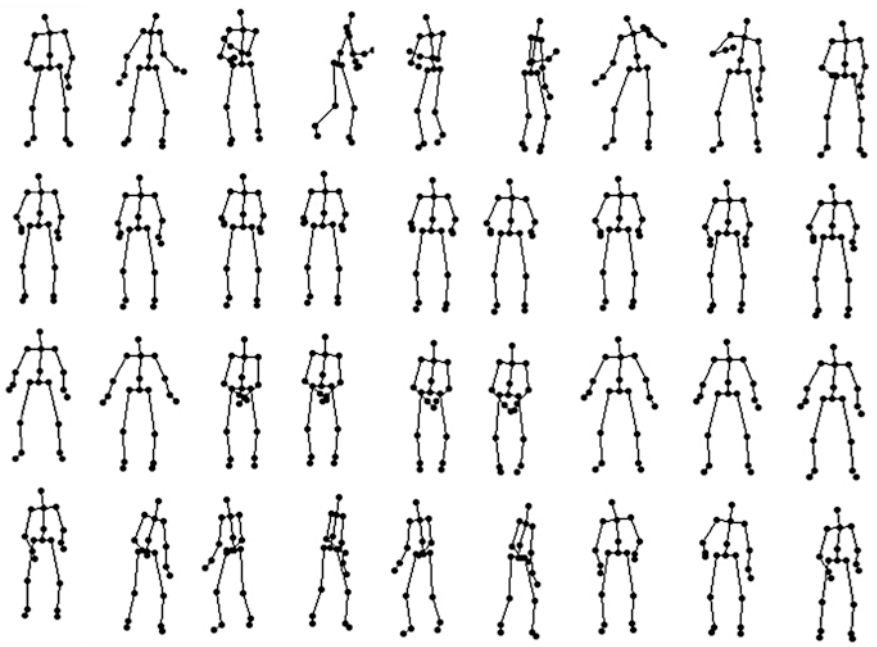
rhythm synonymous. The motor regions create the timing intervals used to pace the body movements. At this point, it is possible to synthesize two points of view and tie both to notions of relative motion perception due to Gunnar Johansson (Chap. 2). What is common to all is that rhythm occurs hierarchically at several levels. The Gestalt principles leading to beats and meters start at the fastest, highest level with each element receiving a beat. At the lower, slower levels, each second or third element receives a beat and the summation of those beats yields a two- or three-beat meter. In thinking about embodied rhythms, we can imagine that the overall dancing movements are built out of the slower motions of the body and torso, the faster movements of the arms and legs, and the still faster motions of the hands. Leman and Naveda (2010) have identified spatiotemporal reference frames, which they term "basic gestures," that match periods of the meter. These repetitive motions link musical properties to body movements. Thus, each level of the motions or gestures can be thought of as being analogous to each level of the beats. Johansson's work demonstrates that our perception of each level of the body motions is relative to the slower movements of other levels. If we did not perceive movements this way, then the rhythmic movements of the hands would seem random because they would be entangled in the movements of all the other body parts. (This is reminiscent of the general problem of auditory perception; namely partitioning the sound wave into coherent sources).

In Johansson's analyses the slower movements of the more massive body parts were subtracted from the faster motions of the lighter body parts in order to see the trajectories of those faster movements. Toiviainen, Luck, and Thompson (2010) mimicked Johansson's lighted dot figures, but instead of people walking or running, they had dancers move rhythmically to music at different tempos. Furthermore, instead of partitioning the motions of one body part from another, they partitioned the overall motion into different "timings" or periodic movements. For example, the slowest motion of a dancer could be a simple back-and-forth swaying. In addition, the dancer could also rotate the torso or bounce up and down at the knees, could also swing the arms at a faster rate, and could also wiggle fingers at even a faster rate. The dance consists of all these movements occurring simultaneously. The motion of the fingers would be the combination of the movement due to swaying, the movement due to the bouncing, the movement due to the swinging arms, and the movement due to the wiggling. The analyses used by Toiviainen et al. (2010) isolated the movements that occurred at the slowest rate, those that occurred at twice that rate, and those that occurred at four times that rate. Adding those movements together makes the original motion reappear. (This is the same way that the quality of a non-changing sound is analyzed, for example, a square wave (see Chap. 5). The amount of energy at the fundamental frequency and at multiples of the fundamental frequency is measured. Playing those frequencies at the derived amounts of energy reproduce the original sound).

A simulation of the various movement levels is shown in the following video. Each column represents a different dancer. The lowest row is the slowest motion and each higher row represents movement that is twice as fast. The top

row is the sum of all the movements (I am partial to the dancer on the far right). It is easy to see that the body parts with the highest mass and inertia move at the slowest rates and the lightest parts move at the highest speeds. The movements are "locked together" in timing to simple ratios of two or possibly three with different music, and I would argue that we perceive those movements relative to a common time base and relative to a common set of motions.

I thank Dr. Toiviainen for kindly providing the movie.



In this single frame, the dots on the dancer's limbs show the location of the lights used to record the movements of the dancers.

Note: The movie is found in the supplementary material

4.2.6 Do Animals Have Rhythm?

Obviously animals make rhythmic movements for locomotion, capturing prey, and so on. Furthermore, birds increase the rhythmic consistency of their song if threatened by interlopers. The research question has been whether animals can learn to synchronize to an external beat. Learning to move in time with the beat is not simple. Children can respond to different tempos far before they can synchronize. Children try to move in synchrony, but they need help to learn to do so. Four- and five-year-olds synchronize poorly, if at all. Synchronization success shows up only after ages seven or eight (Repp, 2013). Gaining the ability to abstract the beat and synchronize a body motion to it can be a long process. The simple rhythms found in infant songs and parental instructions start the process,

and the ability to attend to the levels of rhythm increase with all types of experience (for a broader view of rhythm in animals see Fitch (2018)).

It is still an open question whether animals can synchronize body motions to the beat. It has proven difficult to provide an answer for several reasons.

1. In constructing stimuli to test for synchronization in other species, humans make use of their own beat perceptions. Obviously, we cannot be certain that our perception matches that of other species (remember when discussing camouflage in Chap. 2, we imposed our sensitivities in evaluating the nature and effectiveness of the camouflage, not those of the predators). Probably all species can respond to simple auditory pulses, but most likely lack the cognitive capacity to abstract the beat in more complex sound environments. Moreover, as mentioned above, even human children need extensive practice to match the beat, and without similar training very few species are likely to achieve matching even if they are ultimately capable of doing so.
2. To demonstrate beat matching, the animal must have sufficient control of a limb or body part to effectively synchronize at the correct tempo. Humans can synchronize arm movements at faster tempos than body movements so that it is critical to choose a response that potentially could match the tempo of the beat. It is possible to train a seal to move its head in time with a sound, but not its flipper (Wilson & Cook, 2016).
3. Beat matching in the artificial experimental situations depends on the animal's motivations and attention span. Most of the demonstrations of beat matching involve species of parrots, highly social animals that bond and vocally mimic their owners. The popularity of the videos of Snowball (Patel, Iversen, Bregman, & Schulz, 2009) may have led researchers to limit their search for other species that may beat match, particularly since matching may not occur unless there is extensive training and practice. However, other social animals like dogs do not show beat matching even after extensive training.
4. Beat matching may be far more extensive in the animal's natural environment. Animals may synchronize their calls in social groups so that there are alternating and simultaneous utterances, that is, turn taking. Even bats produce duet-like social calls with one bat responding within 1/3 sec of the end of another's call (Vernes, 2017).

YouTube
Snowball™-Our Dancing Cockatoo (watch?v=N7IZmRnAo6s)
Snowball™-Another one Bites the Dust (watch?v=cJOZp2ZftCw)

Wilson and Cook (2016) make a strong argument that beat matching and other ways of synchronizing body movements to external stimuli is more widespread among animals than previously imagined. They argue that rhythmic

behavior depends on the voluntary control of motor behaviors and the learned coupling of those behaviors to sensory stimuli. If the four constraints listed above are overcome, the entrainment of motor behavior and rhythmic stimuli could be found in a wide range of animals.

Furthermore, there is strong evidence that animals are sensitive to rhythmic tempo. For example, elephant seals can distinguish among individuals based on the tempo and the spectrum of the pulses of their stereotyped mating calls. Each seal has a preferred tempo acquired when young and maintained through adulthood. Across the colony, the mating calls consist mainly of individual pulses presented at rates from roughly 1–3/sec. It is important to note that the preferred tempo for the dominant male can be any value within that range; it is not the fastest or slowest rate. Marthevon, Casey, Reichmuth, and Charriuer (2017) recorded the call of the dominant males in the colony, and during playback the peripheral males made avoidance movements. But, when the authors either increased or decreased the tempo without changing the spectrum, the peripheral males greatly reduced their avoidance and even disregarded the call. It would be interesting to determine how quickly the peripheral males learn the new relevant tempo when a new dominant male emerges.

4.3 Timing

Nearly everything in this book concerns grouping. For the purely static, spatial images in the first three chapters, generalized Gestalt laws attempt to predict which elements combine to form objects or surfaces. For the purely temporal rhythms so far in this chapter, similar Gestalt laws attempt to predict which sound units become the strong and weak beats, thereby creating the meter that brings about the overarching grouping of the sounds to form sources. Here we want to consider two related issues. First, we know that stream segregation and the preferred meter for polyrhythms changes as the tempo is varied. But up to this point we have tacitly assumed that a single rhythm would be perceived identically at all tempos. Here we will consider whether this is true.

Second, space and time are interlocked in several ways. The timing of sensations can determine the nature of the resulting spatial representation. As described below, tactile pulses presented at differing onset-stimulus intervals can lead to the perception of a "hopping" motion on the skin. This hopping or leaping perception has been termed sensory saltation and can also be found for visual and auditory presentation. By reversing our perspective, it is possible to use temporal order judgments to infer how different spatial representations are coordinated.

In addition, spatial configurations change over time. The perceptual problem is to determine the optimal matching of the elements in different rhythms or in each successive visual image, that is, the *correspondence* or constancy problem. Historically, the term perceptual constancy was used to describe the abstraction of visual and auditory objects in spite of diverse orientations, movements, sizes, octaves, timbres, and so on. Here we have used the term correspondence in two ways. First, correspondence refers to the grouping of parts of an interrupted object or source. Second, correspondence refers to the identity of an object or

source in a different context, making it equivalent to constancy. For example, the moving lights on joints are perceived as dancing figures in the demonstrations pioneered by Johansson (Johansson, 1973). The matching problem is simple here because the light points move in predictable arcs with small changes in each image. Tracking an individual firefly in a field of fireflies is far more difficult because the interval between flashes and the motion direction is more erratic. Even when the parts of a simple stimulus move coherently as in the Ternus configuration, the timing between images can change the perceived coherence.

4.3.1 Tempo and Rhythmic Organization

What we need to do is distinguish between two aspects of rhythmic organization. When we initially view a scene or listen to a string of sounds, our initial impressions are global and diffuse. The initial division of the visual array into closed contours is probably based on the principle of uniform connectedness: identical colors, textures, and motions are joined together. Further looking yields finer detail, figure-ground organization and three-dimensional objects. In similar fashion, the initial division of the auditory array into connected sequences would be based on pitch, timbre, and temporal proximity due to stream segregation. Further listening to one stream would differentiate the intervals between successive tones and possibly lead to the emergence of a stable metric. What is therefore common to both looking and listening is the initial splitting of the field into global parts sorted by common Gestalt principles. Following this split, each piece is further analyzed, eliminating alternative possibilities resulting in the final percept.

At the beginning of the chapter, we discussed that rhythms seem most natural between presentation rates of 7.5 elements/sec and 0.5 elements/sec. Beyond these rates, the sense of grouping and timing disappears. Two rhythms are drawn in Table 4.1. In both, there are five tones in a 16-element repeating pattern. The number of stars under each position indicates the metric strength of each position. In the first, all five tones occur at the stronger metric positions for a four-beat strong-weak- strong-weak meter, namely 1, 5, 9, 11, and 13. In the second, the tones occur at weaker metrical positions, namely 1, 4, 8, 10, and 12.

At the fastest tempo of 7.8 elements/sec, there is a sense of three groups, two of single tones and one of multiple tones, but only a very weak sense of repetitive timing. At the two slower tempos of 3.9 and 1.6 elements/sec, the 1-1-3- groups become predominant and there is a strong sense of rhythmic repetition, namely, two single tones followed by a group of three tones. At the two slowest speeds, 1.25 and 0.5 elements/sec, the grouping is far weaker and the elements seem unconnected. To me, there seems to be little difference in the grouping of the metric and non-metric versions.

Here we will consider if the perception of non-metric patterns is the same at different presentation rates. To do this, we present two rhythms at different tempos and ask listeners to determine if the rhythms are the same or different. The experimental strategy is simple: first, present the two rhythms at the same tempo to make sure that listeners correctly judge that the two rhythms are

Table 4.1 Two five-note rhythms constructed in a 16-element sequence. The number of stars indicates the metric strength of each position. The code 1-1-3 indicates that there are two groups of a single element followed by a group of three elements. The spacing between the second single element and the group of three elements determines whether the rhythm is metric or non-metric

	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*	*16*
Metric	X				X				X		X		X			
Non-metric	X			X				X		X		X				
	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
	*		*		*		*		*		*		*		*	
	*				*				*				*			
	*								*							

Sound Files 4.12: The metric rhythm x---x---x-x-x--- and non-metric rhythm x--x---x-x-x---- played at different tempos

Table 4.2 Two pairs of five-note rhythms used to test whether rhythms are perceived in the same way at different tempos

Rhythm	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*	*16*
1-1-1-2-	x			x			x			x		x				
1-2-2-	x			x		x				x		x				
1-1-2-1-	x			x			x		x			x				
1-1-3-	x			x				x		x		x				

Sound Files 4.13: Two pairs of non-metric rhythms shown in Table 4.2 played at different tempos

different, then present the rhythms at different tempos. If listeners can still correctly judge that the rhythms are different that shows that rhythms maintain their organization across tempos, but if listeners cannot judge whether or not the rhythms are different that would suggest that rhythms are perceived in different ways as tempo varies.

Two examples are given in Table 4.2. In the first, the rhythm 1-1-1-2- is compared to the rhythm 1-2-2-. As can be seen in Table 4.2, the only difference between the two occurs at elements six and seven so that 1-1-1-2- is heard as three individual sounds followed by a pair of sounds, while the rhythm 1-2-2- is heard as a single sound followed by two pairs of sounds. If the two rhythms are played at the same tempo, the percentage of different judgments was 67, but if the two were played at different tempos (3.9 tones/sec and 1.8 tones/sec), the percentage of different judgments was 33, less than chance performance. Two-thirds of the listeners thought they were identical. In the second, if the rhythms 1-1-2-1- and 1-1-3- are played at the same tempo it was relatively easy to determine that the rhythms were different, 80 percent. But if the rhythms were played at the different tempos, only 44 percent of the judgments were that the rhythms differed, again less than chance (Handel, 1993).

We can understand these results in terms of the Gestalt grouping principles. Suppose the first rhythm is presented at the faster rate so that the interval between each step of the rhythm is 80 msec. The onset-to-onset interval among

the single tones is 240 msec and the interval between tones heard in a group is 160 msec. At the slower rate the interval between each step is 175 msec so that the onset-to-onset interval between single tones is 525 msec and the interval between tones heard in a group is 350 msec. The interval among tones in a group at the slower tempo is longer than the interval between single tones at the faster tempo. To correctly judge whether the two rhythms are the same, listeners must make a figure-ground reversal so that intervals that formerly signified a group now signify a single tone. These outcomes suggest that this reversal is not easy.

In general, rhythmic organization at slower tempos, 1–3 tones/sec, is more flexible and rhythms are often organized beginning with the initial sound. At faster tempos, however, regardless of the initial sounds, the rhythm is reorganized so that the longest silent interval ends the rhythm. For example, if we started the rhythm 1-1-3- at the eighth element making it 3-1-1-3-1-1-. At a slower tempo listeners could retain the initial 3-1-1-organization, hearing the group of three tones starting the rhythm. But at faster tempos listeners shift to a 1-1-3 rhythmic organization, starting with the two single tones so that the longest silent interval ended the rhythm. I would speculate that at the slower tempos rhythms are not tightly embodied into body movements so that there is greater flexibility in how the rhythm is organized. At the faster tempos, there is a tight connection between the rhythm and body movements so that the organization of the rhythm reflects the neural and muscular constraints imposed by the limbs. These constraints would be more prominent when drummers are tapping syncopated rhythms at fast tempos (Barton, Getz, & Kubovy, 2017).

Sound Files 4.14: Organization of rhythms at different tempos

4.3.2 Sensory Saltation

If a stimulus is presented at one location, and a second one is presented close in time and space, the first stimulus is perceived to shift its location toward the location of the second stimulus. The timing between the two stimuli changes the perceptual space of the body; our spatial representation is a function of the temporal patterning of the sensations. This illusion was first discovered for tactual presentation (Geldard & Sherrick, 1972). What made this illusion so interesting is that if multiple taps were presented to the first location on an arm and followed by a single tap at a second location further down the arm, the taps seemed to hop down the arm. The perception was not a continuous motion as found for apparent motion discussed in Chap. 3, but a series of leaps or hops. This led to calling the illusion the cutaneous rabbit.

To understand the "hopping," it is best to start by considering the "reduced rabbit" paradigm. There are only two stimuli: the first stimulus occurs at one point on the skin and following a variable delay the second stimulus occurs at a different position on the arm. The tactual stimulation at both locations is identical, and in different studies the duration of the stimuli ranges from 5 to 50 msec. The precise characteristics of the stimuli are not critical and the same outcomes occur for physical taps, electric pulses, and even "hot" spots gener-

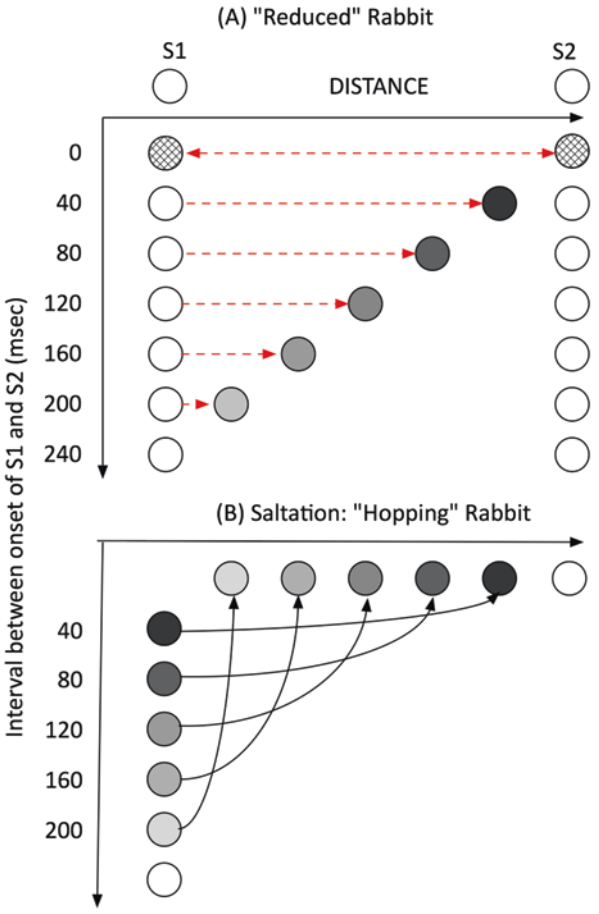


Fig. 4.15 Representation of the "reduced" rabbit and "hopping" rabbit configurations. The reduced rabbit stimulus consists of two pulses at different points on the arm, hand, or torso. (A) If the pulses at S1 and S2 are presented simultaneously, only one pulse is perceived. At the shortest intervals, the pulse at S1 is perceived to be adjacent to S2. As the interval lengthens, the shift in the position of S1 decreases, and at intervals roughly greater than 200 msec no shift occurs. (B) If multiple pulses occur at equal intervals, the shift in position of S1 relative to S2 due to the difference in timing spaces out the pulses, leading to the perception of pulses hopping down the limb or torso

ated by a CO_2 laser (Trojan et al., 2006). (Usually there is a prior stimulus signaling that the trial will begin in about 1 sec, but it does not affect the illusion).

The outcomes are shown in Fig. 4.15A. The two stimuli, S1 and S2, are identical and located a short distance apart. If the interval is greater than 240 msec, the two stimuli are perceived separately, at their actual positions. As the interval decreases, the illusionary displacement of S1 toward S2 increases. Participants report that the strongest perception of movement occurs when onset interval is between 60 and 90 msec. If the stimuli are presented synchronously or at a very short interval (< 25 msec interval), then only one stimulus is perceived at one location. This is illustrated in Panel A, Fig. 4.15.

Now suppose that five pulses are presented at position S1. The first pulse occurs 200 msec before the pulse at S2. The remaining four pulses at S1 occur at 40-msec intervals: 160 msec, 120 msec, 80 msec, and 40 msec before the pulse at S2. Based on the outcomes from the "reduced" rabbit, the initial pulse

that was 200 msec before the pulse at S2 will shift just a short distance toward S2. Each of the following pulses at S1 will shift further toward S2 because each is closer in time to S2. That will lead to the perception of discrete "hops" ending at the position of S2. This transformation of timing differences into spatial distances is shown in Fig. 4.15, Panel B. This is a strong illusion. Cholewiak and Collins (2000) made use of multiple vibrators placed along an arm to create a veridical hopping movement. People were unable to discriminate between the true movement and the illusionary hopping rabbit.

Saltation can be produced in the visual and auditory modalities, and Trojan, Getzmann, Moller, Kleinbohl, and Hölzl (2009) found it possible to create saltation between tactual and auditory stimuli. Nonetheless, there are spatial and temporal limits; the spatial distance beyond which saltation does not occur differs across the body. The saltation area is circular on the hand but elongated on the arm. Most importantly, saltation does not cross the body midline. This differs from tactual apparent motion that easily does cross the midline.

Finally, saltation also reflects the notions of belongingness found for all other perceptual acts. The strength of the hopping movement is reduced if the quality of the pulses or the rhythm of the pulses at S1 is varied. The coherence of the pulses at S1 does matter.

4.3.3 Temporal Order Judgments

Sensory saltation demonstrates the intricate connection between timing and space. The interval between two sensations affects the felt position of those stimuli. A second timing task, the judgment of temporal order, illustrates the somewhat flexible connection between body postures, limb positions, and external space.

In the typical temporal-order judgment task, two stimuli are presented separated by a short period of time, and participants simply judge which was presented first. Across many experiments involving auditory, tactual, and visual stimuli, the minimum asynchrony to correctly judge order is about 20 msec (Hirsh & Sherrick, 1961). The actual values differ due to the specific experimental conditions, but on the average the tactual sense requires the smallest interval for correct judgments. However, just crossing the hands dramatically alters perception of tactile temporal order. Judgments were often inverted at longer stimulus-onset asynchronies (100–200 msec.) so that the incorrect hand was judged to be leading, and correct judgments often required much longer stimulus-onset asynchronies (400 msec–1 sec). This was a finding unique to the tactile modality, because we can cross the feet, arms, and hands (tactile sense organs) in space, but not the eyes or ears (Yamamoto & Kitazawa, 2015). The same effect occurred if one hand and one foot were crossed; in fact, nearly every combination of crossed limbs or fingers worked. Moreover, as David Katz originally commented, tactile signals at the hand due to an extended probe are attributed to the movements at the tip of the probe. To investigate whether physically crossing the arms is necessary or whether stick crossing will suffice, Yamamoto, Moizumi, and Kitazawa

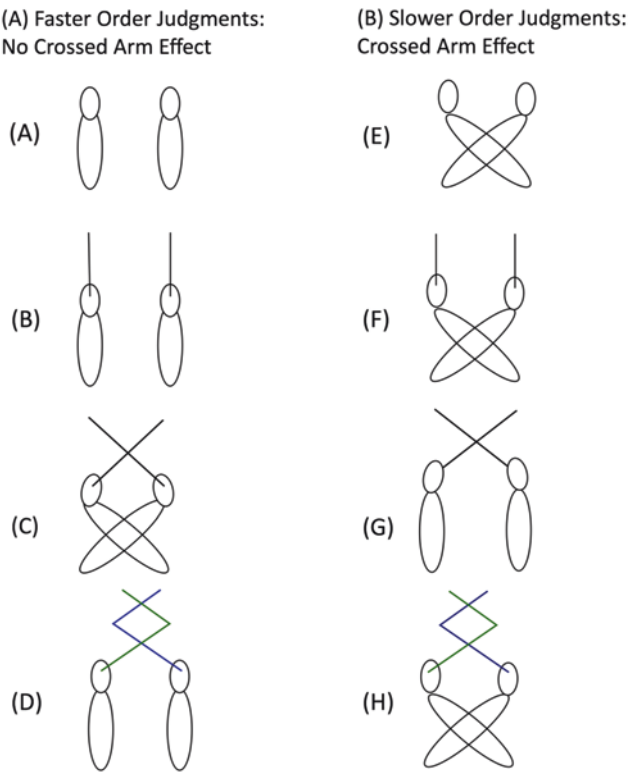


Fig. 4.16 Temporal-order judgments are faster and inversions do not occur if the arm or the arm + stick combination ends up on the ipsilateral side of the body. On this basis, in (C) the crossed sticks compensate for the crossed arms, and in (D) the double angles make the sticks end up on the ipsilateral side. But in (H), the double angles do not compensate for the crossed arms so that the stick end is on the contralateral side

(2005) created several variants of stick shape, arm crossing, and stick crossing shown in Fig. 4.16. In these experiments, participants grasped sticks placed on the vibrators arranged so that the vibrators did not affect the hands directly. The outcomes show that arm crossing (E or F) or stick crossing will bring about the slower-order judgments (G). It is interesting that the important factor determining the order effect is whether the end of the stick ends on the ipsilateral side of the body (C and D) or the contralateral side (G and H).

There are two important outcomes here. First, these results support Katz's contention that the tip of the probe acts like a sensory extension of the limb. In the time order judgment experiments, the probe tip is stationary and the position of the tip relative to the body is critical. But, recent research has emphasized also that movement of a probe allows one to explore objects in the same way as direct contact. If individuals use a wooden probe to touch an object, the perceived location of the touching point was intrinsic to the coordination system of

the arm and probe (Miller et al., 2018). The handheld tools act as a sensory extension (e.g., embodiment) of the user's body. In complementary research but still based on the premise of the purposive nature of touch (Gibson, 1966), Carello and Turvey investigated the mechanical properties of objects that provide the information (invariants in Gibson's terms) necessary to grasp and manipulate those objects (Carello & Turvey, 2017).

Second, the slower and inaccurate judgments for the crossed arm (and stick) experiment suggest one explanation based on the need to coordinate different spatial reference systems. There are at least two spatial representations for touch localizations. The first representation for the localization of the limbs is based on the body, while the second representation for the localization of the limbs is based on the external spatial coordinates. Typically the two are in coordination; the right hand is to the right of the left hand and the body midline, and the reverse is true for the left hand. But, the right hand can cross over the left one or move to the left of the midline and here the two representations must be brought back into registration. One representation has to be remapped into the other. There are several explanations for how this remapping occurs, but all explanations argue for the transition from internal body coordinates to external spatial coordinates for touch. The initial representation with respect to the body is rapidly remapped and transformed within 100–190 msec into the external one. Thus, tactile temporal order depends critically on the process of localizing tactile stimuli in external space, which develops over time.

There are several outcomes that support the idea that the "crossing" effect requires experience to attach the sensations of the skin (or of an external probe) to an external position. Individuals blind from birth do not show a crossover decrement, and for normal-sighted individuals the crossover decrement is much smaller when the hands are crossed behind the back where presumably visual experience is weaker.

We can understand the conversion to the external reference system as part of the need for the multisensory representation of space to be unified (Heed & Azanon, 2014). It would be impossible to act if the felt position of an object did not match the perceived visual position of that same object. The solution to the correspondence problem depends on the unity of the external reference. A different example of the "crossed arm" effect occurs an episode in the BBC/Masterpiece Theater series "Wolf Hall" written by Hilary Mantel. After young Thomas Cromwell picked up a heated tong in one hand, his master told him to put his hands in a watering trough and cross his hands. "Crossing would confuse the pain". Still another example is the crossed hand illusion. Cross one's arms, interlock the fingers by rotating the hands inward, and then rotate both hands upright. At this point the fingers of the left hand are pointing to the right and the fingers of the right hand are pointing to the left. People will then often confuse which fingers belong to each hand. The YouTube video makes this process easy to understand.

YouTube Video
The crossed hand illusion.

4.3.4 Visual Ternus Configuration

In the conventional Ternus display, three identical dots arrayed horizontally are displayed for a set amount of time, usually about 200 msec. The original array disappears, and after a blank interval the same three dots reappear but now offset to the left or right. After the identical blank interval, the original set of dots returns and the alternation continues. The sequence is illustrated in Fig. 4.17.

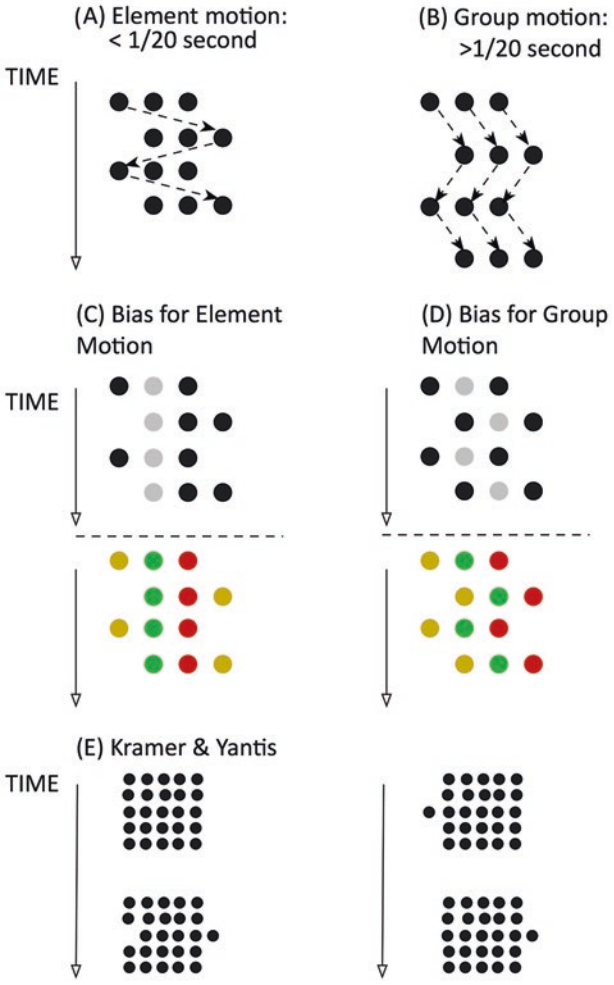


Fig. 4.17 (A) For the Ternus configuration, the perception that one element moves back and forth occurs if the interval between presentations is less than 1/20 sec (50 msec). (B) The perception that all three elements move back and forth as a unit occurs if the presentation interval is greater than 1/20 sec. (C) and (D) It is possible to bias the judgment toward either element or group movement by changing the stimulus configurations, as shown for black and gray dots or red and green dots. (E) The perceived "rigidity" of the dot configuration determines the type of motion. (Adapted from Hein & Moore, 2012; Kramer & Yantis, 1997)

Here, there are two percepts depending on the duration of the blank intervals. If the blank interval is less than 1/20 sec (50 msec) then the outside dot is seen to move to the other side of the two dots that do not appear to move. This has been termed *element motion* (Fig. 4.17A). If the blank interval is longer in duration, then the entire group of three dots is seen to shift to the left or right. This has been termed *group motion* (Fig. 4.17B). As the alternation continues, the dots undergo apparent motion and across a wide range of intervals, the percepts alternate between element and group motion, it is multistable.

Element motion and group motion represent two solutions to the correspondence problem. Several studies have attempted to determine how element similarity determines the blank interval at which the transition between element and group motion occurs. For example, Hein and Moore (2012) varied the order and position among the elements in the alternating stimuli in several ways. In some variations, the initial element of the first stimulus became the final element of the second stimulus so that element motion was fostered and continued at longer intervals (i.e., greater than 1/20 sec.). In the other variations, the second stimulus was identical to the first except offset in space so that group motion was fostered and it emerged at shorter intervals possibly eliminating element motion entirely. Examples based on black/white and color are shown in Fig. 4.17C & D. Apparent motion and the Ternus motion differ with respect to element similarity. Similarity has little effect on apparent motion but does affect the motion in the Ternus display.

Kramer and Yantis (1997) investigated the role of the spatial arrangements in determining the choice of element motion or group motion. Consider the two configurations shown in Fig. 4.17E. In the first, all the horizontal rows line up, suggesting that they form one unit. Only one dot moves and, given the perceived rigidity of the other dots, this sense of unity leads to the perception of element motion. In the second, one of the rows is originally offset to the left and then is shifted to the right in the second stimulus. This weakens the unity of the rows so that group motion increases.

Petersik and Rice (2006) summarize the evolution of explanations in terms of two opposing forces. If the inner elements of the three-element Ternus configurations are identical or seen to be similar, element motion occurs. The identity of the middle element fixes the location of the second configuration. Conversely, if the elements in each of the configurations are seen to be connected, group motion occurs.

Given that there are just two perceptions and that they are mutually exclusive, the Ternus paradigm has been used to investigate other aspects of temporal perception. Harrar and Harris (2007) constructed simplified Ternus configurations using just two stimuli. Both stimuli were either two lights or two tactual activators (i.e., unimodal) or one light and one tactual activator (i.e., bimodal). The same relationship between the stimulus-onset asynchrony and the type of apparent motion was found for all three conditions although the crossover

point between element motion and group motion was roughly 20% longer for the tactile presentation. The crossover point for the visual-tactual presentation equaled that for the visual presentation. Furthermore, as discussed in Chap. 3, the repeated presentation of one stimulus seems to fatigue that stimulus, leading to the increased perception of an alternate stimulus. Following this logic, to fatigue group motion Harrar and Harris (2007) presented the visual Ternus stimuli repeatedly at a long-onset asynchrony for 2.5 minutes. The important finding was that the fatigue procedure shifted the perception for the visual and visual-tactual stimuli toward element motion as expected, but it did not affect the perception of the tactual stimuli. These outcomes lead the authors to argue that there are no general timing mechanisms across modalities. This was the same conclusion from the study of visual and auditory apparent motion discussed in Chap. 3.

Chen (2009) made use of temporal ventriloquism to affect the perceived stimulus-onset asynchrony and thereby affect the shift between element and group motion. As discussed in Chap. 2, two tones that bound two lights (AVVA) seem to increase the asynchrony between the lights, while the same tones interspersed between the lights (VAAV) seem to decrease the asynchrony between the lights. By placing one tone before the first visual stimulus and the second tone after the second stimulus, the stimulus-onset asynchrony interval seemed longer so that the crossover occurred at a shorter interval. Conversely, if the tones were placed between the visual stimuli, the interval seemed shorter so that the crossover occurred at a longer interval. The effect did not occur if there was only one tone, showing once again that multisensory effects happen only if there is a belief that the sensations from each modality come from the same event.

These outcomes reveal how physical properties, spatial arrangements, and timing interact to solve the correspondence problem. To some extent, these aspects are functionally interlocked; it is possible to overcome a change in one property by a compensating change in another.

4.4 Visual Spatial Rhythms

4.4.1 The Visual Grid

It seems quite natural to feel musical rhythms and to experience the visual rhythms of dancers and athletes as they extend in time. However, it seems harder to be aware of the rhythms inherent in static visual designs and friezes. Starting with a small set of distinct patterns, such as shown in Fig. 4.18, complex designs can be made by combining and overlaying these patterns with each other in intricate ways. The patterns might be repeated, alternated, interleaved, rotated, reproduced at varying spatial scales, and/or offset laterally. Any one or combination of these processes can construct the layering characteristic of visual rhythms.

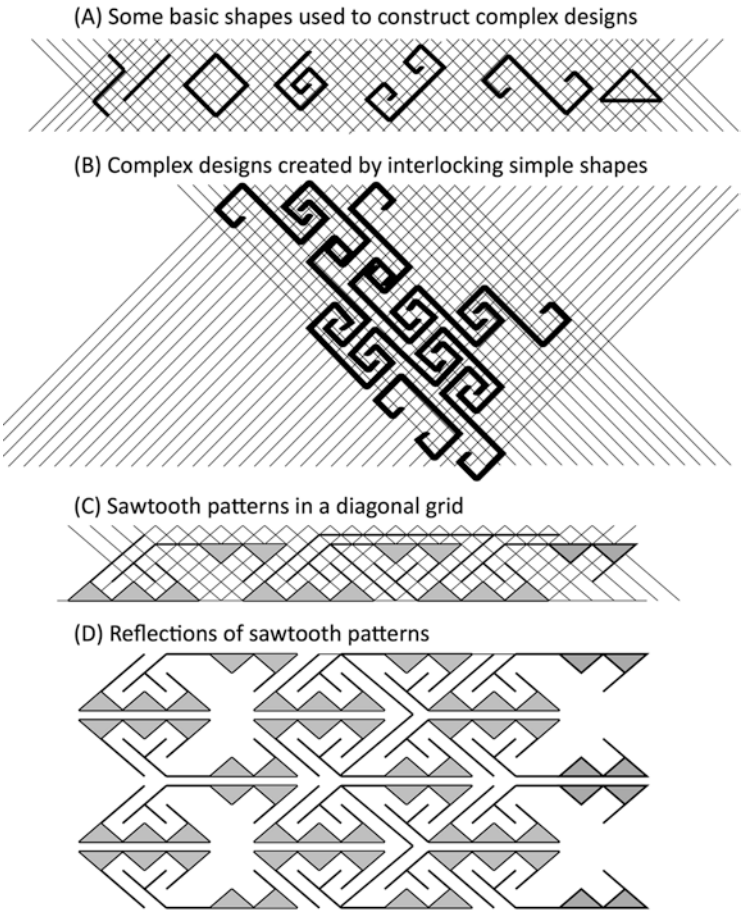


Fig. 4.18 Complex designs can be created on a diagonal grid by combining simple shapes in different groupings. A simple pattern in (C) can be reflected and repeated to create the complex design in (D). (Adapted from Tetlow, 2013)

When we analyzed the organization of auditory rhythms, the highest level could be termed the "metrical grid" and consisted of "dots" that represented each time point. The dots were equally separated in time, and the strong and weak beats represented by "dots" at lower hierarchical levels. By analogy, we can start with a "metrical" spatial grid with equally spaced lines and create a visual rhythmic pattern by darkening some of the metrical lines.

The first visual patterns were probably constructed using only a compass and a straight edge on an underlying grid (Tetlow, 2013). A variant of such a grid uses diagonal lines to fill the space with diamonds or squares at 45°. Now any shape can be described in terms of the grid units travelled before making a 90°

turn as shown in Fig. 4.18A. For example, +4, +4, +4, +4, where + indicates a right turn and – indicates a left turn, yields a small square. The sequence +1, +1+2, +2, +3, +3, +4, +4, and so on yields a square spiral and +1, +2, +3, +4, +10, +4, +3, +2, +1 yields a "C-shape." Mixing left and right turns such as +1, +2, +3, +7, −3, −2, −1 can produce an "S-shape."

What is important about these sorts of patterns is that the size can be varied so that smaller spirals might be embedded in larger ones, one pattern can be reflected or rotated to create different designs, and different patterns can be intermixed to create more elaborate designs. Examples of these possibilities are shown in Fig. 4.18B. Other construction rules can produce "sawtooth" shapes shown in Fig. 4.18C, which could also be used to create larger designs by reflecting or rotating the original pattern and connecting them together. Here, groups of three triangles alternate vertically with groups of two triangles as shown in Fig. 4.18D. What is common to diagonal grids is the ability to vary the size and to interleave patterns to create the final design. These are open to continuous elaboration, much like auditory rhythms.

A second example of the "openness" of visual designs is found in the Cathedral of Palermo described by Garofalo (2017). Smaller hexagons fitted within a larger hexagonal yield a pattern of hexagons and six-sided stars (Fig. 4.19). A third example is the Sierpinski triangle. We start with an equilateral triangle and then embed a single equilateral triangle one half the size of the original. The process is then continued; at the next stage the next three embedded triangles are one quarter the size of the original, and at the following stage, each of the nine embedded triangles are one eighth the size of the original also shown in Fig. 4.19. This process would continue; the triangles at each stage would be one half the size of the triangles at the previous stage. The Sierpinski triangle is a fractal because the identical pattern occurs at increasingly smaller scales. What is important to us is the analogy to metric rhythms. Each level of the metric grid splits the timing interval in half for double meter or thirds for triple meter so that the meter organizes each beat. Here, the overall figure is organized by the constant reduction of the embedded triangles.

4.4.2 Islamic Tiling Patterns

Islamic visual patterns share many properties with Celtic patterns. Islamic patterns are space filling and emphasize equilateral triangles, squares, and equilateral hexagons that can completely fill a space as circles cannot (Critchlow, 1976). In common with the Celtic patterns, the shapes are layered so that they occur in many orientations and sizes and that layering yields the sense of rhythm and repetition.

A good starting point would be the drawing shown in Fig. 4.20. Here are three circles within a larger circle, and within each of the smaller circles is one of the other space-filling shapes.

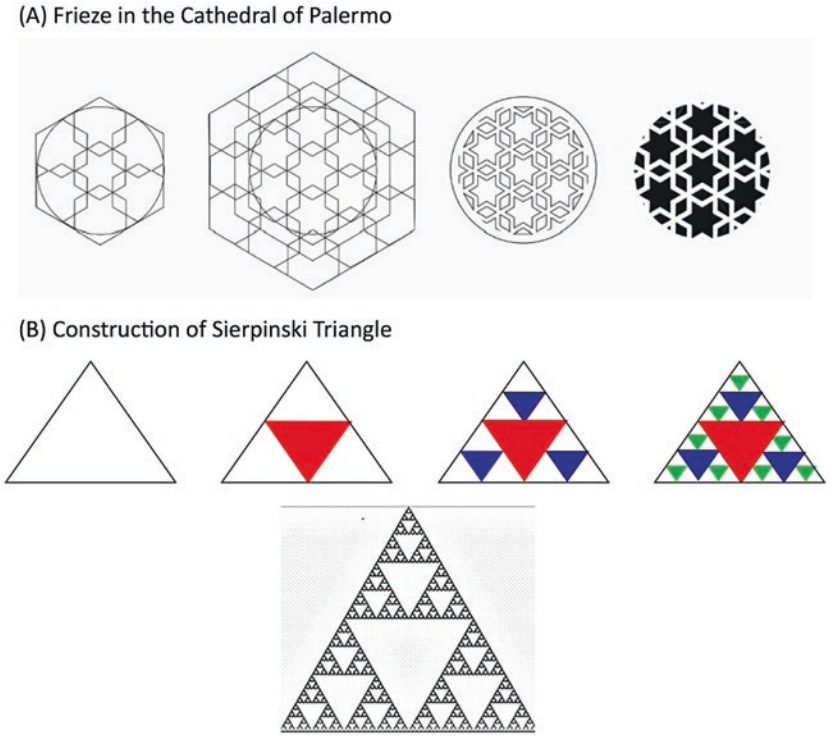


Fig. 4.19 (A) Frieze in the Cathedral of Palermo. In the second panel from the left, three, four, and five smaller hexagons are drawn within a larger hexagon. Due to this embedding, emergent stars result from the overlap of the smaller hexagons. The same kind of superposition of rhythmic lines leads to complex rhythms. (B) In the construction of the Sierpinski triangle, at each stage the embedded equilateral triangles are one half the size of those at the previous level. (Reproduced from Garofalo, 2017. CC license)

For our purposes, the importance of the three shapes is in their ability to be embedded within themselves. As illustrated in Fig. 4.20, the triangles, squares, and hexagons contained within circles can produce all of these space-filling designs at multiple sizes. This is also true for circles, even though there will be gaps in the surface.

Finally, there are combinations of the triangles, squares, and hexagons that fill the entire space. Variants of these combinations form the basis of many patterns found on walls and surfaces. One basic pattern is shown In Fig. 4.21A, and one shading scheme is illustrated in Fig. 4.21B. The same shading scheme is reproduced multiple times in Fig. 4.21C and can give a sense of the rhythm of the design.

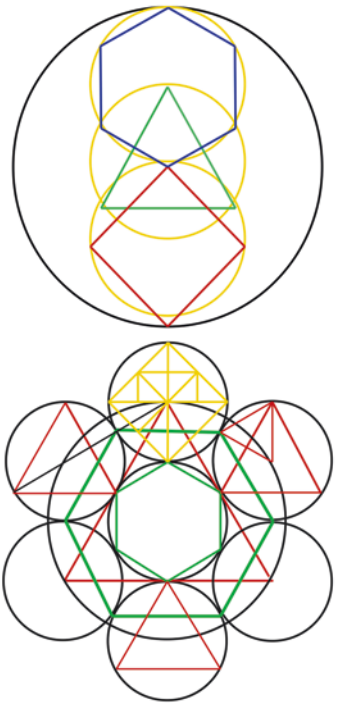


Fig. 4.20 The space-filling squares, triangles, and hexagon are embedded in a circle. Each shape has a particular religious significance. Squares (yellow), hexagons (green), and triangles (red) can be embedded in multiple sizes that yield new shapes making use of sides of other shapes. (Adapted from Critchlow, 1976. Pages 19 & 150)

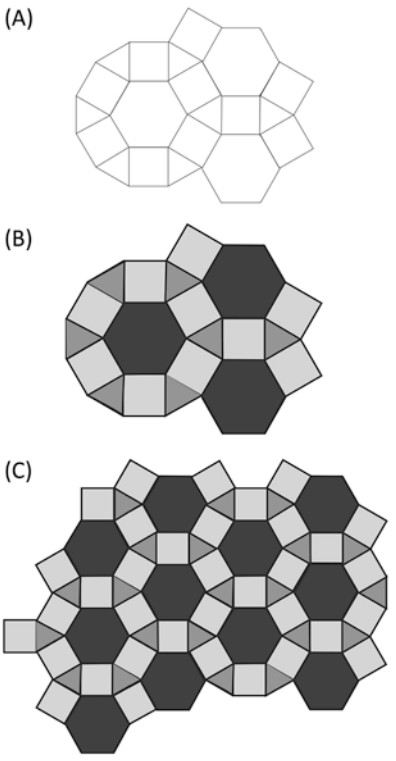


Fig. 4.21 The basic pattern is shown in Figure 4.21A. One possible shading design is illustrated in Figure 4.21B and the same shading design is reproduced multiple times at a smaller scale in Figure 4.21C. (Adapted from Critchlow, 1976. Pages 119 & 123)

4.5 Summary

Rhythms always seem to engage the listener or viewer, and create figures in time and space. Like all other perceptual acts, hearing and seeing rhythms are multilayered. One can zoom in and out; it is possible to react to every beat or to beats widely separated in time, and it is possible to view a narrow spatial field or a wide one. Listeners construct layers to hear the beat and meter even for an isochronous sequence of identical tones, and for more complex polyrhythms and non-metric rhythms. In similar fashion, viewers construct layers to group the dots into larger, more encompassing groups even for a series of equidistant identical dots and for more complex visual patterns that repeat in space. But, musical and visual beats do not simply match auditory or visual sensory features. The beats emerge from the interaction of sensory features and perceptual acts.

References

Barton, S., Getz, L., & Kubovy, M. (2017). Systematic variation in rhythm production as tempo changes. *Music Perception, 34*, 303–312. https://doi.org/10.1525/mp.2017.34.3.303

Bolton, T. L. (1894). Rhythm. *American Journal of Psychology, 6*, 145–238.

Brumm, H., & Slater, P. (2006). Ambient noise, motor fatigue, and serial redundancy in chaffinch song. *Behavioral Ecology and Sociobiology, 60*, 475–481. https://doi.org/10.1007/s00265-006-0188-y

Carello, C., & Turvey, M. T. (2017). Useful dimensions of haptic processing 50 years after "the senses considered as perceptual systems". *Ecological Psychology, 29*(2), 95–121. https://doi.org/10.1080/1040713.2017.1297188

Chen, L. (2009). *Crossmodal temporal capture in visual and tactile apparent motion: Influence of temporal structure and crossmodal grouping* (PhD). Muchen, Germany: Ludwig-Maximilians.

Cholewiak, R. W., & Collins, A. A. (2000). The generation of vibrotactile patterns on a linear array: Influences of body site, time, and presentation mode. *Perception & Psychophysics, 62*, 1220–1235.

Critchlow, K. (1976). *Islamic patterns: An analytical and cosmological approach.* London, UK: Thames & Hudson.

DeLong, L. (2013). *Curves: Flowers, foliates & flourishes in the formal decorative arts.* New York, NY: Bloomsbury USA.

Fitch, W. T. (2018). Four priniciples of biomusicology. In H. Honing (Ed.), *The origins of musicality* (pp. 23–48). Cambridge, MA: MIT Press.

Fox, D., & Weissman, D. (2007). *The great family songbook for piano and guitar.* New York, NY: Black Dog & Leventhal.

Garofalo, V. (2017). *Visual perception and Graphic analysis. The patterns and inlays in the Cathedral of Palermo.* Paper presented at the Inerbationaal and Interdisciplinary Conference IMMAGINI? Image and Imagination between Representation, Communication, Education and Psychology, Brixen, Italy.

Geldard, F. A., & Sherrick, C. E. (1972). The cutaneous "rabbit": A perceptual illusion. *Science, 178*(4057), 178–179.

Gibson, J. J. (1966). *The senses considered as perceptual systems.* Boston, MA: Houghton-Mifflin.
Handel, S. (1984). Using polyrhythms to study rhythm. *Music Perception, 1*, 465–484.
Handel, S. (1993). The effect of tempo and tone duration on rhythm discrimination. *Perception & Psychophysics, 54*, 370–382.
Handel, S., Todd, S., & Zoidis, A. M. (2012). Hierarchical and rhythmic organization in the songs of humpback whales (Megaptera novaeangliae). *Bioacoustics, 21*, 141–156. https://doi.org/10.1080/09524622.2012.668324
Hannon, E. E., & Trehub, S. E. (2005). Metrical categories in infancy and adulthood. *Psychological Science, 16*, 48–55.
Harrar, V., & Harris, L. R. (2007). Multimodal Ternus: Visual, tactile and visio-tactile grouping in apparent movement. *Perception, 36*, 1455–1464. https://doi.org/10.1068/p5844
Heed, T., & Azanon, E. (2014). Using time to investigate space: A review of tactile temporal order judgments as a window onto spatial processing in touch. *Frontiers in Psychology, 5*. https://doi.org/10.3389/fpsyg.2014.00076
Hein, E., & Moore, C. M. (2012). Spatio-temporal priority revisited: The role of feature identity and similarity for object correspondence in apparent motion. *Journal of Experimental Psychology: Human Perception and Performance, 38*(4), 975–988. https://doi.org/10.1037/a0028.197
Hirsh, I. J., & Sherrick, C. E. (1961). Perceived order in different sense modalities. *Journal of Exerimental Psychology, 62*, 423–432.
Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics, 14*, 201–211.
Johansson, G. (1975). Visual motion perception. *Scientific American, 232*, 76–88.
Jones, M. R. (1976). Time, our lost dimension: Toward a new theory of perception, attention, and memory. *Psychological Review, 83*, 323–335.
Kalender, B., Trehub, S. E., & Schellenberg, E. G. (2013). Cross-cultural differences in meter perception. *Psychological Research, 77*, 196–203. https://doi.org/10.1007/s00426-012-0427-y
Kramer, P., & Yantis, S. (1997). Perceptual grouping in space and time: Evidence from the Ternus display. *Perception & Psychophysics, 59*(1), 87–99.
Large, E. W., & Jones, M. R. (1999). The dynamics of attending: How people track time-varying events. *Psychological Review, 106*, 129–159.
Leman, M., & Naveda, L. (2010). Basic gestures as spatiotemporal reference rames for repetitive dance/music patterns in Samba and Charleston. *Music Perception, 28*, 71–91. https://doi.org/10.1525/mp.2010.1.71
Maes, P.-J., Leman, M., Palmer, C., & Wanderley, M. M. (2014). Action-based effects on music perception. *Frontiers in Psychology, 4*, 1–14. https://doi.org/10.3389/fpsyg.2013.01008
Marthevon, N., Casey, C., Reichmuth, C., & Charriuer, I. (2017). Northern elephant seals memorize the rhythm and timbre of their rivals' voice. *Current Biology, 27*, 2065. https://doi.org/10.1016/j.cub.2017.06.005
Miller, L. E., Montroni, L., Koun, E., Salemme, R., Hayward, V., & Farner, A. (2018). Sensing with tools extends intersensory processing beyond the body. *Nature, 561*(13 September), 239–242. https://doi.org/10.1038/s41586-018-0460-0
Mungan, E., Yazici, Z. F., & Kaya, M. (Uğur). (2017). Perceiving boundaries in unfamiliar Turkish Makam music: Evidence for gestalt universals? *Music Perception, 34*, 267–290. https://doi.org/10.1525/mp.2017.34.3.267

Patel, A. D., Iversen, J. R., Bregman, M. R., & Schulz, I. (2009). Experimental evidence for synchronization to a musical beat in a nonhuman animal. *Current Biology, 19*, 827–830. https://doi.org/10.1016/j.cub.2009.03.038

Petersik, J. T., & Rice, C. M. (2006). The evolutions of a perceptual phenomenon: A case history using the Ternus effect. *Perception, 35*, 807–821. https://doi.org/10.1068/p5522

Povel, D., & Essens, P. (1985). Perception of temporal patterns. *Music Perception, 2*, 411–440.

Povel, D., & Okkerman, H. (1981). Accents in equitone sequencies. *Perception & Psychophysics, 30*, 565–572.

Prince, J. B., & Schmuckler, M. A. (2014). The tonal-metric hierarchy: A corpus analysis. *Music Perception, 31*, 254–270. https://doi.org/10.1525/MP.2014.31.254

Repp, B. (2013). Sensorimotor synchronization: A review of recent research (2006–2012). *Psychonomic Bulletin & Review, 20*, 403–452. https://doi.org/10.3758/s13423-012-0371-2

Tetlow, A. (2013). *Celtic pattern: Visual rhythms of the ancient mind* (pp. 26–27). New York, NY: Bloomsbury USA.

Toiviainen, P., Luck, G., & Thompson, M. R. (2010). Embodied meter: Hierarchical eigenmodes in music induced movement. *Music Perception, 28*(1), 59–70. https://doi.org/10.1525/mp.2010.28.1.59

Trojan, J., Getzmann, S., Moller, J., Kleinbohl, D., & Hölzl, R. (2009). Tactile-auditory saltation: Spatiiotemporal integration across sensory modalities. *Neuroscience Letters, 460*, 156–160. https://doi.org/10.1016/j.neulet.2009.05.053

Trojan, J., Stolle, A. M., Kleinbohl, D., Morch, C. D., Arendt-Nielsen, L., & Holzl, R. (2006). The saltation illusion demonstrates integrative processing of spatiotemporal information in thermoceptive and nociceptive networks. *Experimental Brain Research, 170*, 88–96. https://doi.org/10.1007/s00221-005-0190-z

Vernes, S. C. (2017). What bats have to say about speech and language. *Psychonomic Bulletin & Review, 24*, 111–117. https://doi.org/10.3758/s13423-016-1060-3

Wilson, M., & Cook, P. F. (2016). Rhythmic entrainment: Why humans want to, fireflies can't help it, pet birds try, and sea lions have to be bribed. *Psychonomic Bulletin & Review, 23*, 1647–1659. https://doi.org/10.3758/s13423-016-1013-x

Woodrow, H. (1909). A quantitative study of rhythm. *Archives of Psychology, 14*, 1–66.

Yamamoto, S., & Kitazawa, S. (2015). Tactile temporal order. *Scholarpedia, 10*. https://doi.org/10.4249/scholarpedia.8249

Yamamoto, S., Moizumi, S., & Kitazawa, S. (2005). Referral of tactile sensation to the tips of L-shaped sticks. *Journal of Neurophysiology, 93*, 2856–2863. https://doi.org/10.1152/jn.01015.2004

Zwamborn, E., & Whitehead, H. (2017). Repeated call sequences and behavioural context in long-finned pilot whales off Cape Breton, Nova Scotia, Canada. *Bioacoustics, 26*, 169–183. https://doi.org/10.1080/09524622.2016.1233457