Music Perception FREE

Marcus T. Pearce, Queen Mary University of London

https://doi.org/10.1093/acrefore/9780190236557.013.890
Published online: 24 May 2023

Summary

Music perception covers all aspects of psychological and neural processing invoked while listening to music. In order to make sense of a musical stimulus, the perceptual system must infer an internal representation of the structure present in a piece of music, including the attributes of individual events (including pitch, timbre, loudness, and timing), groups of events (such as chords, voices, and phrases), and structural relationships between such groups, so that larger-scale aspects of musical form and thematic structure can be perceived. Such representations are stored in memory at timescales ranging from seconds for echoic memory to decades in the case of long-term memory for music, which consists of schematic knowledge of musical styles, veridical memory for particular familiar pieces of music, and episodic memory for music heard at a particular place and time. Stored representations of music allow the generation of top-down expectations for the attributes of forthcoming events while listening to music, which play a role in the perception of music as it unfolds dynamically in time and also the emotional and aesthetic experience of music. Music is a communicative medium conveying affective meaning from the composer and performer to the listener, via several psychological mechanisms and using a range of cues in the music, some of which are universal, others culture-specific. Individuals show behavioral and physiological effects of listening to music from birth onward and learn the syntactic structure of the musical styles to which they are exposed within their culture, shaping their music perception. Some individuals undertake explicit musical training, which can additionally shape their perception of music, sometimes in fundamental ways. Listening to music can impair performance on concurrent tasks involving working memory due to competing access for resources but can improve performance when listening takes place prior to the task due to its positive effect on affective state. Music is a universal human cultural phenomenon whose complexity requires the activation of a diverse range of perceptual and cognitive mechanisms, making it an interesting target for psychological and neuroscientific investigation.

Keywords: music perception, auditory perception, pitch, timbre, rhythm, auditory scene analysis, memory, expectation, emotion, evolution

Subjects: Affective Science, Biological Foundations of Psychology, Cognitive Psychology/Neuroscience, Developmental Psychology

Introduction

Music is a human cultural domain characterized by structured sequences of sound created and performed using a wide variety of musical instruments including the voice and electronic instruments (sometimes accompanied by a performative component or visual display) and appreciated by an audience, who may be participatory in the music making. Music is a cross-cultural universal, showing both consistency and variation across musical cultures (Mehr et al., 2019), while concrete evidence for music making (and, by inference, music perception) exists in the form of bone flutes dating to approximately 35,000 years ago (Conard et al., 2009).

Although some aspects of music perception rely on general properties of the auditory system, others depend on developmental experience, whereby listeners learn the structure of the musical styles that they experience within their native culture, shaping their music perception. Within a culture, music often serves a communicative function with enculturated listeners recognizing, and sometimes experiencing, the psychological states expressed by the music of their culture.

Music shares many of these characteristics with language, with which it sometimes co-occurs in the form of lyrics. Compared with language, music is not as frequently used for the communication of specific referential meaning and allows for greater range of auditory complexity. For example, a piece of music might simultaneously vary the rhythmic, melodic, harmonic, timbral, and dynamic properties of the sound signals resulting from multiple instruments performed together over long time periods. How listeners make sense of and find pleasure in listening to such an extraordinarily complex auditory signal is the goal of research in the field of music perception, which dates back at least to the 19th century psychologist Hermann von Helmholtz (1863).

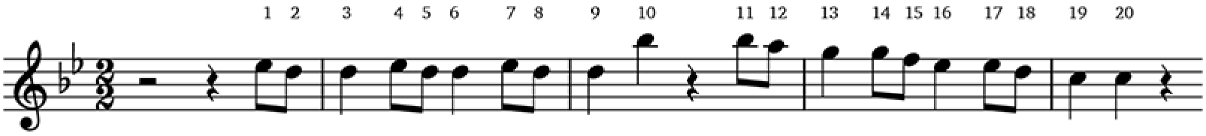
Perceived Musical Attributes

In most music, the basic elements are individual sound events varying in such physical properties as fundamental frequency, spectral content, onset time, duration, amplitude, and amplitude envelope. These physical quantities are represented by the perceptual system in terms of perceived musical *attributes* such as pitch, timbre, timing, and loudness.

Pitch

Pitch is a perceptual indicator of periodicity in the auditory environment that scales as an approximately logarithmic function of frequency and is limited by the frequency response of the human basilar membrane to between approximately 20 and 20,000 Hz, though musical pitch is generally limited to below 5,000 Hz (e.g., the range of a standard piano is 27.5–4,186.01 Hz). The basilar membrane is part of the cochlea, an organ in the inner ear that performs a decomposition of the frequencies present in a sound, known as *place coding*. This is passed on to bundles of fibers in the auditory nerve, which have tuning curves and characteristic frequencies, through various subcortical nuclei to tonotopic organization within auditory cortex in the superior temporal gyri; in addition, pitch is also coded through phase locking of frequencies below about 5,000 Hz with the firing of auditory neurons (Moore, 2013; Plack, 2018).

A



Note number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Pitch height	63	62	62	63	62	62	63	62	62	70	70	69	67	67	65	63	63	62	60	60
Pitch class	3	2	2	3	2	2	3	2	2	10	10	9	7	7	5	3	3	2	0	0
Scale degree	8	7	7	8	7	7	8	7	7	3	3	2	0	0	10	8	8	7	5	5
Pitch interval		-1	0	1	-1	0	1	-1	0	8	0	-1	-2	0	-2	-2	0	-1	-2	0
Pitch contour		-1	0	1	-1	0	1	-1	0	1	0	-1	-1	0	-1	-1	0	-1	-1	0

B

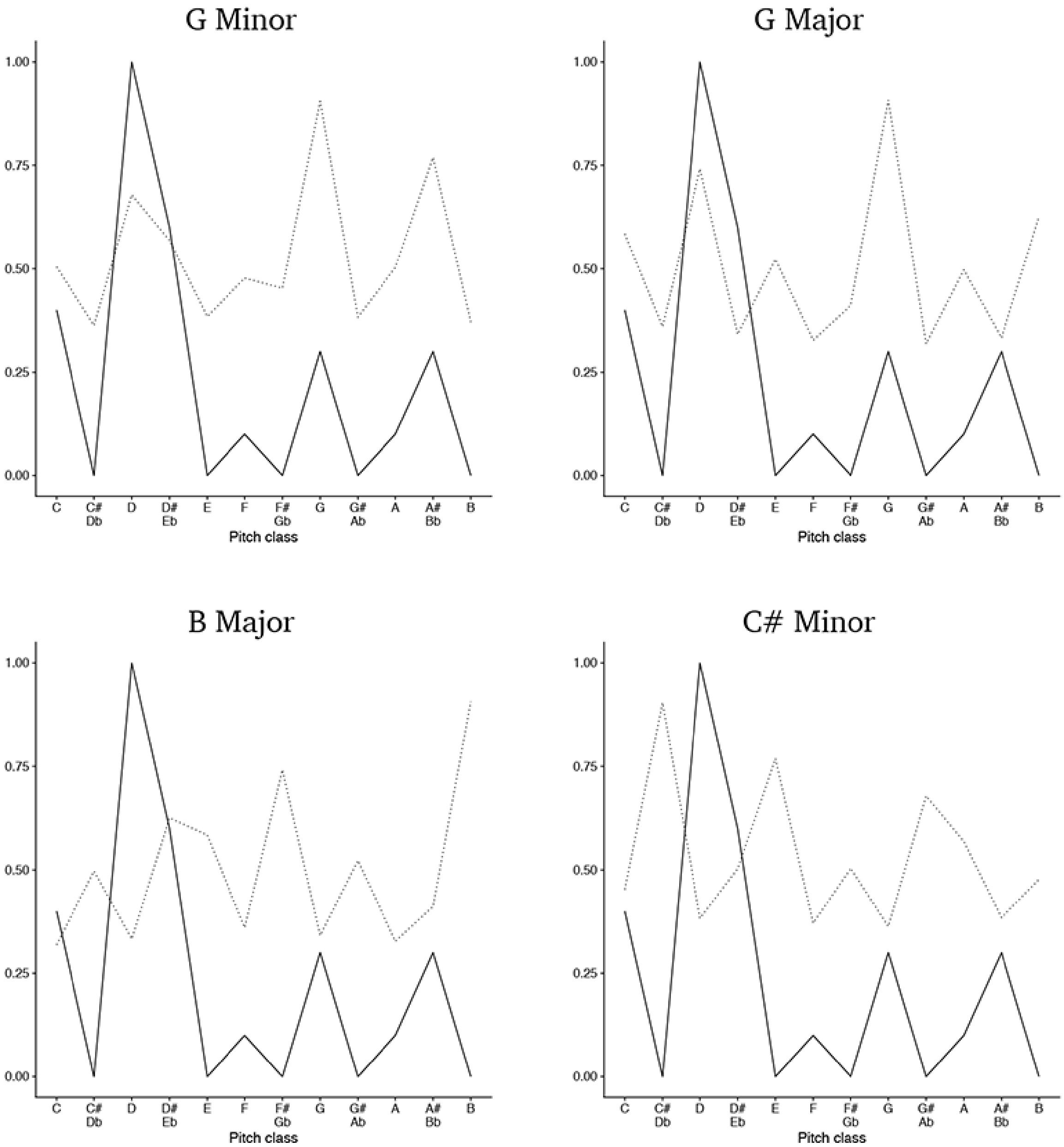


Figure 1. Pitch perception. A: The opening melody from the first movement of Mozart's Symphony No. 40 in G minor accompanied by representations of pitch height (semitones, middle C = 60), pitch class (0 = C, . . ., 7 = G, . . .), scale degree relative to a tonal center (0 = tonic, . . ., 7 = dominant, . . .), pitch interval (in semitones, positive = rising/negative = falling), and contour (-1 = falling, 0 = unison, 1 = rising). B: Inference of Western tonality. The relative duration that each pitch class sounds in the Mozart example from A (solid lines, arbitrary units) plotted against tone profiles corresponding to four keys (dotted lines, [0,1]): The actual key (G minor), the parallel major key (G major), the relative major key (B major), and a distant minor key (C# minor, see Figure 2). Tone profiles reflect listeners' perception of the goodness-of-fit of each pitch class following a strongly key-defining musical context (Krumhansl & Kessler, 1982).

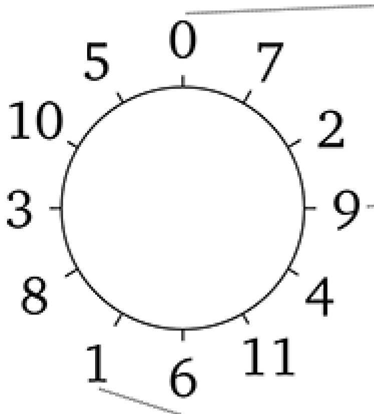
Source: Data plotted from Krumhansl (1990, Table 2.3, p. 37).

The sounds produced by most musical instruments are complex, meaning that they consist of a superposition of *partial frequencies*, usually harmonically related (and known as *harmonics*) at integer multiples of a fundamental frequency, which represents the perceived pitch. Even when the fundamental frequency is missing, it is still perceived as the greatest common divisor of the harmonics, illustrating how the perceptual system attempts to reconstruct a coherent portrait of the sensory environment. This perceptual process appears to depend critically on computations performed in the primary auditory cortex (Zatorre, 1988, 2005).

It has been demonstrated that listeners are sensitive to multiple perceptual representations of pitch (see Figure 1) including: *pitch height*, a continuous representation of pitch from low to high, often measured in semitones or cents (1 semitone = 100 cents), corresponding to a division of the octave (a doubling of frequency) into 12 equally spaced semitones (Rusconi et al., 2006); *pitch class*, a mod 12 reduction of pitch height such that no distinction is made between a pitch in different octaves—the corresponding perceptual correlate is usually referred to as *pitch chroma* (Demany et al., 1985; Warren et al., 2003); melodic *pitch interval*, reflecting the difference in pitch between successive notes in a melody (Edworthy, 1985; Plantinga & Trainor, 2005); and *pitch contour*, reflecting whether the pitch interval rises, falls, or remains the same (Dowling, 1978).

Western listeners also perceive pitch relative to an inferred tonality, consisting of a referent pitch such as the *tonic* and a subset of pitches within the octave represented relative to the referent, some of which have greater perceptual salience or stability than others (Krumhansl, 1990). In the Western musical tradition (see Figures 1 and 2), this generally corresponds to a musical *key* which defines a *scale* (a subset of seven pitch classes within the octave, known as *scale degrees*), with the tonic being the most important and stable scale degree, followed by the fifth (seven semitones above the tonic) and third scale degrees (three or four semitones above the tonic depending on whether the key is minor or major). Listeners enculturated in Western musical styles infer musical keys while listening to music and represent pitches relative to the tonic of the inferred key (Krumhansl & Kessler, 1982): keys are represented by listeners in a psychological space emphasizing relationships between the tonic and fifth scale degrees (see Figure 2) such that keys with tonics differing by an interval of seven semitones (corresponding to the fifth scale degree) are perceived as similar while major keys are also perceived as being similar to the parallel minor (the minor key with the same tonic) and the relative minor (the minor key whose scale contains the same notes but a different tonic). When more than one pitch sounds at the same time, a chord may be perceived and it has been shown that perception of relatedness of different chords by Western listeners follows music theoretical predictions with tonic and dominant chords most similar (Krumhansl, Bharucha, & Castellano, 1982) and similarity also depending on the key membership of the tones making up the chords (Krumhansl, Bharucha, & Kessler, 1982).

Representations of pitch such as chroma, pitch interval, contour, and scale degree allow for abstract equivalence classes in which, for example, melodies that have been transposed to a different pitch height can be recognized as being fundamentally equivalent. These multiple aspects of pitch can be related to one another using geometrical cognitive representations (Krumhansl, 1990; Shepard, 1982). It should also be noted that scale systems in non-Western musical cultures use different scales and sometimes so-called microtones corresponding to frequencies in between Western chromatic pitches (Bozkurt et al., 2014; McBride & Tlusty, 2020). Musicians from a given musical culture tend to perceive melodic pitch intervals categorically according to culture-specific scale systems (Burns & Campbell, 1994; Burns & Ward, 1978; Perlman & Krumhansl, 1996). However, Western listeners are able to learn artificial musical grammars constructed from certain scale systems that deviate from the 12-fold equal division of the octave (Loui et al., 2010; Pelofi & Farbood, 2021).



Major-key tonic pitch class	Number of sharpened or flattened notes	Sharpened or flattened notes	Major Key	Relative Minor Key
0	None		C	A
7	1 sharp	F	G	E
2	2 sharps	F C	D	B
9	3 sharps	F C G	A	F ♯
4	4 sharps	F C G D	E	C ♯
11	5 sharps	F C G D A	B	G ♯
6	6 sharps	F C G D A E	F ♯	D ♯
6	6 flats	B E A D G C	G ♭	E ♭
1	5 flats	B E A D G	D ♭	B ♭
8	4 flats	B E A D	A ♭	F
3	3 flats	B E A	E ♭	C
10	2 flats	B E	B ♭	G
5	1 flat	B	F	D

Figure 2. Relationships between Western keys. The circle of fifths (left) represents the distance between major keys with tonics differing by a perfect fifth (seven semitones). The table (right) can be imagined as being rolled into a cylinder around the circle of fifths. Each major key defines a scale consisting of seven pitch classes which differ by one from the adjacent keys. For example, going clockwise around the scale of C major (pitch class 0) consists of the white keys on the piano whereas G major replaces the F with F#, D additionally replaces C with C#, and so on. Going counterclockwise, F major replaces B with Bb, B major additionally replaces E with Eb, and so on. F# major (six sharps) is enharmonically equivalent to Gb major (six flats). Diametrically opposing keys on the circle of fifths are maximally distant from each other. Each major key has a corresponding relative minor key with the same pitch classes but a different tonic.

There are two interesting cases of extreme pitch perception: *amusia* and *absolute pitch*. Congenital amusia is a developmental condition characterized by impairments of fine-grained pitch perception such as an inability to detect whether two pairs of intervals at different pitch heights have the same or different pitch contour and to discriminate pairs of unfamiliar (but stylistically conventional) melodies that differ by a single out-of-key note (Stewart, 2011). It is usually also accompanied by poor singing accuracy (see Pfordresher, 2022). It has a population prevalence of 1%–4% depending on the diagnostic criterion (Peretz, 2013). Interestingly, these effects generalize to perception of pitch inflections in speech, for example when distinguishing a question from a statement (Liu et al., 2010) even though amusics typically have no real-world language impairment. Amusia appears to involve abnormalities in the right superior temporal and inferior frontal gyri as well as the arcuate fasciculus, a white matter tract connecting the two, based on converging neuroimaging evidence including cortical thickness (Hyde et al., 2007), diffusion-tensor imaging (Loui et al., 2009), functional Magnetic Resonance Imaging (Hyde et al., 2011), and magnetoencephalography (Albouy et al., 2013).

Whereas most adults rely primarily on relative pitch representations such as pitch interval and scale degree for perceiving music, absolute pitch is characterized by an ability to identify the pitch height of a note heard in isolation. Many people with absolute pitch also develop the ability to name pitches and to produce a named pitch without a reference. Absolute pitch is experienced as automatic and effortless and has an estimated prevalence in Europe and North America of less than 0.01% (Deutsch, 2013a) rising to as much as 15% in musicians (Baharloo et al., 2000). Absolute pitch appears to depend on musical experience, in particular early musical training before the age of 5 (Miyazaki, 1988), and is also more prevalent in speakers of tone languages (Deutsch et al., 2006, 2013). Absolute pitch ability is also better for more tonally stable pitches, which occur more frequently, suggesting a causal effect of musical experience (Deutsch et al., 2013; Miyazaki, 1988). However, not all musicians with early training or exposure to tone languages develop absolute pitch and twin studies suggest a significant genetic component (Baharloo et al., 2000; Theusch & Gitschier, 2011), which may relate to pitch memory ability, assessed by distinguishing familiar melodies at their actual pitch from pitch-shifted versions, which is normally distributed in the general population regardless of tone-language experience (Schellenberg & Trehub, 2003, 2008).

Timbre

The most prominent manifestation of musical *timbre* (or tone color) corresponds to the differences perceived between different musical instruments, which are distinguished by infants as young as 6 months (Trainor et al., 2004). Timbre has been defined as "that attribute of auditory sensation which enables a listener to judge that two nonidentical sounds, similarly presented and having the same loudness and pitch, are dissimilar" (American National Standards Institute, 1994, p. 35). While this specifies what timbre is not, rather than what it is (Bregman, 1990), it can be inferred from this definition that timbre is fundamentally perceptual, comparative, and multidimensional. Accordingly, timbre perception has often been investigated by taking pairwise comparisons of instrumental tones—naturally recorded or artificially synthesized—and using multidimensional scaling to extract a dimensional space in which tones that are close in terms of Euclidean distance are also perceptually similar (Caclin et al., 2005; Krumhansl, 1989; McAdams et al., 1995; Wessel, 1979). Generally the results of these experiments demonstrate that instrumental timbre is perceived in terms of three primary dimensions (see Figure 3): first, *attack time*, representing how quickly after the onset of the sound its amplitude envelope reaches a maximum (compare a harpsichord with a bassoon); second, *brightness*, representing the overall balance of high and low frequency spectral content in the sound (compare a trombone with a trumpet); and third, a measure of the variability of spectral content over time, which has proved harder than the first two dimensions to characterize in precise computational terms with any consistency across studies. There is also variation between individuals in the extent to which each dimension is perceptually emphasized (McAdams et al., 1995).

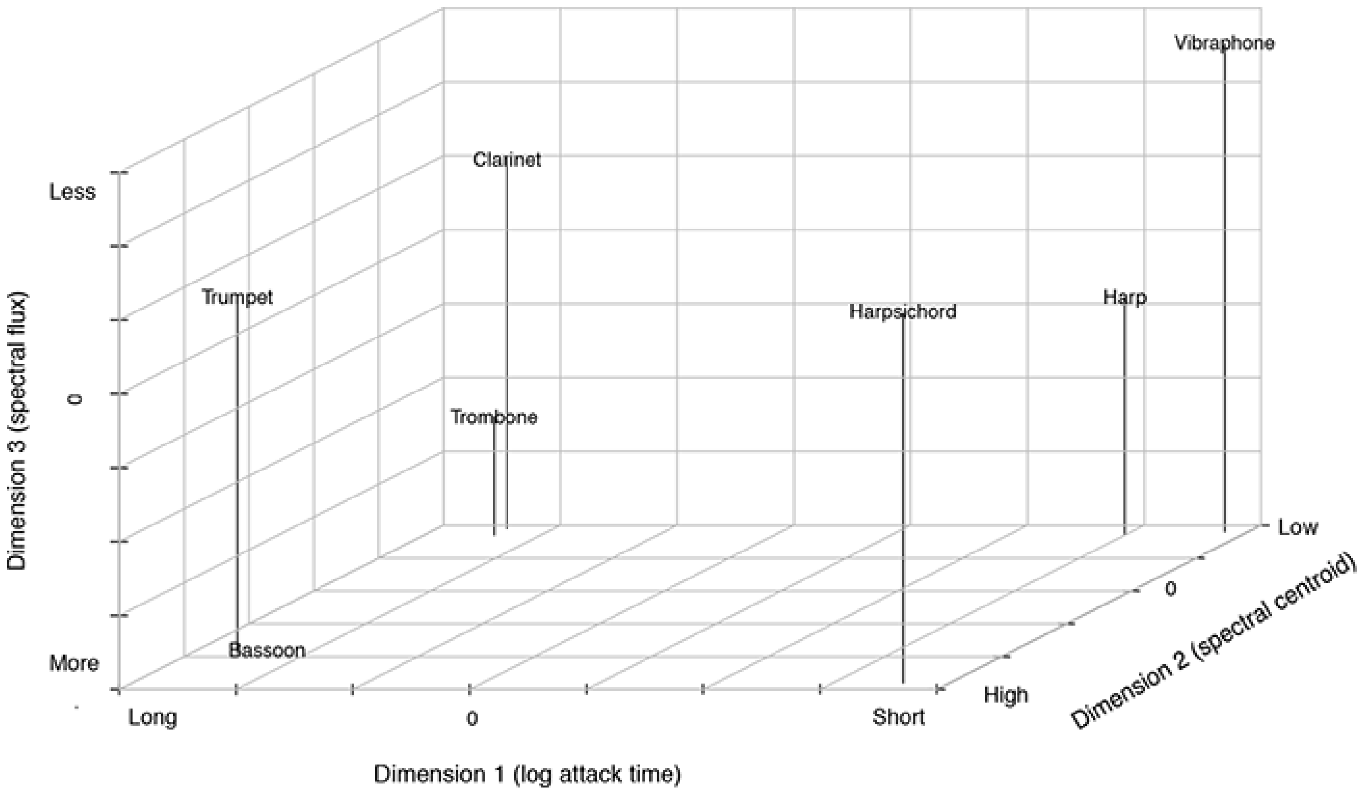


Figure 3. Perceptual timbre space resulting from multidimensional scaling of pairwise similarity ratings of instrumental tones. Selected instruments are shown for illustration and axes are labeled with the corresponding acoustic measures.

Source: Plotted from data reported in McAdams et al. (1995, Table 3, p. 185).

Given that the perception of timbre is influenced by spectral content, it is perhaps not surprising that experimental results have demonstrated significant overlap in psychological processing between pitch and timbre perception (Allen & Oxenham, 2014; Crowder, 1989; Marozeau & de Cheveigné, 2007; Pitt & Crowder, 1992; Steele & Williams, 2006). This is especially true for the spectral dimensions of timbre (Marozeau & de Cheveigné, 2007; Pitt & Crowder, 1992). Furthermore, neural processing of timbre shows significant overlap with the superior temporal cortical regions involved in pitch perception in both neuropsychological (Samson, 2003) and neuroimaging research (Allen et al., 2017).

While the majority of musical timbre research has focused on perception of individual instrumental tones, Alluri and Toiviainen (2010) investigated continuous timbre in polyphonic music with more than one instrument sounding simultaneously by taking ratings of timbral descriptors for 100 short musical excerpts. The results suggested that polyphonic timbre perception can be characterized by three underlying dimensions of activity, brightness, and fullness, each correlated with different acoustic features of the music. As well as allowing the identification and discrimination of instrumental sounds in music, polyphonic timbre can help listeners identify familiar songs from excerpts as short as 100 ms (Schellenberg et al., 1999) and familiar musical styles from excerpts as short as 250 ms (Gjerdingen & Perrott, 2008). Timbre also plays a key role in the perception of orchestrated music, where different instruments are combined for musical effect. Orchestral renditions of musical compositions are perceived as less tense than piano versions due to the timbral segregation of overlapping spectral content into separate perceptual streams, reducing perceived roughness (Paraskeva & McAdams, 1997), and sequences of timbral gestures (sudden or gradual increases or decreases in the number of instruments) having characteristic effects on experiences of emotional intensity (Goodchild & Mcadams, 2018).

Loudness

In addition to their pitch and timbre, musical events may also vary in their sound pressure leading to perceived differences in loudness. The psychophysical relationship between sound pressure and perceived loudness of musical tones is complex, being modulated by factors including tone duration, spectral bandwidth, frequency, and masking by surrounding sounds. The human auditory system shows greatest sensitivity between about 500 and 5,000 Hz and the lowest sensitivity below 500 Hz (i.e., low frequency sounds need to have much higher intensity than higher frequency sounds to be perceived as equally loud; Epstein & Marozeau, 2010; Moore, 2013). All else being equal, however, loudness shows an approximately exponential relationship with sound pressure level (measured in decibels, dB, itself a logarithmic transformation of sound pressure relative to a reference) such that a 10 dB increase produces a doubling of loudness (resulting in an overall power law relationship between sound pressure and loudness). Models of loudness generally assume that it is approximately proportional to the total auditory nerve activity evoked by a sound (Moore, 2013).

Loudness is often varied parametrically over time (i.e., smoothly in small progressive increments or decrements) within a piece of music, increasing to a peak in a crescendo and decreasing to a lull in a decrescendo, with sudden changes used for emphasis due to their perceptual salience. Although crescendi are more frequent and longer-lasting than decrescendi in Western music (Huron, 1991), listeners perceive both equally accurately in relation to the underlying changes of intensity in real musical stimuli (Geringer, 1995). Another important use of loudness in music is in *expressive dynamics* (see Pfordresher, 2022) whereby performers introduce expressive variations of intensity so as to accentuate important musical events or structures (Drake & Palmer, 1993) or communicate to the listener a sense of tension (Granot & Eitan, 2011) or emotional expression (Juslin, 2000).

Timing

Because music unfolds in time, the temporal properties of musical events—and sequences of such events—are fundamental to music perception. Research typically distinguishes four principal aspects of musical timing that impinge on perception: rhythm, meter, tempo, and expressive timing (see Figure 4).

Rhythm refers to the temporal intervals between perceived events commonly conceived as a sequence of interonset intervals (IOIs) or, more abstractly, the ratios between successive IOIs, allowing the same rhythm performed at different speeds to be perceived as equivalent. While Western musical scores convey precise (or quantized) musical timing, musical performances usually deviate from precise timing either intentionally, for expressive effect, or unintentionally, reflecting motor error (which within a certain tolerance is either not perceived or does not detract from the experience of the music). The question then arises of how listeners perceive the equivalence of two musical rhythms, when their precise timing can vary considerably. Research has demonstrated that listeners show a degree of categorical perception (Goldstone & Hendrickson, 2010; Liberman et al., 1957) of IOI ratios between musical intervals such that simple integer ratio IOIs (e.g., 1:1, 1:2, 1:1:2) are perceived, even when the ratios are varied continuously in time (Clarke, 1987; Desain & Honing, 2003; Jacoby & McDermott, 2017; Jacoby et al., 2021; Schulze, 1989). Perceptual boundaries between simple integer ratio rhythms are influenced by the metrical context in which the rhythm is heard (Desain & Honing, 2003).

Meter is a set of recurring temporal periodicities inferred by the perceptual system while listening to music (and other cultural phenomena such as poetry), which may be hierarchically embedded within periodic cycles of different relative durations (London, 2012). This can create a recurring pattern of metrically strong temporal locations, for which the start of lower-level metrical periods coincides with that of higher-level periods, and weaker temporal locations, where there is no such coincidence. In Western music theory, meter is defined by the time signature, which is often used as an underlying temporal framework for musical composition, specifying important embedded time periods such as the bar (or measure) and the tactus (or beat, pulse), usually corresponding to the rate at which one would tap along to the music. Therefore, while meter is often used generatively to create music, it does not exist in the musical surface and listeners must infer its presence and nature. This process of inference involves finding a meter in which strong metrical locations coincide with salient sounding events (Lerdahl & Jackendoff, 1983; Longuet-Higgins & Lee, 1984). Conversely, syncopation is the phenomenon whereby musical events fail to appear in relatively strong metrical temporal positions, especially following an event on a weaker metrical position. Some degree of syncopation can be tolerated by listeners without triggering a reassessment of their metrical interpretation. However, in a task in which listeners tap along to the tactus, highly syncopated rhythms lead to a resetting of the phase of the tapping to coincide with the syncopated events, while syncopated rhythms are also more difficult to reproduce and recognize (Fitch & Rosenfeld, 2007).

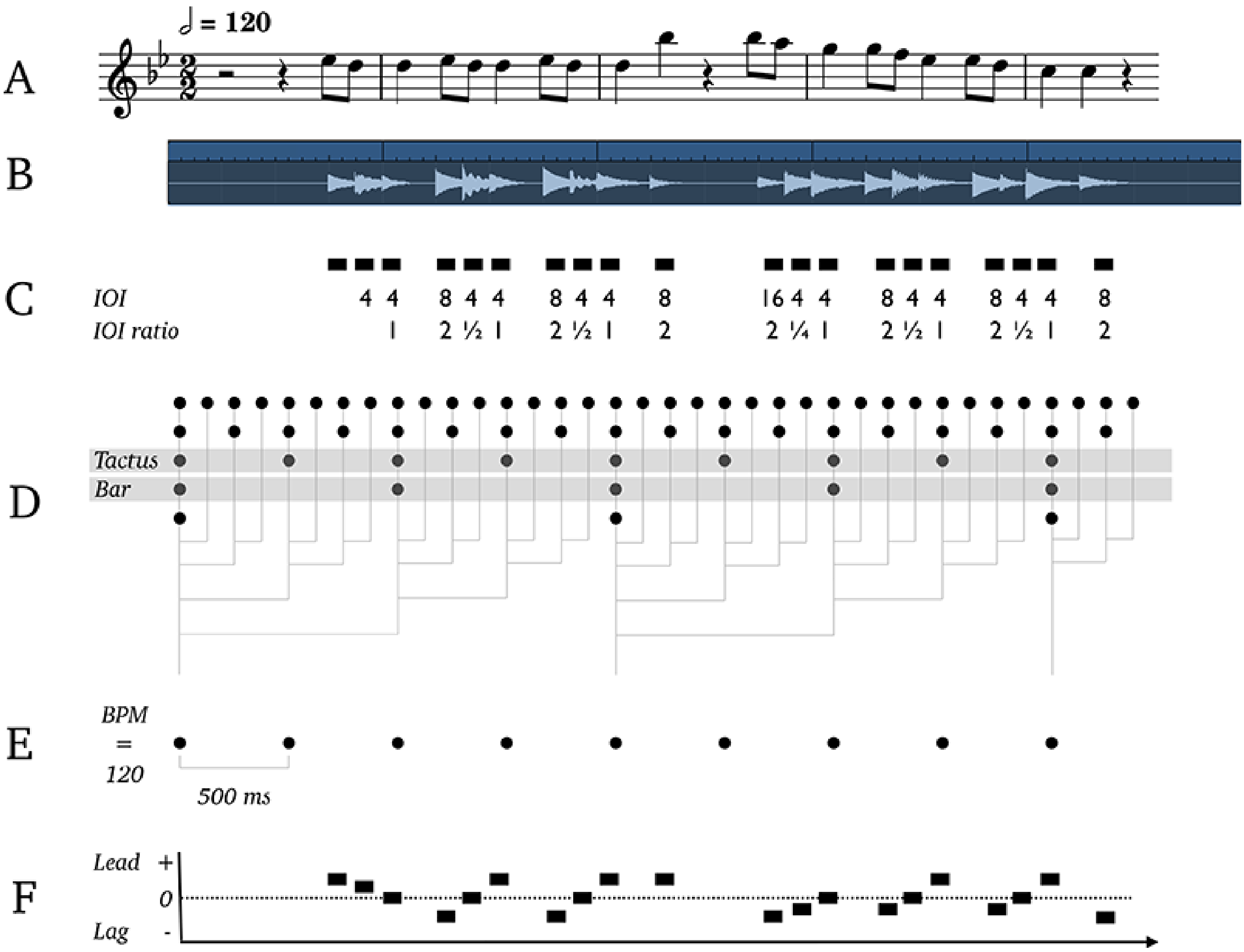


Figure 4. Representations of a musical stimulus (A and B) with corresponding psychological representations of the temporal properties of the stimulus (C–F). Rectangles represent perception of sounding events while circles represent inferred cyclical pulses. A: The score of the opening melody from the first movement of Mozart's Symphony No. 40 in G minor, as might be used by a musician to perform the piece. B: An audio representation of a performance (44,100 kHz, 16 bit). C: The rhythm with notes represented as black rectangles whose left edge is the onset of the note with interonset intervals (IOIs) given in arbitrary temporal units (semibreve or whole note = 32

units) along with ratios between successive IOIs. D: Metrical hierarchy where each black circle represents a pulse at a given metrical level and each horizontal line represents a binary subdivision at the next level of detail from a period of two bars (bottom) to a period of a quaver (or quarter note) with the tactus and bar levels highlighted. E: Tempo represented as the period of the tactus in absolute time units corresponding to a given number of beats per minute (BPM). F: Expressive timing corresponding to temporal deviations of the onset of each note from nonexpressive (deadpan, mechanical) performance timing as shown in C.

Source: With inspiration from Honing (2013, Figure 1, p. 370) and Lerdahl and Jackendoff (1983, Figure 4.9, p. 74).

Metrical inference depends not only on the presence or absence of events at metrically strong temporal locations but also on the salience of those events. Perceptual salience is determined by many factors including, most obviously, loudness. The first event of a rhythm is especially salient and tends to be interpreted as a tactus beat, all other things being equal (Lee, 1991). Furthermore, in the absence of any differences in loudness or pitch, listeners perceive relatively isolated events, the second of a cluster of two events or the first of a cluster of three or more events as being salient (Povel & Essens, 1985). Perceptual salience can also be increased by changes in melodic contour (Hannon et al., 2004) and harmony (Dawe et al., 1994). The psychological mechanisms underlying meter perception have been conceived in three ways (reviewed by van der Weij et al., 2023): first, as the application of symbolic rules (Lerdahl & Jackendoff, 1983; Longuet-Higgins, 1976; Longuet-Higgins & Lee, 1982; Longuet-Higgins & Steedman, 1971; Povel & Essens, 1985; Temperley, 2001); second, as banks of neural oscillators which attune to the salient periodicities present in a musical stimulus (Large & Jones, 1999; Large & Kolen, 1994; Large & Palmer, 2002; Large et al., 2015; McAuley, 1995; Tichko & Large, 2019); third, as a process of probabilistically inferring the most likely metrical interpretation for a given rhythm, given a collection of stored metrical templates, acquired through prior musical experience (Kaplan et al., 2022; Temperley, 2007, 2009; van der Weij et al., 2017).

Although perception of one influences perception of the other, meter and rhythm are distinct both conceptually (meter may be perceived without rhythm, a rhythm may conflict with the meter) and psychologically, involving different neural mechanisms. Perception of metrical rhythms invokes activation in a striato-thalamo-cortical network including the basal ganglia, thalamus, supplementary motor area (SMA), and premotor cortex (Grahn & Brett, 2007; Teki et al., 2011) and is impaired in patients with Parkinson's disease, characterized by degeneration of cells in the substantia nigra (Grahn & Brett, 2009). Perception of nonmetrical rhythms invokes activation in a separate network including the cerebellum and the inferior olive (Teki et al., 2011) and is impaired in patients with cerebellar degeneration (Grube et al., 2010). Results such as these have given rise to the theory that meter perception involves simulation of planned periodic movement in the SMA in order to predict the timing of future beats (Cannon & Patel, 2020; Patel & Iversen, 2014).

Tempo relates metrical periods to absolute timing, commonly measured as the number of occurrences of a particular metrical unit (e.g., the tactus or beat) per minute (beats per minute or BPM). The shortest possible metrical unit has been estimated at 100 ms based on the ability to tap in synchrony with an auditory pulse (Repp, 2003). When asked to tap periodically at a comfortable rate or express preference for stimuli presented at different tempi, listeners show a preferred tempo of about 120–130 BPM (450–500 ms interval). However, this slows with age from a preferred tempo of around 200 BPM (300 ms) for young children to 85 BPM (700 ms) for older adults (McAuley et al., 2006). Preferred tempo also shows an effect of anthropometric factors with taller individuals preferring slower tempi (Dahl et al., 2014; Todd et al., 2007).

Musical performers systematically vary the timing of musical events for expressive effect, usually to accentuate some aspect of musical structure (see Pfordresher, 2022). For example, musicians vary the duration of performed notes to emphasize the metrical structure to listeners (Drake & Palmer, 1993). Furthermore, performers may introduce expressive articulation of musical notes, varying the duration and amplitude envelope of the sound either expressively or according to instructions in a score (e.g., slur, staccato, legato, portato). These aspects of timing are perceived by listeners (Hofmann & Goebl, 2014).

Perception of Musical Structure

A piece of music consists of a collection of potentially overlapping auditory events, each varying in the attributes introduced in "Perceived Musical Attributes." Listeners perceive structure in such collections of events and often this structure is designed by the composer/performer to be perceptible by a suitably enculturated listener. Perception of musical structure is often conceived as a psychological process of grouping events together into larger-scale structures, representing differing degrees of relatedness between events and groups of events that unfold in time. Three kinds of grouping structures are especially important: first, the segregation of simultaneously occurring parallel streams of information within a piece of music, the most prominent example being a melody; second, the fusion of simultaneously sounding events into a single percept, chords being the canonical example; and third, the segmentation of music into sequential groups of events such as motifs, phrases, and sections. Listeners may also perceive larger-scale structures in musical works related to tonal or thematic relationships between different parts of a composition.

The vast majority of research on perception of musical structure has been conducted on Western tonal music. However, there is some research on perception of structure in nontonal Western music (e.g., Dibben, 1999; Krumhansl et al., 1987; Olsen et al., 2016) and non-Western musical cultures (e.g., Ayari & McAdams, 2003; Mungan et al., 2017).

Fusion

At any one point in a piece of music, more than one event may be sounding at the same time. Under certain conditions, these simultaneously sounding auditory events can fuse into a single perceptual object, the most obvious example being a musical chord. Fusion of simultaneous sounds has been shown to be influenced by frequency separation, harmonicity of the individual frequency components, onset and offset asynchrony, common frequency or amplitude modulation, pitch separation, and, as a secondary effect, spatial location via interaural timing differences (Bregman, 1990; Darwin, 1997; Deutsch, 2013b). These effects are consistent with the individual sounds making up the fused percept being generated by the same environmental source.

There has been considerable interest in the question of which sounds fuse to produce *consonant* or *dissonant* percepts, which are respectively pleasing and displeasing to the ear. Rather than being a unitary phenomenon, consonance appears to have at least three contributing components (Harrison & Pearce, 2020). The first is perceptual roughness resulting from interference between the different frequencies present in the fused sound (both fundamental frequencies and harmonics of the individual complex tones involved) which stimulate partially overlapping portions of the basilar membrane, causing a sensation of rapid beating (Hutchinson & Knopoff, 1978; Plomp & Levelt, 1965; von Helmholtz, 1863). The second component relates to the harmonicity or periodicity of the sound, such that sounds having a (possibly incomplete) series of partials that form integer multiples of a fundamental frequency are more harmonic (in the frequency domain) or periodic (in the time domain) and are typically perceived as more consonant (de Cheveigné, 2005; Terhardt, 1974). Finally, cultural experience can have a significant effect due to learned familiarity with particular chords (McLachlan et al., 2013), chord distributions (Harrison & Pearce, 2020), tonal regularities (Johnson-Laird et al., 2012), or levels of harmonicity (McDermott et al., 2016). Strikingly, the Amazonian Tsimané tribe of lowland Bolivia show complete indifference to variations of harmonicity, unlike Western participants and Bolivian city-dwellers, despite being able to perceive harmonicity and showing aversion to roughness just like the other participant groups (McDermott et al., 2016).

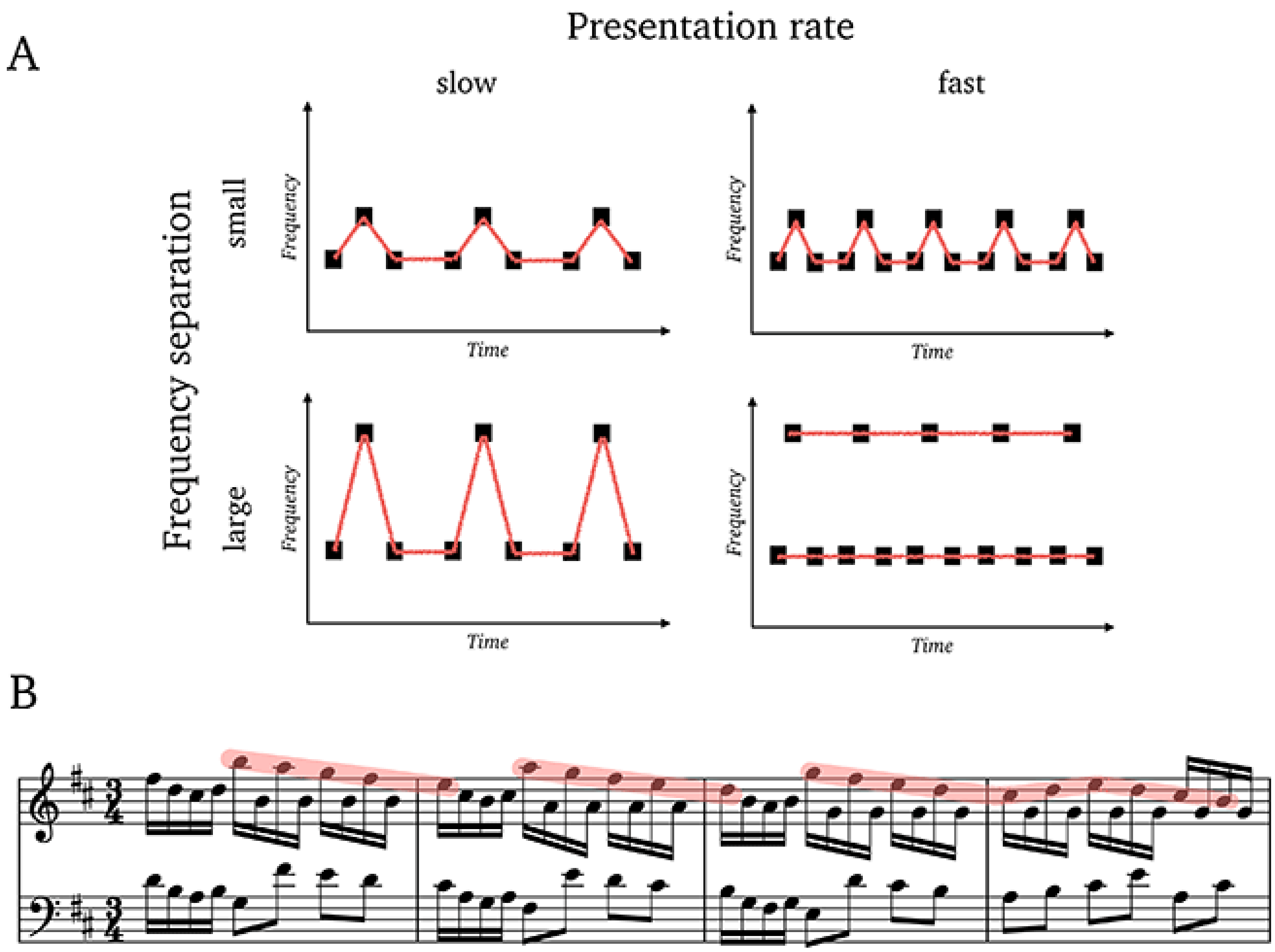


Figure 5. Stream segregation. A: An illustration of auditory streams formed when listening to an ABA pattern formed of low (A) and high (B) tones (van Noorden, 1975). When the presentation rate is slow or the frequency separation between high and low tones is small, a single stream with a characteristic galloping rhythm (ba-da-dum ba-da-dum . . .) is heard. When the presentation rate is sufficiently fast and the frequency separation sufficiently large, two distinct streams are heard, one containing isochronous high tones and the other isochronous low tones. B: An illustration of stream segregation in Prelude 15 from Bach's Well-Tempered Clavier (Book II, bars 9–12).

Stream Segregation

Music may contain multiple parallel streams of information, often but not always corresponding to different performers or instruments. Faced with such a complex auditory stimulus, the perceptual system must integrate some sounds into distinct auditory objects segregating them from other sounds corresponding to other auditory objects (Deutsch, 2013b). This is a specific instance of the phenomenon of auditory scene analysis (Bregman, 1990), which allows the auditory system to make sense of complex auditory scenes by attempting to identify and separate auditory objects corresponding to different sound-producing objects in the environment. Auditory scene analysis has been investigated in two principal ways: first, examining the conditions under which simultaneous sounds fuse into a single percept or segregate into separate percepts (see "Fusion"), and second, examining the conditions under which interleaved auditory sequences segregate into separate streams (see Figure 5). The two phenomena are related since the degree to which two sounds fuse depends on whether or not they already form part of the same sequential stream (Bregman & Pinker, 1978).

The segregation of interleaved auditory sequences is influenced by the rate of presentation and frequency separation (Bregman & Campbell, 1971; Demany, 1982; Miller & Heise, 1950; van Noorden, 1975) as well as timbral similarity (Iverson, 1995; Wessel, 1979) and spatial location (Deutsch, 1979). Stream segregation is a dynamic process in that the auditory system builds up evidence over time before perceiving a segregated stream. Stream segregation and fusion may be seen as competitive processes in which different sequential streams and simultaneous fused objects can compete to capture individual sounds. These bottom-up processes of auditory scene analysis are thought to combine with top-down schematic effects of attention and learning (Dowling, 1973; Dowling et al., 1987). Some components of a sound object may also be perceived as continuations of the preceding auditory context and therefore segregated from the other components of the scene (Bregman, 1990).

Computational models of auditory stream segregation emphasize prediction, assigning auditory events to the stream whose predicted extension provides the closest match (Bo et al., 2013; Elhilali & Shamma, 2008; Skerritt-Davis & Elhilali, 2021). However, these models do not incorporate top-down schematic influences on stream segregation and it is not yet clear whether they can generalize successfully beyond simple examples such as the ABA paradigm (van Noorden, 1975) to complex examples of stream segregation in polyphonic music (see Figure 5). A different approach to modeling musical stream segregation invokes symbolic rules based on temporal and pitch proximity (e.g., Cambouropoulos, 2008), taking inspiration from the observation that auditory scene analysis can be related systematically to Western principles of polyphonic composition (the so-called rules of voice-leading; Huron, 2001). However, such rule-based models have yet to be experimentally evaluated as models of polyphonic music perception.

Sequential Segmentation

As well as being integrated into simultaneous parallel streams and fused auditory objects such as chords, musical events are also grouped into hierarchically embedded sequential segments such as motifs, phrases, sections, and parts (Lerdahl & Jackendoff, 1983). Such segments show parallels with phrases, sentences, paragraphs, or sections in written language and may or may not coincide with metrical periods discussed above (see Figure 6). The perception of such sequential segments is usually conceived as a process of identifying the boundaries between segments. Early models of this perceptual process invoked rules inspired by Gestalt psychology that predict boundaries at points of low temporal proximity (i.e., rests or relatively long events) and low similarity between consecutive musical events in some musical parameter (e.g., pitch, duration, loudness, articulation, or timbre; Cambouropoulos, 2001; Lerdahl & Jackendoff, 1983; Temperley, 2001). There is experimental evidence that perceived segmentation boundaries can be predicted by such principles, especially those related to temporal proximity (Deliège, 1987; Frankland & Cohen, 2004; Peretz, 1989) and timbral or dynamic change (Deliège, 1987).

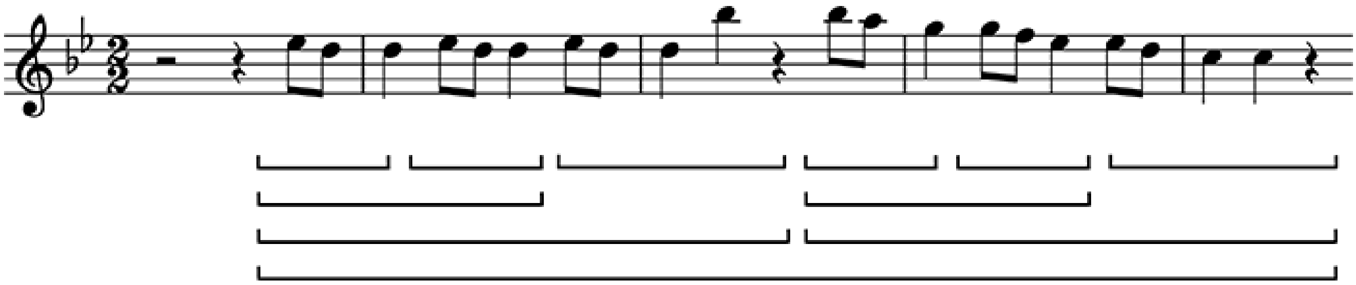


Figure 6. Sequential segmentation of the opening melody from the first movement of Mozart's Symphony No. 40 in G minor.

Source: After Lerdahl and Jackendoff (1983, Figure 3.1).

An alternative theory is that segmentation boundaries are derived from probabilistic prediction of music based on statistical learning, such that the first event in a perceived segment is distinguished from the remaining within-segment events by a lower conditional probability based on the preceding context, which crosses the segment boundary from the event in question to the previous segment. Infants and adults show above-chance recognition of such segments after implicit statistical learning through passive exposure to artificially constructed continuous auditory sequences lacking any other cues to segment boundaries (Frost et al., 2019; Saffran & Kirkham, 2018). The two psychological mechanisms are not necessarily mutually exclusive and have been shown to operate in tandem, enhancing the perception of segments when they coincide and canceling out the perception of segments when they conflict (Tillmann & McAdams, 2004).

The majority of theoretical and empirical research on sequential grouping has focused on melodic rather than polyphonic music and phrase-level sequential segmentation rather than the perception of larger-scale, hierarchical segmentation of music.

Tonal Structure

There is empirical evidence for the perception of tonal structure in music, referring to relationships between musical events (e.g., notes and chords) in terms of their tonal interpretation (e.g., tonic, dominant—see "Pitch"), which are generally thought to be learned through exposure to a musical style. At a very local level, listeners perceive the similarity between pairs of tones or chords in a tonal context asymmetrically, such that similarity is greater when the more tonally stable item is presented first (Bharucha & Krumhansl, 1983; Krumhansl, Bharucha, & Castellano, 1982). Furthermore, when listening to short chord sequences in a tonal context, listeners are surprised by out-of-key chords (Tillmann et al., 1998) and show stronger expectations for chords from related keys than distant keys (Krumhansl, Bharucha, & Castellano, 1982). Within musical phrases, Western listeners show expectations for immediate single chord continuations of incomplete harmonic sequences that are common in the Western tonal idiom to which they have long-term exposure (Bigand et al., 1999) and these effects appear to be stronger than very local effects based on sensory memory (Bigand et al., 2003). At the larger timescale of a pair of phrases (about 10 chords presented over 5 seconds), there is subjective and electrophysiological evidence of greater perceived closure when the second phrase is related to the first via nonsequential tonal dependencies (Koelsch et al., 2013). The vast majority of this research has been conducted on Western diatonic tonal systems and there is less comparable research on the perception of tonal systems from other musical cultures.

Musical Syntax

The tonal relationships introduced in "Tonal Structure" can be conceived as style-specific syntactic dependencies between musical structures. Like syntactic relations in other domains (e.g., natural languages or programming languages), these dependencies can be characterized in terms of the degree of constraint they impose on syntactically well-formed sequences of structures within the style. The generative system capturing these constraints is often referred to as the syntax of a musical style (Pearce & Rohrmeier, 2018; Rohrmeier & Pearce, 2018) and there is a significant tradition of research examining listeners' sensitivity to musical syntax. While there is general agreement that listeners perceive local relationships between adjacent (sequences of) musical structures (see "Tonal Structure" and "Musical Expectations"), several influential theories of music cognition go beyond the local level to hypothesize that listeners infer a cognitive representation of hierarchical relationships between nonadjacent parts of a piece of music resulting in something not entirely unlike a parse tree in natural language processing (Lerdahl & Jackendoff, 1983; Schenker, 1935/1979) and grammatical formalisms derived from natural language processing have been applied to modeling musical syntax (Johnson-Laird, 1991; Rohrmeier, 2011; Steedman, 1996). Just like theories of natural language syntax, theories of musical syntax provide rules for arranging items (sounds and groups of sounds) into their possible combinations within the corresponding musical style. These syntactic rules are generally assumed to be acquired through implicit learning of stylistic regularities during long-term exposure to music in a given tonal idiom.

There is no doubt that hierarchical structure exists in many pieces of music but evidence for the perception of this structure over longer timescales is mixed. Musicians are above chance in choosing the correct reduction (in which structurally important events are retained and unimportant events removed) of a piece of tonal music up to 16 bars long, where correct and incorrect reductions were composed by a skilled musicologist (Dibben, 1994). Although weaker performance was observed by individuals with less musical training on a similar (but slightly easier) task, these listeners did find structure-preserving but harmony-differing pairs of musical fragments slightly more similar than structure-differing but harmony-preserving pairs (Serafine et al., 1989). However, other research has questioned the extent to which listeners can perceive large-scale tonal structure. Even musically trained listeners fail to detect manipulations of tonal structure in rearranged versions of music by Mozart and Handel (Karno & Konečni, 1992; Marvin & Brinkman, 1999) and tension ratings for long chord sequences are dominated by local tonal relations, showing only weak effects of large-scale tonal structure (Bigand & Parncutt, 1999). Furthermore, nonmusicians perceived no differences in expressivity or coherence when pieces of music by Bach and Mozart were rendered with their phrases in reverse order (Tillmann & Bigand, 1996). These examples suggest that further research is required to determine the psychological mechanisms involved in the perception of large-scale tonal structure in music.

Thematic Structure

A striking feature of music is the amount of repetition it contains (Kivy, 2017; Margulis, 2014), often to an extent that would be intolerable in other cultural domains; yet in music, listeners relish the repetition, development, and recapitulation of musical content. Thematic structure in music relates to the repetition (partial or exact) and variation of musical passages over time within a piece of music, such that the individual parts are perceived to make up a coherent and unified whole. Thematic structure can relate to rhythmic as well as pitch content and is orthogonal to tonal structure in principle (nontonal music can possess thematic structure and tonal music need not be strongly thematic) but often coincides with it and has not been carefully delineated from tonal structure in the majority of empirical research.

Research on the perception of thematic structure has proved somewhat inconclusive. Listeners are sensitive to repeated structure in music (Margulis, 2014) and some research has found that listeners perceive original versions of compositions as being more coherent or unified than rearranged versions (Lalitte & Bigand, 2006; Tan & Spackman, 2005). Furthermore, when asked to rearrange sections of music according to perceived coherence, participants show sensitivity to the relative positioning of thematically significant sections (Granot & Jacoby, 2011a, 2011b), though in some cases this is limited to musicians (Deliège et al., 1997). However, other findings point to a relative inability of listeners to distinguish between original and reordered versions of musical compositions (Eitan & Granot, 2008; Karno & Konečni, 1992; Marvin & Brinkman, 1999; Rolison & Edworthy, 2012).

In general, experimental research on large-scale structure in music is complicated by the need to present entire compositions, which is difficult to balance against participant fatigue and the sample sizes required for robust statistical analysis.

Cognitive Processing of Music

The psychological operations described in "Perception of Musical Structure" rely on general cognitive mechanisms such as attention, memory, and expectation.

Memory for Music

Mainstream theories of musical memory posit a number of different stores, operating at different timescales and at different stages of auditory processing, each with different underlying neural substrates (Cowan, 2008; Halpern & Bartlett, 2010; Snyder, 2000, 2016). At the lowest level, *echoic memory* consists of a sensory image covering unprocessed, detailed, and immediate auditory input within a time window of up to a few seconds. Sensory information is then passed into auditory *short-term memory* in a form that integrates different features (pitch, timbre, loudness) and chunks elements together into sequential groups, covering time windows of up to tens of seconds. An example would be *dynamic knowledge* of repeated patterns within a piece of music that are perceived and stored during listening (Huron, 2006). Long-term memory for music has potentially indefinite capacity and temporal span. Different forms of long-term memory have been distinguished based on the kind of knowledge stored (Bharucha, 1987; Huron, 2006): first, *schematic knowledge* of the structural regularities present in the lifetime exposure of an individual to a musical style, which can be compared loosely to implicitly acquired syntactic knowledge of a natural language; second, *veridical knowledge* of particular pieces of familiar music, which are recognized when heard again, often along with semantic details such as the title, composer/performer, and so on; third, *episodic knowledge*, relating to the experience of a particular piece of music at a particular time and place in the past, as tested in recognition memory experiments (Crowder, 1993).

The existence of echoic memory has been demonstrated in experiments that show declining memory for detailed sensory information from an auditory scene (e.g., information about a spatial location) but not for higher-level more abstract information (e.g., the identity of the sounds) over a short retention interval of 4 seconds (Darwin & Turvey, 1972).

Short-term memory for music is typically investigated by asking for same-different judgments of pairs of consecutively presented short melodic fragments or recognition memory judgments for a sequence of melodic fragments with variable intervals between reoccurrences. When there is no delay between the target and a transposed comparison melody, tonality and pitch contour are encoded such that comparison melodies with different intervals are confused with the target if they have the same key and contour whereas stimuli with different contour and key are correctly perceived as different from the target (Dowling, 1978). Somewhat paradoxically, on longer timescales of up to 52 seconds, listeners tend to use more detailed pitch interval representations of melodies in making recognition judgments, perhaps reflecting the intervening experience of other stimuli in different keys (Dowling, 1991). Musicians appear to show slightly better short-term memory for melodies than nonmusicians, perhaps due to enhanced encoding of musical features (Dowling, 1978, 1991) but ageing has only a small impact on performance (Halpern et al., 1995; Meinz, 2000).

Long-term memory for music is typically investigated by asking participants to give recognition memory judgments (old/new) for a set of melodies some of which had been presented 10–30 minutes earlier (old) whereas others had not (new). The melodies used as stimuli may be veridically *familiar*—usually nursery rhymes or festive songs that are well-known within a culture—or *unfamiliar*. The experimental evidence points to recognition judgments being based on a generalized sense of recognizability that combines prior familiarity with experimental oldness. For example, familiar music is more likely to be incorrectly recognized as old than unfamiliar music, unless the familiar and unfamiliar music are presented in separate blocks, allowing a more stringent recognition criterion to be used for familiar stimuli (Bartlett et al., 1995). Furthermore, when asked whether they heard an unfamiliar tune today or yesterday, it is remarkable that participants were more likely to respond “today” for a melody heard three times yesterday than a melody heard once today (McAuley et al., 2004). For familiar tunes, by contrast, performance was much better and correlated with the nameability of the song, suggesting an effect of episodic memory (McAuley et al., 2004) which was also observed in older adults’ long-term memory for familiar music (Bartlett et al., 1995). In general, long-term memory for music does not seem to differ between musicians and nonmusicians (Demorest et al., 2008; Halpern et al., 1995; McAuley et al., 2004) but healthy older adults show slightly impaired long-term memory for music (Bartlett et al., 1995; Halpern et al., 1995).

Musical Expectations

The ability to anticipate the future is critical to many areas of psychological processing, conferring an evolutionary advantage by allowing organisms to tailor and regulate their processing of sensory events, prepare appropriate reactions, and make adaptive choices (Dennett, 1991; Schultz et al., 1997). According to one influential line of theoretical and empirical research, predictions play an important role in perception and action, allowing an organism to infer the environmental causes of sensory events (Clark, 2013; Friston, 2010). Expectations are also thought to play an important role in music perception (Huron, 2006; Krumhansl, 1990; Meyer, 1956; Narmour, 1990), where the listener must track the signal through time as it varies simultaneously along multiple dimensions. Empirical research has focused on expectations for the pitch, timing, and harmony of musical events, addressing two research questions: Where do expectations come from? What are the effects of confirmation and violation of expectations?

Regarding the origins of expectations, there has been debate over whether expectations reflect fixed implications of the local prior musical context or probabilistic regularities acquired through statistical learning over a range of time periods (Huron, 2006; Pearce, 2018). Melodic expectations can be described in terms of simple principles such as pitch proximity (an expectation for a note to form a small pitch interval with its predecessor) or pitch reversal (an expectation for a large pitch interval to be followed by a contour change and a smaller interval; Krumhansl, 1995; Schellenberg, 1997) which have been suggested to represent universal, obligatory properties of the human auditory system (Narmour, 1990). However, music itself also follows these principles (Thompson & Stainton, 1998), perhaps reflecting constraints of performance (Russo & Cuddy, 1999; Tierney et al., 2011), so these regularities could be learned through musical exposure. Consistent with this hypothesis, it has proved possible to simulate accurately listeners' melodic expectations using computational models that generate probabilistic predictions derived from statistical learning, reflecting both lifelong exposure to music and sensitivity to repeated patterns within a piece of music (Pearce, 2018). Harmonic expectations have been accounted for in terms of effects of sensory echoic memory, with expectedness related to the extent of overlap in spectral content between a musical sonority and sonorities held in echoic memory over a period of a few seconds (Bigand et al., 2014). However, in many cases, such sensory effects are confounded with top-down cognitive effects of the kinds acquired through prior statistical learning; experiments in which the two effects are explicitly compared show very little effect of sensory memory (Bigand et al., 2003; Goldman et al., 2021; Sears et al., 2019).

Expectations for the timing of musical events has typically focused on predictions derived from an induced metrical interpretation of a musical stimulus (Large & Palmer, 2002; Palmer & Krumhansl, 1990). However, research has also investigated expectations derived from the learning of rhythmic patterns both within and across stimuli and how these are differentiated from metrical expectations (Bouwer et al., 2020). In corpus analyses of Western tonal music, pitch structure and temporal structure are typically correlated in the sense that tonally stable notes tend to co-occur with metrically stable notes of longer duration (Prince & Schmuckler, 2014) but there is debate in the literature about the extent to which listeners represent and process pitch and timing of musical events dependently. Participants in empirical experiments often appear to process pitch and timing independently (Palmer & Krumhansl, 1987) with pitch expectations taking priority over temporal expectations, unless attention is explicitly directed toward the latter (Prince et al., 2009). However, in experimental situations where the stimuli and task emphasize relationships between pitch and time, participants show unified processing of pitch and time, especially when repeated presentation of the stimuli allows for more sophisticated unified dimensions to develop (Boltz, 1999).

Expectation has been used as a vehicle for examining relationships between the psychological processing of music and language in experiments that present an auditory sequence of musical events simultaneously with a visual sequence of words (Carrus et al., 2013; Fedorenko et al., 2009; Koelsch et al., 2005; Slevc et al., 2009; Steinbeis & Koelsch, 2007). The results suggest interactive effects of musical expectation and linguistic grammaticality on word reading times and event-related potential (ERP) responses to unexpected musical events. This implies that the psychological mechanisms involved in musical expectation and processing of linguistic syntax are not entirely independent (Patel, 2003). However, the precise processes involved in the interaction remain unclear with hierarchical sequence processing, attention, implicit learning, working memory, and cognitive control all having been proposed as candidate resources that are shared between the processing of music and language (Slevc & Okada, 2015). Furthermore, the effects may not be specific to syntax with some studies also finding interactive effects with semantic manipulations of language (Perruchet & Poulin-Charronnat, 2013; Poulin-Charronnat et al., 2005; Steinbeis & Koelsch, 2007).

Regarding the consequences of expectations, it has been proposed that confirmations and violations of expectation should be rewarding and penalizing respectively, since they indicate the success or failure of a listener's predictive model (Huron, 2006). As a result, expectation is thought to bear a close psychological relationship with the emotional and aesthetic experience of music. Musical events that violate a listener's expectations increase arousal thereby creating a sense of tension while events that confirm expectations resolve that tension (Egermann et al., 2013; Gingras et al., 2015; Steinbeis et al., 2006). This fluctuation of tension and arousal is thought to be an important contributor to the aesthetic experience of music (Meyer, 1956) with empirical results suggesting that listeners find music most pleasurable when it provides intermediate degrees of unpredictability (Cheung et al., 2019; Gold et al., 2019).

The Emotional and Aesthetic Experience of Music

Emotional Experience

The emotional and aesthetic experience of music are usually conceived as a process of communication between composers and performers, who express emotional meaning in the musical signal, and listeners, who comprehend the meaning embedded in the signal. Individuals can recognize an emotion expressed by a piece of music, a process usually referred to as *emotion recognition*, and may also feel an emotion induced by the music, referred to as *emotion induction*, which may or may not correspond to the recognized emotion (Juslin & Västfjäll, 2008). Recognized and induced emotions are usually assessed using either categorical (e.g., happiness, sadness, peacefulness, and anger, Ekman, 1992) or dimensional (e.g., valence and arousal, Russell, 1980) conceptions of emotion, which often correspond closely with each other (Vieillard et al., 2008). Although valence (positive vs. negative) is often conceptualized in terms of pleasure, it is important to note that in music—and other artistic domains—pleasure is distinct since, for example, it is perfectly possible to take pleasure in an induced state of sadness. For Western listeners, the expression and recognition of emotion depends on the tempo and tonal mode of the music (Vieillard et al., 2008), with effects of tempo apparent for 5-year-olds and effects of mode emerging in later childhood (Dalla Bella et al., 2001) and also the timing, dynamics, timbre, and articulation of the expressive performance (Juslin, 2000; Pfordresher, 2022; Vieillard et al., 2008). The accuracy of emotion recognition improves only slightly (Castro & Lima, 2014) or negligibly (Bigand et al., 2005) with increases in musical training. There is evidence that the recognition of emotional meaning expressed by music can prime affective evaluation of valenced words, suggesting a link between recognition of emotion in music and language (Goerlich et al., 2012; Koelsch et al., 2004; Tenderini et al., 2022).

There is evidence that emotion induction or mood regulation is an important motivation for many individuals to listen to music (Juslin et al., 2008). There are thought to be several distinct psychological processes underlying emotion induction by music (Juslin, 2019; Juslin & Västfjäll, 2008). These include: responses to extremes or fast changes in basic acoustic features such as loudness, dissonance, or timbre (Blood et al., 1999; Koelsch et al., 2006); emotions arising as a result of behavioral or physiological synchronization with musical beats; conditioned associations between certain pieces of music and valence (e.g., a TV theme tune, Hepper, 1991); musical mimicry of the linguistic or bodily expression of emotion (e.g., slow tempi and descending pitch contours might be associated with sadness); evoked imagery (e.g., a piece of program music might evoke imagery of a storm or of a flock of birds); and musical expectation (discussed in "Musical Expectations"). It is possible for more than one mechanism to be at play while listening to a given piece of music and the invocation of each mechanism may vary between individuals depending on their experience with particular pieces of music and musical enculturation.

There is evidence for some cross-cultural consistency in musical emotion recognition (Balkwill & Thompson, 1999; Balkwill et al., 2004; Fritz et al., 2009) although also divergences, especially for cultures whose musical traditions are more dissimilar (Fritz et al., 2009). There is much less evidence for cross-cultural consistency in emotion induction. Fritz et al. (2009) reported that the isolated Mafa tribe of northwestern Cameroon show valence responses to manipulations of consonance that were conceptually similar (though weaker) than those of Western listeners. However, the opposite has been reported for the Tsimané tribe from the Bolivian rainforest (McDermott et al., 2016) and the Khowar and Kalash tribes native to northwest Pakistan (Lahdelma & Athanasopoulos, 2021). Furthermore, Egermann et al. (2015) found no evidence of any consistency in valence responses between Western listeners and Mbenzele Pygmies from the Congolese rainforest.

Aesthetic Appreciation

Aesthetic appreciation of music is generally considered a distinct (albeit related) psychological process from emotion recognition and induction; for example, listeners often find pleasure in listening to sad songs (Vuoskoski et al., 2012). Aesthetic appreciation involves an experience of pleasure or satisfaction while listening to a piece of music, arising from directing attention to the form and content of a piece of music (including its structure, expressive content, nonexpressive perceptual qualities, narrative-dramatic content, and referential meaning), sometimes accompanied by a judgment of the music based on the experience (Levinson, 2009), which might itself induce so-called aesthetic emotions (Juslin, 2019). The subjective pleasure experienced when listening to music is sometimes accompanied by physiological responses such as chills (de Fleurian & Pearce, 2021) which are associated with activation in neural reward systems, including the ventral striatum (Salimpoor et al., 2011).

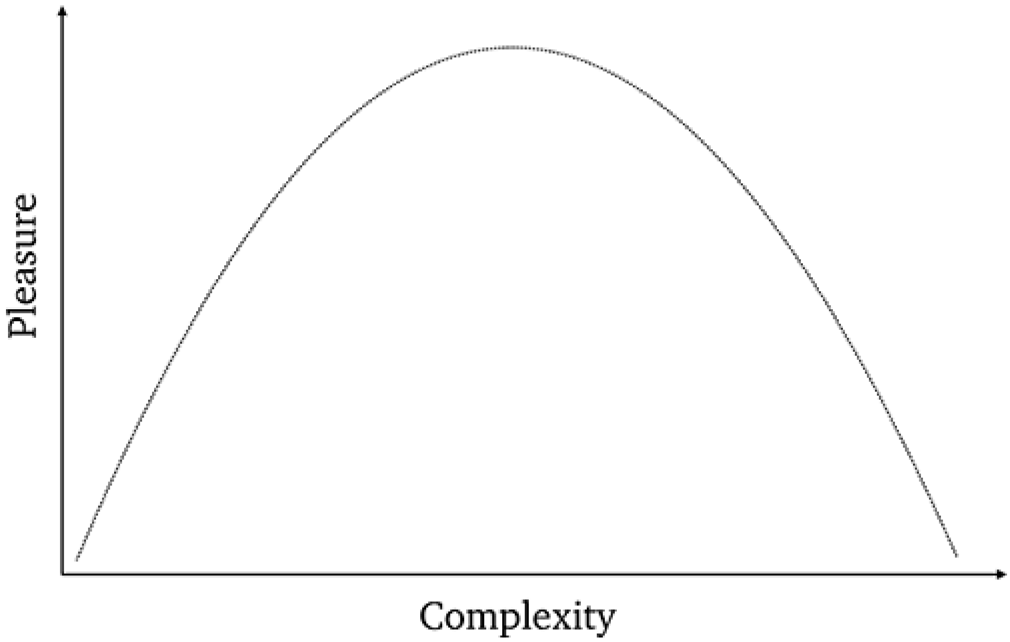


Figure 7. A hypothesized inverted-U shaped relationship (or Wundt curve) between stimulus complexity and experienced pleasure.

Aesthetic experience of music depends on interactions between the structure and content of the music, the psychological makeup of the listener, and the listening context. Regarding the music itself, a significant body of research has examined the hypothesis that pleasure bears an inverted-U shaped relationship with perceived complexity (also known as the Wundt curve, see Figure 7), which is sometimes operationalized in terms of familiarity or predictability (Berlyne, 1960; Wundt, 1874). The empirical evidence is generally consistent with a Wundt curve and it has been suggested that extreme levels of complexity induce suboptimal levels of arousal or provide insufficient interest or learning potential to be pleasurable (Chmiel & Schubert, 2017). Regarding the listener, there is evidence that high trait empathy is related to finding pleasure in listening to sad music (Vuoskoski et al., 2012) while individuals with high openness to experience are more likely to report being moved by music (McCrae, 2007; Nusbaum & Silvia, 2011). As an example of the effects of context, individuals are more likely to choose to listen to a fast and loud rendition of a song after exercising and a slow and soft rendition after relaxing (North & Hargreaves, 2000).

Other approaches to understanding aesthetic appreciation of music have emphasized development of self-identity (Macdonald et al., 2002) or appreciation of musical sources, including the personal attributes of musicians and the sociopolitical, historical, and cultural contexts of music making (Thompson et al., 2023).

The Acquisition and Development of Music Perception

Like language, music is a cultural phenomenon and requires exposure to develop culturally appropriate perception and comprehension. The production and perception of music are more asymmetrical than for language, such that perceptual abilities tend to outweigh production abilities to a greater extent. For this reason, it is important to distinguish enculturation through listening (actively or passively) and through explicit musical training.

Innate Abilities

It is practically impossible to demonstrate conclusively innate psychological faculties that are specific to music perception above and beyond general-purpose auditory perceptual processes. Cross-cultural universality of a psychological faculty related to music perception does not imply that it is innate while abilities that are present in neonatal infants or prenatal fetuses could have been acquired through in utero experience. Indeed, neonates and prenatal fetuses have been shown to produce responses to a TV theme tune that are both specific to that tune and to babies of mothers who watched the corresponding TV show (Hepper, 1991). However, abilities present in neonates can be conceived as candidates for innate musical perceptual abilities, one such ability being the detection of a beat (Winkler et al., 2009).

Musical Enculturation

The effects of enculturation on music perception have been investigated through developmental and cross-cultural research. A number of musical abilities are apparent in infants under 1 year old including the perception of pitch contour (Plantinga & Trainor, 2005), tone chroma (Demany et al., 1985), consonance (Trainor & Heinmiller, 1998), timbre (Trainor et al., 2004), rhythm (Trehub & Thorpe, 1989), and stream segregation (Demany, 1982). Other abilities appear later in childhood presumably reflecting cultural musical experience, possibly in combination with increasing cognitive maturity. Perception of tonality and meter are prime examples of culture-specific developmental components in music perception.

Regarding perception of tonality, 8-month-old infants are equally able to detect changes to a melody regardless of the tonal stability of the replacement tone whereas Western adults are better at detecting out-of-key than within-key replacements (Trainor & Trehub, 1992). Furthermore, while Western adults can better identify mistunings in melodies based on a Western scale than melodies based on a Balinese scale, 6-month-olds perform equally well in either case (Lynch et al., 1990). By the age of 4 or 5 years, however, Western children, like Western adults, more accurately detect changes in tonal than atonal melodies (Trehub et al., 1986) and out-of-key changes than within-key changes (Corrigall & Trainor, 2010). By 7 years, Western children are better at detecting changed within-key notes in a Western melody that violate implied harmony while 5-year-olds do not show such sensitivity (Trainor & Trehub, 1994). Probe-tone studies with Western participants suggest that differentiation of scale degrees based on tonal stability starts at the age of 6–8 years, with increasingly differentiated representations of tonal hierarchy with increasing age up to 11 (Cuddy & Badertscher, 1987; Krumhansl & Keil, 1982; Lamont & Cross, 1994; Speer & Meeks, 1985). Consistent with this, effects of tonality on emotion recognition start to emerge at ages 6–8 (Dalla Bella et al., 2001).

Turning to the perception of meter, Western listeners' temporal expectations given a metrical context show striking similarity to statistical distributions of tones in large corpora of Western music for the corresponding time signatures, giving rise to the suggestion that metrical templates might be acquired through exposure to music (Palmer & Krumhansl, 1990). Furthermore, effects of musical enculturation on meter perception have been demonstrated in empirical research comparing *isochronous meters*, containing binary and ternary subdivisions of metrical time periods (and multiples thereof; e.g., 4/4, 6/8), with *nonisochronous meters*, containing metrical cycles subdivided into a prime number of beats (e.g., 7 or 11) or uneven subdivisions of a metrical cycle (e.g., the Ottoman *aksak*, which subdivides 9 into 2 + 2 + 2 + 3). Unlike isochronous meters, nonisochronous meters are uncommon in Western music but appear frequently in other musical cultures such as Turkish and Indian classical music. While U.S. adults show better discrimination of meter-preserving and meter-violating changes to rhythms in isochronous meters than nonisochronous meters, no such difference is shown by adults from Balkan countries, where nonisochronous meters are common (Hannon & Trehub, 2005a) or U.S. 6-month-olds (Hannon & Trehub, 2005b). U.S. 12-month-olds do show better discrimination for isochronous meters but this was eliminated by 2 weeks of listening to Balkan music, whereas this was not the case for U.S. adults (Hannon & Trehub, 2005b). Subsequent research demonstrates that equivalent processing of isochronous and nonisochronous meters by Turkish adults is only evident for nonisochronous meters that actually appear in Turkish music, with reduced performance for other nonisochronous meters (Hannon et al., 2012). Furthermore, participants show systematic biases toward the temporal ratios present in the music of their own musical cultures on rhythm reproduction and synchronization tasks (Jacoby et al., 2021; Polak et al., 2018).

There is much less research on the effects of enculturation across the adult life span, especially for older adults. However, Halpern et al. (1996) found that older adults (aged 60–80) made greater use of pitch height than tonal relatedness, suggesting some attenuation in learned statistical representations of tonality with age but this experiment used simple artificial stimuli and has not been replicated with more complex and ecologically valid musical stimuli.

Cross-cultural research with adults has typically compared Western participants with samples from one or more non-Western cultures. Demorest et al. (2008) found that recognition memory of American and Turkish participants listening to novel Western, Turkish, and Chinese music (the latter providing a neutral control equally unfamiliar to both cultures) was better for music of the native culture, while Turkish listeners were better for Western than Chinese music suggesting a secondary enculturation effect known as bimusicalism (Wong et al., 2009). The enculturation effect was replicated for 10–11-year-old U.S. children (Morrison et al., 2008) and for isochronous melodies transcribed from the original versions with no rhythmic variation, accompaniment, or expressive performance (Demorest et al., 2016), suggesting that the effect depends primarily on sequential pitch structure. In some cases, enculturation may be related to linguistic rather than musical exposure. For example, while Western listeners tend to hear an alternating sequence of long and short tones as short–long, Japanese listeners tend to hear it as long–short (Iversen et al., 2008). This difference potentially reflects the different positions of function words in Japanese compared to English (e.g., "hon ga" vs. "the book"), an interpretation supported by the finding that the effect emerges at 7–8 months, approximately when understanding of phrasal grouping in language develops (Yoshida et al., 2010).

Musical Training

Certain individuals in many cultures undertake special training to develop their ability to create or perform music (see Pfordresher, 2022). What effect does musical training have on music perception? The vast majority of studies that have addressed this question have compared task performance of musicians and nonmusicians. However, a difference in some aspect of music perception between groups varying in musical training could very plausibly reflect a preexisting perceptual *aptitude* that contributed to an individual becoming a musician; cause and effect are hard to disentangle. Indeed, in a study of monozygotic twins, there was no correlation between intra-pair differences in amount of lifetime musical training (ranging as high as 20,228 hours) and performance on pitch, rhythm, and melody discrimination tasks (Mosing et al., 2014). Further complicating the picture, operational definitions of musical training used as selection criteria for groups of musicians and nonmusicians, usually based on years of musical training, vary widely between studies. The situation has improved with the availability of psychometric measures of musical training and aptitude (Müllensiefen et al., 2014; Wallentin, 2010). However, a final difficulty arises from the fact that musical training and musicianship often covary with other factors such as socioeconomic status and personality (Swaminathan & Schellenberg, 2019) while aptitude is associated with general cognitive abilities (Schellenberg & Weiss, 2013), but these factors are not often controlled for in experimental studies.

In fact, it is surprising that differences between musicians and nonmusicians are often rather subtle or nonexistent in many areas of music perception, including the perception of timbre (McAdams et al., 1995), musical expectation (Bigand et al., 2003; Schellenberg, 1996), emotional experience (Bigand et al., 2005), and long-term memory for music (Demorest et al., 2008; Halpern et al., 1995; McAuley et al., 2004). Musical training is critical for the development of absolute pitch (e.g., Deutsch et al., 2006) and appears to have an effect on short-term memory for music, perhaps due to better encoding of auditory features by musicians (Dowling, 1978; Halpern et al., 1995). The most convincing evidence for effects of musical training comes from the few studies that show instrument- or style-specific effects, which are unlikely to reflect preexisting aptitude. For example, violinists show enhanced evoked electrophysiological responses to violin tones than trumpet tones, while trumpeters show the converse (Pantev et al., 2001), and jazz musicians generate expectations with greater certainty and precision when listening to jazz music than classical musicians (Hansen et al., 2016).

Extramusical Effects of Musical Listening

The question of whether listening to music has any effect on psychological processing in nonmusical tasks has generated a significant body of research (see Kämpfe et al., 2011, for a meta-analysis and Schellenberg & Weiss, 2013, for a review). This is a confusing literature to survey due to multiple conflicting positive, negative, or neutral effects of musical listening in different contexts, suggesting the influence of many factors. However, there are cases where the effects seem to be reliable and plausible underlying mechanisms have been identified.

Background music appears to have a negative effect on memory and reading performance. For example, instrumental music impairs performance in a serial digit recall task, taxing working memory, compared with both silence and white noise, with vocal music (sung in both familiar and unfamiliar languages) having a greater effect (Salame & Baddeley, 1989). It appears likely that music, and particularly vocal music, as a time-varying signal with similarities to speech, interferes with the retention of the digit sequence in the phonological loop component of working memory, which supports mental rehearsal of material to be remembered (Baddeley & Hitch, 2019; Hartley & Hitch, 2022). For similar reasons, music has been found to impair reading comprehension (Kämpfe et al., 2011), especially when it is loud and fast, making it difficult to ignore (Anderson & Fuller, 2010; Thompson et al., 2011). These experiments were run in an otherwise quiet environment, so it remains possible that music might improve performance in noisy environments by masking the noise (Schellenberg & Weiss, 2013).

For other tasks, it has been shown that listening to music can improve subsequent performance through its effects on emotional state, especially by increasing arousal and valence. A good example is the well-known Mozart effect in which performance on visuospatial intelligence tests improves after (not while) listening to the first movement of Mozart's Sonata for Two Pianos in D major (K. 448; Pietschnig et al., 2010; Rauscher et al., 1993). It appears that the effect is not specific to K. 448 but extends to other music (Nantais & Schellenberg, 1999; Schellenberg & Hallam, 2005) and spoken narrative (Nantais & Schellenberg, 1999). In contrast, the effect does not occur for pieces of music which convey sadness (Thompson et al., 2001) or for a version of K. 448 adapted to a minor key and slow tempo (Husain et al., 2002) and disappears when changes in emotional state are accounted for (Thompson et al., 2001). Therefore, the effect is not caused by music per se but by the high valence, high arousal state induced by certain pieces of music and other stimuli. Similar mechanisms produce a positive effect of pretask musical listening on athletic performance where in-task music has also been shown to have an ergogenic (work-enhancing) effect (Karageorghis & Priest, 2012a, 2012b).

Finally, there is some evidence that music can have an implicit influence on consumer behavior. Individuals spend more on food and wine when classical rather than pop music is played (Areni & Kim, 1993; North et al., 2003) and are influenced in their choice of food and wine by the cultural association of the background music (North et al., 1999; Yeoh & North, 2010). Listening to music may even modulate the experience of specific gustatory attributes during the tasting of wine (Spence & Wang, 2015).

Evolution of Musicality

Longstanding interest in the question of an evolutionary role for music has yielded several hypotheses but little concrete evidence for generational change in genetic variation due to natural selection of heritable traits specifically related to music. There is little doubt that music itself, like language, is a cultural phenomenon passed (with variations and developments) through cultural transmission from generation to generation (Repp, 1991), so research has focused on evolution of psychological traits making up the spontaneously developing capacity for making and perceiving music, which has been referred to as *musicality* (Honing et al., 2015). A useful null hypothesis is that musicality is an exaptation, reflecting traits that were evolved in the context of other psychological processes, including language and auditory scene analysis (Pinker, 1997, pp. 528–529). Specific adaptive functions for music have been proposed to include sexual selection (Darwin, 1871; Miller, 2000), group cohesion (Cirelli et al., 2014; Dunbar, 2012; Savage et al., 2021), credible signaling of coalition strength and parental attention (Mehr et al., 2021), parent–infant bonding (Dissanayake, 2000), and developmental enhancement of cognitive, perceptual, and interpersonal skills during the altricial period of parental care (Cross, 2001, 2003).

The fact that musicality consists of a complex set of psychological traits reflecting both polygenic and cultural influences complicates efforts to demonstrate heredity. Another major difficulty is that psychological traits do not fossilize (Lewontin, 1998): it is impossible to assess genetic variation in aspects of musicality or the selection pressures under which they may have evolved during the evolutionary periods in which musicality emerged. Given this, empirical research has attempted to triangulate between different approaches in gathering complementary evidence addressing different questions about musicality (Fitch, 2006; Honing & Ploeger, 2012; Justus & Hutsler, 2005): ontogeny, mechanism, adaptation, and phylogeny (Tinbergen, 1963).

Developmental and cross-cultural research can shed light on the *ontogeny (development)* of musicality, identifying the extent to which individual traits depend on experience. However, because effects of musical experience have been observed *in utero* (Hepper, 1991; Partanen et al., 2013), developmental evidence cannot provide conclusive evidence of genetic determination. Furthermore, universality of a particular trait (such as the perception of infant-directed singing, Trehub et al., 1993) does not necessarily imply heritability. Psychological and neuroscientific research can address the psychological and neural *mechanisms (causation)* involved in musical behaviors, while computational simulations of those behaviors can provide evidence of which traits can be acquired through cultural experience (Justus & Hutsler, 2005, though this doesn't necessarily mean that they *are* acquired through experience). Psychological and ethnomusicological research can identify the current *functions (adaptation)* of musicality in various cultures and societies. However, it is impossible to know the extent to which the current function of musicality overlaps with its function in the context in which it putatively evolved. Finally, comparative research provides a window on the *phylogeny (evolution)* of musicality: homologous traits allow inferences to be drawn about the ancestral precursors of musicality while analogous traits can shed light on mechanism and function. Traits that have no nonhuman homologue or analogue are strong candidates for evolutionary theories of uniquely human aspects of musicality.

Meter perception is one prominent aspect of musicality that has been considered as a candidate trait for evolutionary explanations (Justus & Hutsler, 2005; Patel, 2022). Tempo-flexible beat perception and synchronization (BPS), involving the synchronization of non-sound-producing movements to complex auditory stimuli, is a polygenic trait (Gordon et al., 2021) shown by human neonates (Winkler et al., 2009) but not by other nonhuman primates (Hattori & Tomonaga, 2019; Hattori et al., 2013; Honing et al., 2012, 2018) and cannot easily be accounted for as an exaptation from speech (Patel, 2022). It has been proposed that BPS may be related to the evolution of vocal learning mechanisms representing, in the context of musicality, either an exaptation (Patel et al., 2009) or gene–culture coevolution (Patel, 2021), though examples of BPS in nonvocal learning species such as sea lions (Cook et al., 2013) challenge this hypothesis.

Conclusion

Music perception is a surprisingly rich psychological phenomenon, invoking a wide range of mental representations and processes including sensory processing, perceptual scene analysis, structural inference, syntactic processing, statistical learning, memory, expectation, and affective processing. As a result, music perception provides a broad and revealing window into the workings of the mind and brain, allowing questions of domain specificity and generality to be addressed by comparison with other cultural psychological phenomena such as language. As a cultural domain, music perception shows both cross-cultural consistency and variability, requiring sophisticated analysis of both inherited and acquired psychological components, posing interesting challenges for developmental and evolutionary psychologists. It is perhaps worth remembering that while music is usually considered an auditory phenomenon, it is often associated with visual display (e.g., dance, musical theater, opera, video) and can be readily imagined without auditory input given a suitable memory prompt or, with appropriate training, a visually encoded musical score, providing a fruitful avenue for research on multimodal perception. Much has been learned by generations of psychologists over the 150 years or so since Helmholtz first turned his attention to music perception but, equally, much remains for future generations to understand.

Acknowledgments

I am grateful to Edward Hall, Peter Harrison, Henkjan Honing, Patrik Juslin, Tom Kaplan, Stephen McAdams, Lauren Stewart, and two anonymous reviewers for their feedback on earlier drafts of this article.

Further Reading

Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound*. MIT Press.

Deutsch, D. (Ed.). (2013). *The psychology of music* (3rd ed.). Academic Press.

Hallam, S., Cross, I., & Thaut, M. (Eds.). (2016). *The Oxford handbook of music psychology* (2nd ed.). Oxford University Press.

Halpern, A., & Bartlett, J. (2010). Memory for melodies. In M. R. Jones, A. N. Popper, & R. R. Fay (Eds.), *Music perception* (pp. 233–258). Springer-Verlag.

Honing, H., ten Cate, C., Peretz, I., & Trehub, S. E. (Eds.). (2015). Theme issue “Biology, cognition and origins of musicality.” *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences*, *370*(1664).

Huron, D. (2006). *Sweet anticipation: Music and the psychology of expectation*. MIT Press.

Juslin, P. N., & Sloboda, J. (Eds.). (2011). *Handbook of music and emotion: Theory, research, applications*. Oxford University Press.

Krumhansl, C. L. (1990). *Cognitive foundations of musical pitch*. Oxford University Press.

London, J. (2012). *Hearing in time: Psychological aspects of musical meter* (2nd ed.). Oxford University Press.

Patel, A. D. (2008). *Music, language, and the brain*. Oxford University Press.

Plack, C. (2018). *The sense of hearing* (3rd ed.). Routledge.

Thompson, W. F. (2015). *Music, thought and feeling: Understanding the psychology of music*. Oxford University Press.

References

Albouy, P., Mattout, J., Bouet, R., Maby, E., Sanchez, G., Aguera, P.-E., Daligault, S., Delpuech, C., Bertrand, O., Caclin, A., & Tillmann, B. (2013). Impaired pitch perception and memory in congenital amusia: The deficit starts in the auditory cortex *<https://doi.org/10.1093/brain/awt082>*. *Brain*, *136*, 1639–1661.

Allen, E. J., Burton, P. C., Olman, C. A., & Oxenham, A. J. (2017). Representations of pitch and timbre variation in human auditory cortex <https://doi.org/10.1523/JNEUROSCI.2336-16.2016>. *Journal of Neuroscience*, *37*(5), 1284–1293.

Allen, E. J., & Oxenham, A. J. (2014). Symmetric interactions and interference between pitch and timbre <https://doi.org/10.1121/1.4863269>. *The Journal of the Acoustical Society of America*, *135*, 1371–1379.

Alluri, V., & Toiviainen, P. (2010). Exploring perceptual and acoustical correlates of polyphonic timbre <https://doi.org/10.1525/mp.2010.27.3.223>. *Music Perception*, *27*, 223–242.

American National Standards Institute. (1994). *Psychoacoustic terminology: Timbre*. American National Standards Institute.

Anderson, S. A., & Fuller, G. B. (2010). Effect of music on reading comprehension of junior high school students <https://doi.org/10.1037/a0021213>. *School Psychology Quarterly*, *25*(3), 178–187.

Areni, C. S., & Kim, D. (1993). The influence of background music on shopping behavior: Classical versus top-forty music in a wine store <https://www.acrwebsite.org/volumes/7467/volumes/v20/NA-20>. *Advances in Consumer Research*, *20*, 336–340.

Ayari, M., & McAdams, S. (2003). Aural analysis of Arabic improvised instrumental music (Taqsim) <https://doi.org/10.1525/mp.2003.21.2.159>. *Music Perception*, *21*(2), 159–216.

Baddeley, A. D., & Hitch, G. J. (2019). The phonological loop as a buffer store: An update <https://doi.org/10.1016/j.cortex.2018.05.015>. *Cortex*, *112*, 91–106.

Baharloo, S., Service, S. K., Risch, N., Gitschier, J., & Freimer, N. (2000). Familial aggregation of absolute pitch <https://doi.org/10.1086/301704>. *American Journal of Human Genetics*, *67*, 755–758.

Balkwill, L. L., & Thompson, W. F. (1999). A cross-cultural investigation of the perception of emotion in music: Psychophysical and cultural cues <https://doi.org/10.2307/40285811>. *Music Perception*, *17*(1), 43–64.

Balkwill, L. L., Thompson, W. F., & Matsunaga, R. (2004). Recognition of emotion in Japanese, Western, and Hindustani music by Japanese listeners <https://doi.org/10.1111/j.1468-5584.2004.00265.x>. *Japanese Psychological Research*, *46*(4), 337–349.

Bartlett, J. C., Halpern, A. R., & Dowling, W. J. (1995). Recognition of familiar and unfamiliar melodies in normal aging and Alzheimer's disease <https://doi.org/10.3758/BF03197255>. *Memory & Cognition*, *23*(5), 531–546.

Berlyne, D. E. (1960). *Conflict, arousal and curiosity*. McGraw-Hill.

Bharucha, J. J. (1987). Music cognition and perceptual facilitation: A connectionist framework <https://doi.org/10.2307/40285384>. *Music Perception*, *5*(1), 1–30.

Bharucha, J. J., & Krumhansl, C. L. (1983). The representation of harmonic structure in music: Hierarchies of stability as a function of context <https://doi.org/10.1016/0010-0277(83)90003-3>. *Cognition*, *13*(1), 63–102.

Bigand, E., Delbé, C., Poulin-Charronnat, B., Leman, M., & Tillmann, B. (2014). Empirical evidence for musical syntax processing? Computer simulations reveal the contribution of auditory short-term memory <https://doi.org/10.3389/fnsys.2014.00094>. *Frontiers in Systems Neuroscience*, *8*, 94.

Bigand, E., Madurell, F., Tillmann, B., & Pineau, M. (1999). Effect of global structure and temporal organization on chord processing <https://doi.org/10.1037/0096-1523.25.1.184>. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(1), 184–197.

Bigand, E., & Parncutt, R. (1999). Perceiving musical tension in long chord sequences <https://doi.org/10.1007/s004260050053>. *Psychological Research*, *62*(4), 237–254.

Bigand, E., Poulin-Charronnat, B., Tillmann, B., Madurell, F., & D'Adamo, D. A. (2003). Sensory versus cognitive components in harmonic priming <https://doi.org/10.1037/0096-1523.29.1.159>. *Journal of Experimental Psychology: Human Perception and Performance*, *29*(1), 159–171.

Bigand, E., Vieillard, S., Madurell, F., Marozeau, J., & Dacquet, A. (2005). Multidimensional scaling of emotional responses to music: The effect of musical expertise and of the duration of the excerpts <https://doi.org/10.1080/02699930500204250>. *Cognition and Emotion*, *19*(8), 1113–1139.

Blood, A. J., Zatorre, R. J., Bermudez, P., & Evans, A. C. (1999). Emotional responses to pleasant and unpleasant music correlate with activity in paralimbic brain regions <https://doi.org/10.1038/7299>. *Nature Neuroscience*, *2*(4), 382–387.

Bo, M., Mill, R. W., & Bendixen, A. (2013). Modelling the emergence and dynamics of perceptual organisation in auditory streaming <https://doi.org/10.1371/journal.pcbi.1002925>. *PLoS Computational Biology*, *9*(3), e1002925.

Boltz, M. G. (1999). The processing of melodic and temporal information: Independent or unified dimensions? <https://doi.org/10.1076/jnmr.28.1.67.3121> *Journal of New Music Research*, *28*(1), 67–79.

Bouwer, F. L., Honing, H., & Slagter, H. A. (2020). Beat-based and memory-based temporal expectations in rhythm: Similar perceptual effects, different underlying mechanisms <https://doi.org/10.1162/jocn_a_01529>. *Journal of Cognitive Neuroscience*, *32*, 1221–1241.

Bozkurt, B., Ayangil, R., & Holzapfel, A. (2014). Computational analysis of makam music in Turkey: Review of state-of-the-art and challenges <https://doi.org/10.1080/09298215.2013.865760>. *Journal of New Music Research*, *43*(1), 3–23.

Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound*. MIT Press.

Bregman, A. S., & Campbell, J. (1971). Primary auditory stream segregation and perception of order in rapid sequences of tones <https://doi.org/10.1037/h0031163>. *Journal of Experimental Psychology*, *89*(2), 244–249.

Bregman, A. S., & Pinker, S. (1978). Auditory streaming and the building of timbre <https://doi.org/10.1037/h0081664>. *Canadian Journal of Psychology*, *32*, 19–31.

Burns, E. M., & Campbell, S. L. (1994). Frequency and frequency-ratio resolution by possessors of absolute and relative pitch: Examples of categorical perception <https://doi.org/10.1121/1.411447>. *The Journal of the Acoustical Society of America*, *96*(5 pt 1), 2704–2719.

Burns, E. M., & Ward, W. D. (1978). Categorical perception—phenomenon or epiphenomenon: Evidence from experiments in the perception of melodic musical intervals <https://doi.org/10.1121/1.381737>. *The Journal of the Acoustical Society of America*, *63*(2), 456–468.

Caclin, A., McAdams, S., Smith, B. K., & Winsberg, S. (2005). Acoustic correlates of timbre space dimensions: A confirmatory study using synthetic tones <https://doi.org/10.1121/1.1929229>. *The Journal of the Acoustical Society of America*, *118*(1), 471–482.

Cambouropoulos, E. (2001). The local boundary detection model (LBDM) and its application in the study of expressive timing. In A. Schloss, R. Dannenberg, & P. Driessen (Eds.), *Proceedings of the 2001 International Computer Music Conference, Havana* (pp. 17–22). International Computer Music Association.

Cambouropoulos, E. (2008). Voice and stream: Perceptual and computational modelling of voice separation <https://doi.org/10.1525/mp.2008.26.1.75>. *Music Perception*, *26*(1), 75–94.

Cannon, J. J., & Patel, A. D. (2020). How beat perception co-opts motor neurophysiology <https://doi.org/10.1016/j.tics.2020.11.002>. *Trends in Cognitive Sciences*, *25*(2), 137–150.

Carrus, E., Pearce, M. T., & Bhattacharya, J. (2013). Melodic pitch expectation interacts with neural responses to syntactic but not semantic violations <https://doi.org/10.1016/j.cortex.2012.08.024>. *Cortex*, *49*(8), 2186–2200.

Castro, S. L., & Lima, C. F. (2014). Age and musical expertise influence emotion recognition in music <https://doi.org/10.1525/MP.2014.32.2.125>. *Music Perception*, *32*(2), 125–143.

Cheung, V. K. M., Harrison, P. M. C., Meyer, L., Pearce, M. T., Haynes, J., & Koelsch, S. (2019). Uncertainty and surprise jointly predict musical pleasure and amygdala, hippocampus, and auditory cortex activity <https://doi.org/10.1016/j.cub.2019.09.067>. *Current Biology*, *29*, 4084–4092.e4.

Chmiel, A., & Schubert, E. (2017). Back to the inverted-U for music preference: A review of the literature <https://doi.org/10.1177/0305735617697507>. *Psychology of Music*, *45*, 886–909.

Cirelli, L. K., Einarson, K. M., & Trainor, L. J. (2014). Interpersonal synchrony increases prosocial behavior in infants <https://doi.org/10.1111/desc.12193>. *Developmental Science*, *17*, 1003–1011.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science <https://doi.org/10.1017/S0140525X12000477>. *Behavioral and Brain Sciences*, *36*(03), 181–204.

Clarke, E. F. (1987). Categorical rhythm perception: An ecological perspective. In A. Gabrielsson (Ed.), *Action and perception in rhythm and music* (pp. 19–33). Royal Swedish Academy of Music.

Conard, N. J., Malina, M., & Mu, S. C. (2009). New flutes document the earliest musical tradition in southwestern Germany <https://doi.org/10.1038/nature08169>. *Nature*, *460*(7256), 737–740.

Cook, P. F., Rouse, A., Wilson, M., & Reichmuth, C. (2013). A California sea lion (*Zalophus californianus*) can keep the beat: Motor entrainment to rhythmic auditory stimuli in a non vocal mimic <https://doi.org/10.1037/a0032345>. *Journal of Comparative Psychology*, *127*, 412–427.

Corrigall, K. A., & Trainor, L. J. (2010). Musical enculturation in preschool children: Acquisition of key and harmonic knowledge <https://doi.org/10.1525/mp.2010.28.2.195>. *Music Perception*, *28*, 195–200.

Cowan, N. (2008). What are the differences between long-term, short-term, and working memory? <https://doi.org/10.1016/S0079-6123(07)00020-9> In W. S. Sossin, J.-C. Lacaille, V. F. Castelluci, & S. Belleville (Eds.), *Progress in brain research* (Vol. 169, pp. 323–338). Elsevier.

Cross, I. (2001). Music, cognition, culture, and evolution <https://doi.org/10.1111/j.1749-6632.2001.tb05723.x>. *Annals of the New York Academy of Sciences*, *930*, 28–42.

Cross, I. (2003). Music and evolution: Consequences and causes <https://doi.org/10.1080/0749446032000150906>. *Contemporary Music Review*, *22*(3), 79–89.

Crowder, R. G. (1989). Imagery for musical timbre <https://doi.org/10.1037/0096-1523.15.3.472>. *Journal of Experimental Psychology: Human Perception and Performance*, *15*(3), 472–478.

Crowder, R. G. (1993). Auditory memory. In S. McAdams & E. Bigand (Eds.), *Thinking in sound: The cognitive psychology of human audition* (pp. 111–145). Oxford University Press.

Cuddy, L. L., & Badertscher, B. (1987). Recovery of the tonal hierarchy: Some comparisons across age and levels of musical experience <https://doi.org/10.3758/BF03210493>. *Perception & Psychophysics*, *41*(6), 609–620.

Dahl, S., Huron, D., Brod, G., & Altenmüller, E. (2014). Preferred dance tempo: Does sex or body morphology influence how we groove? <https://doi.org/10.1080/09298215.2014.884144> *Journal of New Music Research*, *43*(2), 214–223.

Dalla Bella, S., Peretz, I., Rousseau, L., & Gosselin, N. (2001). A developmental study of the affective value of tempo and mode in music <https://doi.org/10.1016/S0010-0277(00)00136-0>. *Cognition*, *80*(3), 1–10.

Darwin, C. J. (1997). Auditory grouping. *Trends in Cognitive Sciences*, *1*(9), 327–333.

Darwin, C. J., & Turvey, M. T. (1972). An auditory analogue of the sperling partial report procedure: Evidence for brief auditory storage <https://doi.org/10.1016/0010-0285(72)90007-2>. *Cognitive Psychology*, *267*, 255–267.

Darwin, C. R. (1871). *The descent of man and selection in relation to sex*. John Murray.

Dawe, L. A., Platt, J. R., & Racine, R. J. (1994). Inference of metrical structure from perception of iterative pulses within time spans defined by chord changes <https://doi.org/10.2307/40285755>. *Music Perception*, *12*(1), 57–76.

de Cheveigné, A. (2005). Pitch perception models. In C. J. Plack, A. J. Oxenham, R. R. Fay, & A. N. Popper (Eds.), *Pitch: Neural coding and perception* (pp. 169–233). Springer.

de Fleurian, R., & Pearce, M. T. (2021). Chills in music: A systematic review <https://doi.org/10.1037/bul0000341>. *Psychological Bulletin*, *147*(9), 890–920.

Deliège, I. (1987). Grouping conditions in listening to music: An approach to Lerdahl & Jackendoff's grouping preference rules <https://doi.org/10.2307/40285378>. *Music Perception*, *4*(4), 325–359.

Deliège, I., Mélen, M., Stammers, D., & Cross, I. (1997). Musical schemata in real-time listening to a piece of music <https://doi.org/10.2307/40285715>. *Music Perception*, *14*(2), 117–159.

Demany, L. (1982). Auditory stream segregation in infancy <https://doi.org/10.1016/S0163-6383(82)80036-2>. *Infant Behavior and Development*, *5*, 261–276.

Demany, L., Armand, F., & Demany, L. (1985). The perceptual reality of tone chroma in early infancy <https://doi.org/10.1121/1.391006>. *The Journal of the Acoustical Society of America*, *57*(1984), 57–66.

Demorest, S. M., Morrison, S. J., Beken, M. N., & Jungbluth, D. (2008). Lost in translation: An enculturation effect in music memory performance *<https://doi.org/10.1525/mp.2008.25.3.213>*. *Music Perception*, *25*(3), 213–233.

Demorest, S. M., Morrison, S. J., Nguyen, V. Q., & Bodnar, E. N. (2016). The influence of contextual cues on cultural bias in music memory *<https://doi.org/10.1525/MP.2016.33.5.590>*. *Music Perception*, *33*(5), 590–600.

Dennett, D. (1991). *Consciousness explained*. Little, Brown.

Desain, P., & Honing, H. (2003). The formation of rhythmic categories and metric priming *<https://doi.org/10.1068/p3370>*. *Perception*, *32*(1), 341–366.

Deutsch, D. (1979). Binaural integration of melodic patterns *<https://doi.org/10.3758/BF03199848>*. *Perception & Psychophysics*, *25*(5), 399–405.

Deutsch, D. (2013a). Absolute pitch. In D. Deutsch (Ed.), *The psychology of music* (pp. 141–182). Academic Press.

Deutsch, D. (2013b). Grouping mechanisms in music. In D. Deutsch (Ed.), *The psychology of music* (pp. 183–248). Academic Press.

Deutsch, D., Henthorn, T., & Marvin, E. W. (2006). Absolute pitch among American and Chinese conservatory students: Prevalence differences, and evidence for a speech-related critical period *<https://doi.org/10.1121/1.2151799>*. *The Journal of the Acoustical Society of America*, *119*, 719–722.

Deutsch, D., Li, X., & Shen, J. (2013). Absolute pitch among students at the Shanghai conservatory of music: A large-scale direct-test study *<https://doi.org/10.1121/1.4824450>*. *The Journal of the Acoustical Society of America*, *134*, 3853–3859.

Dibben, N. (1994). The cognitive reality of hierarchic structure in tonal and atonal music *<https://doi.org/10.2307/40285753>*. *Music Perception*, *12*(1), 1–25.

Dibben, N. (1999). The perception of structural stability in atonal music: The influence of salience, stability, horizontal motion, pitch commonality, and dissonance *<https://doi.org/10.2307/40285794>*. *Music Perception*, *16*(3), 265–294.

Dissanayake, E. (2000). *Art and intimacy: How the arts began*. University of Washington Press.

Dowling, W. J. (1973). The perception of interleaved melodies *<https://doi.org/10.1016/0010-0285(73)90040-6>*. *Cognitive Psychology*, *5*, 322–337.

Dowling, W. J. (1978). Scale and contour: Two components in a theory of memory for melodies *<https://doi.org/10.1037/0033-295X.85.4.341>*. *Psychological Review*, *85*, 341–354.

Dowling, W. J. (1991). Tonal strength and melody recognition after long and short delays *<https://doi.org/10.3758/BF03212222>*. *Perception & Psychophysics*, *50*(4), 305–313.

Dowling, W. J., Lung, K. M., & Herrbold, S. (1987). Aiming attention in pitch and time in the perception of interleaved melodies *<https://doi.org/10.3758/BF03210496>*. *Perception & Psychophysics*, *41*(6), 642–656.

Drake, C., & Palmer, C. (1993). Accent structures in music performance *<https://doi.org/10.2307/40285574>*. *Music Perception*, *10*(3), 343–378.

Dunbar, R. I. M. (2012). On the evolutionary function of song and dance. In N. Bannan (Ed.), *Music, language, and human evolution* (pp. 201–214). Oxford University Press.

Edworthy, J. (1985). Interval and contour in melody processing *<https://doi.org/10.2307/40285305>*. *Music Perception*, *2*(3), 375–388.

Egermann, H., Fernando, N., Chuen, L., & McAdams, S. (2015). Music induces universal emotion-related psychophysiological responses: Comparing Canadian listeners to Congolese Pygmies *<https://doi.org/10.3389/fpsyg.2014.01341>*. *Frontiers in Psychology*, *6*, 1341.

Egermann, H., Pearce, M. T., Wiggins, G. A., & McAdams, S. (2013). Probabilistic models of expectation violation predict psychophysiological emotional responses to live concert music *<https://doi.org/10.3758/s13415-013-0161-y>*. *Cognitive, Affective & Behavioral Neuroscience*, *13*(3), 533–553.

Eitan, Z., & Granot, R. Y. (2008). Growing oranges on Mozart's apple tree: "Inner form" and aesthetic judgment *<https://doi.org/10.1525/mp.2008.25.5.397>*. *Music Perception*, *25*, 397–419.

Ekman, P. (1992). An argument for basic emotions *<https://doi.org/10.1080/02699939208411068>*. *Cognition and Emotion*, *6*(3), 169–200.

Elhilali, M., & Shamma, S. (2008). A cocktail party with a cortical twist: How cortical mechanisms contribute to sound segregation *<https://doi.org/10.1121/1.3001672>*. *The Journal of the Acoustical Society of America*, *124*, 3751–3771.

Epstein, M., & Marozeau, J. (2010). Loudness and intensity coding. In C. J. Plack (Ed.), *The Oxford handbook of auditory science: Hearing* (pp. 45–69). Oxford University Press.

Fedorenko, E., Patel, A. D., Casasanto, D., Winawer, J., & Gibson, E. (2009). Structural integration in language and music: Evidence for a shared system *<https://doi.org/10.3758/MC.37.1.1>*. *Memory & Cognition*, *37*(1), 1–9.

Fitch, W. T. (2006). The biology and evolution of music: A comparative perspective *<https://doi.org/10.1016/j.cognition.2005.11.009>*. *Cognition*, *100*, 173–215.

Fitch, W. T., & Rosenfeld, A. J. (2007). Perception and production of syncopated rhythms *<https://doi.org/10.1525/mp.2007.25.1.43>*. *Music Perception*, *25*, 43–59.

Frankland, B. W., & Cohen, A. J. (2004). Parsing of melody: Quantification and testing of the local grouping rules of Lerdahl and Jackendoff's A Generative Theory of Tonal Music *<https://doi.org/10.1525/mp.2004.21.4.499>*. *Music Perception*, *21*(4), 499–543.

Friston, K. J. (2010). The free-energy principle: A unified brain theory? *<https://doi.org/10.1038/nrn2787>* *Nature Reviews Neuroscience*, *11*(2), 127–138.

Fritz, T., Jentschke, S., Gosselin, N., Sammler, D., Peretz, I., Turner, R., Friederici, A. D., & Koelsch, S. (2009). Universal recognition of three basic emotions in music *<https://doi.org/10.1016/j.cub.2009.02.058>*. *Current Biology*, *19*(7), 1–4.

Frost, R., Armstrong, B. C., & Christiansen, M. H. (2019). Statistical learning research: A critical review and possible new directions *<https://doi.org/10.1037/bul0000210>*. *Psychological Bulletin*, *145*(12), 1128–1153.

Geringer, J. M. (1995). Continuous loudness judgments of dynamics in recorded excerpts *<https://doi.org/10.2307/3345789>*. *Journal of Research in Music Education*, *43*(1), 22–35.

Gingras, B., Pearce, M. T., Goodchild, M., Dean, R. T., Wiggins, G., & Mcadams, S. (2015). Linking melodic expectation to expressive performance timing and perceived musical tension *<https://doi.org/10.1037/xhp0000141>*. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(4), 594–609.

Gjerdingen, R. O., & Perrott, D. (2008). Scanning the dial: The rapid recognition of music genres *<https://doi.org/10.1080/09298210802479268>*. *Journal of New Music Research*, *37*(2), 93–100.

Goerlich, K. S., Witteman, J., Schiller, N. O., van Heuven, V. J., Aleman, A., & Martens, S. (2012). The nature of affective priming in music and speech *<https://doi.org/10.1162/jocn_a_00213>*. *Journal of Cognitive Neuroscience*, *24*, 1725–1741.

Gold, B. P., Pearce, M. T., Mas-herrero, E., Dagher, A., & Zatorre, R. J. (2019). Predictability and uncertainty in the pleasure of music: A reward for learning? *<https://doi.org/10.1523/JNEUROSCI.0428-19.2019>* *Journal of Neuroscience*, *39*(47), 9397–9409.

Goldman, A., Harrison, P. M. C., Jackson, T., & Pearce, M. T. (2021). Reassessing syntax-related ERP components using popular music chord sequences: A model-based approach *<https://doi.org/10.1525/MP.2021.39.2.118>*. *Music Perception*, *39*, 118–144.

Goldstone, R. L., & Hendrickson, A. T. (2010). Categorical perception *<https://doi.org/10.1002/wcs.26>*. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*, 69–78.

Goodchild, M., & Mcadams, S. (2018). Perceptual processes in orchestration *<https://doi.org/10.1093/oxfordhb/9780190637224.013.10>*. In E. I. Dolan & A. Rehding (Eds.), *The Oxford handbook of timbre* (pp. 496–524). Oxford University Press.

Gordon, R. L., Ravignani, A., Bruno, J. H., Robinson, C. M., Scartozzi, A., Embalabala, R., Niarchou, M., 23andMe Research Team, Cox, N. J., & Creanza, N. (2021). Linking the genomic signatures of human beat synchronization and learned song in birds *<https://doi.org/10.1098/rstb.2020.0329>*. *Philosophical Transactions of the Royal Society B*, *376*, 20200329.

Grahn, J. A., & Brett, M. (2007). Rhythm and beat perception in motor areas of the brain *<https://doi.org/10.1162/jocn.2007.19.5.893>*. *Journal of Cognitive Neuroscience*, *19*(5), 893–906.

Grahn, J. A., & Brett, M. (2009). Impairment of beat-based rhythm discrimination in Parkinson's disease *<https://doi.org/10.1016/j.cortex.2008.01.005>*. *Cortex*, *45*(1), 54–61.

Granot, R. Y., & Eitan, Z. (2011). Musical tension and the interaction of dynamic auditory parameters *<https://doi.org/10.1525/mp.2011.28.3.219>*. *Music Perception*, *28*, 219–245.

Granot, R. Y., & Jacoby, N. (2011a). Musically puzzling I: Sensitivity to overall structure in the sonata form? *<https://doi.org/10.1177/1029864911409508>* *Musicae Scientiae*, *15*, 365–386.

Granot, R. Y., & Jacoby, N. (2011b). Musically puzzling II: Sensitivity to overall structure in a Haydn E-minor sonata *<https://doi.org/10.1177/1029864911423146>*. *Musicae Scientiae*, *16*, 67–80.

Grube, M., Cooper, F. E., Chinnery, P. F., & Griffiths, T. D. (2010). Dissociation of duration-based and beat-based auditory timing in cerebellar degeneration *<https://doi.org/10.1073/pnas.0910473107>*. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(25), 11597–11601.

Halpern, A. R., & Bartlett, J. C. (2010). Memory for melodies *<https://doi.org/10.1007/978-1-4419-6114-3>*. In M. Riess Jones, R. R. Fay, & A. N. Popper (Eds.), *Music perception* (Vol. 36, pp. 233–258). Springer.

Halpern, A. R., Bartlett, J. C., & Dowling, W. J. (1995). Aging and experience in the recognition of musical transpositions *<https://doi.org/10.1037/0882-7974.10.3.325>*. *Psychology and Aging*, *10*(3), 325–342.

Halpern, A. R., Kwak, S., Bartlett, J. C., & Dowling, W. J. (1996). Effects of aging and musical experience on the representation of tonal hierarchies *<https://doi.org/10.1037/0882-7974.11.2.235>*. *Psychology and Aging*, *11*(2), 235–246.

Hannon, E. E., Snyder, J. S., & Krumhansl, C. L. (2004). The role of melodic and temporal cues in perceiving musical meter *<https://doi.org/10.1037/0096-1523.30.5.956>*. *Journal of Experimental Psychology: Human Perception and Performance*, *30*(5), 956–974.

Hannon, E. E., Soley, G., & Ullal, S. (2012). Familiarity overrides complexity in rhythm perception: A cross-cultural comparison of American and Turkish listeners *<https://doi.org/10.1037/a0027225>*. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(3), 543–548.

Hannon, E. E., & Trehub, S. E. (2005a). Metrical categories in infancy and adulthood *<https://doi.org/10.1111/j.0956-7976.2005.00779.x>*. *Psychological Science*, *16*, 48–55.

Hannon, E. E., & Trehub, S. E. (2005b). Tuning in to musical rhythms: Infants learn more readily than adults *<https://doi.org/10.1073/pnas.0504254102>*. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(35), 12639–12643.

Hansen, N. C., Vuust, P., & Pearce, M. T. (2016). “If you have to ask, you’ll never know”: Effects of specialised stylistic expertise on predictive processing of music *<https://doi.org/10.1371/journal.pone.0163584>*. *PLOS ONE*, *11*, e0163584.

Harrison, P. M. C., & Pearce, M. T. (2020). Simultaneous consonance in music perception and composition *<https://doi.org/10.1037/rev0000169>*. *Psychological Review*, *127*(2), 216–244.

Hartley, T., & Hitch, G. J. (2022). Working Memory *<https://doi.org/10.1093/acrefore/9780190236557.013.768>*. In *Oxford Research Encyclopedia of Psychology*. Oxford University Press.

Hattori, Y., & Tomonaga, M. (2019). Rhythmic swaying induced by sound in chimpanzees (*Pan troglodytes*) *<https://doi.org/10.1073/pnas.1910318116>*. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(2), 936–942.

Hattori, Y., Tomonaga, M., & Matsuzawa, T. (2013). Spontaneous synchronized tapping to an auditory rhythm in a chimpanzee *<https://doi.org/10.1038/srep01566>*. *Scientific Reports*, *3*, 1566.

Hepper, P. G. (1991). An examination of fetal learning before and after birth *<https://doi.org/10.1080/03033910.1991.10557830>*. *The Irish Journal of Psychology*, *12*, 95–107.

Hofmann, A., & Goebl, W. (2014). Production and perception of legato, portato, and staccato articulation in saxophone playing <https://doi.org/10.3389/fpsyg.2014.00690>. *Frontiers in Psychology*, *5*, 690.

Honing, H. (2013). Structure and interpretation of rhythm in music <https://doi.org/10.1016/B978-0-12-381460-9.00009-2>. In D. Deutsch (Ed.), *The psychology of music* (pp. 369–404). Academic Press.

Honing, H., Bouwer, F. L., Prado, L., & Merchant, H. (2018). Rhesus monkeys (*Macaca mulatta*) sense isochrony in rhythm, but not the beat: Additional support for the gradual audiomotor evolution hypothesis <https://doi.org/10.3389/fnins.2018.00475>. *Frontiers in Neuroscience*, *12*, 475.

Honing, H., Cate, C., Peretz, I., & Trehub, S. E. (2015). Without it no music: Cognition, biology and evolution of musicality <https://doi.org/10.1098/rstb.2014.0088>. *Philosophical Transactions of the Royal Society B*, *370*, 20140088.

Honing, H., Merchant, H., Haden, G. P., Prado, L., & Bartolo, R. (2012). Rhesus monkeys (*Macaca mulatta*) detect rhythmic groups in music, but not the beat <https://doi.org/10.1371/journal.pone.0051369>. *PLOS ONE*, *7*(12), e51369.

Honing, H., & Ploeger, A. (2012). Cognition and the evolution of music: Pitfalls and prospects <https://doi.org/10.1111/j.1756-8765.2012.01210.x>. *Topics in Cognitive Science*, *4*, 513–524.

Huron, D. (1991). The ramp archetype: A score-based study of musical dynamics in 14 piano composers <https://doi.org/10.1177/0305735691191003>. *Music Perception*, *19*, 33–45.

Huron, D. (2001). Tone and voice: A derivation of the rules of voice-leading from perceptual principles <https://doi.org/10.1525/mp.2001.19.1.1>. *Music Perception*, *19*(1), 1–64.

Huron, D. (2006). *Sweet anticipation: Music and the psychology of expectation*. MIT Press.

Husain, G., Thompson, W. F., & Schellenberg, E. G. (2002). Effects of musical tempo and mode on arousal, mood, and spatial abilities <https://doi.org/10.1525/mp.2002.20.2.151>. *Music Perception*, *20*(2), 151–171.

Hutchinson, W., & Knopoff, L. (1978). The acoustic component of Western consonance <https://doi.org/10.1080/09298217808570246>. *Interface*, *7*, 1–29.

Hyde, K. L., Lerch, J. P., Zatorre, R. J., Griffiths, T. D., Evans, A. C., & Peretz, I. (2007). Cortical thickness in congenital amusia: When less is better than more <https://doi.org/10.1523/JNEUROSCI.3039-07.2007>. *Journal of Neuroscience*, *27*(47), 13028–13032.

Hyde, K. L., Zatorre, R. J., & Peretz, I. (2011). Functional MRI evidence of an abnormal neural network for pitch processing in congenital amusia <https://doi.org/10.1093/cercor/bhq094>. *Cerebral Cortex*, *21*, 292–299.

Iversen, J. R., Patel, A. D., & Ohgushi, K. (2008). Perception of rhythmic grouping depends on auditory experience <https://doi.org/10.1121/1.2973189>. *The Journal of the Acoustical Society of America*, *124*, 2263.

Iverson, P. (1995). Auditory stream segregation by musical timbre: Effects of static and dynamic acoustic attributes <https://doi.org/10.1037/0096-1523.21.4.751>. *Journal of Experimental Psychology: Human Perception and Performance*, *21*(4), 751–763.

Jacoby, N., & McDermott, J. H. (2017). Integer ratio priors on musical rhythm revealed cross-culturally by iterated reproduction <https://doi.org/10.1016/j.cub.2016.12.031>. *Current Biology*, *27*, 359–370.

Jacoby, N., Polak, R., Grahn, J. A., Cameron, D. J., Lee, K. M., Undurraga, E. A., Huanca, T., Thalwitzer, T., Doumbia, N., Margulis, E. H., Wong, P. C. M., Jure, L., Rocamora, M., Fujii, S., Savage, P. E., Ajimi, J., Konno, R., Oishi, S., Jakubowski, K., . . . Mcdermott, J. (2021). Universality and cross-cultural variation in mental representations of music revealed by global comparison of rhythm priors *<https://doi.org/10.31234/osf.io/b879v>*. PsyArXiv.

Johnson-Laird, P. N. (1991). Jazz improvisation: A theory at the computational level. In P. Howell, R. West, & I. Cross (Eds.), *Representing musical structure* (pp. 291–325). Academic Press.

Johnson-Laird, P. N., Kang, O. E., & Leong, Y. C. (2012). On musical dissonance *<https://doi.org/10.1525/mp.2012.30.1.19>*. *Music Perception*, *30*(1), 19–35.

Juslin, P. N. (2000). Cue utilization in communication of emotion in music performance: Relating performance to perception *<https://doi.org/10.1037/0096-1523.26.6.1797>*. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(6), 1797–1813.

Juslin, P. N. (2019). *Musical emotions explained: Unlocking the secrets of musical affect*. Oxford University Press.

Juslin, P. N., Liljeström, S., Västfjäll, D., Barradas, G., & Silva, A. (2008). An experience sampling study of emotional reactions to music: Listener, music, and situation *<https://doi.org/10.1037/a0013505>*. *Emotion*, *8*(5), 668–683.

Juslin, P. N., & Västfjäll, D. (2008). Emotional responses to music: The need to consider underlying mechanisms *<https://doi.org/10.1017/S0140525X08005293>*. *Behavioral and Brain Sciences*, *31*(5), 559–621.

Justus, T. C., & Hutsler, J. J. (2005). Fundamental issues in the evolutionary psychology of music: Assessing innateness and domain specificity *<https://doi.org/10.1525/mp.2005.23.1.1>*. *Music Perception*, *23*(1), 1–27.

Kämpfe, J., Sedlmeier, P., & Renkewitz, F. (2011). The impact of background music on adult listeners: A meta-analysis *<https://doi.org/10.1177/0305735610376261>*. *Psychology of Music*, *39*(4), 424–448.

Kaplan, T., Cannon, J. J., Jamone, L., & Pearce, M. T. (2022). Modeling enculturated bias in entrainment to rhythmic patterns *<https://doi.org/10.1371/journal.pcbi.1010579>*. *PLoS Computational Biology*, *18*(9), e1010579.

Karageorghis, C. I., & Priest, D. (2012a). Music in the exercise domain: A review and synthesis (Part I) *<https://doi.org/10.1080/1750984X.2011.631026>*. *International Review of Sport and Exercise Psychology*, *5*(1), 44–66.

Karageorghis, C. I., & Priest, D. (2012b). Music in the exercise domain: A review and synthesis (Part II) *<https://doi.org/10.1080/1750984X.2011.631027>*. *International Review of Sport and Exercise Psychology*, *5*(1), 67–84.

Karno, M., & Konečni, V. J. (1992). The effects of structural interventions in the first movement of Mozart's Symphony in G minor K. 550 on aesthetic preference *<https://doi.org/10.2307/40285538>*. *Music Perception*, *10*, 63–72.

Kivy, P. (2017). On the recent remarriage of music to philosophy *<https://doi.org/10.1111/jaac.12402>*. *Journal of Aesthetics and Art Criticism*, *75*(4), 429–438.

Koelsch, S., Fritz, T., von Cramon, D. Y., Müller, K., & Friederici, A. D. (2006). Investigating emotion with music: An fMRI study *<https://doi.org/10.1002/hbm.20180>*. *Human Brain Mapping*, *27*, 239–250.

Koelsch, S., Gunter, T. C., Wittfoth, M., & Sammler, D. (2005). Interaction between syntax processing in language and in music: An ERP study *<https://doi.org/10.1162/089892905774597290>*. *Journal of Cognitive Neuroscience*, *17*(10), 1565–1577.

Koelsch, S., Kasper, E., Sammler, D., Schulze, K., Gunter, T. C., & Friederici, A. D. (2004). Music, language and meaning: Brain signatures of semantic processing *<https://doi.org/10.1038/nn1197>*. *Nature Neuroscience*, *7*(3), 302–307.

Koelsch, S., Rohrmeier, M. A., Torrecuso, R., & Jentschke, S. (2013). Processing of hierarchical syntactic structure in music *<https://doi.org/10.1073/pnas.1300272110>*. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(38), 15443–15448.

Krumhansl, C. L. (1989). Why is musical timbre so hard to understand? In S. Nielzen & O. Olsson (Eds.), *Structure and perception of electroacoustic sound and music* (pp. 43–53). Elsevier.

Krumhansl, C. L. (1990). *Cognitive foundations of musical pitch*. Oxford University Press.

Krumhansl, C. L. (1995). Effects of musical context on similarity and expectancy. *Systematische Musikwissenschaft*, *3*(2), 211–250.

Krumhansl, C. L., Bharucha, J. J., & Castellano, M. A. (1982). Key distance effects on perceived harmonic structure in music *<https://doi.org/10.3758/BF03204269>*. *Perception & Psychophysics*, *32*(2), 96–108.

Krumhansl, C. L., Bharucha, J. J., & Kessler, E. J. (1982). Perceived harmonic structure of chords in three related musical keys *<https://doi.org/10.1037/0096-1523.8.1.24>*. *Journal of Experimental Psychology: Human Perception and Performance*, *8*(1), 24–36.

Krumhansl, C. L., & Keil, F. C. (1982). Acquisition of the hierarchy of tonal functions in music *<https://doi.org/10.3758/BF03197636>*. *Memory & Cognition*, *10*(3), 243–251.

Krumhansl, C. L., & Kessler, E. J. (1982). Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys *<https://doi.org/10.1037/0033-295X.89.4.334>*. *Psychological Review*, *89*(4), 334–368.

Krumhansl, C. L., Sandell, G. J., & Sergeant, D. C. (1987). The perception of tone hierarchies and mirror forms in twelve-tone serial music *<https://doi.org/10.2307/40285385>*. *Music Perception*, *5*(1), 31–78.

Lahdelma, I., & Athanasopoulos, G. (2021). Sweetness is in the ear of the beholder: Chord preference across United Kingdom and Pakistani listeners *<https://doi.org/10.1111/nyas.14655>*. *Annals of the New York Academy of Sciences*, *1502*, 72–84.

Lalitte, P., & Bigand, E. (2006). Music in the moment? Revisiting the effect of large scale structures *<https://doi.org/10.2466/pms.103.3.811-828>*. *Perceptual and Motor Skills*, *103*, 811–828.

Lamont, A., & Cross, I. (1994). Children's cognitive representations of musical pitch *<https://doi.org/10.2307/40285754>*. *Music Perception*, *12*(1), 27–55.

Large, E. W., Herrera, J. A., & Velasco, M. J. (2015). Neural networks for beat perception in musical rhythm *<https://doi.org/10.3389/fnsys.2015.00159>*. *Frontiers in Systems Neuroscience*, *9*, 159.

Large, E. W., & Jones, M. R. (1999). The dynamics of attending: How people track time-varying events *<https://doi.org/10.1037/0033-295X.106.1.119>*. *Psychological Review*, *106*(1), 119–159.

Large, E. W., & Kolen, J. F. (1994). Resonance and the perception of musical meter *<https://doi.org/10.1080/09540099408915723>*. *Connection Science*, *6*(2–3), 177–208.

Large, E. W., & Palmer, C. (2002). Perceiving temporal regularity in music *<https://doi.org/10.1016/S0364-0213(01)00057-X>*. *Cognitive Science*, *26*(1), 1–37.

Lee, C. S. (1991). The perception of metrical structure: Experimental evidence and a model. In P. Howell, R. West, & I. Cross (Eds.), *Representing musical structure* (pp. 59–127). Academic Press.

Lerdahl, F., & Jackendoff, R. (1983). *A generative theory of tonal music*. MIT Press.

Levinson, J. (2009). The aesthetic appreciation of music *<https://doi.org/10.1093/aesthj/ayp043>*. *British Journal of Aesthetics*, *49*(4), 415–425.

Lewontin, R. (1998). The evolution of cognition: Questions we will never answer. In D. Scarborough & S. Sternberg (Eds.), *An invitation to cognitive science, volume IV: Methods, models and conceptual issues* (2nd ed., pp. 107–132). MIT Press.

Liberman, A. M., Harris, K. S., Howard, S., Griffith, B. C., & Hill, M. (1957). The discrimination of speech sounds within and across phoneme boundaries *<https://doi.org/10.1037/h0044417>*. *Journal of Experimental Psychology*, *54*(5), 358–368.

Liu, F., Patel, A. D., Fourcin, A., & Stewart, L. (2010). Intonation processing in congenital amusia: Discrimination, identification and imitation *<https://doi.org/10.1093/brain/awq089>*. *Brain*, *133*(6), 1682–1693.

London, J. (2012). *Hearing in time* (2nd ed.). Oxford University Press.

Longuet-Higgins, H. C. (1976). Perception of melodies *<https://doi.org/10.1038/263646a0>*. *Nature*, *263*, 646–653.

Longuet-Higgins, H. C., & Lee, C. (1982). The perception of musical rhythms *<https://doi.org/10.1068/p110115>*. *Perception*, *11*, 115–128.

Longuet-Higgins, H. C., & Lee, C. S. (1984). The rhythmic interpretation of monophonic music *<https://doi.org/10.2307/40285271>*. *Music Perception*, *1*(4), 424–441.

Longuet-Higgins, H. C., & Steedman, M. J. (1971). On interpreting Bach. In B. Meltzer & D. Michie (Eds.), *Machine intelligence 6* (pp. 221–241). Edinburgh University Press.

Loui, P., Alsop, D., & Schlaug, G. (2009). Tone deafness: A new disconnection syndrome? *<https://doi.org/10.1523/JNEUROSCI.1701-09.2009>* *Journal of Neuroscience*, *29*(33), 10215–10220.

Loui, P., Wessel, D. L., & Kam, C. L. H. (2010). Humans rapidly learn grammatical structure in a new musical scale *<https://doi.org/10.1525/mp.2010.27.5.377>*. *Music Perception*, *27*(5), 377–388.

Lynch, M. P., Eilers, R. E., Oller, D. K., & Urbano, R. C. (1990). Innateness, experience, and music perception *<https://doi.org/10.1111/j.1467-9280.1990.tb00213.x>*. *Psychological Science*, *1*, 272–276.

Macdonald, R., Hargreaves, D. J., & Miell, D. (Eds.). (2002). *Musical identities*. Oxford University Press.

Margulis, E. H. (2014). *On repeat: How music plays the mind*. Oxford University Press.

Marozeau, J., & de Cheveigné, A. (2007). The effect of fundamental frequency on the brightness dimension of timbre <https://doi.org/10.1121/1.2384910>. *The Journal of the Acoustical Society of America*, *121*(1), 383–387.

Marvin, E. W., & Brinkman, A. (1999). The effect of modulation and formal manipulation on perception of tonic closure by expert listeners <https://doi.org/10.2307/40285801>. *Music Perception*, *16*(4), 389–407.

McAdams, S., Winsberg, S., Donnadieu, S., de Soete, G., & Krimphoff, J. (1995). Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes <https://doi.org/10.1007/BF00419633>. *Psychological Research*, *58*(3), 177–192.

McAuley, J. D. (1995). *Perception of time as phase: Toward an adaptive oscillator model of rhythmic pattern processing*. Indiana University.

McAuley, J. D., Jones, M. R., Holub, S., Johnston, H. M., & Miller, N. S. (2006). The time of our lives: Life span development of timing and event tracking <https://doi.org/10.1037/0096-3445.135.3.348>. *Journal of Experimental Psychology: General*, *135*(3), 348–367.

McAuley, J. D., Stevens, C., & Humphreys, M. S. (2004). Play it again: Did this melody occur more frequently or was it heard more recently? The role of stimulus familiarity in episodic recognition of music <https://doi.org/10.1016/j.actpsy.2004.02.001>. *Acta Psychologica*, *116*, 93–108.

McBride, J. M., & Tlusty, T. (2020). *Cross-cultural data shows musical scales evolved to maximise imperfect fifths* <https://doi.org/10.48550/arXiv.1906.06171>. ArXiv.

McCrae, R. R. (2007). Aesthetic chills as a universal marker of openness to experience <https://doi.org/10.1007/s11031-007-9053-1>. *Motivation and Emotion*, *31*, 5–11.

McDermott, J. H., Schultz, A. F., Undurraga, E. A., & Godoy, R. A. (2016). Indifference to dissonance in native Amazonians reveals cultural variation in music perception <https://doi.org/10.1038/nature18635>. *Nature*, *25*(7613), 21–25.

McLachlan, N., Marco, D., Light, M., & Wilson, S. (2013). Consonance and pitch <https://doi.org/10.1037/a0030830>. *Journal of Experimental Psychology: General*, *142*(4), 1142–1158.

Mehr, S. A., Krasnow, M. M., & Hagen, E. H. (2021). Origins of music in credible signaling <https://doi.org/10.1017/S0140525X20000345>. *Behavioral and Brain Sciences*, *44*, e60.

Mehr, S. A., Singh, M., Knox, D., Ketter, D. M., Pickens-Jones, D., Atwood, S., Lucas, C., Jacoby, N., Egner, A. A., Hopkins, E. J., Howard, R. M., Hartshorne, J. K., Jennings, M. V, Simson, J., Bainbridge, C. M., Pinker, S., Donnell, T. J. O., Krasnow, M. M., & Glowacki, L. (2019). Universality and diversity in human song <https://doi.org/10.1126/science.aax0868>. *Science*, *366*, eaax0868.

Meinz, E. J. (2000). Experience-based attenuation of age-related differences in music cognition tasks <https://doi.org/10.1037/0882-7974.15.2.297>. *Psychology and Aging*, *15*(2), 297–312.

Meyer, L. B. (1956). *Emotion and meaning in music*. University of Chicago Press.

Miller, G. A., & Heise, G. A. (1950). The trill threshold <https://doi.org/10.1121/1.1906663>. *The Journal of the Acoustical Society of America*, *22*(5), 637–638.

Miller, G. F. (2000). Evolution of human music through sexual selection <https://doi.org/10.7551/mitpress/5190.003.0025>. In N. L. Wallin, B. Merker, & S. Brown (Eds.), *The origins of music* (pp. 329–360). MIT Press.

Miyazaki, K. (1988). Musical pitch identification by absolute pitch possessors <https://doi.org/10.3758/BF03207484>. *Perception & Psychophysics*, *44*(6), 501–512.

Moore, B. C. J. (2013). *An introduction to the psychology of hearing* (6th ed.). Brill.

Morrison, S. J., Demorest, S. M., & Stambaugh, L. A. (2008). Enculturation effects in music cognition: The role of age and music complexity <https://doi.org/10.1177/0022429408322854>. *Journal of Research in Music Education*, *56*(2), 118–129.

Mosing, M. A., Madison, G., Pedersen, N. L., Kuja-halkola, R., & Ullén, F. (2014). Practice does not make perfect: No causal effect of music practice on music ability <https://doi.org/10.1177/0956797614541990>. *Psychological Science*, *25*(9), 1795–1803.

Müllensiefen, D., Gingras, B., Musil, J., & Stewart, L. (2014). The musicality of non-musicians: An index for assessing musical sophistication in the general population <https://doi.org/10.1371/journal.pone.0089642>. *PLOS ONE*, *9*(2), e89642.

Mungan, E., Yazıcı, Z. F., & Kaya, M. (2017). Perceiving boundaries in unfamiliar Turkish makam music: Evidence for Gestalt universals? <https://doi.org/10.1525/mp.2017.34.3.267> *Music Perception*, *34*(3), 267–290.

Nantais, K. M., & Schellenberg, E. G. (1999). The Mozart effect: An artifact of preference <https://doi.org/10.1111/1467-9280.00170>. *Psychological Science*, *10*(4), 370–373.

Narmour, E. (1990). *The analysis and cognition of basic melodic structures: The implication-realisation model*. University of Chicago Press.

North, A. C., & Hargreaves, D. J. (2000). Musical preferences during and after relaxation and exercise <https://doi.org/10.2307/1423460>. *American Journal of Psychology*, *113*(1), 43–67.

North, A. C., Hargreaves, D. J., & McKendrick, J. (1999). The influence of in-store music on wine selections <https://doi.org/10.1037/0021-9010.84.2.271>. *Journal of Applied Psychology*, *84*(2), 271–276.

North, A. C., Shilcock, A., & Hargreaves, D. J. (2003). The effect of musical style of restaurant customers' spending <https://doi.org/10.1177/0013916503254749>. *Environment & Behavior*, *35*(5), 712–718.

Nusbaum, E. C., & Silvia, P. J. (2011). Shivers and timbres: Personality and the experience of chills from music <https://doi.org/10.1177/1948550610386810>. *Social Psychological and Personality Science*, *2*(2), 199–204.

Olsen, K. N., Dean, R. T., & Leung, Y. (2016). What constitutes a phrase in sound-based music? A mixed-methods investigation of perception and acoustics <https://doi.org/10.1371/journal.pone.0167643>. *PLOS ONE*, *11*(12), e0167643.

Palmer, C., & Krumhansl, C. L. (1987). Independent temporal and pitch structures in determination of musical phrases <https://doi.org/10.1037/0096-1523.13.1.116>. *Journal of Experimental Psychology: Human Perception and Performance*, *13*(1), 116–126.

Palmer, C., & Krumhansl, C. L. (1990). Mental representations for musical meter <https://doi.org/10.1037/0096-1523.16.4.728>. *Journal of Experimental Psychology: Human Perception and Performance*, *16*(4), 728–741.

Pantev, C., Roberts, L. E., Schulz, M., Engelien, A., & Ross, B. (2001). Timbre-specific enhancement of auditory cortical representations in musicians. *NeuroReport*, *12*(1), 169–174.

Paraskeva, S., & McAdams, S. (1997). Influence of timbre, presence/absence of tonal hierarchy and musical training on the perception of musical tension and relaxation schemas. In T. Rikakis & P. Cook (Eds.), *Proceedings of the 1997 international computer music conference, Thessaloniki*. International Computer Music Association.

Partanen, E., Kujala, T., Näätänen, R., Liitola, A., Sambeth, A., & Huotilainen, M. (2013). Learning-induced neural plasticity of speech processing before birth <https://doi.org/10.1073/pnas.1302159110>. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(37), 15145–15150.

Patel, A. D. (2003). Language, music, syntax and the brain <https://doi.org/10.1038/nn1082>. *Nature Neuroscience*, *6*(7), 674–681.

Patel, A. D. (2021). Vocal learning as a preadaptation for the evolution of human beat perception and synchronization <https://doi.org/10.1098/rstb.2020.0326>. *Philosophical Transactions of the Royal Society B*, *376*, 20200326.

Patel, A. D. (2022). Human musicality and gene-culture coevolution: Ten concepts to guide productive exploration. In E. H. Margulis, D. Loughridge, & P. Loui (Eds.), *The science-music borderlands: Reckoning with the past, imagining the future*. MIT Press.

Patel, A. D., & Iversen, J. R. (2014). The evolutionary neuroscience of musical beat perception: The Action Simulation for Auditory Prediction (ASAP) hypothesis <https://doi.org/10.3389/fnsys.2014.00057>. *Frontiers in Systems Neuroscience*, *8*, 1–14.

Patel, A. D., Iversen, J. R., Bregman, M. R., & Schultz, I. (2009). Experimental evidence for synchronization to a musical beat in a nonhuman animal <https://doi.org/10.1016/j.cub.2009.03.038>. *Current Biology*, *19*(10), 827–830.

Pearce, M. T. (2018). Statistical learning and probabilistic prediction in music cognition: Mechanisms of stylistic enculturation <https://doi.org/10.1111/nyas.13654>. *Annals of the New York Academy of Sciences*, *1423*, 378–395.

Pearce, M. T., & Rohrmeier, M. A. (2018). Musical syntax II: Empirical perspectives <https://doi.org/10.1007/978-3-662-55004-5_26>. In R. Bader (Ed.), *Springer handbook of systematic musicology* (pp. 487–505). Springer.

Pelofi, C., & Farbood, M. M. (2021). Asymmetry in scales enhances learning of new musical structures <https://doi.org/10.1073/pnas.2014725118>. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(31), e2014725118.

Peretz, I. (1989). Clustering in music: An appraisal of task factors <https://doi.org/10.1080/00207594.1989.10600040>. *International Journal of Psychology*, *24*(2), 157–178.

Peretz, I. (2013). The biological foundations of music: Insights from congenital amusia. In D. Deutsch (Ed.), *The psychology of music* (pp. 551–564). Academic Press.

Perlman, M., & Krumhansl, C. L. (1996). An experimental study of internal interval standards in Javanese and Western musicians *<https://doi.org/10.2307/40285714>*. *Music Perception*, *14*(2), 95–116.

Perruchet, P., & Poulin-Charronnat, B. (2013). Challenging prior evidence for a shared syntactic processor for language and music *<https://doi.org/10.3758/s13423-012-0344-5>*. *Psychonomic Bulletin & Review*, *20*, 310–317.

Pfordresher, P. (2022). Music Performance *<https://doi.org/10.1093/acrefore/9780190236557.013.492>*. In *Oxford Research Encyclopedia of Psychology*. Oxford University Press.

Pietschnig, J., Voracek, M., & Formann, A. K. (2010). Mozart effect–Shmozart effect: A meta-analysis *<https://doi.org/10.1016/j.intell.2010.03.001>*. *Intelligence*, *38*(3), 314–323.

Pinker, S. (1997). *How the mind works*. W. W. Norton.

Pitt, M. A., & Crowder, R. G. (1992). The role of spectral and dynamic cues in imagery for musical timbre *<https://doi.org/10.1037/0096-1523.18.3.728>*. *Journal of Experimental Psychology: Human Perception and Performance*, *18*(3), 728–738.

Plack, C. J. (2018). *The sense of hearing* (3rd ed.). Taylor & Francis.

Plantinga, J., & Trainor, L. J. (2005). Memory for melody: Infants use a relative pitch code *<https://doi.org/10.1016/j.cognition.2004.09.008>*. *Cognition*, *98*(1), 1–11.

Plomp, R., & Levelt, W. J. M. (1965). Tonal consonance and critical bandwidth *<https://doi.org/10.1121/1.1909741>*. *The Journal of the Acoustical Society of America*, *38*, 548–560.

Polak, R., Jacoby, N., Fischinger, T., Goldberg, D., Holzapfel, A., & London, J. (2018). Rhythmic prototypes across cultures: A comparative study of tapping synchronization *<https://doi.org/10.1525/mp.2018.36.1.1>*. *Music Perception*, *36*, 1–23.

Poulin-Charronnat, B., Bigand, E., Madurell, F., & Peereman, R. (2005). Musical structure modulates semantic priming in vocal music *<https://doi.org/10.1016/j.cognition.2004.05.003>*. *Cognition*, *94*(3), B67–B78.

Povel, D.-J. J., & Essens, P. J. (1985). Perception of temporal patterns *<https://doi.org/10.2307/40285311>*. *Perception*, *2*(4), 411–440.

Prince, J. B., & Schmuckler, M. A. (2014). The tonal-metric hierarchy: A corpus analysis *<https://doi.org/10.1525/mp.2014.31.3.254>*. *Music Perception*, *31*(3), 254–270.

Prince, J. B., Thompson, W. F., & Schmuckler, M. A. (2009). Pitch and time, tonality and meter: How do musical dimensions combine? *<https://doi.org/10.1037/a0016456>* *Journal of Experimental Psychology: Human Perception and Performance*, *35*(5), 1598–1617.

Rauscher, F. H., Shaw, G. L., & Ky, K. N. (1993). Music and spatial task performance *<https://doi.org/10.1038/365611a0>*. *Nature*, *365*, 611.

Repp, B. H. (1991). Some cognitive and perceptual aspects of speech and music. In J. Sundberg, L. Nord, & R. Carlsen (Eds.), *Music, language, speech and brain* (pp. 257–258). Palgrave.

Repp, B. H. (2003). Rate limits in sensorimotor synchronization with auditory and visual sequences: The synchronization threshold and the benefits and costs of interval subdivision *<https://doi.org/10.1080/00222890309603156>*. *Journal of Motor Behavior*, *35*(4), 355–370.

Rohrmeier, M. A. (2011). Towards a generative syntax of tonal harmony *<https://doi.org/10.1080/17459737.2011.573676>*. *Journal of Mathematics and Music*, *5*(1), 35–53.

Rohrmeier, M. A., & Pearce, M. T. (2018). Musical syntax I: Theoretical perspectives *<https://doi.org/10.1007/978-3-662-55004-5_25>*. In R. Bader (Ed.), *Springer handbook of systematic musicology* (pp. 473–486). Springer.

Rolison, J. J., & Edworthy, J. (2012). The role of formal structure in liking for popular music *<https://doi.org/10.1525/mp.2012.29.3.269>*. *Music Perception*, *29*(3), 269–284.

Rusconi, E., Kwan, B., Giordano, B. L., & Umilta, C. (2006). Spatial representation of pitch height: The SMARC effect *<https://doi.org/10.1016/j.cognition.2005.01.004>*. *Cognition*, *99*, 113–129.

Russell, J. A. (1980). A circumplex model of affect *<https://doi.org/10.1037/h0077714>*. *Journal of Personality and Social Psychology*, *39*(6), 1161–1178.

Russo, F. A., & Cuddy, L. L. (1999). A common origin for vocal accuracy and melodic expectancy: Vocal constraints *<https://doi.org/10.1121/1.425870>*. *The Journal of the Acoustical Society of America*, *105*(2), 1217.

Saffran, J. R., & Kirkham, N. Z. (2018). Infant statistical learning *<https://doi.org/10.1146/annurev-psych-122216-011805>*. *Annual Review of Psychology*, *69*, 181–203.

Salame, P., & Baddeley, A. (1989). Effects of background music on phonological short-term memory *<https://doi.org/10.1080/14640748908402355>*. *Quarterly Journal of Experimental Psychology*, *41A*, 107–122.

Salimpoor, V. N., Benovoy, M., Larcher, K., Dagher, A., & Zatorre, R. J. (2011). Anatomically distinct dopamine release during anticipation and experience of peak emotion to music *<https://doi.org/10.1038/nn.2726>*. *Nature Neuroscience*, *14*, 257–262.

Samson, S. (2003). Neuropsychological studies of musical timbre *<https://doi.org/10.1196/annals.1284.016>*. *Annals of the New York Academy of Sciences*, *999*, 144–151.

Savage, P. E., Loui, P., Tarr, B., & Schachner, A. (2021). Music as a coevolved system for social bonding *<https://doi.org/10.1017/S0140525X20000333>*. *Behavioural and Brain Sciences*, *44*, 1–22.

Schellenberg, E. G. (1996). Expectancy in melody: Tests of the implication-realisation model *<https://doi.org/10.1016/0010-0277(95)00665-6>*. *Cognition*, *58*(1), 75–125.

Schellenberg, E. G. (1997). Simplifying the implication-realisation model of melodic expectancy *<https://doi.org/10.2307/40285723>*. *Music Perception*, *14*(3), 295–318.

Schellenberg, E. G., & Hallam, S. (2005). Music listening and cognitive abilities in 10- and 11-year-olds: The blur effect *<https://doi.org/10.1196/annals.1360.013>*. *Annals of the New York Academy of Sciences*, *1060*, 202–209.

Schellenberg, E. G., Iverson, P., & McKinnon, M. C. (1999). Name that tune: Identifying popular recordings from brief excerpts <https://doi.org/10.3758/BF03212973>. *Psychonomic Bulletin & Review*, *6*(4), 641–646.

Schellenberg, E. G., & Trehub, S. E. (2003). Good pitch memory is widespread <https://doi.org/10.1111/1467-9280.03432>. *Psychological Science*, *14*(3), 262–266.

Schellenberg, E. G., & Trehub, S. E. (2008). Is there an Asian advantage for pitch memory? <https://doi.org/10.1525/mp.2008.25.3.241> *Music Perception*, *25*(3), 241–252.

Schellenberg, E. G., & Weiss, M. W. (2013). Music and cognitive abilities. In D. Deutsch (Ed.), *The psychology of music* (pp. 499–550). Academic Press.

Schenker, H. (1979). *Free composition* (E. Oster, Ed. and Trans.). Longman. (Original work published 1935)

Schultz, W., Dayan, P., & Montague, R. (1997). A neural substrate of prediction and reward <https://doi.org/10.1126/science.275.5306.1593>. *Science*, *275*, 1593–1599.

Schulze, H. (1989). Categorical perception of rhythmic patterns <https://doi.org/10.1007/BF00309270>. *Psychological Research*, *51*, 10–15.

Sears, D. R. W., Pearce, M. T., Spitzer, J., Caplin, W. E., & McAdams, S. (2019). Expectations for tonal cadences: Sensory and cognitive priming effects <https://doi.org/10.1177/1747021818814472>. *Quarterly Journal of Experimental Psychology*, *72*, 1422–1438.

Serafine, M. L., Glassman, N., & Overbeeke, C. (1989). The cognitive reality of hierarchic structure in music <https://doi.org/10.2307/40285440>. *Music Perception*, *6*(4), 397–430.

Shepard, R. N. (1982). Structural representations of musical pitch. In D. Deutsch (Ed.), *Psychology of music* (pp. 343–390). Academic Press.

Skerritt-Davis, B., & Elhilali, M. (2021). Computational framework for investigating predictive processing in auditory perception <https://doi.org/10.1016/j.jneumeth.2021.109177>. *Journal of Neuroscience Methods*, *360*, 109177.

Slevc, L. R., & Okada, B. M. (2015). Processing structure in language and music: A case for shared reliance on cognitive control <https://doi.org/10.3758/s13423-014-0712-4>. *Psychonomic Bulletin & Review*, *22*, 637–652.

Slevc, L. R., Rosenberg, J. C., & Patel, A. D. (2009). Making psycholinguistics musical: Self-paced reading time evidence for shared processing of linguistic and musical syntax <https://doi.org/10.3758/16.2.374>. *Psychonomic Bulletin & Review*, *16*(2), 374–381.

Snyder, B. (2000). *Music and memory: An introduction*. MIT Press.

Snyder, B. (2016). Memory for music. In S. Hallam, I. Cross, & M. Thaut (Eds.), *The Oxford handbook of music psychology* (2nd ed., pp. 167–180). Oxford University Press.

Speer, J. R., & Meeks, U. P. (1985). School children's perception of pitch in music <https://doi.org/10.1037/h0094200>. *Psychomusicology*, *5*, 49–56.

Spence, C., & Wang, Q. J. (2015). Wine and music (II): Can you taste the music? Modulating the experience of wine through music and sound <https://doi.org/10.1186/s13411-015-0043-z>. *Flavour*, *4*, 33.

Steedman, M. J. (1996). The blues and the abstract truth: Music and mental models. In A. Garnham & J. Oakhill (Eds.), *Mental models in cognitive science* (pp. 305–318). Erlbaum.

Steele, K. M., & Williams, A. K. (2006). Is the bandwidth for timbre invariance only one octave? <https://doi.org/10.1525/mp.2006.23.3.215> *Music Perception*, *23*(3), 215–220.

Steinbeis, N., & Koelsch, S. (2007). Shared neural resources between music and language indicate semantic processing of musical tension-resolution patterns <https://doi.org/10.1093/cercor/bhm149>. *Cerebral Cortex*, *18*(5), 1169–1178.

Steinbeis, N., Koelsch, S., & Sloboda, J. A. (2006). The role of harmonic expectancy violations in musical emotions: Evidence from subjective, physiological and neural responses <https://doi.org/10.1162/jocn.2006.18.8.1380>. *Journal of Cognitive Neuroscience*, *18*(8), 1380–1393.

Stewart, L. (2011). Characterizing congenital amusia <https://doi.org/10.1080/17470218.2011.552730>. *Quarterly Journal of Experimental Psychology*, *64*(4), 625–638.

Swaminathan, S., & Schellenberg, E. G. (2019). Music training and cognitive abilities: Associations, causes, and consequences <https://doi.org/10.1093/oxfordhb/9780198804123.013.26>. In M. H. Thaut & D. A. Hodges (Eds.), *The Oxford handbook of music and the brain* (pp. 645–670). Oxford University Press.

Tan, S.-L., & Spackman, M. P. (2005). Listeners' judgments of the musical unity of structurally altered and intact musical compositions <https://doi.org/10.1177/0305735605050648>. *Psychology of Music*, *33*, 133–153.

Teki, S., Grube, M., Kumar, S., & Griffiths, T. D. (2011). Distinct neural substrates of duration-based and beat-based auditory timing <https://doi.org/10.1523/JNEUROSCI.5561-10.2011>. *Journal of Neuroscience*, *31*(10), 3805–3812.

Temperley, D. (2001). *The cognition of basic musical structures*. MIT Press.

Temperley, D. (2007). *Music and probability*. MIT Press.

Temperley, D. (2009). A unified probabilistic model for polyphonic music analysis <https://doi.org/10.1080/09298210902928495>. *Journal of New Music Research*, *38*(1), 3–18.

Tenderini, M. S., de Leeuw, E., Eilola, T. M., & Pearce, M. T. (2022). Reduced cross-modal affective priming in the L2 of late bilinguals depends on L2 exposure <https://doi.org/10.1037/xlm0000889>. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *48*(2), 284–303.

Terhardt, E. (1974). Pitch, consonance, and harmony <https://doi.org/10.1121/1.1914648>. *The Journal of the Acoustical Society of America*, *55*(5), 1061–1069.

Theusch, E., & Gitschier, J. (2011). Absolute Pitch Twin Study and Segregation Analysis <https://doi.org/10.1375/twin.14.2.173>. *Twin Research and Human Genetics*, *14*(2), 173-178.

Thompson, W. F., Bullot, N. J., & Margulis, E. H. (2023). The psychological basis of music appreciation: Structure, self, source <https://doi.org/10.1037/rev0000364>. *Psychological Review*, *30*(1), 260–284.

Thompson, W. F., Schellenberg, E. G., & Husain, G. (2001). Arousal, mood, and the Mozart effect <https://doi.org/10.1111/1467-9280.00345>. *Psychological Science*, *12*(3), 248–251.

Thompson, W. F., Schellenberg, E. G., & Letnic, A. K. (2011). Fast and loud background music disrupts reading comprehension <https://doi.org/10.1177/0305735611400173>. *Psychology of Music*, *40*(6), 700–708.

Thompson, W. F., & Stainton, M. (1998). Expectancy in Bohemian folk song melodies: Evaluation of implicative principles for implicative and closural intervals <https://doi.org/10.2307/40285766>. *Music Perception*, *15*(3), 231–252.

Tichko, P., & Large, E. W. (2019). Modeling infants perceptual narrowing to musical rhythms: Neural oscillation and Hebbian plasticity <https://doi.org/10.1111/nyas.14050>. *Annals of the New York Academy of Sciences*, *1453*(1), 125–139.

Tierney, A. T., Russo, F. A., & Patel, A. D. (2011). The motor origins of human and avian song structure <https://doi.org/10.1073/pnas.1103882108>. *Proceedings of the National Academy of Sciences of the United States of America*, *108*, 15510–15515.

Tillmann, B., & Bigand, E. (1996). Does formal musical structure affect perception of musical expressiveness? <https://doi.org/10.1177/0305735696241002> *Psychology of Music*, *24*, 3–17.

Tillmann, B., Bigand, E., & Pineau, M. (1998). Effects of global and local contexts on harmonic expectancy <https://doi.org/10.2307/40285780>. *Music Perception*, *16*, 99–118.

Tillmann, B., & McAdams, S. (2004). Implicit learning of musical timbre sequences: Statistical regularities confronted with acoustical (dis)similarities <https://doi.org/10.1037/0278-7393.30.5.1131>. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(5), 1131–1142.

Tinbergen, N. (1963). On aims and methods of ethology <https://doi.org/10.1111/j.1439-0310.1963.tb01161.x>. *Zeitschrift für Tierpsychologie*, *20*, 410–433.

Todd, N. P. M., Cousins, R., & Lee, C. S. (2007). The contribution of anthropometric factors to individual differences in the perception of rhythm <https://doi.org/10.18061/1811/24478>. *Empirical Musicology Reviews*, *2*(1), 1–13.

Trainor, L. J., & Heinmiller, B. M. (1998). The development of evaluative responses to music: Infants prefer to listen to consonance over dissonance <https://doi.org/10.1016/S0163-6383(98)90055-8>. *Infant Behavior and Development*, *21*(1), 77–88.

Trainor, L. J., & Trehub, S. E. (1992). A comparison of infants' and adults' sensitivity to Western musical structure <https://doi.org/10.1037/0096-1523.18.2.394>. *Journal of Experimental Psychology: Human Perception and Performance*, *18*(2), 394–402.

Trainor, L. J., & Trehub, S. E. (1994). Key membership and implied harmony in Western tonal music: Developmental perspectives <https://doi.org/10.3758/BF03213891>. *Perception & Psychophysics*, *56*(2), 125–132.

Trainor, L. J., Wu, L., & Tsang, C. D. (2004). Long-term memory for music: Infants remember tempo and timbre <https://doi.org/10.1111/j.1467-7687.2004.00348.x>. *Developmental Science*, *7*(3), 289–296.

Trehub, S. E., Cohen, A. J., Thorpe, L. A., & Morrongiello, B. A. (1986). Development of the perception of musical relations: Semitone and diatonic structure <https://doi.org/10.1037/0096-1523.12.3.295>. *Journal of Experimental Psychology: Human Perception and Performance*, *12*(3), 295–301.

Trehub, S. E., & Thorpe, L. A. (1989). Infants' perception of rhythm: Categorization of auditory sequences by temporal structure <https://doi.org/10.1037/h0084223>. *Canadian Journal of Psychology*, *43*(2), 217–229.

Trehub, S. E., Unyk, A. M., & Trainor, L. J. (1993). Adults identify infant-directed music across cultures <https://doi.org/10.1016/0163-6383(93)80017-3>. *Infant Behavior and Development*, *16*, 193–211.

van der Weij, B., Pearce, M. T., & Honing, H. (2017). A probabilistic model of meter perception: Simulating enculturation <https://doi.org/10.3389/fpsyg.2017.00824>. *Frontiers in Psychology*, *8*, 824.

van der Weij, B., Pearce, M. T., & Honing, H. (2023). Computational modelling of rhythm perception and the role of enculturation. In D. Shanahan, J. A. Burgoyne, & I. Quinn (Eds.) *Oxford Handbook of Music and Corpus Studies*. Oxford University Press.

van Noorden, L. P. A. S. (1975). *Temporal coherence in the perception of tone sequences*. Eindhoven University of Technology.

Vieillard, S., Peretz, I., Gosselin, N., Khalfa, S., Gagnon, L., & Bouchard, B. (2008). Happy, sad, scary and peaceful musical excerpts for research on emotions <https://doi.org/10.1080/02699930701503567>. *Cognition and Emotion*, *22*(4), 720–752.

von Helmholtz, H. (1863). *On the sensations of tone as a physiological basis for the theory of music*. Dover.

Vuoskoski, J. K., Thompson, W. F., McIlwain, D., & Eerola, T. (2012). Who enjoys listening to sad music and why? <https://doi.org/10.1525/mp.2012.29.3.311> *Music Perception*, *29*(3), 311–317.

Wallentin, M. (2010). The musical ear test, a new reliable test for measuring musical competence <https://doi.org/10.1016/j.lindif.2010.02.004>. *Learning and Individual Differences*, *20*(3), 188–196.

Warren, J. D., Uppenkamp, S., Patterson, R. D., & Griffiths, T. D. (2003). Separating pitch chroma and pitch height in the human brain <https://doi.org/10.1073/pnas.1730682100>. *Proceedings of the National Academy of Sciences of the United States of America*, *100*, 10038–10042.

Wessel, D. L. (1979). Timbre space as a musical control structure <https://doi.org/10.2307/3680283>. *Computer Music Journal*, *3*, 45–52.

Winkler, I., Háden, G. P., Ladinig, O., Sziller, I., & Honing, H. (2009). Newborn infants detect the beat in music <https://doi.org/10.1073/pnas.0809035106>. *Proceedings of the National Academy of Sciences of the United States of America*, *106*, 2468–2471.

Wong, P. C. M., Roy, A., & Margulis, E. H. (2009). Bimusicalism: The implicit dual enculturation of cognitive and affective systems <https://doi.org/10.1525/mp.2009.27.2.81>. *Music Perception*, *27*, 81–88.

Wundt, W. M. (1874). *Grundzüge der physiologischen Psychologie*. Wilhelm Engelmann.

Yeoh, J. P. S., & North, A. C. (2010). The effects of musical fit on choice between two competing foods <https://doi.org/10.1177/102986491001400107>. *Musicae Scientiae*, *14*, 165–180.

Yoshida, K. A., Iversen, J. R., Patel, A. D., Mazuka, R., Nito, H., Gervain, J., & Werker, J. F. (2010). The development of perceptual grouping biases in infancy: A Japanese-English cross-linguistic study <https://doi.org/10.1016/j.cognition.2010.01.005>. *Cognition*, *115*, 356–361.

Zatorre, R. J. (1988). Pitch perception of complex tones and human temporal lobe function <https://doi.org/10.1121/1.396834>. *The Journal of the Acoustical Society of America*, *84*, 566–572.

Zatorre, R. J. (2005). Finding the missing fundamental <https://doi.org/10.1038/4361093a>. *Nature*, *436*, 1093–1100.

Related Articles

Auditory Attention

Music Performance