Trends in
Cognitive Sciences

CellPress

Review

Core systems of music perception

Samuel A. Mehr [1,2,*]

Human musicality is supported by two distinct systems of representation: one for tonal perception, which contextualizes pitch input in reference to a hierarchy of tones; and one for metrical perception, which contextualizes temporal input in reference to a hierarchy of rhythmic groupings. Growing evidence suggests that the two systems are universal, automatic, encapsulated, and relatively early-developing. But like speech perception, and unlike several other perceptual systems, they appear to be uniquely human. The systems of tonal and metrical perception form a foundational structure for musicality that, when combined with the processing of other acoustical information (e.g., timbre or auditory scenes), and applied in conjunction with other cognitive domains, yields a human psychology of music.

Highlights

Two complementary systems underlie human musicality: tonal and metrical perception.

Tonal and metrical perception show evidence for automaticity, encapsulation, and relatively early development.

Despite the variability of music worldwide, tonal and metrical perception appear to be universal.

Similar to systems underlying language, but unlike those underlying other cognitive domains, tonal and metrical perception are uniquely human.

Tonal and metrical perception form the basis for human musicality.

The puzzle of musicality

> *The theory of Gestalt-qualities began with the attempt to answer a question: What is melody? First and most obvious answer: the sum of the individual notes which make up the melody. But opposed to this is the fact that the same melody may be made up of quite different groups of notes, as happens when the self-same melody is transposed into different keys… the essence of melody must reside in a sum of special sensations which as note-sensations accompany the notes.*
>
> [Christian von Ehrenfels, *On Gestalt-Qualities* (1937)]

When a human listens to music, the voice, instrument, or recording sends vibrations into the ear. These musical vibrations are transmitted, via the eardrum, through a tiny network of bone, muscle, fluid, and flesh, to eventually be converted into electrical signals by hair cells in the organ of Corti, located on the basilar membrane, and sent up the cochlear nerve to the brain.

This is already an odd process, but what happens next is odder. The listener may move, be moved, become excited or sad, experience chills, or fall asleep. They may want to listen again, perhaps together with others, or join in, making music themselves.

How does this work? Music perception obviously is more than the absorption of vibrations: it involves 'an active structuring of information in forms not explicitly present in the external signal' [1]. How are musical bits converted into structured information? How does the mind make sense of music?

I show that two distinct systems of representation are the fundamental components underlying music perception (Figure 1). These are: (i) a system for tonal perception that takes as input pitch information and represents tones hierarchically, in terms of relative stability; and (ii) a system for metrical perception that takes as input temporal sequences and represents them hierarchically in terms of the relative strength of their locations in time.

Tonal and metrical representations have been extensively studied (non-exhaustively [1–9]), but the degree of their **universality** (see Glossary), **automaticity**, **encapsulation**, and early appearance

[1]School of Psychology, University of Auckland, Auckland, New Zealand
[2]Child Study Center, Yale University, New Haven, CT, USA

*Correspondence:
sam@auckland.ac.nz (S.A. Mehr).

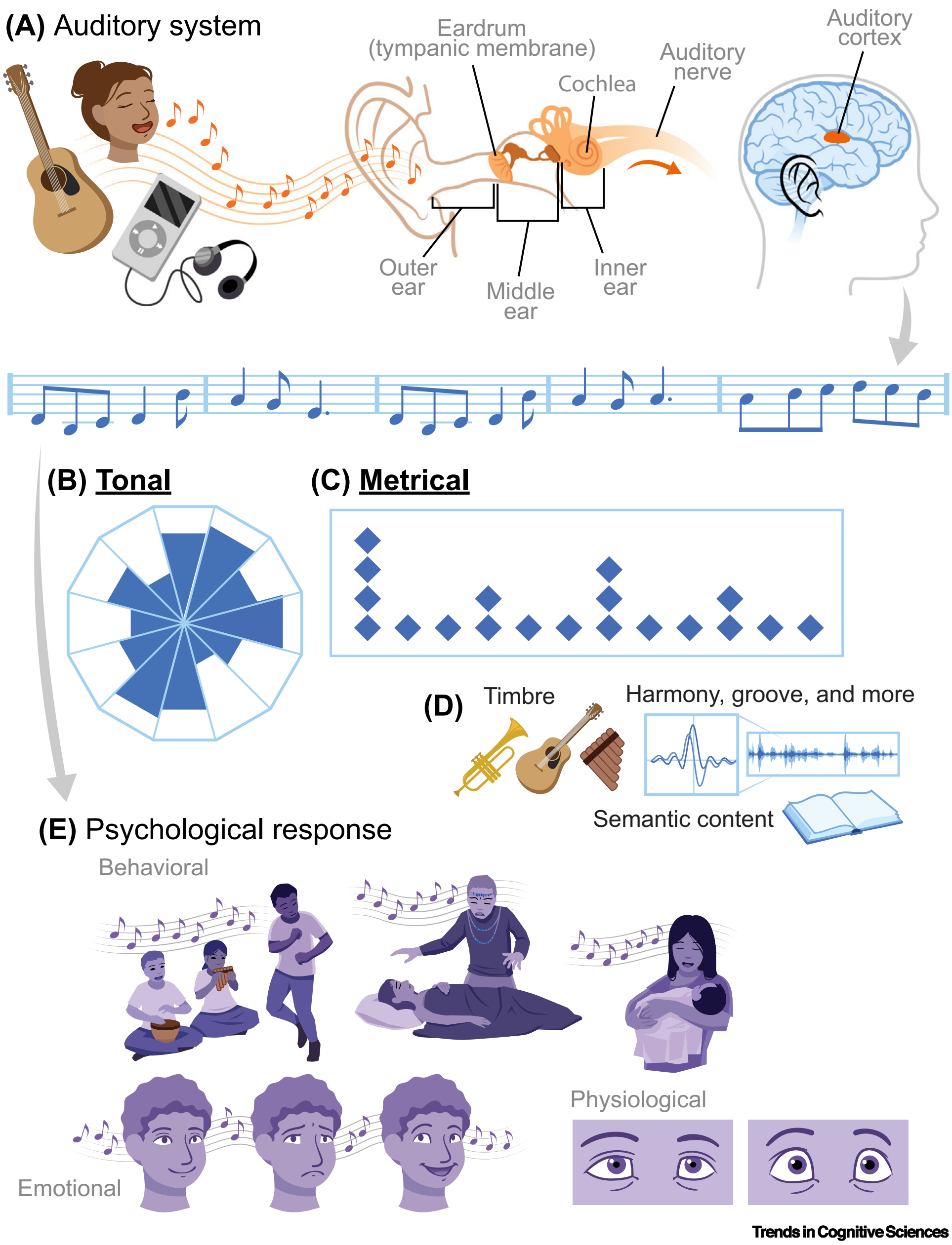

Figure 1. From unstructured acoustic input to structured musical representations. (A) Sound waves containing musical information enter the ear; the auditory system converts this vibrating air into neuronal signals, which are processed in auditory cortex. Human music perception is specialized to process this 'raw data' via (B) tonal perception, where pitch information (events delineated by changes in frequency, i.e., higher or lower sounds) is interpreted in the context of the relative stability of tones, depicted by a polar histogram of stability levels; and (C) metrical perception, where temporal information (events delineated by their timing and temporal patterns) is interpreted in the context of the relative strength of locations in an underlying beat, depicted by a dotplot of strength levels. (D) Other musical information, such as lyrics and timbral characteristics, complement these two primary percepts. (E) Together, tonal and metrical representations combine with each other and with other information (from external sources and from within the acoustical signal) to yield musical understanding, characterized by a variety of psychological responses to music.

Glossary

Amusia: congenital, developmental, or acquired disorder affecting the processing of pitch information. Colloquially, amusia is referred to as 'tone-deafness'.

Auditory scene analysis: grouping of individual elements of auditory environments into discrete perceptual units (i.e., objects and streams), namely, sound segregation. Auditory scene analysis is considered the fundamental psychological process of audition across species.

Automaticity: property of cognition wherein a process is undertaken via implicit knowledge, without requiring intervention or explicit understanding; this property is a spectrum, with some processes high on it (e.g., the startle reflex) and some not (e.g., high-level verbal reasoning).

Beat: the pulse, a periodic, repetitive stimulus that forms the basic rhythmic foundation of music; it is usually isochronous (i.e., with an immutable period length).

Cocktail party problem: the problem of auditory scene analysis in natural auditory environments where many sounds co-occur but the listener must focus on only one of them, as at a cocktail party; the phenomenon includes both sound segregation and the direction of attention to the sound source of interest.

Encapsulation: property of cognition wherein a process has some degree of impenetrability, in that it is unalterable by beliefs, desires, or knowledge, and does not rely on information from other processes; this property is a spectrum, with some processes high on it (e.g., depth perception in a visual field) and some not (e.g., high-level mathematical reasoning).

Groove: property of some forms of music that prompts dance and other rhythmic movement from listeners and/or performers.

Melodic contour: directional pattern of a melody relative to its tonal content, as opposed to an absolute pitch level. For example, in English, a melodic contour might be described as going 'up' and then 'down'.

Metrical hierarchy: high-level organization of rhythmic information in music, involving multiple levels of beat strength, such that some beats are consistently heard as stronger than in ontogeny are not yet widely appreciated in the cognitive sciences. Meanwhile, comparative research has revealed a strong degree of human uniqueness. All these characteristics are signatures of foundational perceptual systems in our species.

Two systems of musical representation

Speech, our most common vocalization, shares some features with music (e.g., both contain spectral modulations), but some differences are immediately evident in the treatment of frequency and temporal information by the two modalities. When reading the previous sentence aloud, if you double the duration of the word 'treatment', altering the temporal pattern of the sentence; or raise the frequency of your voice while saying it, altering the pitch pattern of the sentence; the core content of the sentence is mostly preserved (in nontonal languages).

This is not true of music, where altering the duration or frequency of a tone alters the identity of a melody. More drastically, removing pitch information but preserving durations (i.e., speaking lyrics aloud rather than singing them) or doing the reverse (i.e., singing the tones in order with irregular rhythms) leaves the song close to incomprehensible. The reliance of music on frequency and temporal information also contrasts with **auditory scene analysis** [10], which incorporates frequency and temporal information, but progresses whether or not a scene includes tones or rhythms.

Thus, a key distinction between music perception and the perception of other sounds is that music perception relies fundamentally on pitch and duration information. Music perception does not simply prioritize these features in audition, however; it instead uses them in a special fashion.

Tonal perception

The extraction of pitch (the percept corresponding to fundamental frequency) from natural sounds is computationally complex, but is achieved across species [11]. Rather than attending to the absolute frequency of pitches in a melody, adults typically use relative pitch information to process music, automatically encoding information concerning **melodic contour** [12]. For example, whether *Happy Birthday* is sung with a low or high voice, it is understood as the same song.

While tones in a melody are heard relative to one another, they are not interchangeable. In music, humans hear pitch in the context of a **tonal hierarchy**, wherein some tones are more stable than others, and tones are interpreted in relation to one another in terms of stability [5,13] (Figure 2A–C; and Audio S1–S4 in the supplemental information online).

The hierarchical representation of pitch in a tonal hierarchy is an automatic, core feature of music perception. The evidence takes many forms. The simplest test is to ask participants to rate the stability of a **probe tone**, relative to a melodic context, on a rating scale [14]. On this and similar tasks, listeners demonstrate substantial inter-rater agreement [14,15], across ages [16] and musical styles [17], responding fastest to probes that conform to an implied tonal hierarchy [18].

In melodic expectancy tasks, listeners are asked to rate whether a note 'fits' or 'should come next' in an incomplete melody (Figure 2C), rather than rating stability in isolation. Here too, listeners show consensus in ratings [19] and singing [20]. Their expectations are also detectable via evoked response potentials [21]. While melodic expectancy can be predicted, in part, on the basis of characteristics unrelated to tonal structure, such as intervallic proximity [22], their results are largely consistent with non-expectation-based tasks.

Implicit tests show that listeners may be unaware of their tonal expectations, suggesting automaticity. For example, in a restoration task, melodic tones are replaced with static. Listeners 'fill in'

others (e.g., the first event in a repeating rhythmic group).
Musical surface: lowest level of event information in music (e.g., a sequence of notes).
Pitch chroma: group of pitch levels that are separated by an octave (i.e., a doubling of frequency), also known as a pitch class. Colloquially, members of the same pitch chroma share a note name; all the Fs on a piano have the same chroma.
Probe tone: tone in perceptual experiments that is compared to a context; on a given trial of a probe–tone experiment, a melody is played followed by a probe tone, which is compared with the melody.
Tonal hierarchy: high-level organization of pitch information in music, involving multiple levels of stability, such that some tones are heard as more stable than others (e.g., the tonal center, which is the most stable tone in a melody or other musical example).
Universality: property of a species, wherein a given phenomenon (here, a cognitive process) is expected to appear in some form across all typically developing individuals.

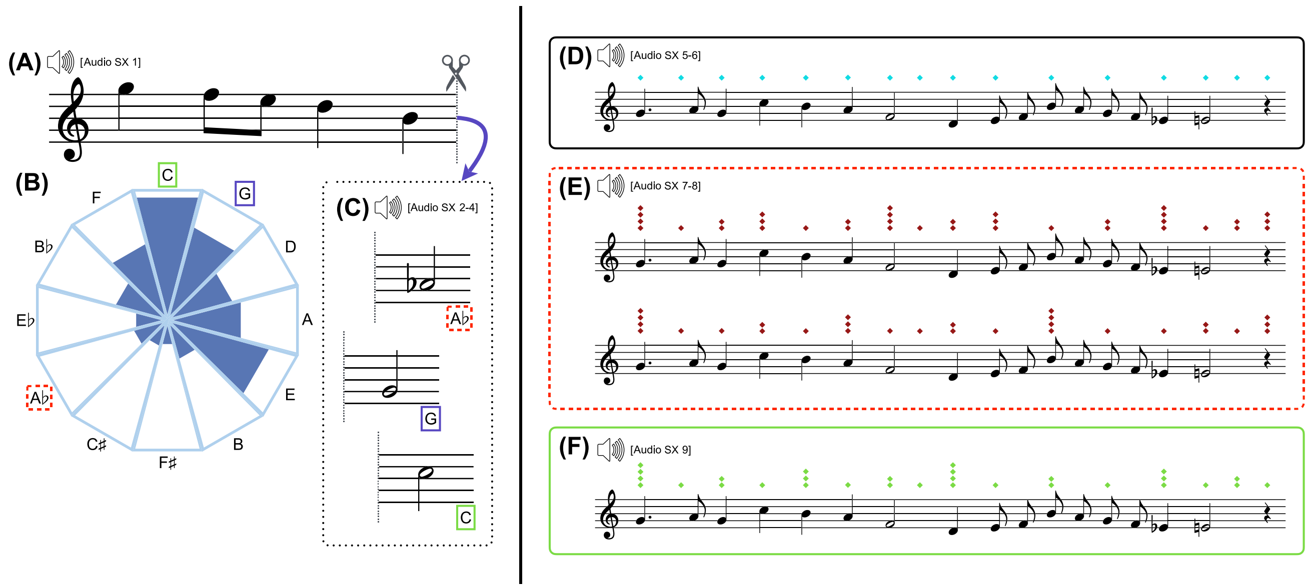

Figure 2. Demonstrations of tonal and metrical perception. To demonstrate tonal perception, listen to the incomplete melody in (A) (Audio S1 in the supplemental information online) and imagine, hum, or sing aloud the tone you expect to complete it. A schematic polar histogram of stability priors is in (B); the blue bars suggest the perceived stability of potential tones (e.g., the tone E has relatively high stability, whereas the tone F# does not). In Western music, tones an octave apart are typically heard as members of the same category, or **pitch chroma**, hence the circular coordinates in (B) (see [120], including Figure 1 therein). Three potential completions of the melody are presented in (C), with examples of low stability, such as the tone A-flat; moderate stability, such as the tone G; and high stability, such as the tone C, which is the tonal center. The complete melody, with each potential ending, can be heard in Audio S2–S4 in the supplemental information online. While each listener may have a different set of priors (i.e., a different distribution of shaded bars in the polar histogram) based on their musical experience, some general cross-listener regularities are expected. To demonstrate metrical perception, listen to the melody in (D), first without any rhythmic accompaniment (Audio S5 in the supplemental information online), and second, with rhythmic accompaniment where all beats are equally weighted (as with a metronome; Audio S6 in the supplemental information online). The diamonds above the staff notation indicate the location of the beat. This melody conforms to only one primary meter but is composed to allow for some metrical ambiguity. Two unusual rhythmic accompaniments are presented in (E); listen to these in Audio S7 and S8 in the supplemental information online. While these are metrical, in that they contain patterned information about the relative strength of each beat (indicated by the stacked diamonds, where a larger stack denotes higher strength), many would consider their meters to be incongruent with the melody. By contrast, the rhythmic accompaniment in (F) does conform to the implied meter of the melody, and is audibly better-fitting (Audio S9 in the supplemental information online). Note that stability or congruence should not be confused with 'musical quality' or any form of aesthetic preference.

tones consistent with a tonal context [23]. When listeners detect 'oddballs' or 'wrong notes', performance is highest when stimuli are presented in a tonal context [24,25]. A particularly elegant test played a melody continuously, modulating through 24 keys while inserting oddballs. Participants noticed them most when they contrasted with the current tonal context of the melody, even though the same tones were used throughout [26].

In recall tasks, listeners' memory errors imply automatic processing of tonal information: tones closely related in a tonal hierarchy are most often confused with one another [27,28]. Such automaticity may also be reflected by differential neural processing of tonal versus atonal melodic content [29].

Tonal hierarchies also facilitate the processing of other information. Temporal information is most reliably understood in tonally stable contexts [30]; intonation judgments are more accurate in closely related chord changes than in distant ones [31]; and children's judgments of vowels are most accurate in tonal contexts [32]. Conversely, processing is impaired when tonal expectations are uncertain [33] and, in sparse auditory contexts, such as isolated tones played non-musically, listeners interpret tones in relation to a hierarchy even when no hierarchy is present [34].

All this is wholly different from how pitch information is used in other aspects of auditory perception [10], speech [35], and other features of music [13]. The contrast is observable at the neural level, with substantively different processing for language and music [36,37]. Recent evidence suggests similar dissociations within pitch processing, with a differentiated neural code between hierarchical tonal perception and other aspects of pitch [38]. These phenomena, together with longstanding evidence from studies of auditory disorders (Box 1), suggest the encapsulation of tonal perception.

Metrical perception

Whereas tonal perception involves stability information with respect to a tonal context, the units of metrical perception are different: duration information with respect to a **beat** [3,4,39,40]. The beat forms a map that events can be placed on, the fundamental rhythmic element of music that Jackendoff calls the **musical surface** [1]. In this context, the beat is a point in time with no duration, the unit to which we synchronize when tapping along to music. The beat provides a palette on which metrical perception is built, providing the fundamental context for how we process rhythm in music: a **metrical hierarchy**, where specific beats on the musical surface vary in terms of their relative strength [1,39].

Thus, metrical perception is subtly different from tonal perception: while a given tone always has some degree of strength in a tonal hierarchy, a specific rhythmic pattern has no corresponding characteristic. Rather, the location of the pattern in the metrical grid determines its strength, because individual beats vary in terms of their relative strength within a meter. Rhythmic stress, used across cultures in conjunction with, or in opposition to, metrical information [41], can also alter perceived strength, but it is not a prerequisite for metrical perception; location in a metrical grid is the most informative characteristic concerning strength in the hierarchy.

Box 1. Disordered auditory perception yields clues to the encapsulation of tonal perception

Research pioneered by Isabelle Peretz demonstrates that certain neurological impairments produce deficits specific to music perception. For example, in **amusia**, pitch perception is impaired, while speech perception is typically left intact [121]. Such musical disorders provide evidence for the encapsulation of tonal or metrical perception.

Some case studies point toward pitch perception being separable from tonal perception. In one study [120], G.L.'s brain damage spared his pitch perception and pitch memory ability, but he was unable to accurately judge the stability of a tone in a melodic context. Instead, he reliably chose the less stable of two tones when predicting how a melody should conclude. He also showed no preference for tonal melodies over atonal ones, in stark contrast to control participants [122].

A similar pattern was found in C.N., a lesioned patient whose non-musical cognitive abilities were preserved, but whose music-processing abilities were substantially impaired; her lowest score was on a test evaluating the well formedness of a melody in terms of its tonal hierarchy [123]. The pitch perception deficit in congenital amusia also does not impair tonal music production: when asked to sing, amusics produce melodies that conform to a tonal hierarchy [124].

Addressing the topic from a converse perspective, that is, general cognitive deficits that preserve music-specific perceptual abilities, also provides clues of encapsulation. Severe cognitive deficits have been reported in patients with dementia that nonetheless preserve the ability to detect oddballs in melodic organization [125]. A skilled harpsichordist whose semantic dementia left him unable to comprehend common nouns could nonetheless perform music from memory, altering the music in tonally appropriate fashions in each performance; when sight-reading unfamiliar music, his embellishments of the notated music also conformed to the underlying tonality [126]. That musical expertise can be preserved in the face of neurodegeneration [127] also raises possibilities of clinical applications of music in dementia care.

Of course, interpreting case studies can be difficult, in that tasks designed to assess music perception ability can tap a variety of perceptual phenomena, or more general abilities than tonal or metrical perception. Furthermore, specific musical disorders may provide more direct evidence for encapsulation than broader disorders (e.g., dementia or autism spectrum disorder), especially insofar as double dissociations between tonal and metrical perception can be detected in individuals with these conditions.

In sum, the classical case-study approach to probe the structure of the mind continues to produce tantalizing clues to the nature of music perception.

Figure 2D–F and Audio S5–S9 in the supplemental information online illustrate the phenomenon, arguably the most important high-level feature of how humans hear rhythm in music. Notably, metrical perception in music differs substantially from how rhythms are heard in speech and other sounds [1,35]; while speech does contain stronger and weaker events, forming groupings [42], they are qualitatively different than those in musical meters, in that they typically occur without isochrony or repetition.

What is the evidence for metrical perception? Most simply, when music listeners tap to the beat, their taps imply a representation of multiple beat levels: they hear some beats as stronger than others, and agree on which ones are which [43,44]. This hierarchical representation of beats is a hallmark of metrical perception [8] and is reflected by participants' actions: rhythmic tapping is most precise on strong beats [45], with the highest degree of stability in tapping (i.e., higher precision/lower timing variability) in metrically structured sequences, as opposed to irregular ones [46]. The format of metrical perception appears to be specific to the auditory domain, however; such effects are only elicited by an auditory stimulus, not a visual one [45].

As with tonal perception, errors in recall tasks also imply the representation of metrical information. Rhythmic patterns with clear metrical information are better recalled than those without [47], and the errors listeners make when imitating a rhythm by tapping imply a pattern of '… [attempting] to fit the presented sequences into one or another internal structure' [48]; that is, accuracy is substantially higher when the source rhythm follows simple ratios, and drops when it does not. Similarly, when different rhythmic patterns preserve the same hierarchy of strong and weak beats, listeners confuse them with one another [49].

Neural evidence also supports the existence of metrical perception. When listening to a two-note pattern with an ambiguous meter, interpretable either as 'DUN-dun [break]' or 'dun-DUN [break]', magnetoencephalography shows different responses in the beta frequency band as a function of which tone participants imagine as the strong beat [50]. In clearly metrical music, gamma-band activity reflects strong beats, even when no note is played on the strong beat [51]. This pattern has also been found via electroencephalography, where beat-related evoked responses were found at a different frequency than meter-related evoked responses [52]. Evoked responses to omitted beats are even detectable in newborn infants [53], although it is not yet known if this reflects an early representation of meter (Box 2).

Like tonal perception, metrical perception appears to occur automatically and with a degree of encapsulation. None of the results described above require formal music training, and listeners are often not explicitly aware of their metrical representations. Just as one cannot force oneself to perceive an incongruent tonal center in music, one cannot force oneself to perceive an incongruent metrical grouping (Figure 2). These characteristics, also evident for Gestalt event segmentation in auditory scene analysis [10], are also reflected by the sense of rhythmic groupings in the absence of any grouping information at all [54].

However, metrical perception is likely more flexible than tonal perception in that different meters are nested within the metrical grid: one can represent the same meter at one of multiple levels (e.g., in Figure 2, try tapping your foot at the strongest beat level or the second-strongest). This has no analog in tonal perception. Moreover, while most music globally is isochronous, meters can be implied even in the presence of substantial deviations from isochrony [55]. Nevertheless, this flexibility is not unlimited; for example, hearing two levels of meter simultaneously is difficult, even with training [56].

Box 2. Development of tonal and metrical perception

What is the ontogeny of tonal and metrical perception? At one hypothetical extreme, young infants might be priorless, hearing pitch or beat equipotentially. At the other extreme, young infants might be predisposed to expect specific structures, with weightings toward common tonalities or meters.

On a moderate position, for tonal perception, infants are predisposed to attend to a few categories of tones that occur regularly (e.g., the tonal center and closely related tones), but without prespecified weights. Those might be drawn from a uniform prior, updating as musical input accrues. For metrical perception, some predispositions might be assumed (e.g., the existence of a beat-based metrical grid), while others might not (e.g., evenly spaced groupings).

Some evidence supports the moderate position, with early-developing seeds of tonal and metrical perception that are shaped through childhood. Eight-month-olds detect oddballs in melodies regardless of their tonal characteristics, whereas adults' performance is weakest without tonal cues [25]; and adults detect mistunings better in tonal than in atonal contexts, but 6-month-olds perform comparably with both [128]. However, infant tonal perception is not fully equipotential: 9-month-olds detected oddballs best when the structure of the task shared a property with tonal hierarchies (i.e., an unequally-stepped scale), even if the melodies were atonal [129]. While young children show relatively stable tonal intuitions in the same direction as adults [130], the strength of those intuitions increases over time [16].

Some metrical predispositions are detectable in neonates [131,132] and through infancy [133], with preferences for simple rhythmic intervals in infancy [134] and childhood [135], and with higher-level abilities continuing to develop through early childhood [56]. Similar to tonal perception, metrical perception is evidently shaped by enculturation, with 6-month-old infants detecting oddballs in complex-meter music that American adults miss (the adults being more accustomed to simple meters; [75]). Twelve-month-olds failed a similar task, but succeeded after exposure to complex-metered music, whereas adults did not [136]. In general, metrical perception appears to develop more gradually than does tonal perception [137], not reaching adult-like levels until adolescence [138].

These bodies of work leave many questions open. Few studies have been replicated, and most study Western music in Western infants and children, raising questions about reproducibility and generalizability. Few studies test neonates to assess the initial state of the perceptual phenomena. In addition, the specificity of the results is not fully understood: that music perceptions become adult-like through childhood may reflect the self-evident fact that children become better research participants as they grow up.

The systems are likely to be universal

The production of music is universal, reported by anthropologists throughout a representative sample of human cultures [57]. Many of its common features, such as instrumental accompaniment, steady beat, and rising/falling contours, are widespread, as has been described by ethnomusicologists [58]. Concurrently, psychological responses to music reveal universals in emotional and behavioral contexts [59], with links between the forms music takes and its common behavioral functions [60]. Such findings do not imply universality of the structural components of music perception, but they show its plausibility: the perceptual systems of the people making music in a foreign culture are apparently aligned with the perceptual systems of the people hearing them.

Principal questions, then, are what exactly those perceptual systems contain, and whether they share the processing phenomena described here. Direct evidence is limited, at present, but several lines of work support the hypothesis that tonal and metrical perception are universal psychological systems [4,5].

Universals in tonal perception

When Western-acculturated listeners evaluate tonality in unfamiliar, foreign music, they uniformly report hearing tonal hierarchies, with high cross-listener consistency [57]. Thirty musicians from a variety of backgrounds, including experts in ethnomusicology and music theory, rated whether unfamiliar, foreign songs contained a clear tonal center. They answered 'yes' 97.8% of the time; they agreed with one another on the pitch level of the center(s); and their responses were predictable by a key-finding algorithm developed for Western classical music.

However, these data do not show how the producers of the songs perceived them, and they are confounded by the Western musical experiences of the listeners. Stronger evidence comes from studies applying a listening paradigm across cultures. In now-classic experiments, participants from multiple societies completed probe–tone tasks. The results suggested universality: Indian and Western listeners rated the same tones as highest fitting in the context of Indian rāgas [61] and Balinese music [62]. Similar results have been found in cross-cultural melodic expectancy tasks [63,64]. While probe–tone ratings do show some variability in cross-cultural comparisons, the variability is modest relative to the degree of agreement across cultures [65]. Similarly, cross-cultural differences are evident, but small, in the estimation of interval sizes across expert Javanese and Western-trained musicians [66].

Analyses of the patterns of tones present in music globally, combined with experimental tests of music production across cultures, also point to the universality of tonal perception. Finnish folk music, Saami songs, and Korean court music have similar patterns of transitional probabilities between tones, which may support tonal expectations [63,67]. Similar 'musical *n*-gram' analysis showed that the frequency distribution of melodic intervals approximately follows a power law, globally, such that only three small intervals (unison, major second, and minor third) account for the majority (73%) of observed intervals [57].

Similar patterns are reflected in cross-cultural perceptual experiments. When Tsimane and Western listeners heard pairs of tones related via a tonal context, people from both societies were more likely to mistake them for a single tone than when they were not [68]. When attempting to sing back pairs of tones, even when making substantial errors, both Tsimane and Western participants sung pitches that revealed the use of a logarithmic scale, common within tonal systems [69] and paralleling those appearing in music globally [57].

Universals in metrical perception

Metrical perception has been studied cross-culturally less often than has tonal perception, but here too the evidence is suggestive of universality. While ethnomusicologists have noted substantial cross-cultural variability in meters [58], this variability is underlain by regularities in the distribution of rhythmic groupings found globally. Analysis of sequential pairs of note durations (rhythmic intervals) showed that the most common durational ratios are unisons, where both notes have the same duration; or small integer ratios (e.g., 1:2, 2:1, 1:3, 3:1) [57], according with production biases revealed experimentally [70].

This property likely leads to cross-cultural metrical regularities. Indeed, when expert raters analyzed the meters present in the same corpus, they had high agreement concerning the metrical information present in the song [57]. However, musical experience shapes the representation of rhythm [71] and, thus, it is unknown whether this result merely reflects raters' mostly Western backgrounds.

Cross-cultural experiments on meter suggest a more general phenomenon. Chinese and German musicians display similar evoked response potentials while listening to Chinese music [72]; metrical phrase boundaries in Turkish music are detected across cultures [73]; and Tunisian and French participants make similar metrical inferences when listening to each other's music, despite synchronizing to the music somewhat differently [74]. Developmental evidence also suggests universality: 6-month-old North American infants detected metrical disruptions in both Western and Balkan rhythms [75].

Another compelling form of evidence comes from iterated tapping tasks. Therein, participants hear a rhythm and attempt to reproduce it by tapping; their reproduction is played to new participants, the process repeats, and a consensus interpretation of the rhythm is eventually analyzed. Participants spontaneously generate metrical grids even when the seed rhythm is random or ambiguous [76,77]. The consensus rhythms in these cases tend to be related by integers, similar to songs produced worldwide [57], suggesting predispositions toward metrical groupings.

The linchpin comes when this approach is applied across many cultures, where less-frequent consensus rhythmic ratios (e.g., a 3:3:2 rhythm, present in Botswana and Mali, but not China or the Tsimane) embellish a pattern of universality: every society studied shows consensus responses at simple metrical groupings (e.g., 1:1:1 or 1:1:2 rhythms, found in all groups studied) [78]. This result mirrors rhythmic bigrams found cross-culturally, where a few possible rhythmic bigrams (1:1, 2:1, 3:1) accounted for the majority (86%) of rhythms observed globally [57].

The systems are uniquely human

The search for signatures of musicality in many nonhuman species is longstanding [4,9]. Methods range from description of spontaneous music-like behavior, probing the productive behaviors of nonhuman animals [79]; to direct experimentation with musical stimuli, probing the cognitive abilities of nonhuman animals in the domain of sound [80]; to the neuroscience of auditory perception, probing the brain circuitry of nonhuman animals [81,82].

A superficial cross-species comparison of vocalization and audition suggests broad similarities between humans and other species. Many species produce song-like vocalizations and listen to them, as in birds, gibbons, whales, seals and other pinnipeds, and fish [83]. Some species alter their vocalizations in response to noise, solving the **cocktail party problem** in a similar fashion to humans [84]. Nonhuman vocalizations also share some general properties with music, such as the use of simple integer ratios [85,86], and several species can discriminate musical examples from one another [87].

However, when experiments test interests in, and responses to, sounds, they reveal sharp differences between humans and other species. Cotton-top tamarins show no preference between a screeching sound, which is often aversive to humans, and amplitude-matched white noise, which is not [88]. Tamarins also prefer silence to music [89]. Gorillas showed indications of stress, restlessness, or aversion during music listening, relative to silence or rainforest sounds [90]. The rhythmic motions that chimpanzees make in response to piano sounds [91] may similarly suggest aversion [92]. These surprising examples point to the risk of overinterpreting music-like behavior in nonhumans (Box 3).

Nonhuman animals do not appear to represent tones hierarchically

There are some similarities in the production and perception of pitch across species. Macaques and humans display similar segmentation of tone sequences [93]; bullfinches can reproduce melodies by rote, similar to humans [94]; and many species can extract pitch from complex harmonic tones [11]. These processing similarities likely underlie the ability of many species to distinguish sound patterns from one another, as in tamarins, which distinguish scrambled from unscrambled sequences of tones [95]; or macaques, which distinguish consonant chords from dissonant ones [96].

Nevertheless, such successful discrimination abilities do not rely on human-like pitch perception. When experiments test this narrower phenomenon, the pattern of results is messier, with limited evidence for relative pitch perception and none for tonal perception. Capuchins and rats are unable to generalize a simple melody to other keys [97], and neither are European starlings (Box 3). Macaques that succeed at discriminating ascending from descending pitch patterns require tens of thousands of training trials and, even then, they perform far below ceiling [98]. A similar pattern has been observed in ferrets [99], and budgerigars showed no preference for tonal versus nontonal piano melodies [100].

Box 3. Bird 'song' is a misnomer

Colloquially, songbirds are a musical species: they produce complex vocalizations that sound musical; they transmit those vocalizations to their young, which learn to produce them; and so on. One is tempted to assume birdsong is analogous to music.

A remarkable study of European starlings calls this assumption into question. While starlings can reliably perceive pitch, they fail a test with a very low bar: recognizing a melody they have heard hundreds of times.

Bregman and colleagues [80] trained starlings to peck on one side of an apparatus when hearing an ascending melody, and on the other when hearing a descending melody. The birds performed over 90% correct, but when the melodies were transposed, preserving relational and tonal information but changing the key, performance dropped to chance.

Why? One wonders if the birds focused on absolute rather than relative pitches, but subsequent experiments showed a more surprising explanation. Altering the timbre of the melodies, presented at the same pitch level but with different spectral patterns (analogous to a change from flute to clarinet), drove recognition performance to chance again. Last, when the birds listened to versions of the melodies altered with noise-vocoding, which preserves the spectral shape of the melodies but removes pitch information, recognition approached 90% correct again.

This represents a dramatic difference between how humans and starlings perceive tones: the pitches in starling songs are, apparently, not terribly important to the starlings. When a melody was transposed by merely one semitone, starlings behaved as though the melody was completely unfamiliar; and yet, when listening to noise-vocoded melodies, which, lacking pitch, do not sound to humans like melodies at all — the starlings recognized them easily.

Is starling auditory perception unusual? We do not yet know, but many experiments typically interpreted to demonstrate absolute pitch perception in songbirds may reflect a processing bias for spectral information over pitch information [139]. In principle, this phenomenon could underlie auditory perception in many other nonhuman species.

This finding illustrates the danger of assuming nonhuman animals perceive the world as we do: in fact, the auditory percepts of European starlings, and potentially many other species, are different from ours. If this finding is repeated in other species, it would provide strong evidence that the predisposition to hear tones in reference to a tonal center is uniquely human, because the prioritization of pitch information in auditory processing is a prerequisite for tonal perception.

In general, the representation of pitch appears to be less flexible in nonhumans than in humans. For example, the recognition by starlings of conspecific songs is reduced to chance levels when small manipulations of pitch or timbre are introduced [101]. Similarly, there is only suggestive evidence of the perception of octave equivalence in rats and dolphins [102], and a single nonhuman primate [103] (but see discussion in [4]).

It is not yet known why the mechanisms of pitch processing are so different between humans and nonhumans, although substantive differences are apparent at the neural level [82], and, while anatomical similarities in cochlear partitioning across species yield similar overall structure (such as the logarithmic perception of pitch), humans appear to be more sensitive to fine-grained pitch changes than do other species [104]. Nevertheless, it appears evident that tonal perception is far from the norm in nonhuman species.

Some nonhuman animals entrain to a beat, but none apparently represent meter

The rhythmic behaviors of nonhumans are well studied, perhaps because the behaviors are surprising, impressive, or humorous. The best-known examples involve individual animals synchronizing with music, such as Snowball, a sulfur-crested cockatoo [79,105] or Ronan, a California sea lion [106]; and a corpus study of home videos of 'dancing animals' [107].

As with the pitched sounds of complex birdsongs, one is tempted to jump to conclusions of high-level musicality. However, the perception of a beat and entrainment to it imply only the low-level perception of periodicity, far from the human phenomenon of metrical perception.

Indeed, of those animals that have been shown to entrain to a beat, no evidence has been reported for rhythmic behaviors requiring metrical perception, even though most musical stimuli used to test beat synchronization have obvious metrical structure (e.g., Queen's *Another one bites the dust*, which Snowball attempts to synchronize with; see supplementary video in [79]). Similarly, beat-related but non-metrical synchrony behaviors have been reported in chimpanzees [108] and macaques [109]. This lack of metrical production behavior is not due to an inability to produce different types of motion in response to music, as humans do (e.g., tapping to all beats while nodding only on strong beats). Snowball, for example, produces a variety of head motions while synchronizing to a beat [105], but not in response to metrical groupings.

Two further means by which nonhuman animals might demonstrate sensitivity to metrical representations are in their spontaneous rhythmic production and in their ability to distinguish different auditory stimuli from one another on the basis of metrical information. There is little evidence for metrical representation in either case. Chimpanzees spontaneously drum on roots, but their drumming bouts have no underlying beat, let alone metrical information [110]. Gibbon calls can be repetitive and isochronous but display no evidence for metrical groupings [111]. Non-metrical rhythmic behavior has also been replicated in the lab: budgerigars can learn to tap to an audiovisual, metronomic stimulus, but their tapping lacks grouping structure [112]; as does bonobo tapping [113].

Notably, nonhuman animals have difficulty distinguishing stimuli that differ in rhythmic information. Starlings fail to discriminate tonal patterns that contain a beat from those that do not [114]. In an electroencephalography experiment, rhesus macaques showed sensitivity to isochrony but not meter [115]. When nonhumans do succeed at discrimination in stimuli that include meters, their perception is relatively inflexible and likely relies on nonrhythmic features of pitch, timbre, or volume [87,116].

Last, some evidence suggests that the rhythm production of several species includes categorical prototypes that are differentiable by their rhythmic events, such as nightingales and zebra finches [86] and lemurs [117]. However, here too, the ability falls short of metrical perception, because such temporal events often occur in integer ratios without implying the presence of metrical groupings, which in music typically involve multiple nested hierarchies [8]. Such groupings may simply not be of interest to nonhuman listeners. For example, starlings ignore grouping information in favor of temporal rate when making perceptual judgments [114], despite their ability to positively distinguish isochronous rhythms from triplet rhythms. And rhesus monkeys show increased evoked response potentials when hearing music that contains similar timing information to music they previously heard [118], but no evidence yet connects such responses to grouping-level information.

Concluding remarks

> *Take the case of a baby being spoken to by her mother. The baby starts to imitate her mother's voice. However, she does not insert into the imitation the squeaks of her cradle that have been occurring at the same time. Why not? A physical record of what she has heard would include them. Somehow she has been able to reject the squeak as not being part of the perceptual 'object' formed by her mother's voice. In doing so, the infant has solved a scene analysis problem in audition.*
>
> [– Albert Bregman, *Auditory Scene Analysis: The Perceptual Organization of Sound* (1990)]

Whereas many species come into the world prepared to process their acoustic environment, human auditory scene analysis is unusual in that it contains numerous specializations for the

Outstanding questions

What are the origins of hierarchical processing in music? Ideas from biological and cultural evolution, oscillator models of perception, and predictive coding do not yet show any consensus for why musicality is supported by tonal and metrical representations.

Which aspects of tonal and metrical perception are universal? Too few cultures and musical examples have been studied to determine the answer.

What structure do the predispositions for hierarchical music perception take, and how independent are they of one another? It appears unlikely that humans have innate preferences for particular tonal or metrical patterns, but we may be predisposed toward certain categories of hierarchies.

What precursors to tonal and metrical perception exist across taxa? Studies of nonhuman species reveal some music-like abilities, but there is scant evidence of tonal or metrical perception outside of our species. When and how humans' high-level auditory perception diverged from other animals is unknown.

How much musical experience is required for the emergence of music perception? The developmental trajectory of tonal and metrical processing has not yet been fully described. In particular, experiments that can isolate innate perceptual phenomena from those requiring auditory experience would be valuable.

What are the genetic correlates of musicality? Relatively little is understood about how our genomic architecture gives rise to the neural circuitry underlying music perception, in part because of the cost and complexity of genome-wide association studies; the Musicality Genomics Consortium[i] aims to solve this problem.

How reproducible and generalizable is research on music perception? Similar to many areas of the cognitive sciences, research on the psychology of music is often based on small, underpowered studies; thus, a key question for the field surrounds the reliability and robustness of key findings.

Box 4. On nontonal, non-metrical aspects of musicality

Two reviewers of this paper noted that tonal and metrical representations cannot be the whole story of music perception. They are correct. Many aspects of human musicality are not explained by such representations, and readers should take care not to assume otherwise.

Indeed, many of the most visceral experiences of listening to music do not rely on tonal or metrical representations. For example, **groove** relies on both beat and syncopation [8], which together can interact with metrical information, but a groove percept can in principle be induced by non-metrical music. The intricate texturing of timbres used across many genres, such as renaissance polyphony, modern film scoring, or group singing in traditionally living societies, can be fully independent of tonal or metrical information. The well-documented mutual intelligibility of musical behavior and musical emotion across cultures operates separately from tonal and metrical characteristics of music [59]. In addition, the systems of representation discussed in this paper probably will only go so far to explain, in Pinker's words, 'the most blazingly obvious feature of music — people enjoy it' [140].

Domain-general aspects of musicality may take advantage of the tonal and metrical structure of music, such as the cycle of listeners forming predictions about the music they hear and those predictions being met, or not [7]; but such processes could operate regardless of the presence or absence of tonal or metrical structure. Similarly, oscillatory neurodynamics have been proposed as a mechanism for the emergence of high-level structure in music (e.g., [6]) and could, in principle, apply to a variety of other structures in auditory perception.

Nevertheless, the evidence reviewed here documents the fundamental nature of tonal and metrical representations in human auditory perception. This immediately raises questions regarding how other aspects of human musicality rely on or interact with them.

processing of language and music (e.g., [2,10,83]). Bregman's infant not only isolates her mother's voice from the squeaks of the cradle, but then decodes the sounds the voice produces, processing it into syntactic information, inferring emotion information from prosody, and so on, via ordinary speech processing [35].

The constituent parts of music perception are less well understood, perhaps because music is often assumed to be inaccessible to science, too complex a behavior for empirical inquiry, or 'less important of a behavior' than speech or vision (to quote a reviewer of this paper, who does not hold this view, but has encountered it). That the two key principles of music perception reviewed here appear to be so pervasive in our species, and not other species, raises the possibility that our auditory system is designed by natural selection to decode song and other forms of music into hierarchically structured tonal and metrical information, in a fashion that is distinct from the mechanisms of audition that are specialized for speech.

These specializations for music perception are limited in their scope [3]. Neither tonal perception nor metrical perception alone can account for a variety of other high-level aspects of musicality (Box 4). Such phenomena may be facilitated by the interaction of tonal and metrical representations with each other, with other acoustical phenomena, and with other domains of human cognition.

Future research (see Outstanding questions) may determine the extent to which tonal and metrical perception accurately describe the mind's tasks of processing music, how these develop, how universal and uniquely human they in fact are, and more, while also using new methodological approaches to improve the reproducibility and generalizability of research on the psychology of music [119]. Such a research program may also eventually explain aspects of musicality that have long intrigued cognitive scientists, such as esthetics, grammars, and the interplay between music, language, and other core domains of cognition.

Acknowledgments

The preparation of this article was supported by the US National Institutes of Health (DP5OD024566); the Royal Society of New Zealand Te Apārangi (RDF-UOA2103 and MFP-UOA2133); and the 2023 New Zealand Prime Minister's MacDiarmid

Emerging Scientist Prize. I thank the organizers of, and participants at, the First Bali Workshop on Music and Brain, which sparked discussions that shaped the ideas presented here (especially with Stella Christie, Xiaoqin Wang, Andrew Oxenham, and Isabelle Peretz). Many thanks also to Sabine Deviche, who provided illustrations. Last, I am grateful to Nori Jacoby, Jonathan Kominsky, Isabelle Peretz, Jenny Saffran, Richard Aslin, and the members of The Music Lab for their feedback on drafts of this paper; to Joel Snyder, Stefan Koelsch, and an anonymous reviewer for their suggestions; and to Lindsey Drayton for her generous editorial work, advice, and support.

Declaration of interests

None declared by author.

Supplemental information

Supplemental information to this article can be found online at https://doi.org/10.1016/j.tics.2025.05.013.

Resources

[i]www.mcg.uva.nl/musicgens

References

1. Jackendoff, R. (1994) *Consciousness and the Computational Mind*, MIT Press
2. Peretz, I. and Morais, J. (1989) Music and modularity. *Contemp. Music. Rev.* 4, 279–293
3. Peretz, I. and Coltheart, M. (2003) Modularity of music processing. *Nat. Neurosci.* 6, 688
4. Patel, A.D. (2008) *Music, Language, and the Brain*, Oxford University Press
5. Krumhansl, C.L. (2004) The cognition of tonality – as we know it today. *J. New Music Res.* 33, 253–268
6. Large, E.W. *et al.* (2016) A neurodynamic account of musical tonality. *Music. Percept.* 33, 319–331
7. Koelsch, S. *et al.* (2019) Predictive processes and the peculiar case of music. *Trends Cogn. Sci.* 23, 63–77
8. Snyder, J.S. *et al.* (2024) Theoretical and empirical advances in understanding musical rhythm, beat and metre. *Nat. Rev. Psychol.* 3, 449–462
9. Honing, H. (2019) *The Evolving Animal Orchestra. In Search of What Makes Us Musical*, MIT Press
10. Bregman, A.S. (1990) *Auditory Scene Analysis: The Perceptual Organization of Sound*, MIT Press
11. Oxenham, A.J. (2013) The perception of musical tones. In *The Psychology of Music* (Deutsch, D., ed.), pp. 1–33, Elsevier
12. Trainor, L.J. *et al.* (2002) Automatic and controlled processing of melodic contour and interval information measured by electrical brain activity. *J. Cogn. Neurosci.* 14, 430–442
13. Peretz, I. (2006) The nature of music from a biological perspective. *Cognition* 100, 1–32
14. Krumhansl, C.L. and Shepard, R.N. (1979) Quantification of the hierarchy of tonal functions within a diatonic context. *J. Exp. Psychol. Hum. Percept. Perform.* 5, 579–594
15. Vuvan, D.T. *et al.* (2011) Probing the minor tonal hierarchy. *Music. Percept.* 28, 461–472
16. Krumhansl, C.L. and Keil, F.C. (1982) Acquisition of the hierarchy of tonal functions in music. *Mem. Cogn.* 10, 243–251
17. Vuvan, D.T. and Hughes, B. (2021) Probe tone paradigm reveals less differentiated tonal hierarchy in rock music. *Music. Percept.* 38, 425–434
18. Janata, P. and Reisberg, D. (1988) Response-time measures as a means of exploring tonal hierarchies. *Music. Percept.* 6, 161–172
19. Palmer, C. and Krumhansl, C.L. (1987) Independent temporal and pitch structures in determination of musical phrases. *J. Exp. Psychol. Hum. Percept. Perform.* 13, 116–126
20. Fogel, A.R. *et al.* (2015) Studying musical and linguistic prediction in comparable ways: The melodic cloze probability method. *Front. Psychol.* 6, 1718
21. Maess, B. *et al.* (2001) Musical syntax is processed in Broca's area: an MEG study. *Nat. Neurosci.* 4, 540–545
22. Schellenberg, E.G. *et al.* (2002) Expectancy in melody: tests of children and adults. *J. Exp. Psychol. Gen.* 131, 511–537
23. DeWitt, L.A. and Samuel, A.G. (1990) The role of knowledge-based expectations in music perception: evidence from musical restoration. *J. Exp. Psychol. Gen.* 119, 123–144
24. Trehub, S.E. *et al.* (1986) Development of the perception of musical relations: semitone and diatonic structure. *J. Exp. Psychol. Hum. Percept. Perform.* 12, 295–301
25. Trainor, L.J. and Trehub, S.E. (1992) A comparison of infants' and adults' sensitivity to Western musical structure. *J. Exp. Psychol. Hum. Percept. Perform.* 18, 394–402
26. Janata, P. *et al.* (2003) Online detection of tonal pop-out in modulating contexts. *Music. Percept.* 20, 283–305
27. Dewar, K.M. *et al.* (1977) Recognition memory for single tones with and without context. *J. Exp. Psychol. Hum. Learn. Mem.* 3, 60–67
28. Curtis, M.E. and Bharucha, J.J. (2009) Memory and musical expectation for tones in cultural context. *Music. Percept.* 26, 365–375
29. Fernández-Rubio, G. *et al.* (2022) Magnetoencephalography recordings reveal the spatiotemporal dynamics of recognition memory for complex versus simple auditory sequences. *Commun. Biol.* 5, 1–12
30. Lebrun-Guillaud, G. and Tillmann, B. (2007) Influence of a tone's tonal function on temporal change detection. *Percept. Psychophys.* 69, 1450–1459
31. Justus, T.C. and Bharucha, J.J. (2001) Modularity in musical processing: the automaticity of harmonic priming. *J. Exp. Psychol. Hum. Percept. Perform.* 27, 1000–1011
32. Schellenberg, E.G. *et al.* (2005) Children's implicit knowledge of harmony in Western music. *Dev. Sci.* 8, 551–566
33. Hansen, N.C. *et al.* (2021) Predictive uncertainty underlies auditory boundary perception. *Psychol. Sci.* 32, 1416–1425
34. Shepard, R.N. and Jordan, D.S. (1984) Auditory illusions demonstrating that tones are assimilated to an internalized musical scale. *Science* 226, 1333–1334
35. Zatorre, R.J. *et al.* (2002) Structure and function of auditory cortex: music and speech. *Trends Cogn. Sci.* 6, 37–46
36. Chen, X. *et al.* (2023) The human language system, including its inferior frontal component in 'Broca's area,' does not support music perception. *Cereb. Cortex* 33, 7904–7929
37. Norman-Haignere, S. *et al.* (2015) Distinct cortical pathways for music and speech revealed by hypothesis-free voxel decomposition. *Neuron* 88, 1281–1296
38. Sankaran, N. *et al.* (2024) Encoding of melody in the human auditory cortex. *Sci. Adv.* 10, eadk0010
39. London, J. (2004) *Hearing in Time: Psychological Aspects of Musical Meter*, Oxford University Press
40. Huron, D.B. (2006) *Sweet Anticipation: Music and the Psychology of Expectation*, MIT Press

41. London, J. *et al.* (2017) Rhythm histograms and musical meter: a corpus study of Malian percussion music. *Psychon. Bull. Rev.* 24, 474–480
42. Kotz, S.A. *et al.* (2018) The evolution of rhythm processing. *Trends Cogn. Sci.* 22, 896–910
43. Drake, C. *et al.* (2000) Tapping in time with mechanically and expressively performed music. *Music. Percept.* 18, 1–23
44. Snyder, J. and Krumhansl, C.L. (2001) Tapping to ragtime: cues to pulse finding. *Music. Percept.* 18, 455–489
45. Patel, A.D. *et al.* (2005) The influence of metricality and modality on synchronization with a beat. *Exp. Brain Res.* 163, 226–238
46. Keller, P.E. and Repp, B.H. (2005) Staying offbeat: sensorimotor syncopation with structured and unstructured auditory sequences. *Psychological Research* 69, 292–309
47. Povel, D.-J. and Essens, P. (1985) Perception of temporal patterns. *Music. Percept.* 2, 411–440
48. Povel, D.-J. (1981) Internal representation of simple temporal patterns. *J. Exp. Psychol. Hum. Percept. Perform.* 7, 3–18
49. Palmer, C. and Krumhansl, C.L. (1990) Mental representations for musical meter. *J. Exp. Psychol. Hum. Percept. Perform.* 16, 728–741
50. Iversen, J.R. *et al.* (2009) Top-down control of rhythm perception modulates early auditory responses. *Ann. N. Y. Acad. Sci.* 1169, 58–73
51. Snyder, J.S. and Large, E.W. (2005) Gamma-band activity reflects the metric structure of rhythmic tone sequences. *Cogn. Brain Res.* 24, 117–126
52. Nozaradan, S. *et al.* (2012) Selective neuronal entrainment to the beat and meter embedded in a musical rhythm. *J. Neurosci.* 32, 17572–17581
53. Winkler, I. *et al.* (2009) Newborn infants detect the beat in music. *Proc. Natl. Acad. Sci. U. S. A.* 106, 2468–2471
54. Ongchoco, J.D.K. and Scholl, B.J. (2022) Scaffolded attention in time: 'Everyday hallucinations' of rhythmic patterns from regular auditory beats. *Atten. Percept. Psychophys.* 84, 332–340
55. Jakubowski, K. *et al.* (2022) Aesthetics of musical timing: culture and expertise affect preferences for isochrony but not synchrony. *Cognition* 227, 105205
56. Nave-Blodgett, J.E. *et al.* (2021) Hierarchical beat perception develops throughout childhood and adolescence and is enhanced in those with musical training. *J. Exp. Psychol. Gen.* 150, 314–339
57. Mehr, S.A. *et al.* (2019) Universality and diversity in human song. *Science* 366, eaax0868
58. Nettl, B. (2015) *The Study of Ethnomusicology: Thirty-Three Discussions*, University of Illinois Press
59. Singh, M. and Mehr, S.A. (2023) Universality, domain-specificity and development of psychological responses to music. *Nat. Rev. Psychol.* 2, 333–346
60. Yurdum, L. *et al.* (2023) Universal interpretations of vocal music. *Proc. Natl. Acad. Sci. U. S. A.* 120, e2218593120
61. Castellano, M.A. *et al.* (1984) Tonal hierarchies in the music of north India. *J. Exp. Psychol. Gen.* 113, 394–412
62. Kessler, E.J. *et al.* (1984) Tonal schemata in the perception of music in Bali and in the west. *Music Percept. Interdiscip. J.* 2, 131–165
63. Krumhansl, C.L. *et al.* (2000) Cross-cultural music cognition: cognitive methodology applied to North Sami Yoiks. *Cognition* 76, 13–58
64. Eerola, T. *et al.* (2009) Expectancy in Sami Yoiks revisited: the role of data-driven and schema-driven knowledge in the formation of melodic expectations. *Music. Sci.* 13, 231–272
65. Justus, T. *et al.* (2019) Remembering melodies from another culture: Turkish and American listeners demonstrate implicit knowledge of musical scales. *Anal. Approaches World Music* 7, 1–20
66. Perlman, M. and Krumhansl, C.L. (1996) An experimental study of internal interval standards in Javanese and Western musicians. *Music. Percept.* 14, 95–116
67. Nam, U. (1998) Pitch distributions in Korean court music: evidence consistent with tonal hierarchies. *Music Percept. Interdiscip. J.* 16, 243–247
68. McPherson, M.J. *et al.* (2020) Perceptual fusion of musical notes by native Amazonians suggests universal representations of musical intervals. *Nat. Commun.* 11, 2786
69. Jacoby, N. *et al.* (2019) Universal and non-universal features of musical pitch perception revealed by singing. *Curr. Biol.* 29, 3229–3243.e12
70. Repp, B.H. *et al.* (2013) Systematic distortions in musicians' reproduction of cyclic three-interval rhythms. *Music. Percept.* 30, 291–305
71. Polak, R. *et al.* (2018) Rhythmic prototypes across cultures: a comparative study of tapping synchronization. *Music. Percept.* 36, 1–23
72. Nan, Y. *et al.* (2006) The perception of musical phrase structure: across-cultural ERP study. *Brain Res.* 1094, 179–191
73. Mungan, E. *et al.* (2017) Perceiving boundaries in unfamiliar Turkish Makam music: evidence for Gestalt universals? *Music. Percept.* 34, 267–290
74. Drake, C. and Heni, J.B.E. (2003) Synchronizing with music: intercultural differences. *Ann. N. Y. Acad. Sci.* 999, 429–437
75. Hannon, E.E. and Trehub, S.E. (2005) Metrical categories in infancy and adulthood. *Psychol. Sci.* 16, 48–55
76. Jacoby, N. and McDermott, J.H. (2017) Integer ratio priors on musical rhythm revealed cross-culturally by iterated reproduction. *Curr. Biol.* 27, 359–370
77. Ravignani, A. *et al.* (2016) Musical evolution in the lab exhibits rhythmic universals. *Nat. Hum. Behav.* 1, 0007
78. Jacoby, N. *et al.* (2024) Commonality and variation in mental representations of music revealed by a cross-cultural comparison of rhythm priors in 15 countries. *Nat. Hum. Behav.* 8, 846–877
79. Jao Keehn, R.J. *et al.* (2019) Spontaneity and diversity of movement to music are not uniquely human. *Curr. Biol.* 29, R621–R622
80. Bregman, M.R. *et al.* (2016) Songbirds use spectral shape, not pitch, for sound pattern recognition. *Proc. Natl. Acad. Sci.* 113, 1666–1671
81. Song, X. *et al.* (2016) Complex pitch perception mechanisms are shared by humans and a New World monkey. *Proc. Natl. Acad. Sci. U. S. A.* 113, 781–786
82. Norman-Haignere, S.V. *et al.* (2019) Divergence in the functional organization of human and macaque auditory cortex revealed by fMRI responses to harmonic tones. *Nat. Neurosci.* 22, 1057–1060
83. Mehr, S.A. *et al.* (2021) Origins of music in credible signaling. *Behav. Brain Sci.* 44, e60
84. Hohl, L. *et al.* (2025) Galápagos yellow warblers differ in behavioural plasticity in response to traffic noise depending on proximity to road. *Anim. Behav.* 222, 123119
85. Doolittle, E.L. *et al.* (2014) Overtone-based pitch selection in hermit thrush song: unexpected convergence with scale construction in human music. *Proc. Natl. Acad. Sci. U. S. A.* 111, 16616–16621
86. Roeske, T.C. *et al.* (2020) Categorical rhythms are shared between songbirds and humans. *Curr. Biol.* 30, 3544–3555
87. Celma-Miralles, A. and Toro, J.M. (2020) Non-human animals detect the rhythmic structure of a familiar tune. *Psychon. Bull. Rev.* 27, 694–699
88. McDermott, J. and Hauser, M. (2004) Are consonant intervals music to their ears? Spontaneous acoustic preferences in a nonhuman primate. *Cognition* 94, B11–B21
89. McDermott, J. and Hauser, M. (2007) Nonhuman primates prefer slow tempos but dislike music overall. *Cognition* 104, 654–668
90. Robbins, L. and Margulis, S.W. (2014) The effects of auditory enrichment on gorillas. *Zoo Biol.* 33, 197–203
91. Hattori, Y. and Tomonaga, M. (2020) Rhythmic swaying induced by sound in chimpanzees (*Pan troglodytes*). *Proc. Natl. Acad. Sci. U. S. A.* 117, 936–942
92. Bertolo, M. *et al.* (2021) Sound-induced motion in chimpanzees does not imply shared ancestry for music or dance. *Proc. Natl. Acad. Sci. U. S. A.* 118, e2015664118
93. Micheyl, C. *et al.* (2005) Perceptual organization of tone sequences in the auditory cortex of awake macaques. *Neuron* 48, 139–148
94. Nicolai, J. *et al.* (2014) Human melody singing by bullfinches (*Pyrrhula pyrrula*) gives hints about a cognitive note sequence processing. *Anim. Cogn.* 17, 143–155
95. Neiworth, J.J. *et al.* (2017) Artificial grammar learning in tamarins (*Saguinus oedipus*) in varying stimulus contexts. *J. Comp. Psychol.* 131, 128–138

96. Izumi, A. (2000) Japanese monkeys perceive sensory consonance of chords. *J. Acoust. Soc. Am.* 108, 3073–3078
97. D'Amato, M.R. (1988) A search for tonal pattern perception in *Cebus* monkeys: why monkeys can't hum a tune. *Music. Percept.* 5, 453–480
98. Brosch, M. *et al.* (2004) Macaque monkeys discriminate pitch relationships. *Cognition* 91, 259–272
99. Yin, P. *et al.* (2010) Do ferrets perceive relative pitch? *J. Acoust. Soc. Am.* 127, 1673–1680
100. Wagner, B. *et al.* (2020) Is consonance attractive to budgerigars? No evidence from a place preference study. *Anim. Cogn.* 23, 973–987
101. Bregman, M.R. *et al.* (2012) Stimulus-dependent flexibility in non-human auditory pitch processing. *Cognition* 122, 51–60
102. Hoeschele, M. *et al.* (2015) Searching for the origins of musicality across species. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 370, 20140094
103. Wright, A.A. *et al.* (2000) Music perception and octave generalization in rhesus monkeys. *J. Exp. Psychol. Gen.* 129, 291–307
104. Walker, K.M. *et al.* (2019) Across-species differences in pitch perception are consistent with differences in cochlear filtering. *eLife* 8, e41626
105. Patel, A.D. *et al.* (2009) Experimental evidence for synchronization to a musical beat in a nonhuman animal. *Curr. Biol.* 19, 827–830
106. Cook, P. *et al.* (2013) A California sea lion (*Zalophus californianus*) can keep the beat: motor entrainment to rhythmic auditory stimuli in a non vocal mimic. *J. Comp. Psychol.* 127, 412–427
107. Schachner, A. *et al.* (2009) Spontaneous motor entrainment to music in multiple vocal mimicking species. *Curr. Biol.* 19, 831–836
108. Fuhrmann, D. *et al.* (2014) Synchrony and motor mimicking in chimpanzee observational learning. *Sci. Rep.* 4, 5283
109. Takeya, R. *et al.* (2017) Predictive and tempo-flexible synchronization to a visual metronome in monkeys. *Sci. Rep.* 7, 6127
110. Babiszewska, M. *et al.* (2015) Social, contextual, and individual factors affecting the occurrence and acoustic structure of drumming bouts in wild chimpanzees (*Pan troglodytes*). *Am. J. Phys. Anthropol.* 156, 125–134
111. Raimondi, T. *et al.* (2023) Isochrony and rhythmic interaction in ape duetting. *Proc. R. Soc. B Biol. Sci.* 290, 20222244
112. Hasegawa, A. *et al.* (2011) Rhythmic synchronization tapping to an audio–visual metronome in budgerigars. *Sci. Rep.* 1, 120
113. Large, E.W. and Gray, P.M. (2015) Spontaneous tempo and rhythmic entrainment in a bonobo (Pan paniscus). *J. Comp. Psychol.* 129, 317–328
114. Samuels, B. *et al.* (2021) European starlings (*Sturnus vulgaris*) discriminate rhythms by rate, not temporal patterns. *J. Acoust. Soc. Am.* 149, 2546–2558
115. Honing, H. *et al.* (2018) Rhesus monkeys (*Macaca mulatta*) sense isochrony in rhythm, but not the beat: additional support for the gradual audiomotor evolution hypothesis. *Front. Neurosci.* 12, 475
116. ten Cate, C. *et al.* (2016) Can birds perceive rhythmic patterns? A review and experiments on a songbird and a parrot species. *Front. Psychol.* 7, 730
117. Gregorio, C.D. *et al.* (2021) Categorical rhythms in a singing primate. *Curr. Biol.* 31, R1379–R1380
118. Bianco, R. *et al.* (2024) Neural encoding of musical expectations in a non-human primate. *Curr. Biol.* 34, 444–450
119. Hilton, C.B. and Mehr, S.A. (2022) Citizen science can help to alleviate the generalizability crisis. *Behav. Brain Sci.* 45, e21
120. Cohen, A.J. (2000) Development of tonality induction: plasticity, exposure, and training. *Music Percept. Interdiscip. J.* 17, 437–459
121. Peretz, I. (2016) Neurobiology of congenital amusia. *Trends Cogn. Sci.* 20, 857–867
122. Peretz, I. (1993) Auditory atonalia for melodies. *Cogn. Neuropsychol.* 10, 21–56
123. Steinke, W.R. *et al.* (1997) Dissociation of musical tonality and pitch memory from nonmusical cognitive abilities. *Can. J. Exp. Psychol.* 51, 316–335
124. Weiss, M.W. and Peretz, I. (2022) Improvisation is a novel tool to study musicality. *Sci. Rep.* 12, 12595
125. Cuddy, L.L. and Duffin, J. (2005) Music, memory, and Alzheimer's disease: is music recognition spared in dementia, and how can it be assessed? *Med. Hypotheses* 64, 229–235
126. Weinstein, J. *et al.* (2011) Preserved musical semantic memory in semantic dementia. *Arch. Neurol.* 68, 248–250
127. Tun, S.M. *et al.* (2022) A professional musician with sight-reading difficulties: a case study. *Alzheimers Dement.* 18, e059696
128. Lynch, M.P. and Eilers, R.E. (1992) A study of perceptual development for musical tuning. *Percept. Psychophys.* 52, 599–608
129. Trehub, S.E. *et al.* (1999) Infants' and adults' perception of scale structure. *J. Exp. Psychol. Hum. Percept. Perform.* 25, 965–975
130. Lamont, A. and Cross, I. (1994) Children's cognitive representations of musical pitch. *Music. Percept.* 12, 27–55
131. Edalati, M. *et al.* (2023) Rhythm in the premature neonate brain: very early processing of auditory beat and meter. *J. Neurosci.* 43, 2794–2802
132. Saadatmehr, B. *et al.* (2025) Auditory rhythm encoding during the last trimester of human gestation: from tracking the basic beat to tracking hierarchical nested temporal structures. *J. Neurosci.* 45, e0398242024
133. Lenc, T. *et al.* (2023) Infants show enhanced neural responses to musical meter frequencies beyond low-level features. *Dev. Sci.* 26, e13353
134. Hannon, E.E. *et al.* (2011) Constraints on infants' musical rhythm perception: effects of interval ratio complexity and enculturation. *Dev. Sci.* 14, 865–872
135. Nave, K. *et al.* (2024) The development of rhythmic categories as revealed through an iterative production task. *Cognition* 242, 105634
136. Hannon, E.E. and Trehub, S.E. (2005) Tuning in to musical rhythms: Infants learn more readily than adults. *Proc. Natl. Acad. Sci. U. S. A.* 102, 12639–12643
137. Nave-Blodgett, J.E. *et al.* (2021) Auditory superiority for perceiving the beat level but not measure level in music. *J. Exp. Psychol. Hum. Percept. Perform.* 47, 1516–1542
138. Nave, K.M. *et al.* (2023) Sustained musical beat perception develops into late childhood and predicts phonological abilities. *Dev. Psychol.* 59, 829–844
139. Patel, A.D. (2017) Why doesn't a songbird (the European starling) use pitch to recognize tone sequences? The informational independence hypothesis. *Comp. Cogn. Behav. Rev.* 12, 19–32
140. Pinker, S. (2021) Sex and drugs and rock and roll. *Behav. Brain Sci.* 44, e109