MIT Press • Open Encyclopedia of Cognitive Science

Music Cognition

Josh H. McDermott[1,2]

[1]Massachusetts Institute of Technology, [2]Harvard University

MIT Press

Published on: Jul 21, 2026

URL: https://oecs.mit.edu/pub/x1cmpaio

Music is a form of communication in which one or more people produce structured sound sequences. When perceived by others, these sound signals can cause the experience of emotions and the impulse to move in synchrony with the sound, and can often be understood by the receiver in a way that allows them to reproduce a similar signal themselves. Humans learn the structural tendencies of the music of their culture over development, and most individuals have some degree of perceptual and productive competence, although there is substantial variation in abilities and taste. Knowledge of music's structural tendencies allows the listener to anticipate the trajectory of a musical piece and allows music to be represented and stored in compressed form. Music has superficial similarities to language, but current evidence suggests that it recruits perceptual and cognitive brain circuits that are distinct from those supporting speech and language. Although music is practiced in all known cultures, the science of music cognition has historically been oriented towards ideas that are prevalent in academic traditions surrounding Western music. Recent scholarship has begun to address the diversity of music and music perception around the world.

History

Music was of great interest to the pioneering psychophysicists of the 1800s. Research in that era was constrained by stimulus generation and measurement, with early observations relying on tuning forks and other physical means of generating sound. Despite these limitations, much of what we know about musical acoustics and related aspects of auditory perception has its roots in this era. Hermann von Helmholtz's treatise *On the Sensations of Tone* is perhaps the best-known scientific product of this period. One influential contribution was to attempt to link aesthetic judgments of combinations of notes (considered *consonant* or *dissonant* to Western listeners, depending on the note combination) to the acoustic and perceptual consequences of the constituent frequencies contained in musical notes.

Music was relatively neglected during the behaviorist era of psychology, but the subsequent cognitive revolution gave rise to foundational work on music cognition in the 1970s and 1980s, aided by the ability to record and play back audio stimuli. Research in this period used behavioral experiments to characterize mental representations of musical structure—in particular, how they differed from literal representations of the musical sound stimulus. The best of this work has held up well, although it was Western centric in focus. The 1990s and 2000s brought influential studies documenting the developmental trajectory of musical competencies, along with widespread uptake of brain imaging methods [see Neuroaesthetics]. The prospect of functionally localizing the brain basis of cognitive abilities and representations led to a focus on the domain specificity of music. The 2010s and 2020s were marked by a growing appreciation of the diversity of music around the world, fueled by quantitative studies of musical corpora and by attempts to rigorously conduct comparisons of music perception across different cultures.

Core concepts

Musical structure

Music begins as a representation in the mind of a human music producer who then produces sound by singing, playing an instrument, or other means. In most musical systems, music is structured in pitch and time, among other variables, and is composed of elements that are at least partially discretized. For instance, notes are usually drawn from a fairly small set of possible pitches (e.g., 5 to 7 notes per octave, defining a *scale*) and tend to be positioned on a temporal grid of discrete times rather than being at arbitrary pitches or times. Musical elements are usually also temporally organized to form periodic rhythms, with repetition of motifs being common. Every culture's music exhibits strong structural tendencies, such that some structures are much more probable than others. Some of these tendencies are fairly common across cultures (for instance, the tendency for adjacent notes in melodies to be separated by small, rather than large, pitch steps), whereas others are culture specific (for instance, the particular scales from which notes tend to be drawn and the particular rhythms that are commonly used). Notably, the structure of music is distinct from that of other communication signals. Speech, for instance, is not organized into periodic rhythms or discrete pitch classes and does not typically feature intentional phrase repetition (Patel, 2008). Music is also distinct from speech in tending to convey emotion and mood rather than specific semantic content [see Emotions, Moods, and Affective States].

Perceptual and cognitive representations of music

Musical sound is decoded by the brain of the listener. There are many reasons to think that the decoded representation is symbolic in nature, consisting of relatively discrete elements akin to notes, chords, or melodic phrases. One computational function of music perception and cognition is thus to recover this symbolic representation from sound (loosely akin to inferring a score-like representation from audio, although the inferred representation likely deviates very substantially from an actual score). Part of the problem is to infer the individual acoustic events that are often superimposed in the sound signal (e.g., notes from different instruments or voices), a process that leverages known perceptual grouping effects of harmonic frequency structure, temporal synchrony, and repetition (Darwin, 1997). Another part of the problem is to infer the symbolic representation intended by the producer despite potential production errors and other sources of uncertainty. Human listeners exhibit well-documented perceptual biases that presumably help them to correctly estimate symbolic content despite production errors. For instance, humans are biased to hear rhythms with time intervals defined by low-integer ratios, tending to perceive such ratios even when the stimulus deviates from them (Desain & Honing, 2003). One property of the decoded representation is that it allows the receiver to produce their own rendition of the heard music (e.g., by singing back a heard melody), with the receiver's perceptual biases likely serving to stabilize such renditions.

A second candidate computational function of music cognition is to predict upcoming musical events. The enjoyment of music is often proposed to relate to this ability (Huron, 2006), as it allows people to form expectations that are confirmed or violated, making music interesting to listen to. The formation of predictions relies on knowledge of the statistical tendencies of musical structure. For instance, if Western listeners hear a particular chord, they implicitly know that some notes are likely to subsequently occur and that others are not (Krumhansl & Kessler, 1982). In Western music, expectations are driven in part by an implicit understanding of the musical *key*, in which a particular note of the scale—the tonic—serves as a reference point at which melodies tend to end.

Accurate prediction of upcoming musical events likely facilitates musical coordination when multiple individuals sing or play instruments together. However, such coordination is not limited to musicians, with most humans being able to synchronously tap along to a periodic musical rhythm (Repp, 2005). Even when listeners are not actually moving, auditory stimuli that convey a regular beat tend to activate motor circuits in the brain (Grahn & Brett, 2007).

Mental representations of music omit details of heard music in ways that suggest the influence of memory constraints. For instance, if played two successive notes, listeners can discriminate the pitch interval between the notes, demonstrating that the interval is encoded. However, upon hearing a longer sequence of notes, listeners can often only discriminate the direction of pitch changes between notes (the *contour*) rather than the fine-grained pitch intervals (Dowling & Fujitani, 1971). The fine-grained pitch intervals of the melody are instead summarized by aggregate properties of the note pitches, such as the musical key. Together, the melodic contour and sense of key provide a compressed representation that is plausibly more efficient to store in memory but that retains some of the structure needed to form expectations of upcoming notes.

Learning and expertise

Knowledge of music is believed to be acquired via implicit learning from incidental experience with a culture's music over the course of an individual's development. Evidence of such learning comes from two types of findings. First, adults tend to be more sensitive to culture-specific musical structure than infants or young children. For instance, adults notice when a note in a melody falls outside the scale implied by the rest of the melody (producing the sense of a "sour note" played by mistake), whereas infants do not (Trainor & Trehub, 1992). Moreover, adults are in some cases worse than infants at differentiating structures that are less common in their culture's music (Hannon & Trehub, 2005). Second, many basic musical cognitive capacities appear to be qualitatively similar in individuals without explicit training in music (nonmusicians) as in trained musicians. For instance, musicians and nonmusicians exhibit similar expectations of what is likely to occur next given an excerpt of music (Bigand & Poulin-Charronnat, 2006). Musicians are nonetheless often quantitatively better at music-related tasks than nonmusicians (e.g., being better able to detect small changes in pitch or rhythm; Micheyl et al., 2006).

The neural basis of music

Current evidence suggests that music perception and cognition depend on brain circuits that are somewhat specific to music rather than co-opting circuits used for other perceptual or cognitive functions. First, some parts of the auditory cortex respond much more to music than to other types of sounds (Norman-Haignere et al., 2015). Second, musical abilities can be compromised, either via brain damage or by congenital deficits, without much noticeable effect on other abilities (Peretz & Coltheart, 2003). However, the detailed function of music-specific brain circuitry remains unclear as of now, and it is unclear whether the initial processing of musical elements (e.g., to extract pitch or timing) is specific to music. It also remains unclear whether music-specific neural circuits are present from birth or instead develop from experience with music.

Questions, controversies, and new developments

Moving beyond Western music

One ongoing challenge for the science of music is to develop an account of music cognition that does not just apply to listeners of Western music. Past research has often revolved around structural properties of music that are prevalent in Western music, in particular, classical Western music. This situation has begun to be addressed with two types of empirical research. One involves surveying music from around the world and over history and documenting regularities and variation in musical structure (Mehr et al., 2019). In parallel, cross-cultural experiments have probed for similarities and differences in the perceptual traits of listeners from different cultures. Such work has thus far yielded some evidence for universal musical traits, in particular, perceptual biases favoring rhythms composed of time intervals related by a low-integer ratio (Jacoby et al., 2024). However, experiments have also revealed examples of striking cross-cultural variation, for instance, in preferences for some note combinations over others (McDermott et al., 2016).

Moving beyond the score

Music research has also been preferentially directed towards variables represented in the traditional Western musical score, which predominantly notates pitch and timing. Less is known about the other variables that describe music, many of which are traditionally referred to as *timbre* (defined as encompassing all aspects of a sound other than pitch and loudness). The many choices made by a music producer in turning a score into actual music—the choice of instruments, singing style, environmental acoustic effects such as reverberation, and, nowadays, a huge assortment of digital audio effects—have an enormous effect on what a listener feels when hearing the music (and indeed, whether they want to listen at all). Studying such effects seems essential for a full understanding of the human musical experience.

Broader connections

Another promising direction for the coming decade seems likely to be the development of working computational models of human music cognition. A complete account of music requires being able to build models that exhibit human-like musical competence—giving human-like judgments when presented with musical audio—and that can reproduce the developmental trajectory of human music perception and cognition in response to cumulative exposure to music. Such models are newly approachable because of recent advances in artificial intelligence. Machine learning models can now be trained to make music, learning from examples of existing music to sequentially generate musical events given past musical context (Huang et al., 2019) [see Transformers; Recurrent Neural Networks]. Such models provide novel approaches to music composition, but because they generate predictions of upcoming musical events from past events, they also provide an avenue via which candidate models of human music cognition could be built.

Further reading

- Deutsch, D. (Ed.). (2013). *The psychology of music* (3rd ed.). Academic Press.
- Huron, D. (2006). *Sweet anticipation: Music and the psychology of expectation*. MIT Press.

References

- Bigand, E., & Poulin-Charronnat, B. (2006). Are we "experienced listeners"? A review of the musical capacities that do not depend on formal musical training. *Cognition*, *100*(1), 100-130. https://doi.org/10.1016/j.cognition.2005.11.007

  ↩
- Darwin, C. J. (1997). Auditory grouping. *Trends in Cognitive Sciences*, *1*(9), 327-333. https://doi.org/10.1016/S1364-6613(97)01097-8

  ↩
- Desain, P., & Honing, H. (2003). The formation of rhythmic categories and metric priming. *Perception*, *32*(3), 341-365. https://doi.org/10.1068/p3370

  ↩
- Dowling, W. J., & Fujitani, D. S. (1971). Contour, interval, and pitch recognition in memory for melodies. *The Journal of the Acoustical Society of America*, *49*(2), 524-531. https://doi.org/10.1121/1.1912382

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

  ↩

- Hannon, E. E., & Trehub, S. E. (2005). Metrical categories in infancy and adulthood. *Psychological Science*, *16*(1), 48-55. https://doi.org/10.1111/j.0956-7976.2005.00779.x

  ↩
- Huang, C.-Z. A., Vaswani, A., Uszkoreit, J., Simon, I., Hawthorne, C., Shazeer, N., Dai, A. M., Hoffman, M. D., Dinculescu, M., & Eck, D. (2019). Music transformer. *arXiv*. https://doi.org/10.48550/arXiv.1809.04281

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

  ↩
- Jacoby, N., Polak, R., Grahn, J. A., Cameron, D. J., Lee, K. M., Godoy, R., Undurraga, E. A., Huanca, T., Thalwitzer, T., Doumbia, N., Goldberg, D., Margulis, E. H., Wong, P. C. M., Jure, L., Rocamora, M., Fujii, S., Savage, P. E., Ajimi, J., Konno, R., . . . McDermott, J. H. (2024). Commonality and variation in mental representations of music revealed by a cross-cultural comparison of rhythm priors in 15 countries. *Nature Human Behaviour*, *8*(5), 846-877. https://doi.org/10.1038/s41562-023-01800-9

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

  ↩
- 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. *Nature*, *535*, 547-550. https://doi.org/10.1038/nature18635

  ↩
- 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., O'Donnell, T. J., Krasnow, M. M., & Glowacki, L. (2019). Universality and diversity in human song. *Science*, *366*(6468), eaax0868. https://doi.org/10.1126/science.aax0868

  ↩
- Micheyl, C., Delhommeau, K., Perrot, X., & Oxenham, A. J. (2006). Influence of musical and psychoacoustical training on pitch discrimination. *Hearing Research*, *219*(1-2), 36-47. https://doi.org/10.1016/j.heares.2006.05.004

  ↩

- Norman-Haignere, S., Kanwisher, N., & McDermott, J. H. (2015). Distinct cortical pathways for music and speech revealed by hypothesis-free voxel decomposition. *Neuron*, *88*(6), 1281-1296. https://doi.org/10.1016/j.neuron.2015.11.035

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

  ↩
- Peretz, I., & Coltheart, M. (2003). Modularity of music processing. *Nature Neuroscience*, *6*(7), 688-691. https://doi.org/10.1038/nn1083

  ↩
- Repp, B. H. (2005). Sensorimotor synchronization: A review of the tapping literature. *Psychonomic Bulletin & Review*, *12*(6), 969-992. https://doi.org/10.3758/bf03206433

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

  ↩