



Published in final edited form as:

Curr Opin Neurobiol. 2008 August ; 18(4): 452–463. doi:10.1016/j.conb.2008.09.005.

Music perception, pitch, and the auditory system

Josh H. McDermott and Andrew J. Oxenham

Department of Psychology, University of Minnesota

Abstract

The perception of music depends on many culture-specific factors, but is also constrained by properties of the auditory system. This has been best characterized for those aspects of music that involve pitch. Pitch sequences are heard in terms of relative, as well as absolute, pitch. Pitch combinations give rise to emergent properties not present in the component notes. In this review we discuss the basic auditory mechanisms contributing to these and other perceptual effects in music.

Introduction

Music involves the manipulation of sound. Our perception of music is thus influenced by how the auditory system encodes and retains acoustic information. This topic is not a new one, but recent methods and findings have made important contributions. We will review these along with some classic findings in this area. Understanding the auditory processes that occur during music can help to reveal why music is the way it is, and perhaps even provide some clues as to its origins. Music also provides a powerful stimulus with which to discover interesting auditory phenomena; these may in turn reveal auditory mechanisms that would otherwise go unnoticed or underappreciated.

We will focus primarily on the role of pitch in music. Pitch is one of the main dimensions along which sound varies in a musical piece. Other dimensions are important as well, of course, but the links between basic science and music are strongest for pitch, mainly because something is known about how pitch is analyzed by the auditory system. Timbre [1] and rhythm [2], for instance, are less well linked to basic perceptual mechanisms, although these represent interesting areas for current and future research.

Pitch

Pitch is the perceptual correlate of periodicity in sounds. Periodic sounds by definition have waveforms that repeat in time (Fig. 1a). They typically have harmonic spectra (Fig. 1b), the frequencies of which are all multiples of a common fundamental frequency (F_0). The F_0 is the reciprocal of the period – the time it takes for the waveform to repeat once. The F_0 need not be the most prominent frequency of the sound, however (Fig. 1b), or indeed even be physically present. Although most voices and instruments produce a strong F_0 , many audio playback devices, such as handheld radios, do not have speakers capable of reproducing low frequencies, and so the fundamental is sometimes not present when music is listened to. Sounds with a particular F_0 are generally perceived to have the same pitch despite potentially having very different spectra.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Most research on pitch concerns the mechanisms by which the pitch of an individual sound is determined [3]. The pitch mechanism has to determine whether a sound waveform is periodic, and with what period. In the frequency domain, periodicity is signaled by harmonic spectra. In the time domain, periodicity is revealed by the waveform autocorrelation function, which contains regular peaks at time lags equal to multiples of the period (Fig. 1c). Both frequency and time domain information are present in the peripheral auditory system. The filtering that occurs in the cochlea provides a frequency-to-place, or “tonotopic,” mapping that breaks down sound according to its frequency content. This map of frequency, established in the cochlea, is maintained to some degree throughout the auditory system up to and including primary auditory cortex [4]. Periodicity information in the time domain is maintained through the phase-locking of neurons, although the precision of phase-locking deteriorates at each successive stage of the auditory pathway [5]. The ways in which these two sources of information are used remains controversial [6,7]. Although pitch mechanisms are still being studied and debated, there is recent evidence for cortical neurons beyond primary auditory cortex that are tuned to pitch [8]. These are distinguished from neurons that are merely frequency-tuned by being tuned to the F0 of a complex tone, irrespective of its exact spectral composition.

Pitch Relations Across Time – Relative Pitch

When listening to a melody, we perceive much more than just the pitch of each successive note. In addition to these individual pitches, which we will term the absolute pitches of the notes, listeners also encode how the pitches of successive notes relate to each other – for instance, whether a note is higher or lower in pitch than the previous note, and perhaps by how much. Relative pitch is intrinsic to how we perceive music. We readily recognize a familiar melody when all the notes are shifted upwards or downwards in pitch by the same amount (Fig. 2a), even though the absolute pitch of each note changes [9]. This ability depends on a relative representation, as the absolute pitch values are altered by transposition. Relative pitch is presumably also important in intonation perception, in which meaning can be conveyed by a pitch pattern (e.g. the rise in pitch that accompanies a question in spoken English), even though the absolute pitches of different speakers vary considerably.

The existence of relative pitch perception may seem unsurprising given the relational abilities that characterize much of perception. However, standard views of the auditory system might lead one to believe that absolute pitch would dominate perception, as the tonotopic representations that are observed from the cochlea [10] to the auditory cortex [4] make absolute, rather than relative, features of a sound’s spectrum explicit. Despite this, relative pitch abilities are present even in young infants, who seem to recognize transpositions of melodies just as do adults [11]. Relative pitch may thus be a feature inherent to the auditory system, in any case not requiring extensive experience or training to develop.

Relative pitch – behavioral evidence

One of the most salient aspects of relative pitch is the direction of change (up or down) from one note to the next, known as the contour (Fig. 2b). Most people are good at encoding the contour of a novel sequence of notes, as evidenced by the ability to recognize this contour when replicated in a transposed melody (Fig. 2a) [12,13]. Recent evidence indicates that contours can also be perceived in dimensions other than pitch, such as loudness and brightness [14]. A pattern of loudness variation, for instance, can be replicated in a different loudness range, and can be reliably identified as having the same contour. It remains to be seen whether contours in different dimensions are detected and represented by a common mechanism, but their extraction appears to be a general property of the auditory system.

In contrast to their general competence with contours, people tend to be much less accurate at recognizing whether the precise pitch intervals (Fig. 2b) separating the notes of an unfamiliar

melody are preserved across transposition. If listeners are played a novel random melody, followed by a second melody that is shifted to a different pitch range, they typically are unable to tell if the intervals between notes have been altered so long as the contour does not change [12], particularly if they do not have musical training [15]. This has led to a widely held distinction between contour and interval information in relative pitch. A priori it seems plausible that the difficulty in encoding intervals might simply be due to their specification requiring more information than that of a contour; in our view this hypothesis is difficult to reject given current evidence.

Discrimination thresholds for the pitch interval between two notes are measured by presenting listeners with two pairs of sequential tones, one after the other, with the pitch interval larger in one pair than the other. The listener's task is to identify the larger interval. To distinguish this ability from mere frequency discrimination, the lower note of one interval is set to be higher than that of the other interval, such that the task can only be performed via the relative pitch interval between the notes. Thresholds obtained with this procedure are typically on the order of a semitone in listeners without musical training [16]. A semitone is the smallest amount by which musical intervals normally differ (Fig. 3a). This suggests that the perceptual difference between neighboring intervals (e.g. major and minor thirds; Fig. 3b) is probably not very salient for many listeners, at least in an isolated context, as it is very close to threshold.

Intervals defined by simple integer ratios (Fig. 3b) have a prominent role in Western music. The idea that they enjoy a privileged perceptual status has long had popularity, but supporting evidence has been elusive. Interval discrimination is no better for "natural" intervals (e.g. the major third and fourth, defined by 5:4 and 4:3 ratios) than for unnatural (e.g. 4.5 semitones, approximately 13:10) [16]. One exception to this is the octave. Listeners seem more sensitive to deviations from the octave than to deviations from adjacent intervals, both for simultaneously [17] and sequentially [18,19] presented tones. The same method yields weak and inconsistent effects for other intervals [20], suggesting that the octave is unique in this regard. This special status dovetails with the prevalence of the octave in music from around the globe; apart from the fifth, other intervals are not comparably widespread with much consistency. There is, however, one report that changing a "natural" interval (a perfect fifth or fourth) to an unnatural one (a tritone) is more salient to infant listeners than the reverse change [21], so this issue remains somewhat unresolved.

Despite the large thresholds characterizing interval discrimination, and despite the generally poor short-term memory for the pitch intervals of arbitrary melodies, interval differences on the order of a semitone are critically important in most musical contexts. For melodies obeying the rules of tonal music (see next paragraph), a pitch-shifted version containing a note that violates these rules (for instance, by being outside of the scale; Fig. 3a) is highly noticeable, even though such changes are often only a semitone in magnitude. A mere semitone change to two notes can turn a major scale into a minor scale (Fig. 3a), which in Western music can produce a salient change in mood (minor keys often being associated with sadness and tension). Intervals are also a key component of our memory for familiar melodies, which are much less recognizable if only the contour is correctly reproduced [12,14]. Moreover, electrophysiological evidence suggests that interval changes to familiar melodies are registered by the auditory system even when listeners are not paying attention, as indexed by the mismatch negativity response [22].

The ability to encode intervals in melodies seems intimately related to the perception of tonal structure. Musical systems typically use a subset of the musically available notes at any given time, and generally give special status to a particular note within that set [23]. In Western music this note is called the tonic (in the key of C major, depicted in Fig. 3a, C is the tonic). Different notes within the pitch set are used with different probabilities, with the tonic occurring most

frequently and with longer durations. In Western music this probability distribution defines what is known as the key; a melody whose pitch distribution follows such tendencies is said to be tonal. Listeners are known to be sensitive to these probabilities, and use them to form expectations for what notes to expect (Fig. 3c) [24], expectations that are a critical aspect of musical tension and release [25,26].

In many musical systems the pitch sets that are commonly used are defined by particular patterns of intervals between notes. Thus, if presented with five notes of the major scale, a Western listener will have expectations for what other notes are likely to occur, even if they have not yet been played, because only some of the remaining available notes have the appropriate interval relations with the observed notes. Listeners thus internalize templates for particular pitch sets that are common in the music of their culture (see Fig. 3c for two examples from Western music). These templates are relative pitch representations in that the tonic can be chosen arbitrarily; the structural roles of other pitches are then determined by the intervals that separate them from the tonic.

Most instances in which interval alterations are salient to listeners involve violations of tonal structure – the alteration introduces a note that is inappropriate given the pitch set that the listener expects. Interval changes that substitute another note within the same scale, for instance, are often not noticed [27]. Conversely, manipulations that make tonal structure more salient, such as lengthening the test melodies, make interval changes easier to detect [13]. It thus seems that pitch intervals are not generally retained with much accuracy, but can be readily incorporated into the tonal pitch structures that listeners learn via passive exposure [28]. Interval perception in other dimensions of sound has been little explored [29], so it remains unclear whether there are specialized mechanisms for representing pitch intervals, but the precise interval perception that is critical to music seems to depend greatly on matching observed pitch sequences to learned templates.

Neural mechanisms of relative pitch

Evidence from neuropsychology has generally been taken as suggestive that contour and intervals are mediated by distinct neural substrates [30,31], with multiple reports that brain damage occasionally impairs interval information without having much effect on contour perception. Such findings are, however, also consistent with the idea that the contour is simply more robust to degradation. Alternatively, anatomical segregation could be due to separate mechanisms for contour and tonality perception [32], the latter of which seems to be critical to many interval tasks. It has further been suggested that interval representations might make particular use of left temporal regions [33], although some studies find little or no evidence for this pattern of lateralization [31,34].

Functional imaging studies in healthy human subjects indicate that pitch changes activate temporal regions, often in the right hemisphere [35-38], with one recent report that the right hemisphere is unique in responding to fine-grained pitch changes on the order of a semitone or so [39]. Such results could merely indicate the presence of pitch selective neurons [40-42], rather than direction selective units per se, but studies of brain damaged patients indicate that deficits specific to pitch change direction derive from right temporal lesions, whereas the ability to simply detect pitch changes (without having to identify their direction) is not much affected [43-45].

An impaired ability to discriminate pitch change direction is thought to at least partially characterize tone-deafness, officially known as congenital amusia. Tests of individuals who claim not to enjoy or understand music frequently reveal elevated thresholds for pitch direction discrimination [46-48], although the brain differences that underlie these deficits remain unclear [49]. At the other end of the spectrum, trained musicians are generally better than non-

musicians at relative pitch tasks like interval and contour discrimination [15,16]. However, they also perform better on basic frequency discrimination [50] and other psychoacoustic tasks [51]. Perceptual differences related to musical training therefore do not seem specific to relative pitch, a conclusion consistent with the many structural and functional brain differences evident in musicians [52,53].

There is thus some evidence for a locus for relative pitch in the brain, which when damaged can impair music perception. The mechanisms by which this pitch information is extracted remain poorly characterized. Demany and Ramos recently reported psychophysical evidence in humans for frequency-shift detectors [54], which might constitute such a mechanism. When presented with a random “chord” of pure tone components, listeners had trouble saying whether a subsequently presented tone was part of the chord. But if the subsequent tone was slightly offset from one of the chord components, listeners reported hearing a pitch shift, and could discriminate upward from downward shifts. Because listeners could not hear the individual tones, performance could only be explained with a mechanism that responds specifically to frequency shifts.

Physiological investigations of frequency sweeps have demonstrated direction selectivity to FM in the auditory cortex [55], but responses to sequences of discrete tones paint a more complex picture. Playing one pure tone before another can alter the response to the second tone [56-58], but such stimuli do not normally produce direction-selective responses. Brosch and colleagues recently reported direction-selective responses to discrete tone sequences in primary and secondary auditory cortex, in monkeys trained to respond to downward pitch shifts [59]. Their animals were trained for two years on the task, and responses were only observed to the downward (rewarded) direction, so the responses are probably not representative of the normal state of the auditory system. The neural mechanisms of relative pitch thus remain poorly understood. It is worth noting that relative pitch perception appears to be less natural for nonhuman animals than for humans, as animals trained to recognize melodies typically generalize to transpositions only with great difficulty [60,61]. There is a pronounced tendency for nonhuman animals to attend to absolute rather than relative features of sounds [62], and it is still unclear whether this reflects differences in basic perceptual mechanisms.

Relative, absolute, and perfect pitch

Our dichotomy of absolute and relative pitch omits another type of pitch perception that has received much attention – that which is colloquially known as perfect pitch. To make matters more confusing this type of pitch perception is often referred to as absolute pitch in the scientific literature. Perfect pitch refers specifically to the ability to attach verbal labels to a large set of notes, typically those of the chromatic scale. It is a rare ability (roughly 1 in 10,000 people have it), and has little relevance to music perception in the average listener, but it does have some very interesting properties. The reader is referred to other reviews of this phenomenon for more details [63,64].

Perhaps in part because of the rarity of perfect pitch, it is often claimed that the typical human listener has poor memory for absolute pitch. In fact, all that most humans clearly lack is the ability to label pitches at a fine scale, being limited to the usual 7 +/- 2 number of categories [65]. This categorization limit need not mean that the retention of the absolute pitch of a sound (by which we mean the perceptual correlate of its estimated F0) is limited. Studies of song memory indicate that the average person encodes absolute pitch in memory with fairly high fidelity, at least for highly over-learned stimuli. These experiments have asked subjects to either sing back well-known popular songs [66], or to judge whether a rendition of such a tune is in the correct key or shifted up or down by a small amount [67]. Subjects produce the correct absolute pitch more often than not when singing, and can accurately judge a song’s correct

pitch to within a semitone, indicating that absolute pitch is retained with reasonably high fidelity. Both these studies utilized stimuli that subjects had heard many times and that were thus engrained in long-term memory. Absolute pitch is also retained in short-term memory to some extent, particularly if there are no intervening stimuli [68]. Unsurprisingly, the absolute pitch of stimuli heard a single time decays gradually in short-term memory [69]; similar measurements have not been made for relative pitch. Because relative and absolute pitch abilities have never been quantitatively compared, there is little basis for claiming superiority of one over the other. The typical person encodes both types of pitch information, at least to some extent.

Representation of Simultaneous Pitches - Chords and Polyphony

One of the interesting features of pitch is that different sounds with different pitches can be combined to yield a rich array of new sounds. Music takes full advantage of this property, as the presence of multiple simultaneous voices in music is widespread [70]. This capability may in fact be one reason why pitch has such a prominent role in music, relative to many other auditory dimensions [71]. An obvious question involves what is perceived when multiple pitches are played at once. Do people represent multiple individual pitches, or do they instead represent a single sound with aggregate properties?

Examples of music in which multiple sequences of tones are heard as separate “streams” are commonplace, especially in Western music. This is known as polyphony, and is believed to take advantage of well-known principles of auditory scene analysis [72]. Electrophysiological evidence indicates that for simple two-part polyphony, even listeners without musical training represent two concurrent melodies as separate entities [73]. The study in question made use of the mismatch negativity (MMN), a response that occurs to infrequent events. Changes made to one of two concurrent melodies on 25% of trials produced an MMN, even though a change made to the stimulus as a whole occurred 50% of the time, not infrequently enough to produce a MMN. This indicates that the two melodies are represented as distinct at some level of the auditory system. This ability probably does not extend indefinitely; behavioral studies of polyphony perception indicate that once there are more than three voices present, even highly experienced listeners tend to underestimate their number, suggesting limitations on how many things can be represented at once [74].

In other contexts, sequences of multiple simultaneous notes are not intended to be heard as separate streams but rather as fused chords; this is known as homophony. What is represented in the auditory system when listeners hear a chord? The ability to name the notes comprising a chord generally requires considerable practice, and is a major focus of ear training in music, suggesting that chords may not naturally be represented in terms of their component notes. However, to our knowledge there have been no psychophysical tests of this issue with music-like stimuli. MEG experiments again suggest that two simultaneously presented pitches are encoded separately, as a change to either one can produce a MMN [75], but combinations of more than two notes have not been examined.

Related issues have been addressed with artificial “chords” generated from random combinations of pure tones. When presented with such stimuli followed by a probe tone, listeners are unable to tell if the probe tone was contained in the chord for chords containing more than three or four tones, even when the tones are sufficiently far apart so as to preclude peripheral masking [54]. The tones seem to fuse together into a single sound, even when the frequencies are not harmonically related as in a periodic sound. It remains unclear to what extent these findings will extend to the perception of musical chords composed of multiple complex tones.

The representation of chords thus represents an interesting direction for future research. The issue has implications for theories of chord quality and chord relations, some of which postulate that the notes and intervals comprising a chord are individually represented [25,76]. If chord notes are not individually represented, chord properties must be conveyed via aggregate acoustic properties that are as yet unappreciated.

Consonance

Consonance is perhaps the most researched emergent property that occurs in chords. To Western listeners, certain combinations of notes, when played in isolation, seem pleasant (consonant), whereas others seem unpleasant (dissonant). Of course, the aesthetic response to an interval or chord is also a function of the musical context; with appropriate surroundings, a dissonant interval can be quite pleasurable, and often serves important musical functions. However, in isolation, Western listeners consistently prefer some intervals to others, and this depends little on musical training [77]. This effect is often termed sensory consonance to distinguish it from the more complicated effects of context [78].

The possible innateness of these preferences has been the subject of much interest and controversy. Infants as young as 2 months of age appear to prefer consonant intervals to dissonant ones [79-81]. This is consistent with an innate account, although it is difficult to rule out effects of incidental music exposure that all infants surely have, even while still in the womb. There is an unfortunate dearth of cross-cultural studies testing whether preferences for consonance are universal [77,82], but the prevalence of dissonant intervals in some other musical cultures [70] could indicate that such preferences may be learned, or at least easily modifiable. The apparent absence of consonance preferences in some species of nonhuman primates is also evidence that the preference is not an inevitable consequence of the structure of the auditory system [83]. On the other hand, in at least some cultures in which conventionally dissonant intervals are common, they appear to be used to induce tension [84], much as they are in Western music. It thus remains conceivable that there is a component of the response to dissonance that is universal, and that its prevalence in other cultures is due to alternative uses for tension in music.

Even if the preference for consonance is not predetermined by the auditory system, it might nonetheless be constrained by sensory factors. For instance, acoustic properties of note combinations could determine classes of intervals and chords that are perceptually similar. These classes might then be used in different ways in music, which could in turn determine their aesthetic value. Proposals for the psychoacoustic basis of consonance date back to the Greeks, who noted that consonant pairs of notes have fundamental frequencies that are related by simple integer ratios.

Most contemporary accounts of consonance hold that ratios are not the root cause, but rather are correlated with a more fundamental acoustic variable. The most prevalent such theory is usually attributed to Helmholtz, who contended that consonance and dissonance are distinguished by a phenomenon known as beating. Beating occurs whenever two simultaneous tones are close but not identical in frequency (Fig. 4a). Over time the tones drift in and out of phase, and the amplitude of the summed waveform waxes and wanes. Helmholtz noted that the simple ratios by which consonant notes are related cause many of their frequency components to overlap exactly (Fig. 4b), and contended that consonance results from the absence of beats. In contrast, a dissonant interval tends to have many pairs of components that are close but not identical in frequency, and that beat (Fig. 4c). The waveforms of consonant and dissonant intervals accordingly reveal substantial differences in the degree of amplitude modulation (Fig. 4d&e). Helmholtz argued that these fluctuations give rise to the perception of roughness, and that roughness distinguishes dissonance from consonance.

Quantitative tests of the roughness account of consonance have generally been supportive [85-87] – calculated roughness minima occur at intervals with simple integer ratios, and maxima at intervals with more complex ratios, such as the minor second and tritone (also known as the augmented fourth). Neural responses to consonant and dissonant intervals are also consistent with a role for roughness. Auditory cortical neurons tend to phase-lock to amplitude fluctuations in a sound signal, so long as they are not too fast [88]. These oscillatory firing patterns may be the physiological correlate of roughness, and have been observed in response to dissonant, but not consonant musical intervals in both monkeys and humans [89].

Although the roughness account of consonance has become widely known, an alternative theory, originating from ideas about pitch perception, also has plausibility. Many have noted that because consonant intervals are related by simple ratios, the frequency components that result from combining a consonant pair of notes are a subset of those that would be produced by a single complex tone with a lower fundamental frequency [78,90]. As a result, a consonant combination of notes has an approximately periodic waveform, generally with a period in the range where pitch is perceived. This is readily seen in the waveform autocorrelation function (Fig. 4f). Dissonant intervals, in contrast, because of the complex ratio generally relating the two F0s, are only consistent with an implausibly low F0, often below the lower limit of pitch (of around 30 Hz) [91], and accordingly do not display signs of periodicity in their autocorrelation (Fig. 4g).

As with roughness, there are physiological correlates of this periodicity. Tramo and colleagues recorded responses to musical intervals in the auditory nerve of cats, computing all-order interspike interval histograms (equivalent to the autocorrelation of the spike train) [92]. Periodic responses were observed for the consonant intervals tested, but not the dissonant intervals, and a measure of periodicity computed from these histograms correlated with consonance ratings in humans.

Despite being qualitatively different mechanistic explanations, both roughness and periodicity provide plausible psychoacoustic correlates of consonance, and there is currently no decisive evidence implicating either account. There are two case reports in which damage to auditory cortex produced abnormal consonance perception [93,94], but such findings are equally compatible with both accounts, as both roughness and periodicity perception may depend on the auditory cortex. The two theories make similar predictions in many conditions – narrowly spaced frequency components yield beats, but also low periodicity, as they are consistent only with an implausibly low F0. However, the theories are readily dissociated in nonmusical stimuli, which, for instance, can vary in roughness while having negligible periodicity, or in periodicity while having negligible roughness. Discussion of these two acoustic factors also neglects the role of learning and enculturation in consonance/pleasantness judgments, which is likely to be important (perhaps even dominant). Further experiments quantitatively comparing predictions of the various acoustic models to data are needed to resolve these issues. Determining the factors underlying consonance could help to determine its origins [95]. If there is an innate component to the consonance preferences commonly observed in Westerners, one might expect it to be rooted in acoustic properties of stimuli with greater ecological relevance, such as distress calls. A better understanding of the acoustic basis of consonance will aid the exploration of such links.

Summary and Concluding Remarks

The processing of pitch combinations is essential to the experience of music. In addition to perceiving the individual pitches of a sequence of notes, we encode and remember the relationships between the pitches. Listeners are particularly sensitive to whether the pitch increases or decreases from one note to the next. The precise interval by which the pitch changes

is important in music, but is not readily perceived with much accuracy for arbitrary stimuli; listeners seem to use particular learned pitch frameworks to encode intervals in musical contexts. There is strong evidence for the importance of temporal lobe structures in these aspects of pitch perception, though the neural mechanisms are not understood in any detail at present. The perception of chords formed from multiple simultaneously presented pitches remains understudied. In some cases it is clear that multiple pitches can be simultaneously represented; in others, listeners appear to perceive an aggregate sound. What makes these sounds consonant or dissonant is still debated, but there are several plausible accounts involving acoustic factors that are extracted by the auditory system. These factors could contribute to aesthetic responses, although enculturation is probably critically important as well.

Other aspects of pitch relations are less clearly linked to auditory mechanisms. The assignment of structural importance to particular pitches in a piece (tonality; Fig. 3c) is widespread in music, and is partly determined by learned templates of interval relations. However, there is also evidence that listeners are sensitive to pitch distributions in pieces from unfamiliar musical systems, tracking the likelihood of different pitches and forming expectations on their basis even when the interval patterns are not familiar [96-98]. The mechanism for tracking these probabilities could plausibly be supra-modal, monitoring statistical tendencies for any sort of stimulus, pitch-based or not. It would be interesting to test whether similar expectations could be set up with stimuli that vary along other dimensions, implicating such a supra-modal process. Another common feature of pitch structuring in music is the tendency to use scales with unequal steps (see Fig. 3a for examples). It has been proposed that this may serve to uniquely identify particular pitch sets [99], but also that unequal step sizes might be more accurately encoded by the auditory system [100]. There are thus open questions regarding how much of musical pitch structure is constrained by audition per se; these represent interesting lines of future research.

Annotations

Ref 3. An up-to-date survey of all aspects of pitch research.

Ref 11. A recent study demonstrating sensitivity to relative pitch in infants.

Ref 14. This study suggests that contours are not unique to pitch, but rather are perceived in other dimensions of sound as well.

Ref 24. A review of past and current research on tonality in music.

Ref 39. A functional imaging study reporting that the right, but not the left, auditory cortex is sensitive to pitch changes on the order of those used in music.

Ref 50. This study provides a striking example of the perceptual differences that are routinely observed in musicians compared to non-musicians.

Ref 54. A psychoacoustics study providing evidence for frequency change detectors in the auditory system.

Ref 59. This study provides the first report of neurons in auditory cortex tuned to the direction of discrete pitch changes, albeit in animals extensively trained to discriminate pitch direction.

Ref 71. Authoritative treatment of all things related to music and language.

Ref 75. The latest in a series of studies using a clever electrophysiological technique to demonstrate that the auditory system represents multiple simultaneous pitches as distinct entities.

Acknowledgements

The authors thank Matt Woolhouse for supplying Figure 3c, and Laurent Demany and Lauren Stewart for helpful comments on the manuscript. This work was supported by NIH grant R01 DC 05216.

References

1. Risset, JC.; Wessel, DL. Exploration of timbre by analysis and synthesis. In: Deutsch, D., editor. *The Psychology of Music*. Academic Press; 1999. p. 113-169.
2. London, J. *Hearing in time: Psychological aspects of musical meter*. Oxford: Oxford University Press; 2004.
3. Plack, CJ.; Oxenham, AJ.; Fay, RR.; Popper, AN. *Pitch: Neural Coding and Perception*. Fay, RR.; Popper, AN., editors. New York: Springer; 2005.
4. Kaas JH, Hackett TA, Tramo MJ. Auditory processing in primate cerebral cortex. *Current Opinion in Neurobiology* 1999;9:164–170. [PubMed: 10322185]
5. Joris PX, Schreiner CE, Rees A. Neural processing of amplitude-modulated sounds. *Physiological Reviews* 2004;84:541–577. [PubMed: 15044682]
6. Heinz MG, Colburn HS, Carney LH. Evaluating auditory performance limits: I. One-parameter discrimination using a computational model for the auditory nerve. *Neural Computation* 2001;13:2273–2316. [PubMed: 11570999]
7. Lu T, Liang L, Wang X. Temporal and rate representations of time-varying signals in the auditory cortex of awake primates. *Nature Neuroscience* 2001;4:1131–1138.
8. Bendor D, Wang X. Cortical representations of pitch in monkeys and humans. *Current Opinion in Neurobiology* 2006;16:391–399. [PubMed: 16842992]
9. Attneave F, Olson RK. Pitch as a medium: A new approach to psychophysical scaling. *American Journal of Psychology* 1971;84:147–166. [PubMed: 5566581]
10. Ruggero MA. Responses to sound of the basilar membrane of the mammalian cochlea. *Current Opinion in Neurobiology* 1992;2:449–456. [PubMed: 1525542]
11. Plantinga J, Trainor LJ. Memory for melody: Infants use a relative pitch code. *Cognition* 2005;98:1–11. [PubMed: 16297673]
12. Dowlow WJ, Fujitani DS. Contour, interval, and pitch recognition in memory for melodies. *Journal of the Acoustical Society of America* 1970;49:524–531. [PubMed: 5541747]
13. Edworthy J. Interval and contour in melody processing. *Music Perception* 1985;2:375–388.
14. McDermott JH, Lehr AJ, Oxenham AJ. Is relative pitch specific to pitch? *Psychological Science*. 2008In press
15. Cuddy LL, Cohen AJ. Recognition of transposed melodic sequences. *Quarterly Journal of Experimental Psychology* 1976;28:255–270.
16. Burns EM, Ward WD. Categorical perception - phenomenon or epiphenomenon: evidence from experiments in the perception of melodic musical intervals. *Journal of the Acoustical Society of America* 1978;63:456–468. [PubMed: 670543]
17. Demany L, Semal C. Dichotic fusion of two tones one octave apart: Evidence for internal octave templates. *Journal of the Acoustical Society of America* 1988;83:687–695. [PubMed: 3351127]
18. Demany L, Semal C. Harmonic and melodic octave templates. *Journal of the Acoustical Society of America* 1991;88:2126–2135. [PubMed: 2269728]
19. Demany L, Armand F. The perceptual reality of tone chroma in early infancy. *Journal of the Acoustical Society of America* 1984;76:57–66. [PubMed: 6747112]
20. Demany L, Semal C. Detection of inharmonicity in dichotic pure-tone dyads. *Hearing Research* 1992;61:161–166. [PubMed: 1526889]

21. Schellenberg EG, Trehub S. Frequency ratios and the perception of tone patterns. *Psychonomic Bulletin & Review* 1994;1:191–201.
22. Trainor LJ, McDonald KL, Alain C. Automatic and controlled processing of melodic contour and interval information measured by electrical brain activity. *Journal of Cognitive Neuroscience* 2002;14:430–442. [PubMed: 11970802]
23. Herzog, G. Song: Folk song and the music of folk song. In: Leach, M., editor. *Funk & Wagnall's Standard Dictionary of Folklore, Mythology, and Legend*. 2. Funk & Wagnall's Co; 1949. p. 1032-1050.
24. Krumhansl CL. The cognition of tonality - as we know it today. *Journal of New Music Research* 2004;33:253–268.
25. Lerdahl, F. *Tonal pitch space*. New York: Oxford University Press; 2001.
26. Bigand E, Parncutt R, Lerdahl F. Perception of musical tension in short chord sequences: The influence of harmonic function, sensory dissonance, horizontal motion, and musical training. *Perception & Psychophysics* 1996;58
27. Dowling WJ. Scale and contour: Two components of a theory of memory for melodies. *Psychological Review* 1978;85:341–354.
28. Dowling, WJ. Melodic information processing and its development. In: Deutsch, D., editor. *The Psychology of Music*. Academic Press; 1982.
29. McAdams S, Cunibule JC. Perception of timbre analogies. *Philosophical Transactions of the Royal Society, London, Series B* 1992;336:383–389.
30. Liegeois-Chauvel C, Peretz I, Babai M, Laguitton V, Chauvel P. Contribution of different cortical areas in the temporal lobes to music processing. *Brain* 1998;121:1853–1867. [PubMed: 9798742]
31. Schuppert M, Munte TM, Wieringa BM, Altenmuller E. Receptive amusia: evidence for cross-hemispheric neural networks underlying music processing strategies. *Brain* 2000;123:546–559. [PubMed: 10686177]
32. Peretz I. Auditory atonalia for melodies. *Cognitive Neuropsychology* 1993;10:21–56.
33. Peretz I, Coltheart M. Modularity of music processing. *Nature Neuroscience* 2003;6:688–691.
34. Stewart L, Overath T, Warren JD, Foxton JM, Griffiths TD. fMRI evidence for a cortical hierarchy of pitch pattern processing. *PLoS ONE* 2008;3:e1470. [PubMed: 18231575]
35. Zatorre RJ, Evans AC, Meyer E. Neural mechanisms underlying melodic pitch perception and memory for pitch. *Journal of Neuroscience* 1994;14:1908–1919. [PubMed: 8158246]
36. Zatorre RJ, Belin P. Spectral and temporal processing in human auditory cortex. *Cerebral Cortex* 2001;11:946–953. [PubMed: 11549617]
37. Patterson RD, Uppenkamp S, Johnsrude IS, Griffiths TD. The processing of temporal pitch and melody information in auditory cortex. *Neuron* 2002;36:767–776. [PubMed: 12441063]
38. Overath T, Cusack R, Kumar S, von Kriegstein K, Warren JD, Grube M, Carlyon RP, Griffiths TD. An information theoretic characterization of auditory encoding. *PLoS Biology* 2007;5:e288. [PubMed: 17958472]
39. Hyde KL, Peretz I, Zatorre RJ. Evidence for the role of the right auditory cortex in fine pitch resolution. *Neuropsychologia* 2008;46:632–639. [PubMed: 17959204]
40. Penagos H, Melcher JR, Oxenham AJ. A neural representation of pitch salience in nonprimary human auditory cortex revealed with functional magnetic resonance imaging. *Journal of Neuroscience* 2004;24:6810–6815. [PubMed: 15282286]
41. Zatorre RJ. Pitch perception of complex tones and human temporal-lobe function. *Journal of the Acoustical Society of America* 1988;84:566–572. [PubMed: 3170948]
42. Bendor D, Wang X. The neuronal representation of pitch in primate auditory cortex. *Nature* 2005;426:1161–1165. [PubMed: 16121182]
43. Johnsrude IS, Penhune VB, Zatorre RJ. Functional specificity in the right human auditory cortex for perceiving pitch direction. *Brain* 2000;123:155–163. [PubMed: 10611129]
44. Tramo MJ, Shah GD, Braida LD. Functional role of auditory cortex in frequency processing and pitch perception. *Journal of Neurophysiology* 2002;87:122–139. [PubMed: 11784735]
45. Tramo MJ, Cariani PA, Koh CK, Makris N, Braida LD. Neurophysiology and neuroanatomy of pitch perception: auditory cortex. *Annals of the New York Academy of Science* 2005;1060:148–174.

46. Peretz I, Ayotte J, Zatorre RJ, Mehler J, Ahad P, Penhune VB, Jutras B. Congenital amusia: a disorder of fine-grained pitch discrimination. *Neuron* 2002;33:185–191. [PubMed: 11804567]
47. Foxtton JM, Dean JL, Gee R, Peretz I, Griffiths TD. Characterization of deficits in pitch perception underlying ‘tone deafness’. *Brain* 2004;127:801–810. [PubMed: 14985262]
48. Stewart L, von Kriegstein K, Warren JD, Griffiths TD. Music and the brain: disorders of musical listening. *Brain* 2006;129:2533–2553. [PubMed: 16845129]
49. Stewart L. Fractionating the musical mind: insights from congenital amusia. *Current Opinion in Neurobiology*. 2008submitted
50. Micheyl C, Delhommeau K, Perrot X, Oxenham AJ. Influence of musical and psychoacoustical training on pitch discrimination. *Hearing Research* 2006;219:36–47. [PubMed: 16839723]
51. Oxenham AJ, Fligor BJ, Mason CR, Kidd G. Informational masking and musical training. *Journal of the Acoustical Society of America* 2003;114:1543–1549. [PubMed: 14514207]
52. Gaser C, Schlaug G. Brain structures differ between musicians and nonmusicians. *Journal of Neuroscience* 2003;23:9240–9245. [PubMed: 14534258]
53. Wong PCM, Skoe E, Russo NM, Dees T, Kraus N. Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nature Neuroscience* 2007;10:420–422.
54. Demany L, Ramos C. On the binding of successive sounds: Perceiving shifts in nonperceived pitches. *Journal of the Acoustical Society of America* 2005;117:833–841. [PubMed: 15759703]
55. Tian B, Rauschecker JP. Processing of frequency-modulated sounds in the lateral auditory belt cortex of the rhesus monkey. *Journal of Neurophysiology* 2004;92:2993–3013. [PubMed: 15486426]
56. Brosch M, Schreiner CE. Sequence selectivity of neurons in cat primary auditory cortex. *Cerebral Cortex* 2000;10:1155–1167. [PubMed: 11073865]
57. Brosch M, Scheich H. Tone-sequence analysis in the auditory cortex of awake macaque monkeys. *Experimental Brain Research* 2008;184:349–361.
58. Weinberger NM, McKenna TM. Sensitivity of single neurons in auditory cortex to contour: Toward a neurophysiology of music perception. *Music Perception* 1988;5:355–390.
59. Selezneva E, Scheich H, Brosch M. Dual time scales for categorical decision making in auditory cortex. *Current Biology* 2006;16:2428–2433. [PubMed: 17174917]
60. D’Amato MR. A search for tonal pattern perception in cebus monkeys: Why monkeys can’t hum a tune. *Music Perception* 1988;5:453–480.
61. Hulse SH, Cynx J, Humpal J. Absolute and relative pitch discrimination in serial pitch perception by birds. *Journal of Experimental Psychology: General* 1984;113:38–54.
62. Brosch, M.; Scheich, H. Neural representation of sound patterns in the auditory cortex of monkeys. In: Ghazanfar, AA., editor. *Primate Audition: Behavior and Neurobiology*. CRC Press; 2002.
63. Zatorre RJ. Absolute pitch: A model for understanding the influence of genes and development on neural and cognitive function. *Nature Neuroscience* 2003;6:692–695.
64. Levitin DJ, Rogers SE. Absolute pitch: perception, coding, and controversies. *Trends in Cognitive Sciences* 2005;9:26–33. [PubMed: 15639438]
65. Miller GA. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review* 1956;101:343–352. [PubMed: 8022966]
66. Levitin DJ. Absolute memory for musical pitch: Evidence from the production of learned melodies. *Perception and Psychophysics* 1994;56:414–423. [PubMed: 7984397]
67. Schellenberg E, Trehub SE. Good pitch memory is widespread. *Psychological Science* 2003;14:262–266. [PubMed: 12741751]
68. Deutsch D. Tones and numbers: Specificity of interference in immediate memory. *Science* 1970;168:1604–1605. [PubMed: 5420547]
69. Clement S, Demany L, Semal C. Memory for pitch versus memory for loudness. *Journal of the Acoustical Society of America* 1999;106:2805–2811. [PubMed: 10573896]
70. Jordania, J. *Who Asked the First Question? The Origins of Human Choral Singing, Intelligence, Language, and Speech*. Tbilisi: Logos; 2006.
71. Patel, AD. *Music, Language, and the Brain*. Oxford: Oxford University Press; 2008.
72. Bregman, AS. *Auditory Scene Analysis: The Perceptual Organization of Sound*. Cambridge, MA: MIT Press; 1990.

73. Fujioka T, Trainor LJ, Ross B, Kakigi R, Pantev C. Automatic encoding of polyphonic melodies in musicians and nonmusicians. *Journal of Cognitive Neuroscience* 2005;17:1578–1592. [PubMed: 16269098]
74. Huron D. Voice denumerability in polyphonic music of homogeneous timbres. *Music Perception* 1989;6:361–382.
75. Fujioka T, Trainor LJ, Ross B. Simultaneous pitches are encoded separately in auditory cortex: an MMNm study. *Neuroreport* 2008;19:361–366. [PubMed: 18303582]
76. Cook ND, Fujisawa TX. The psychophysics of harmony perception: Harmony is a three-tone phenomenon. *Empirical Musicology Review* 2006;1:106–124.
77. Butler JW, Daston PG. Musical consonance as musical preference: A cross-cultural study. *Journal of General Psychology* 1968;79:129–142. [PubMed: 5672277]
78. Terhardt E. Pitch, consonance, and harmony. *Journal of the Acoustical Society of America* 1974;55:1061–1069. [PubMed: 4833699]
79. Trainor LJ, Tsang CD, Cheung VHW. Preference for sensory consonance in 2- and 4-month-old infants. *Music Perception* 2002;20:187–194.
80. Trainor LJ, Heinmiller BM. The development of evaluative responses to music: Infants prefer to listen to consonance over dissonance. *Infant Behavior and Development* 1998;21:77–88.
81. Zentner MR, Kagan J. Perception of music by infants. *Nature* 1996;383:29. [PubMed: 8779711]
82. Maher TF. “Need for resolution” ratings for harmonic musical intervals: A comparison between Indians and Canadians. *Journal of Cross Cultural Psychology* 1976;7:259–276.
83. McDermott J, Hauser M. Are consonant intervals music to their ears? Spontaneous acoustic preferences in a nonhuman primate. *Cognition* 2004;94:B11–B21. [PubMed: 15582619]
84. Vassilakis P. Auditory roughness as a means of musical expression. *Selected Reports in Ethnomusicology* 2005;12:119–144.
85. Plomp R, Levelt WJM. Tonal consonance and critical bandwidth. *Journal of the Acoustical Society of America* 1965;38:548–560. [PubMed: 5831012]
86. Hutchinson W, Knopoff L. The acoustical component of western consonance. *Interface* 1978;7:1–29.
87. Kameoka A, Kuriyagawa M. Consonance theory. *Journal of the Acoustical Society of America* 1969;45:1451–1469. [PubMed: 5803168]
88. Fishman YI, Reser DH, Arezzo JC, Steinschneider M. Complex tone processing in primary auditory cortex of the awake monkey. I. Neural ensemble correlates of roughness. *Journal of the Acoustical Society of America* 2000;108:235–246. [PubMed: 10923888]
89. Fishman YI, Volkov IO, Noh MD, Garell PC, Bakken H, Arezzo JC, Howard MA, Steinschneider M. Consonance and dissonance of musical chords: neural correlates in auditory cortex of monkeys and humans. *Journal of Neurophysiology* 2001;86:2761–2788. [PubMed: 11731536]
90. Rameau, JP. *Treatise on Harmony*. New York: Dover Publications, Inc; 1722/1971.
91. Pressnitzer D, Patterson RD, Krumholz K. The lower limit of melodic pitch. *Journal of the Acoustical Society of America* 2001;109:2074–2084. [PubMed: 11386559]
92. Tramo MJ, Carians PA, Delgutte B, Braida LD. Neurobiological foundations for the theory of harmony in Western tonal music. *Annals of the New York Academy of Science* 2001;930:92–116.
93. Tramo MJ, Bharucha JJ, Musiek FE. Music perception and cognition following bilateral lesions of auditory cortex. *Journal of Cognitive Neuroscience* 1990;2:195–212.
94. Peretz I, Blood AJ, Penhune VB, Zatorre RJ. Cortical deafness to dissonance. *Brain* 2001;124:928–940. [PubMed: 11335695]
95. McDermott J. The evolution of music. *Nature* 2008;453:287–288. [PubMed: 18480798]
96. Castellano MA, Bharucha JJ, Krumhansl CL. Tonal hierarchies in the music of North India. *Journal of Experimental Psychology: General* 1984;113:394–412. [PubMed: 6237169]
97. Krumhansl CL, Louhivuori J, Toiviainen P, Jarvinen T, Eerola T. Melodic expectancy in Finnish folk hymns: Convergence of behavioral, statistical, and computational approaches. *Music Perception* 1999;17:151–196.

98. Krumhansl CL, Toivanen P, Eerola T, Toivianen P, Jarvinen T, Louhivuori J. Cross-cultural music cognition: cognitive methodology applied to North Sami yoiks. *Cognition* 2000;76:13–58. [PubMed: 10822042]
99. Balzano, GJ. The pitch set as a level of description for studying musical pitch perception. In: Clynes, M., editor. *Music, mind and brain: the neuropsychology of music*. Plenum; 1982. p. 321-351.
100. Trehub SE, Schellenberg E, Kamenetsky SB. Infants' and adults' perception of scale structure. *Journal of Experimental Psychology: Human Perception and Performance* 1999;25:965–975. [PubMed: 10464941]

Glossary

Pure tone

a tone with a sinusoidal waveform, consisting of a single frequency

Complex tone

any periodic tone whose waveform is not sinusoidal, consisting of multiple discrete frequencies

Harmonic

a pure tone whose frequency is an integer multiple of another frequency

F₀

fundamental frequency. This is defined as the inverse of the period of a periodic sound, or equivalently as the greatest common factor of a set of harmonically related frequencies.

Octave

a frequency interval corresponding to a doubling in frequency

Chromatic scale

the scale from which the notes of Western music are drawn; the octave is divided into 12 steps that are equally spaced on a logarithmic scale (Fig. 3a).

Semitone

the basic unit with which frequency is measured in musical contexts; a twelfth of an octave. It is the spacing between adjacent notes on the chromatic scale.

Timbre

those aspects of a sound not encompassed by its loudness or its pitch, including the perceptual effect of the shape of the spectrum, and of the temporal envelope.

Roughness

the perceptual correlate of amplitude modulation at moderate frequencies, e.g. between 20 and 200 Hz. Proposed to be a cause of sensory dissonance.

Contour

the sequence of up/down changes between adjacent notes of a melody.

Interval

the pitch difference between two notes.

Chord

a combination of two or more notes played simultaneously.

Polyphony

music that combines multiple melodic lines intended to retain their identity rather than fusing into a single voice.

Homophony

music in which two or more parts move together, their relationships creating different chords.

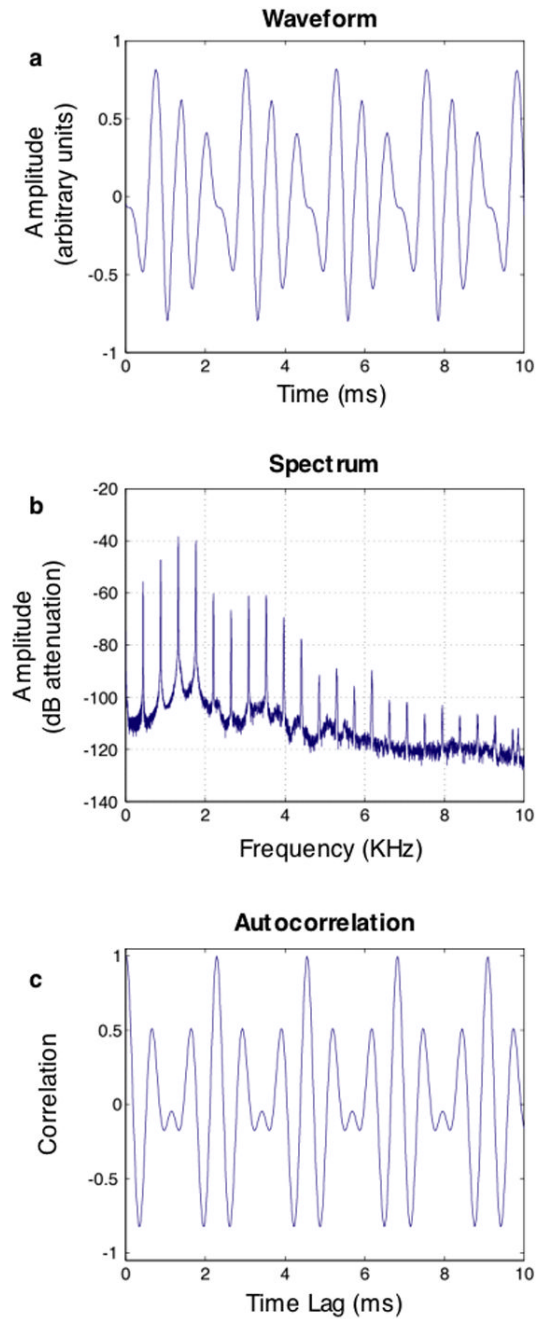


Figure 1. Waveform, spectrum, and autocorrelation function for a note (the A above middle C, with an F0 of 440 Hz) played on an oboe

a) Excerpt of waveform. Note that the waveform repeats every 2.27 ms, which is the period. b) Spectrum. Note the peaks at integer multiples of the F0, characteristic of a periodic sound. In this case the F0 is physically present, but the second, third, and fourth harmonics actually have higher amplitudes. c) Autocorrelation. The correlation coefficient is always 1 at a lag of 0 ms, but because the waveform is periodic, correlations close to 1 are also found at integer multiples of the period (2.27, 4.55, 6.82, and 9.09 ms in this example). Note that whereas evaluating periodicity in the waveform requires examining every point, the autocorrelation

makes periodicity more explicit – it is signified by high correlations at regularly spaced intervals.

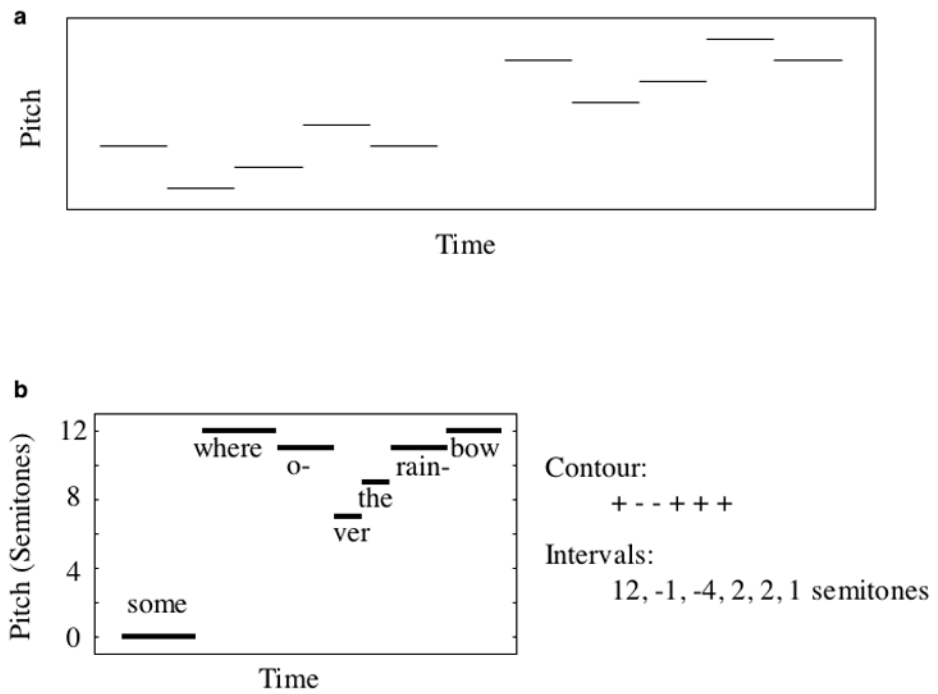


Figure 2. Relative pitch

a) Transposition. Figure depicts two five-note melodies. The second melody is shifted upwards in pitch. In this case both the contour and the intervals are preserved. b) Contour and intervals of “Somewhere Over the Rainbow”.

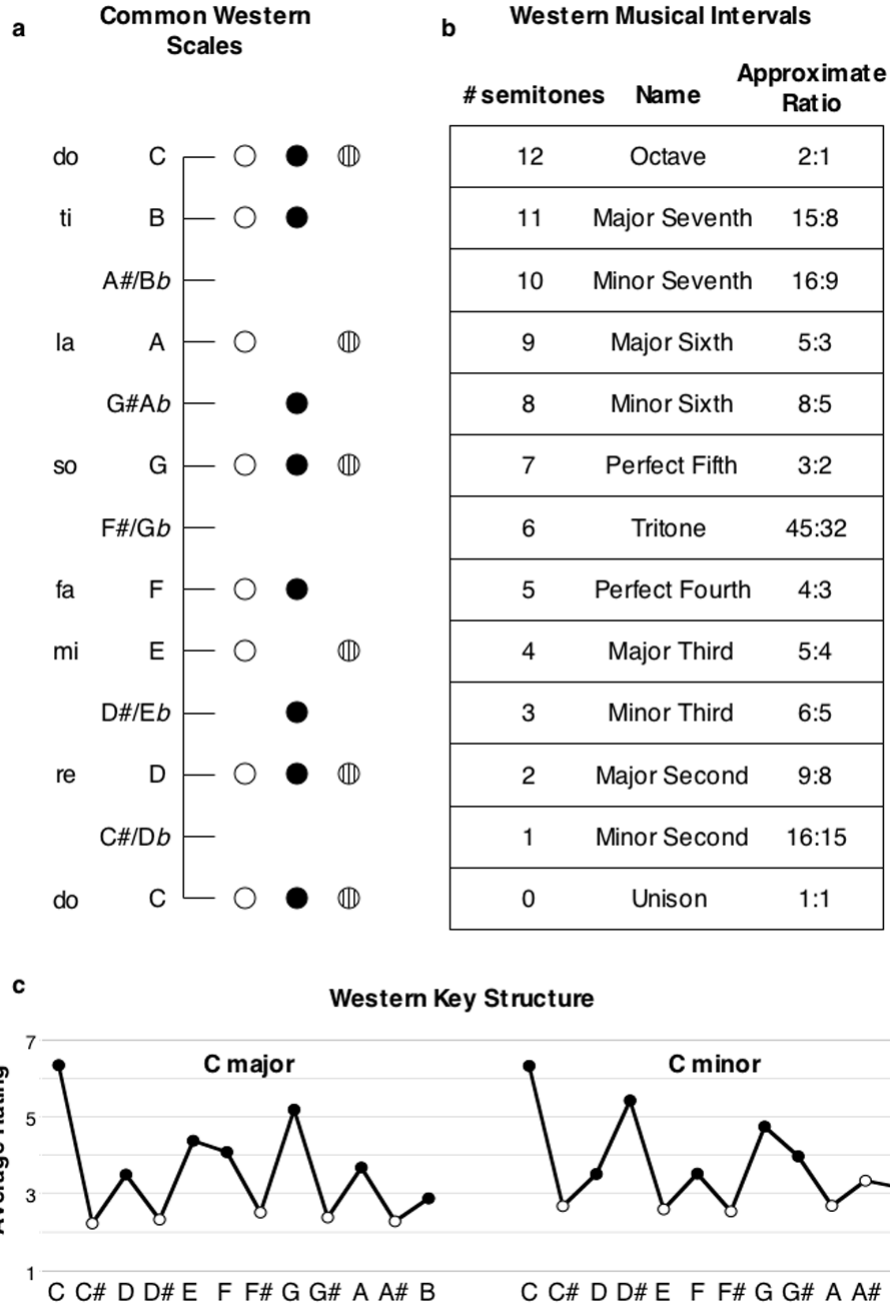


Figure 3. Scale and interval structure

a) Scales. In Western music the octave is divided up into 12 semitones, equally spaced on a logarithmic frequency scale. All musical scales are drawn from this set of notes, which are given letter names. Three common scales are depicted, in the key of C: the major scale (outline circles); the harmonic minor scale (solid circles); a pentatonic scale (circles with lines). The solfege note names for the major scale are also included on the left. Other musical systems use different scales, but usually share some properties of Western scales - notes are typically not separated by less than a semitone, and the steps separating notes are usually not all equal. b) Musical intervals derived from the chromatic scale. In the equal-tempered scale that is common nowadays, the ratios in the rightmost column are only approximate. c) Tonal hierarchy. The

figure plots results of probe-tone experiments from Krumhansl and colleagues. Listeners were played a melody in the key of C major or C minor, after which they were asked to rate the appropriateness of different probe tones. Notes outside the scale (white circles) are rated as inappropriate. The tonic (C) is most appropriate, followed by the other notes of the major triad (E and G for the major key; D# and G for the minor key), followed by the other notes of the key. The particular key structure shown here (i.e. the pattern of intervals between the notes) is specific to Western music, but qualitatively similar structures appear to be used and perceived in many other musical systems as well.

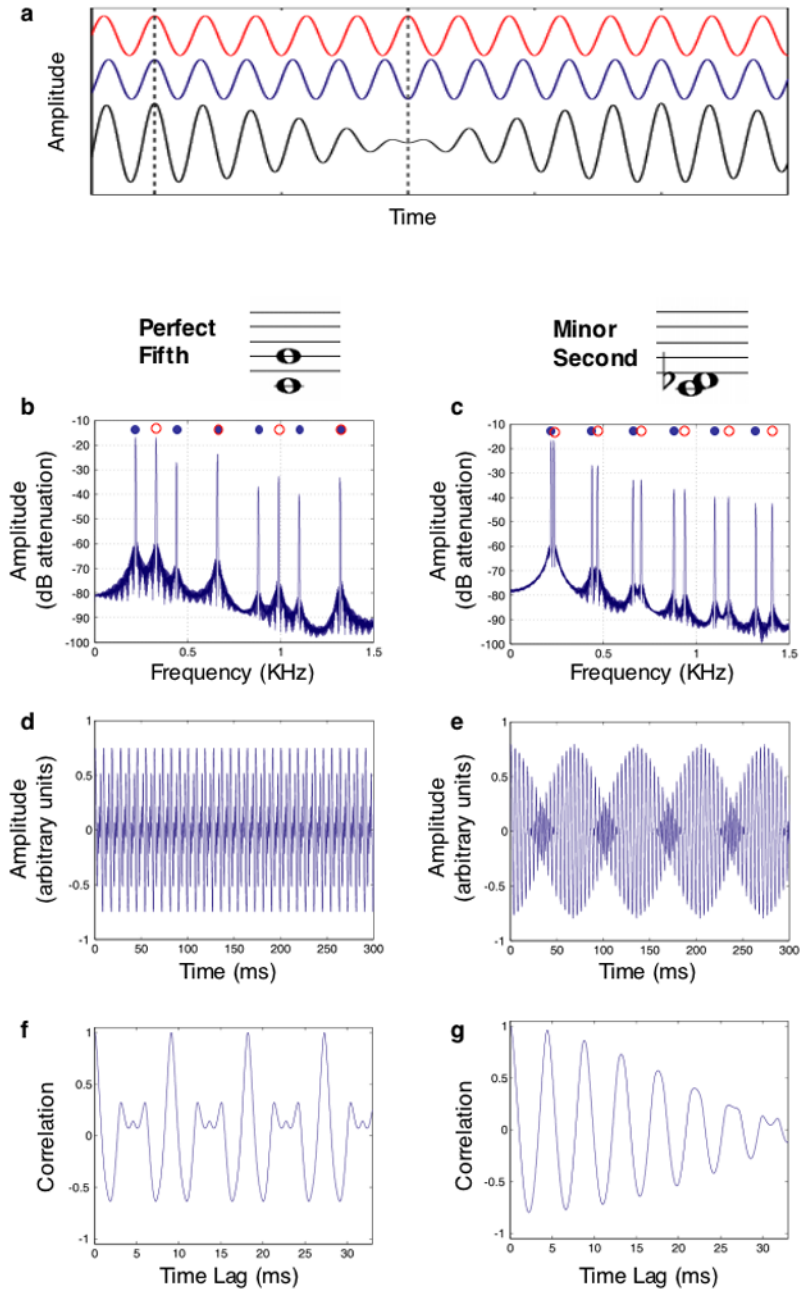


Figure 4. Acoustics of consonance and dissonance

a) Beating. The red and blue waveforms differ in frequency by 10%, and so shift in and out of phase over time. The first dashed line denotes a point where they are in phase; the second line denotes a point where they are out of phase. The black waveform is the sum of the red and blue, and exhibits beats – amplitude fluctuations that result from the changing phase relations of the two components. b) Spectrum for a perfect fifth harmonic interval (a consonant interval in Western music). Blue and red circles indicate the frequency components of the two notes of each interval. Note that for the perfect fifth, the components of the two notes either exactly coincide, or are far apart. c) Spectrum for a minor second harmonic interval, considered dissonant in Western music. Note that the frequency components tend to be close in frequency but not exactly overlapping. d) Waveform of perfect fifth. e) Waveform of minor second. Note

amplitude modulation, produced by pairs of components that beat. f) Autocorrelation of waveform of perfect fifth. Note the regular peaks near 1 – the signature of periodicity. g) Autocorrelation of waveform of minor second. Note the absence of regular peaks.