Music and the Brain

Psyche Loui
Department of Music, Center for Brain and Cognitive Health
Northeastern University
360 Huntington Ave, Boston, MA 02115
p.loui@northeastern.edu

Alexander Belden
Department of Biology, Center for Brain and Cognitive Health
Northeastern University
360 Huntington Ave, Boston, MA 02115
belden.a@husky.neu.edu

Teaser: Activity and connectivity throughout the human brain enable the complex experience of music

Running Title: Music and the Brain

Key Words: cognitive, neuroscience, neuropsychology, psychology, pitch, rhythm, harmony, melody, timbre, absolute pitch, amusia, tone deafness

Technical Committee(s): Psychological and Physiological Acoustics; Music
In recent years, the cognitive neuroscience of music has captured increasing interest from scientists, scholars, and the public alike. Part of this interest comes from methodological advancements to examine the living human brain. Another source of interest comes from increased awareness of the value of interdisciplinary research. Researchers come from diverse backgrounds, ranging from neurobiology to music education. This brings about a diversity of ideas. Finally, interest comes from the possibility that findings may translate towards better tools for music therapy, something that is being applied to an increasing variety of neurological and psychiatric disorders. Even for the healthy brain, there is a push towards using music to improve mood and cognition both in adulthood and in development.

This article reviews recent advances in the cognitive neuroscience of music, with special attention to the cognitive neuroscience of pitch, rhythm, harmony, and melody. We begin with a brief introduction of the tools in use to examine musical functions in the brain with both spatial and temporal accuracy and precision. This is followed by a brief overview of brain functions that enable perception of musical pitch, timbre, rhythm, harmony, and melody. Finally, we examine the role of expectation and reward as a guiding principle for why humans appreciate music.

**Tools and Principles**

To detect brain activity, the best currently-available methods trade off in spatial resolution, temporal resolution, and invasiveness (Figure 1A). Since music perception and cognition studies primarily require awake and behaving human subjects, the majority of cognitive neuroscience studies in music have used noninvasive methods. These include electroencephalography (EEG), structural and functional Magnetic Resonance Imaging (MRI), magnetoencephalography (MEG), and neuropsychological testing on special populations, as well as noninvasive brain-stimulation methods to perturb normal brain activity reversibly. Here we review methodologies and findings as they directly relate to music. For a more comprehensive overview or for a more general introduction into human neuroanatomy, the reader is directed to more general texts in cognitive neuroscience (e.g. Purves et al., 2008).
Since its discovery in the 1920s, EEG continues to be a relatively low-cost, efficient technique for recording brain activity with high temporal resolution. The technique involves fixing an array of electrodes on the surface of the scalp. The electrodes register changes in local field potentials that come from neural populations that respond to stimulus events. Event-Related Potential (ERP) is an application of the technique that allows researchers to link specific patterns of brain electrical potentials to stimulus events, by repeatedly presenting stimuli of interest while recording EEG, and then averaging the EEG data across the repeated stimulus presentations (Figure 1B).

EEG can also register activity from way stations in the auditory brainstem. This auditory brainstem response (ABR) is particularly accurate at discriminating between different sounds. As it is a stimulus-driven response, the ABR resembles the stimulus itself, and this stimulus-brain resemblance is taken as a neural marker of the fidelity with which the auditory brainstem codes for sounds. Importantly, the fidelity of ABR in encoding sounds is higher in musically trained participants (Kraus et al., 2010). The musicians’ advantage in neural encoding, as indexed with the ABR, has been observed for a variety of sounds including speech as well as music, and in older adults as well as in children.

Although EEG can resolve fine-grained temporal details in brain activity, it is relatively limited in its ability to locate the source of the neural response in the brain, or spatial resolution. Some boost in spatial resolution comes from recording from magnetoencephalography (MEG), which records the magnetic fluctuations that accompany electrical changes in the activity of neural populations. MEG provides the same temporal resolution as EEG, but increases in spatial resolution especially in activity that originates from inwards folds away from the surface of the brain (called sulci), due to geometric constraints of the arrangement of sensors on the scalp. Thus, music researchers that are interested in auditory sources of brain activity are able to map these sources with increased spatiotemporal resolution.

While EEG and MEG offer good temporal resolution, structural and functional MRI offer superior spatial resolution. Functional MRI captures the oxygenation level of blood, as required by neural activity (Figure 1C). Structural MRI includes anatomical and diffusion images, among others. Anatomical
images are effective at comparing relative volume, cortical thickness, and surface area of cortical and subcortical structures in grey matter cell bodies (neurons), as well as identifying any lesions such as due to stroke or traumatic brain injury. Diffusion images are useful for visualizing the white matter pathways, which consist of bundles of axons that connect the neuronal cell bodies in the brain.

**Pitch**

Pitch is a basic building block of music. It is the perceptual attribute of sound that most closely maps on the fundamental frequency (F0). This psychological attribute of pitch ranges from low to high, and two sounds can have the same pitch despite having energy at different frequencies, as long as the F0 is the same. Musical training seems to hone a finer-grained ability to discriminate between small differences in pitch, as classical musicians have frequency discrimination thresholds that are six times smaller than that of non-musicians (Micheyl et al., 2006).

While different harmonics with the same F0 can have different timbres, they nevertheless share the same sensation of pitch. In the most extreme case, a pitch can still be heard even when there is no energy at the F0 altogether, as long as energy is present at multiple harmonics. This phenomenon, known as virtual pitch, has been both a challenge and a test case for neural models of pitch coding. Specifically, a neural mechanism that codes for pitch must yield an output similar to the F0 even when no energy is present at the F0, as long as harmonics of the F0 are present.

Recording from the auditory cortex of marmosets (*Callithrix jacchus*), Bendor and Wang (2005) found that a small population of neurons in the anterolateral border of the primary auditory cortex responded selectively to both the F0 and harmonic complexes of overtones above the F0, thus providing a neural correlate of pitch constancy. Follow-up studies showed that this pitch constancy is likely accomplished using a combination of spectral and temporal cues. By building on these cues, and then recombining pitches hierarchically to form melodies and harmonies, the neural coding of pitch provides the basis for higher-order coding of musical structure in the brain.
Amusia

One way to understand how pitch processing works in the brain is to look into individuals who have impairments in pitch processing ability. Congenital amusia, also known as tone-deafness, is a lifelong musical disorder that prevents individuals from developing skills in pitch perception and production, despite no apparent deficits in speech or hearing, or cognitive ability (Ayotte et al., 2002). It is most commonly identified by using a neuropsychological test known as the Montreal Battery for Evaluation of Amusia (MBEA, Peretz et al., 2003). People with congenital amusia, who could not consciously report the directions of pitched intervals, could nevertheless produce (by singing) pairs of tones with above-chance accuracy in pitch interval direction (Loui et al., 2008). This dissociation between perception and production ability suggests that that there may be multiple paths towards auditory processing.

Studies from multiple neuroimaging methods have shown differences in auditory as well as auditory-motor brain processes that are linked to congenital amusia. Using structural MRI and cortical thickness measures, Hyde et al (2007) observed differences in frontal lobe as well as superior temporal lobe of amusics, specifically in the superior temporal gyrus (STG) and the inferior frontal gyrus (IFG) as shown in Figure 2.

The simultaneous disruption of temporal lobe and frontal lobe regions, specifically the STG and IFG, may suggest further difficulties in memory, learning, or auditory-motor integration of pitch information. A parsimonious explanation for these simultaneously observed deficits was that white matter connectivity between the STG and IFG could be disrupted in congenital amusics, leading to abnormal neuronal development or migration at the endpoints of this connection. This was observed in a diffusion imaging study, in which the arcuate fasciculus, which connects the temporal and frontal regions of the brain, was smaller in volume in people with congenital amusia (Loui et al., 2009).

<Figure 2 about here>
Other studies have employed MEG in combination with structural MRI to provide further support for a right frontotemporal deficit in connectivity (Albouy et al., 2013), while ERP work suggests that congenital amusia may be fundamentally an issue with the lack of awareness of pitch information (Peretz et al., 2009). Taken together, results from congenital amusia provide support for a crucial role of the pathway between frontal and temporal lobes, probably mostly in the right hemisphere, that enables the tight coupling between pitch perception and production.

While congenital amusia refers to the lifelong deficit of musical abilities, acquired amusia refers to the loss of musical ability resulting from brain damage. Lesion analyses show that disconnections in multiple pathways are common among those with acquired amusia. Those with lesions covering multiple white matter pathways are least likely to recover from acquired amusia after a stroke, whereas those who recover from acquired amusia are more likely to have damage in one pathway, while sparing others (Sihvonen et al., 2017). These findings provide insight into possible targets for neurorehabilitation after a stroke or other brain injury. As stroke is a leading cause of long-term disability in older adults, rehabilitating musical functions in those suffering from the aftermath of a stroke will be key to improving quality of life in these affected individuals (Norton et al., 2008).

**Absolute Pitch**

While amusia is a deficit in pitch perception and production ability, Absolute Pitch (AP) seems ostensibly to reflect the opposite. People with AP have the ability to identify the pitch class of musical notes without an external reference (Ward, 1999). In addition to being more common in musically-trained individuals, especially those who started musical training before the age of seven, AP runs in families and is more common among those with East Asian descent. It is especially common in those of East Asian heritage who speak tone languages fluently, suggesting that the ability is associated with early language experience (Deutsch et al., 2009). Even among people with AP, there is a range of pitch identification ability. While some AP possessors are able to name any note in any timbre, other individuals have AP only for the instruments they play (Miyazaki, 1989). Due to these intriguing interactions between genetic
and environmental factors, AP is an ideal model for understanding the influences of genes and the environment.

The neural substrates that enable automatic pitch categorization likely come from the planum temporale, a region within the aforementioned STG (Schlaug et al., 1995), which is exceptionally larger in the left hemisphere of AP musicians, presenting as a more leftwardly-asymmetric brain in MRIs (Schlaug et al., 1995). In addition to being larger in volume, the left STG is also better connected in AP musicians relative to their non-AP counterparts, and pitch categorization accuracy is correlated with white matter volume of connections identified from STG (Loui et al., 2011). Functional MRI results point to a distributed network of enhanced activity throughout the brain in AP musicians, albeit with results centering around the left STG (Loui et al., 2012). Thus, it appears that both specific brain structure and general network-level brain functioning are special in AP possessors.

While AP is presumed to be rare, occurring in less than 1% of the general population (Ward, 1999), most listeners possess some absolute memory for pitch, as shown by being able to produce familiar songs at the right starting pitch after repeated listening (Levitin et al., 1994). However, the enhanced categorization ability seems relatively rare, and specific to a unique population. This has led some researchers to ask whether the AP possessors’ tendency to categorize pitch might be thought of as a savant-like ability as seen in some individuals with autism (Mottron et al., 2013). In that regard, musicians with and without AP were tested on subclinical traits of autism using the autism spectrum quotient (Dohn et al., 2012). Results showed that while AP possessors scored higher than non-AP counterparts in the autism quotient, they were still well lower than would meet criterion for autism spectrum disorders. Crucially, AP possessors showed some higher scores in imagination but no differences from controls in social and communicative subscales. Taken together, AP could be considered an enhanced perceptual categorization ability, likely subserved by a network of regions centering around the structurally altered superior temporal lobe. Whether this brain network and its supported functions can be trained in the lab, or in the practice room, remains an active area of both psychological and pedagogical research.
Timbre

While the sensation of pitch is correlated with the F0 of periodic sounds, the sensation of timbre is an emergent property of spectral and temporal characteristics of sounds. The temporal envelope of a sound, especially the time between its onset and its peak amplitude (“attack time”), is a strong determinant of the perceptual attribute of “bite.” While attack time is a feature of the temporal envelope, spectral centroid is a feature of the spectral envelope and is computed as the weighted average of the frequency of all harmonics present, giving rise to its “brightness.” Classic studies have found that spectral centroid and attack time are two orthogonal dimensions that account for much of the variance in judgments in sound quality (Wessel, 1979). The third and most salient dimension is both spectral and temporal in nature is spectral flux, which is the change in spectral centroid over time (McAdams, 2013). Because these dimensions of sound are clearly defined and orthogonal to each other, both perceptually and in terms of their underlying acoustic attributes, they are useful both for musical expression and for creating well-controlled experimental stimuli in order to understand how the brain processes timbre.

An ERP study compared the three dimensions of timbre (attack time, spectral centroid, and spectral flux) directly using the mismatch negativity (MMN), a negative ERP that occurs around 200 ms after any small deviations from the sound context. The MMNs elicited by the different dimensions were indeed separate in their underlying neural sources, suggesting separate dimensions of processing in auditory memory and providing convergent results with the psychophysical data (Caclin et al., 2006). An fMRI experiment showed that modulations in spectral shape elicited right-lateralized brain activity in the superior temporal sulcus (STS), an area immediately below the STG (Warren et al., 2005). This right-lateralized STS activity supports the idea that right hemisphere is relatively tuned to spectral changes, whereas the left hemisphere is more tuned to fine-grained temporal structure (Flinker et al., 2019).

Since timbre is an important cue towards the perception of auditory objects, the study of “dystimbria,” or the impairment of spectral and temporal analysis without loudness or pitch processing difficulty (Griffiths et al., 2007), can provide a window into how the brain perceives objects from the
world of sound. On the other hand, people with exceptionally fine-grained training in listening to timbres, such as piano tuners, have enhanced grey matter in the superior temporal structures as well as the hippocampal complex, which is crucial for learning and memory (Teki et al., 2012). These grey matter differences are correlated with the duration of one’s career in piano tuning, and not with actual age, suggesting that there are adaptations in brain structure that come with advanced musical experience, even within the specific task of listening for timbre in order to tune a musical instrument.

Rhythm

Beat

Beat is the basic building block of musical rhythm which allows for the synchronization of musical events. Beat not only synchronizes musical events to one another, but also synchronizes our brains to the rhythm of musical stimuli. When presented with a beat, EEG recordings show rhythmic activity at the beat frequency (Nozarandan et al., 2011). This neural sense of beat is produced through the coupled oscillation of auditory and motor pathways, including motor areas of the brain (Grahn and Brett, 2007), which could be responsible for the ability to predict a pulse even in situations where there is no spectral energy produced on the beat itself (Tal et al., 2017). The motor system contributions to beat perception are particularly apparent in classical musicians and in patients with Parkinson’s disease. Parkinson’s patients, who have multiple motor deficits, are impaired at discriminating beat-based rhythms, but not rhythms that lack an underlying beat (Grahn and Brett, 2009). This suggests an important role of the motor system in the detection and generation of internal beats.

The coupling between auditory and motor systems when confronted with beat-based stimuli might explain why we feel “groove,” which is defined as the pleasurable urge to move to music. Music that is rated as high in groove elicits larger motor-evoked potentials (electrical fluctuations measured from motor neurons in the hand and arm) in musicians than low-groove music or noise (Stupacher et al., 2013). One feature that determines the sensation of groove is syncopation, which can be defined as a slight
violation of an expected and beat-based rhythm. However, this is not to say that higher syncopation necessarily means a higher sense of groove. In fact, an inverse U-shaped relationship between groove and syncopation has been observed, with medium-syncopated music tending to elicit a stronger desire to move than either high- or low-syncopated music (Witek et al., 2014).

In addition to being pleasurable, it is well known that rhythm can tune large populations of people, motivating them to movement. Recent research is addressing whether synchronous, rhythmic movement can affect social behavior. In a sample of 14-month old infants, parents holding the infant faced an experimenter and moved either synchronously (on-beat) or asynchronously (off-beat) from one another. Infants proceeded to be more helpful towards the experimenter after the synchronous condition as compared to the asynchronous condition (Cirelli et al., 2014). This remarkable finding demonstrates the social value of synchronizing to a beat, even during infancy.

**Meter**

Meter refers to the hierarchical organization of beats into recurring groups. Meter usually consists of groupings of two, three, or four beats, though in some cases rarer meters that contain five, seven, or even larger or variable sets of beats may occur. Regardless of the number of beats, there is a tendency to have an accent on the first beat of a given measure (often with secondary accents later on for meters containing four or more beats). This tendency in humans to accentuate metrical beats is so strong that EEG recordings have shown an enhancement of the beat frequency associated with the “first” beat of an imagined meter, even with an unaccented isochronous beat stimulus (Nozaradan et al., 2011).

**Harmony and Melody**

While melodies are pitches presented sequentially over time, harmony refers to the simultaneous presentation of pitches over time. Music theorists and scientists alike have attempted to define a space that represents how we conceptualize tonal harmony. Insight came from human subjective ratings from
the probe tone experiment, in which a melodic context is presented, followed by a tone, and subjects’ task is to rate how well the final tone fit the preceding melody. These probe tone profiles matched the relative importance of pitches within a tonal context as dictated by principles of music theory. Applying dimensionality-reduction algorithms on these probe tone data yields the empirically derived tonal space. Based on mathematical modeling of the empirical data, Krumhansl and colleagues (1990) found that the best geometric solution of tonal space is in the shape of a torus (Figure 3). The toroidal representation is effective at capturing the close relative distances between neighboring keys within the circle of fifths from music theory. It also captures the further distance between parallel minors than between relative minors, as is observed in empirical ratings data from the probe tone paradigm.

<Figure 3 about here>

Given that the torus describes our mental representation of Western tonal music, it was possible to compose continuous modulating melodies and harmonies that smoothly navigate the surface of the torus. Janata et al. (2002) traced brain activity in an fMRI study as participants listened to these continuously modulating melodies. Results showed that a region of the frontal lobe, the ventromedial prefrontal cortex (vmPFC), was consistently responsive to modulating melodies: crucially, contiguous voxels in the vmPFC were active as the melody changed keys to contiguous parts of the torus, suggesting that the brain was tracking tonal movement in these regions.

While the continuous perception of harmony is important, the violation of harmonic expectation has also lent insight into how the brain processes harmony. In response to unexpected chords within a chord progression, an Early Right Anterior Negativity (ERAN), which is a negative waveform peaking at approximately 200 ms after the onset of the unexpected chord (Koelsch et al., 2000). The ERAN indexes our expectation that music follows a known syntactic structure. Interestingly, even new music that we learn rapidly within the course of an hour can elicit the ERAN, suggesting that the neural generators of the ERAN can flexibly and rapidly learn to integrate new sounds and sound patterns given their statistical context in the environment (Loui et al., 2009).
The brain structures that generate the ERAN are in the left inferior frontal gyrus (IFG), which is sensitive to linguistic syntax (Levitin and Menon, 2003). Patients with IFG lesions show behavioral deficits in processing musical structure that are coupled with diminished or altered ERAN (Sammler et al., 2011). These results suggest that areas of the brain that used to be thought of as language-specific regions, such as the left IFG, are in fact processing syntactic structure in music as well. This lends credence to the idea that music and language processing interact in the brain specifically for the processing of syntactic structure, as articulated by Patel’s Shared Syntactic Integration Resource Hypothesis (SSIRH) (Patel, 2010).

**Prediction and Reward**

The musical features reviewed above are acoustic devices that ultimately provide the groundwork for us to make predictions about events in the immediate future. We expect sounds to occur on or at even subdivisions of the beat, and accents to occur at specific beats within a given meter; we also expect certain pitch intervals and harmonies to follow others. The systematic fulfillment and violations of expectations have long been posited to contribute to emotional content and perceived meaning in music (Meyer, 1956). For example, while beat and meter establish rhythmic expectations, syncopation is a violation of expectations that can lead to an even richer music listening experience. This balance between predictability and prediction violations is critical to the enjoyment of music. Through prolonged exposure, humans acquire knowledge of, and predictions for, musical features that are common in our culture.

While our emotional responses to music can be complex, multidimensional, and context-dependent, recent insights show that linking music to brain’s the reward system can be fruitful in revealing why humans appreciate music. Although music is a relatively abstract interplay of sound patterns, it can trigger the release of dopamine, a neurochemical that is released during the experience of rewards such as food, sex, and recreational drugs (Ferreri et al., 2019). Listening to music that we value activates the striatum, a dopamine-rich set of brain regions, and the striatum is correlated in activity with
the auditory cortex (Salimpoor et al., 2013). This functional coupling between the auditory cortex and reward-sensitive regions is an especially compelling account of why humans love music. In effect, music might be considered an auditory channel towards areas of the brain that are sensitive to reward.

Further support for this view comes from the finding that white matter connectivity between the auditory cortex and the emotion and social processing regions of the brain are larger among people who experience “chills” when listening to intensely pleasurable music (Sachs et al., 2016). In contrast, people who are insensitive to the pleasures of music, in a condition known as musical anhedonia, have diminished physiological and neural sensitivity to the rewards of music listening, despite normal reward responses to non-musical reward tasks such as gambling (Martinez-Molina et al., 2016). Although individuals differ in reward sensitivity to music, acquired musical anhedonia is very rare, even among people with focal brain damage (Belfi et al., 2017). This means that in the vast majority of people recovering from neurological and/or psychiatric disorders, such as depression or dementia, music could be used in therapeutic interventions due to its reliable engagement of the reward system, which in turn guides motivated behavior.

Conclusions

The value of the studies reviewed above comes not only from satisfying intellectual curiosity of how music works and why we have music, but also in the hopes that a deeper understanding of the tools and principles of music will enable the design of better musical interventions for a variety of applications, from recovery from neurological and/or psychiatric disorders to enhancing the optimal function of the healthy developing brain.

References


Figure captions

Figure 1. A. Approximate spatial and temporal resolution of neural recording methods. B. EEG methods provide good temporal but limited spatial resolution. Top: Early Right Anterior Negativity, from Loui et al (2005). Bottom: Birds-eye view of the EEG recording setup. C. fMRI methods provide good spatial but lower temporal resolution. Top: Brain activity in auditory cortices, as observed during music listening, from Loui et al., 2012. Bottom: MRI setup.

Figure 2. Model brain with brain regions that correspond to the regions discussed in this article. Colored locations are approximate landmarks in the brain. Red = superior temporal gyrus (STG), yellow = inferior frontal gyrus (IFG), green = ventromedial prefrontal cortex (vmPFC), orange = nucleus accumbens. Blue = arcuate fasciculus.

Figure 3. The torus is a good approximation of our mental representation of Western music. Left: two-dimensional tonal space. The letters denote musical keys, with upper-case letters denoting major keys and lower-case letters denoting minor keys. A chord progression in A major, shown in this example, elicits activity near the A major region, while suppressing activity near its dissimilar keys such as E-flat major and d-sharp minor. Right: three-dimensional tonal space, in the shape of a torus. This can be derived by wrapping the two-dimensional tonal space in the left-to-right direction as indicated by the black arrows, and then wrapping the resulting tube again in a circular direction as indicated by the orange arrow.