

UC Irvine

UC Irvine Electronic Theses and Dissertations

Title

The Neuroanatomical and Evolutionary Basis of Music

Permalink

<https://escholarship.org/uc/item/5996s3r3>

Author

PAYNE, JOHN

Publication Date

2016

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA,
IRVINE

The Neuroanatomical and Evolutionary Basis of Human Musical Experience

DISSERTATION

submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in Cognitive Neuroscience

by

John Phillip (Jack) Payne

Dissertation Committee:
Professor Gregory Hickok, Chair
Professor Charlie Chubb
Professor Christopher Dobrian

2016

TABLE OF CONTENTS

	Page
LIST OF FIGURES	iii
LIST OF TABLES	iv
ACKNOWLEDGEMENTS	v
CURRICULUM VITAE	vi
ABSTRACT OF THE DISSERTATION	vii
Chapter 1 Introduction: The Biological and Physical Basis of Music Processing	1
Chapter 2 Subjective Emotional Ratings of Music	23
Chapter 3 Prosodic Information is Conveyed by Spectral and Temporal Modulations in Speech	36
Chapter 4 Functional Neuroanatomical Differences between Language and Music	45
Chapter 5 Conclusion: The Role of Music In Human Evolution	60

LIST OF FIGURES

	Page
Figure 1.1 Bach BWV 100	14
Figure 2.1 Aggregated responses to all stimuli	26
Figure 2.2 Averaged responses to all stimuli	26
Figure 2.3 Average responses for each genre	28
Figure 2.4 Divisions of stimuli used for emotional content analysis	29
Figure 2.5 Spectral power for each emotional group	29
Figure 3.1 Modulation heat map for Angry Speech	39
Figure 3.2 Modulation heat map for Happy Speech	39
Figure 3.3 Difference map for the two conditions	40
Figure 4.1 Conjunction Analysis: Sentences - Music	50
Figure 4.2 Conjunction Analysis: Music - Music with Violations	51
Figure 4.3 Conjunction Analysis: Sentences - Scrambled Sentences	52

LIST OF TABLES

		Page
Table 2.1	Correlations between ratings and RMS	27
Table 2.2	Effect of Lyrics on Valence and Arousal	28
Table 2.3	Musical features and variance correlations	30
Table 2.4	Musical features and rating correlations	31

ACKNOWLEDGEMENTS

Music is an extremely multi-faceted perceptual experience. Attempting to explain its existence is a difficult undertaking, studying it from multiple angles paints a clearer picture about why it is so prevalent in human societies. Although it is pleasurable, it's not simply there for enjoyment. It can serve as a brain training system, help to improve our ability to communicate with others, modulate our emotions, and attract mates. Studying it from a neuroscientific perspective these last 5 years has been a very enriching experience. I'm lucky to have had the opportunity, while somehow being allowed to pursue a musical career with other parts of my time.

I would like to thank my brilliant, incredibly patient advisor, Greg Hickok. I would also like to thank the members of my committee, Charlie Chubb for his ability to impart enthusiasm for research into his advice, and Chris Dobrian for his unique, engaging perspective on my work. Having such intelligent people lend their insight has helped a huge deal as I've gradually coalesced these many disparate trains of thought into a dissertation.

My fellow lab members also deserve praise for their perspectives and guidance over the duration of my PhD program.

CURRICULUM VITAE

John Phillip (Jack) Payne

- 2007 B.A. in Cognitive Neuroscience, Macalester College
- 2007-2009 Research Analyst, Separation Kinetics, Inc
- 2010-2016 Teaching Assistant, Cognitive Sciences, University of California, Irvine
- 2014 M.S. in Cognitive Neuroscience, University of California, Irvine
- 2016 Ph.D. in Cognitive Neuroscience, University of California, Irvine

FIELD OF STUDY

The Neuroscience of Music and Language Processing

ABSTRACT OF THE DISSERTATION

The Neuroanatomical and Evolutionary Basis of Music

By

Jack Payne

Doctor of Philosophy in Cognitive Neuroscience

University of California, Irvine, 2016

Professor Gregory Hickok, Chair

Music exists in a variety of forms and is nearly ubiquitous in modern society but the reasons for its prevalence are poorly understood. Current attempts to answer this question are lacking in substantive explanatory power. Taking a wider view of music helps to clarify the reasons that it has developed. Musical activity engages an extensive network of brain regions and musical training confers a host of behavioral, motor, emotional and social benefits onto the practitioner. The fact that it engages such a wide swath of the brain also means that musical experiences can be more deeply encoded in memory. Although there are a number of shared resources for music and language processing, at later stages in the processing network, a different set of neuroanatomical regions are recruited for the two systems. Prosodic contour can carry information from a speaker to a listener via a separate set of neuroanatomical channels than other parts of language; it also carries cues about the emotional state of the speaker. The genetic links between creativity, musical proclivity and mental illness provide a critical key for understanding the role of music in human evolution. Interpreting evidence from dis-

parate scientific fields creates a better understanding of music's role in shaping the human experience.

CHAPTER 1

INTRODUCTION

Music is a fundamental part of the human experience. Music itself comes in many forms - whether it is experienced in a mother's voice, in a concert hall, around a campfire, in religious setting or in a sweaty dance club, music appears in a remarkably wide range of human situations. In each case, its role on the surface may vary, but the fundamental impact is the same; music manipulates our emotional states.

Music has existed for a sizable portion of human history. The oldest known instrument is a bone flute that dates back to circa 43,000 years ago (Münzel 2002). The physical capabilities for coordinated movement and sound production were available to pre-historic humans, so it is not unreasonable to hypothesize that a music-like experience was present for humans prior to widespread use of specialized tone producing tools (Kimura 1979; Thelen 1983). The capabilities needed for coordinated movement exist in all animals, suggesting that this ability emerged in a distant common ancestor. How and why musical abilities emerged is still a matter of extensive debate (Fitch 2006).

Scientists have puzzled over the relationship of language and music for a long time. The precise nature of the two systems' relationship has been studied from a variety of angles - from theorizing, to behavioral methodology to brain imaging. Some theorists posit that music and language emerged from a single musical protolanguage (Masataka 2009, Kirby 2011; Fitch 2009). Others posit that music emerged relatively recently as a spandrel of language ability, itself emerging from larger informational processing capacity of the brain (Pinker 1999).

The musical protolanguage hypothesis (Fitch 2009; Kirby 2011; Masataka 2009) states that music and language share a common ancestor, which was comprised of attributes of both systems that later diverged. One of the core supports for this hypothesis are that both music and language are primarily auditory streams and thus use overlapping brain regions for the bulk of their processing (Sammler 2009; Patel 2010; Patel 2012; Bidelman 2013). Auditory nerve, brain stem, and primary cortical auditory regions are necessarily shared by the two systems but at later stages of the processing streams, the two systems diverge (Rogalsky 2011). Investigating how and where processing for language and music overlap or separate helps elucidate the roles of the two systems in modern humans. Another key piece of evidence supporting this theory is that the two systems exist on similar time-scales and have hierarchical organizational structures. Phonemes, syllables, words and sentences map somewhat well to notes, groups, and phrases in music.

The other famous explanation for music is that it emerged as a spandrel after evolution selected for proto-humans with more neurological faculties for language (Pinker 1999). Although on the surface this explanation appears complete, it fails to account for the wide swath of abilities on which music draw and the benefits that musical training can convey onto individuals. Pitch tracking, speech-in-noise, verbal memory, executive function, second language learning and general auditory perceptual ability are among some of the benefits that musical training confers on the practitioner (Bialystock 2009; Csilbi 2013; Franklin 2008; Musacchia 2007; Patel 2007; Trainor 2009; Strait 2012).

Early Auditory Regions

The ability of an organism to gather information about its environment is beneficial to its longevity even at the unicellular level (Libert 2007). Largely, sensory systems increase our ability to perform adaptive behaviors.

Sound exists as longitudinal waves in the air and is characterized by modulations in the temporal and spectral domains. When sound waves strike the tympanic membrane, a cascade of activity begins: the hair cells on the inner ear transduce the sonic energy into electrochemical impulses (Mountcastle 1998). The basilar membrane itself has a tonotopic organization which is maintained through the primary auditory regions of the cortex (Da Costa 2015; Romani 1982). The fact that the experience of sound is organized tonotopically at such an early stage of the auditory pathway suggests that the ability to differentiate between sounds of different frequencies is extremely adaptive.

Although humans generally rely on their visual systems as their primary method of gathering information about their environment, the auditory system is the only other way that humans can sense objects at a distance. Visual information is more detailed but has a smaller field of view; objects outside of our field of vision are invisible to us and our visual field only covers a relatively small portion of the space around us. The auditory system complements the drawbacks of the visual system (Mountcastle 1998). Whereas the visual system has a narrow focus and fine detail, the auditory system can take in information from the entirety of the environment. Humans are capable of placing objects in space based on several auditory features: the binaural temporal and level differences for left-right space and the head-related transfer function for up-down and

front-back space (Hafter 1992; Spagnol 2013). The inferior colliculus is a low-level brain region with a specific function - the auditory startle and orienting response (Mountcastle 1998). The presence of this function at such a phylogenetically old region suggests that this ability is strongly selected for - without the auditory system to guide attention to a salient object, all the fine detail of the visual system is rendered useless. These spatial cues are coarse but powerful; they aid in orienting our visual system for finer inspection. In addition to spatial cues, the spectral content of sounds itself contains important information- there are some constants in naturally occurring sounds.

- I. Sounds with large amounts of low frequency energy are associated with larger events, which can represent threats to an organism.
- II. Sounds with less low frequency energy are associated with smaller organisms, which are generally less threatening.

Even within our species itself, people with lower voices are perceived as larger and more dominant (Puts 2006). Composers often take advantage of this fact when creating sounds for movies. Extreme levels of low frequency content serve the function of orienting the viewer's attention to the screen and the intense volume levels (much greater than what occurs naturally) can overwhelm the viewer. In modern music, particularly hip hop and EDM, huge amounts of bass content are used to accomplish the same effect on listeners and dancers. Baselines and beats are often described as "huge"- this is no accident, but a natural consequence of the environment in which humans emerged. Music with higher relative bass drum volume has been found to induce better tempo entrainment and more dancing activity overall; the relationship between music and dance is mediated by low-frequency content (Van Dyck 2013).

There is evidence for loud (>90dB (A) SPL) music being particularly pleasurable, sometimes called the “rock and roll threshold” (Todd 2000). This experience is mediated by vestibular stimulation and can even occur in people lacking a cochlea (Todd 2000). There is evidence that vestibular receptors can affect auditory evoked potentials, suggesting that the relationship between these two systems is closer than previously thought (Todd 2014).

As the neural impulse passes up to the cortex, a further refinement of tonotopic activity occurs; neurons in primary auditory areas respond selectively to resolved harmonics in a way that closely parallels behavior (Da Costa 2015; Norman-Haignere 2013). After primary auditory areas, the neural impulse propagates to other brain regions. Their functions are manifold- directing attention, creating auditory objects, evaluating the salience of the perceived sound, creating and tracking rhythmic models, tracking the pitch of the auditory object, identifying a speaker and parsing linguistic information are but a few of the tasks that the brain performs (Hickok 2000, 2009; Mountcastle 1998).

Rhythm processing recruits both the motor and auditory systems. There is a growing body of work surrounding the basal ganglia in particular. The basal ganglia play a crucial role in motor activity, emotional regulation and the perception of rhythm (Grahn 2009; Nelson 2014). Rhythms with a stronger beat selectively activate the the basal ganglia (Leow 2014). It has been found that BG activity was relatively constant across variations of complexity, suggesting that BG is not sensitive to the difficulty of beat perception task, merely the presence of it. (Kung 2013). BG damage is also associated with impairment of the perception of rhythm in speech (Kotz 2015).

EEG investigations of beat perception show that there is a direct connection between acoustic information and neural activity for beat perception (Nozaradan 2012, 2013). Populations of neurons in the brain are able to synchronize and enhance activity to sounds that occur at regular intervals. Pattern detection is one of the key functions of the human brain- it is performed even when subjects are engaged in a distractor task (Barascud 2016). Recent work has attempted to hijack this skill at implicit learning to use humans as predictors of stock markets albeit with limited results (Worrall 2009).

What is music

Dividing music from language can be theoretically complex, and as the definition broadens, so does the possibility of misinterpretations. Edgar Varèse defined music as “organized sound,” and this definition stands up well to scrutiny; questions of taste and cultural differences are not a problem using this definition. The nature of the organization does not matter, nor the provenance of the organizational structure. One issue with this definition is that it potentially subsumes language itself. Although language can exist in the auditory, visual, and haptic domains, the prototypical use of language is also organized sound. Despite this failing, using “organized sound” as the foundational definition of music is the least problematic way to define it. Other potential definitions for music impart substantial amounts of culturally specific information that is not appropriate for encompassing the entire range of musical experiences. Using the definition of organized sound implicates that many other sounds will fit within the definition. Is a jackhammer music? Is the sweeping of a broom music? Is the sound of traffic music? There

are certainly organizing principles in these sounds and rhythmic regularities at a wide variety of levels. Although at first blush these sounds may not appear musical, given the proper context, they can be very effective (i.e. “Stomp”). The presence of structure in the sounds provides the opportunity to interpret the sounds as musical. Intentionality factors into this definition, but it can exist either in the producing or receiving side.

Like many languages, music also has a writing system, although that writing system is relatively younger. In the same way that written language is a condensed form of speech, written music is a condensed form of music. The general structure is present on the page, but many other aspects can be changed without losing the core of the piece. Trained musicians can read music and react to atonal events, however the amplitude of these signals (i.e. the P600) is weaker than for unexpected linguistic events (Schön 2002, 2005). Tempos can be changed, notes can be exaggerated in time, pitch, or vibrato, even the instrument used can change while still using the same piece of music. Yo-Yo Ma’s rendition of the first Bach Cello prelude (BWV 1007), if fully transcribed, would look remarkably different from what is provided by Bach. Conversely, if played exactly as written on the page, the piece might not not carry create the same emotional impact on listeners as other, less precise renditions (Juslin 2003). Indeed, there may be underlying timing rules for how to promote expressivity in music (Desain 1991).

The sonic vocabulary

Part of the organization of music is the in its vocabulary of sounds. Some aspects of this organization are based upon physical and biological principles whereas

other aspects are determined by cultural traditions. Consonance and dissonance are driven partly at the level of the physical waveform (Gill 2009). Waveforms that are multiples of each other, or exist as whole number fractions of each other are found in a wide range of tuning systems.

Oscillations in the brain seem to be present during many events. Despite a massive amount of research into oscillations, whether or not these oscillations are artifacts of other processes or if they actually have powerful implications is as yet unclear. However, dissonance presents case where ill-aligned oscillations in an auditory waveform are rated as unpleasant. In an EEG experiment, it was found that consonant chords elicit more gamma activity than matched dissonant chords (Park 2010). Musical waveforms themselves, in an electric form, are represented in lower brain regions, the brain is sensitive to harmonicity at a very low level (Bidelman 2009). Octave equivalence occurs across all humans; tones with frequencies that are factors of 2 apart are perceived as belonging to the same pitch class (Patel 2003). Rhesus monkeys can show octave equivalence for simple melodies as well (Wright 2000), but other chickadees fail this test (Hoeschele 2013).

Different musical traditions have different divisions of the octave to make meaningful scales. The Western musical tradition divides the octave into 12 equally spaced tones, although in most of the scales used only 7 tones are labeled as consonant to the scale. The division of the octave into pentatonic or heptatonic groups seems to be quite common across a range of musical traditions and is perhaps based on the number of items available in working memory (Patel, 2003; Miller 1956; Cowan 2012). Color per-

ception presents a nearly direct correlate to dividing up a continuous spectrum into relatively discrete categories.

Despite the ability to differentiate up to 240 divisions of an octave, penta- and heptatonic scales often emerge in a wide variety of cultural traditions (Gill 2009). Additionally, the spectral profile of these scales is remarkably consistent, suggesting a biological basis for the organization of pitch (Gill 2009). Indian Ragas, which divide the octave based on whole number fractions, play notes selected from 22 possible, but within those 22 possible tones, only 7 are used in any given raga and the ordering of tones used is strict - the variations in one scale to another can be as small as a microtonal shift in one *swara*, or tone. In the Melakarta system of Indian ragas, there are 72 possible ragas, which are analogous to scales in Western music (Kirthika 2012). In Western music, there are generally 24 scales used - a major and minor scale with variations on the 12 possible tonic centers. In both the major and minor scales, the only intervals present are based on semitones, mathematically defined above. There are other scales used in Western music that organize the twelve tones with some larger or smaller intervals, but their use is rarer than the major and minor scales.

These commonly found octave divisions dovetail nicely to the number of items available to working memory, which is often described as 7 ± 2 items (Miller 1956, Cowan 2012) and also aligns fairly closely with the range of words for colors across languages (Berlin and Kay, 1969).

There is a large body of work that asks the question of how much impact the linguistic divisions have on the perceptual experience, commonly called linguistic relativity or the Sapir-Whorf hypothesis (Whorf, 2012; Hammer 2014; Dolscheid 2013). Arguing

the validity of linguistic relativity is beyond the scope of this paper, it will suffice to say that humans have a tendency to group items into categories. The precise divisions of those categories are largely irrelevant, but the question of how items from those categories are reorganized and blended together can have a large perceptual impact.

Prosody

The role of prosodic contour is perhaps under-appreciated as a carrier of meaning. Prosodic contour provides important information about the meaning of an utterance. In English, any sentence can be changed from a statement to a question by the use of a rising tone at the end of the utterance (barring the Australian accent which does this more regularly). Different emphases on individual words highlights their relative importance in the utterance and can even distort the meaning completely. Additionally, the overall contour can provide information about the emotional state of a speaker that is not present in the semantic meaning itself (Pichon 2013; Smith 2014).

The earliest forms of written language are between 4500 and perhaps 8000 years old, depending on the stringency of the definition (Sampson 2015). Even today, not all humans can read and write. There are advantages to written language (i.e. longevity, wider range of distribution) but there are some limitations as well (i.e. lack of prosody). Before written language emerged, histories were preserved in auditory forms, notably the traditions of epic poetry and song (Rose 1992). The use of multi-level schemes to improve memory encoding and recall lends support to this hypothesis (Kirchoff 2012).

Some languages, such as Mandarin, use tonal information to convey semantic meaning. The presence of tonal semantic content means that sung words can be interpreted as having different meanings; however the the precise tone of a word can be changed to fit the contour of the song and listeners use contextual cues to properly parse the words. Other languages use stress and accent to convey different meaning; Japanese and Norwegian are two such examples. In Japanese, for example, a single accent changes the meaning of the word mother (“okāsan”) to Mr. Oka (“okasan”). English and most European languages rely on pronunciation and syntactical rules to convey meaning, thus allowing prosodic information to be relatively separated from the semantic meaning of an utterance.

Despite this theoretical delineation of prosody and language, some scientists posit that a substantial portion of meaning is conveyed through nonverbal means, or even that language emerged from something resembling pure prosody (Knapp 2013; Altenmüller 2013). “Motherese” is a cross culturally stable phenomenon wherein a mother will exaggerate aspects of linguistic rules in the presence of an infant. This behavior appears to be instinctive and can be triggered not only by a variety of cues. It’s widely understood that this form of speech plays a role in language acquisition (Golinkoff 2015). Interestingly, “motherese” is characterized by its “sing-song” nature and the sounds themselves can be described as pleasant or happy. There are clues embedded in the system that certain characteristics of sounds have stable effects on human emotions; the minor third seems to occur in sad speech as well as music although the directionality of this relationship remains to be determined (Day-O’Connell 2013; Quinto 2013). Strong bonds between mother and child help the child grow;

strengthening positive associations early on aids with the creation of these bonds, so it stands to reason that “motherese” would have the attributes of sounds associated with positive emotions - more research is needed on this topic.

Memory and Music

Music itself is deeply encoded in memory. In order to understand why, consider the nature of neural activity at the level of single neurons and large networks.

- I. Any neuron has an instantaneous probability of firing from zero to one.
- II. Once a neuron begins firing, it follows a set pattern (the action potential).
- III. Connections between neurons can be created and strengthened.
- IV. Connections between neurons can be excitatory or inhibitory.
- V. Larger networks have more neurons active simultaneously.
- VI. Networks can operate in both a temporally or spatially structured way.

With these basic attributes of neuronal activity it becomes possible to explain the effectiveness of mnemonic devices and the fact that music tends to be more memorable. Mnemonic devices work by providing a more richly structured version of the items that the subject wants to memorize. They achieve this by a variety of means, either by expanding, contracting or otherwise modifying the words (e.g. “ROY G. BIV” for the color spectrum, the “alphabet song” for letters, etc etc). There are a huge variety of ways to use mnemonic devices, but the same strategy is present across all of them. Enrich the items by providing another way to encode them and it becomes easier for

people to recall those items in the future. Several of the commonly used mnemonic tools are musical or prosodic in nature.

In the same way, music creates deep memories because of its multi-modal nature. Rhythm, melody and the lyrics themselves are all potential “points of entry” to recalling the circuit for the song. So presuming that there are more “points of entry” to activate the memory circuit, having a more widely embedded circuit should indicate a greater propensity to be reactivated. The memories themselves are more deeply encoded further because of their potential for intense emotional associations.

Auditory Illusions

There has been a large body of work on visual illusions and their implications for the way that the brain processes visual information. The primary implication for this research is that visual perception is *constructive*. Organisms do not perceive “truth” in the pure sense, but a representation of the environment that aids to guide evolutionarily useful behavior (Hoffman 2015) - the auditory system is no different in this regard.

The first constriction, as with vision, is that there is a limited bandwidth of sensory ability, generally described as 20Hz to 20KHz (Mountcastle 1998). Pitch perception is another constriction, and moving up the ladder, rather than hearing groups of tones as independent frequencies, humans hear groups of pitches as chords.

There is a limit to how many individual notes can be perceived within a second. Although humans can respond to sounds of very short duration, music employs note lengths of .125 to .5 seconds on average. At the high end of the tempo spectrum (i.e.

240 bpm), an eighth note is only .125 seconds long, and a sixteenth is again half that, .0625 seconds, which corresponds to 16Hz. This is about as fast as music gets- it becomes perceptually difficult to hear those sounds as individual notes above around 8Hz - additionally notes of power pitch are slower than notes with higher pitch (Broze 2013). As always, with training, it is possible to become more accustomed to music at these speeds, but even with that there is a point at which the oscillations will be perceived as pitch.

Bach used pitch grouping cues to create different “streams” within his music. He often would create the percept of two separate streams in one interleaved set of notes by differentially grouping the tones- this phenomenon is called *melodic fission* (Davis 2006).



BWV 100 with the different streams highlighted in different colors.

Although there is only one series of single notes played by a single instrument, the perception of two separate streams quickly emerges. The perceptual principle of grouping based on similarity in pitch is hijacked by a composer to engage the listener. There are many other well known auditory illusions and several audiovisual illusions. Several of the involve pitch perception, but there are other types as well.

The Shepard tone, or Shepard-Risset tone illusion, employs a few attributes of auditory processing to create an illusion of an endlessly rising tone. Octave equivalence plays a role as does pitch perception. As a musical pitch is usually a complex tone (as compared to a simple sine wave tone), the Shepard tone maintains a set vol-

ume curve as each individual tone rises in frequency. The loudest part of the sound remains in a single octave, so that as each individual tone rises, it starts quiet, grows louder for a time and then fades out again. Through this, the conflicting perceptions of a constantly rising tone and a steady tone are present in the signal.

The Risset beat is a similar illusion to the Shepard tone, but it uses a drum beat rather than a tone. It appears to accelerate while simultaneously maintaining a constant tempo. The relative volumes of the different drum beats present modulates over time in the same way that the tones in the Shepard tone modulate over time.

The “missing fundamental” is another illusion that shows the way our auditory perception is constructive (Smith, 1978). The perception of complex tones as single pitches comes into play here as well. In this case, a sound is presented to the listener wherein the lowest fundamental frequency has been filtered out. However, listeners do not notice that the fundamental is missing - the tones present are interpreted as a harmonic series with the implied fundamental. This illusion partly explains why many consumers do not seem to mind listening to music on audio systems without full spectral content; the strength of their auditory system can make up for the weakness of audio systems.

The McGurk effect has been studied as a means to show that language perception relies on visual information as well as auditory information (Matchin 2014). By mixing audio from one phoneme with video from another phoneme, the percept can be changed to be in between the two (i.e. combining “ba, ba” and “ga, ga” produces the percept of “da, da”). This is an instance of high level informational integration affecting low level perceptions. The conflicting visual information modifies the auditory percept.

Another simple illusion is created every time a person hears their own voice on a recording. Although people are accustomed to the sound of their own voice from the inside of their own head, when others hear it, they do not experience the resonances of the skull and the rest of the body of the speaker. It can be very jarring to hear the recording because of the lack of these acoustic resonances.

The existence of these illusions and other perceptual phenomena support the idea that auditory perception is constructive in the same way that vision is. Humans interpret the sounds in our environment to create a working mental model to guide our behavior. The way that humans experience those sounds is not necessarily a truthful representation of those sounds. Although the energy that reaches the eardrum is nothing but a stream of pressure differentials, the brain imbues these changes in air pressure with deep meanings.

Interlude

The subsequent chapters of this dissertation will discuss several experiments built around question of how the brain processes music, particularly with regards to its relationship with the limbic system and emotions.

REFERENCES

1. Areni, C. S., & Kim, D. (1993). The influence of background music on shopping behavior: classical versus top-forty music in a wine store. *NA-Advances in Consumer Research* Volume 20.
2. Altenmüller, E., Schmidt, S., & Zimmermann, E. (Eds.). (2013). *The evolution of emotional communication: From sounds in nonhuman mammals to speech and music in man*. OUP Oxford.
3. Barascud, N., Pearce, M. T., Griffiths, T. D., Friston, K. J., & Chait, M. (2016). Brain responses in humans reveal ideal observer-like sensitivity to complex acoustic patterns. *Proceedings of the National Academy of Sciences*, 201508523.
4. Bialystok, E., & DePape, A. M. (2009). Musical expertise, bilingualism, and executive functioning. *Journal of Experimental Psychology: Human Perception and Performance*, 35(2), 565.
5. Bidelman, G.M. and A. Krishnan, *Neural Correlates of Consonance, Dissonance, and the Hierarchy of Musical Pitch in the Human Brainstem*. *Journal of Neuroscience*, 2009. **29**(42): p. 13165-13171.
6. Bidelman, G. M., Hutka, S., & Moreno, S. (2013). Tone language speakers and musicians share enhanced perceptual and cognitive abilities for musical pitch: evidence for bidirectionality between the domains of language and music. *PLoS One*, 8(4), e60676.
7. Blood, A. J., Zatorre, R. J., Bermudez, P., & Evans, A. C. (1999). Emotional responses to pleasant and unpleasant music correlate with activity in paralimbic brain regions. *Nature neuroscience*, 2(4), 382-387.
8. Blood, A. J., & Zatorre, R. J. (2001). Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion. *Proceedings of the National Academy of Sciences*, 98(20), 11818-11823.
9. Broze, Y., & Huron, D. (2013). Is higher music faster? Pitch–speed relationships in Western compositions. *Music Perception: An Interdisciplinary Journal*, 31(1), 19-31.
10. Caldwell, G. N., & Riby, L. M. (2007). The effects of music exposure and own genre preference on conscious and unconscious cognitive processes: A pilot ERP study. *Consciousness and cognition*, 16(4), 992-996.
11. Cowan, N. (2012). *Working memory capacity*. Psychology press.
12. Cross, I. (2003). *Music and Biocultural Evolution*. *The cultural study of music: a critical introduction*, 19.
13. Csibi, A. (2013). *The Influence of Musical Training on L2 Intonation, Phonology and Verbal Memory*.
14. Da Costa, S., Saenz, M., Clarke, S., & Van Der Zwaag, W. (2015). Tonotopic gradients in human primary auditory cortex: concurring evidence from high-resolution 7 T and 3 T fMRI. *Brain topography*, 28(1), 66-69.

15. Danion, F., Varraine, E., Bonnard, M., & Pailhous, J. (2003). Stride variability in human gait: the effect of stride frequency and stride length. *Gait & Posture*, 18(1), 69-77.
16. Davis, S. (2006). Implied polyphony in the solo string works of JS Bach: A case for the perceptual relevance of structural expression. *Music Perception: An Interdisciplinary Journal*, 23(5), 423-446.
17. Day-O'Connell, J. (2013). Speech, Song, and the Minor Third. *Music Perception: An Interdisciplinary Journal*, 30(5), 441-462.
18. Desain, P., & Honing, H. (1991). Towards a calculus for expressive timing in music. *Computers in Music Research*, 3, 43-120.
19. Dolscheid, S., Shayan, S., Majid, A., & Casasanto, D. (2013). The Thickness of Musical Pitch Psychophysical Evidence for Linguistic Relativity. *Psychological Science*, 24(5), 613-621.
20. Falk, D. (2004). Prelinguistic evolution in early hominins: Whence motherese?. *Behavioral and Brain Sciences*, 27(04), 491-503.
21. Fitch, W. T. (2009). Musical protolanguage: Darwin's theory of language evolution revisited. Guest-post on [www.st-andrews.ac.uk](http://www.st-andrews.ac.uk/~wtsf/DarwinLanguageEvolution.html). Retrieved from web site: <http://www.st-andrews.ac.uk/~wtsf/DarwinLanguageEvolution.html>.
22. Fitch, W. T. (2006). The biology and evolution of music: A comparative perspective. *Cognition*, 100(1), 173-215.
23. Franklin, M. S., Moore, K. S., Yip, C. Y., Jonides, J., Rattray, K., & Moher, J. (2008). The effects of musical training on verbal memory. *Psychology of Music*.
24. Fritz, T., Jentschke, S., Gosselin, N., Sammler, D., Peretz, I., Turner, R., ... & Koelsch, S. (2009). Universal recognition of three basic emotions in music. *Current biology*, 19(7), 573-576.
25. Gill K. Z., Purves D. (2009) A Biological Rationale for Musical Scales. *PLoS ONE* 4(12): e8144. doi:10.1371/journal.pone.0008144
26. Golinkoff, R. M., Can, D. D., Soderstrom, M., & Hirsh-Pasek, K. (2015). (Baby) Talk to Me The Social Context of Infant-Directed Speech and Its Effects on Early Language Acquisition. *Current Directions in Psychological Science*, 24(5), 339-344.
27. Grahn, J. A. (2009). The role of the basal ganglia in beat perception. *Annals of the New York Academy of Sciences*, 1169(1), 35-45.
28. Hafter, E. R., Saberi, K., Jensen, E. R., & Briolle, F. (1992). Localization in an echoic environment. *Adv Biosci*, 83, 555-561.
29. Hammer, K. (2014). Thinking and Speaking in Two Languages. *Sociolinguistic Studies*, 8(2), 327.
30. Hickok, G. (2009). The functional neuroanatomy of language. *Physics of life reviews*, 6(3), 121-143.
31. Hickok, G., & Poeppel, D. (2004). Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition*, 92(1), 67-99.

32. Hickok, G., & Poeppel, D. (2000). Towards a functional neuroanatomy of speech perception. *Trends in cognitive sciences*, 4(4), 131-138.
33. Hoeschele, M., Weisman, R. G., Guillette, L. M., Hahn, A. H., & Sturdy, C. B. (2013). Chickadees fail standardized operant tests for octave equivalence. *Animal cognition*, 16(4), 599-609.
34. Hoffman, D. D., Singh, M., & Prakash, C. (2015). The interface theory of perception. *Psychonomic bulletin & review*, 22(6), 1480-1506.
35. Honing, H., & Ploeger, A. (2012). Cognition and the evolution of music: Pitfalls and prospects. *Topics in Cognitive Science*, 4(4), 513-524.
36. Huron, D. (2001). Is music an evolutionary adaptation?. *Annals of the New York Academy of sciences*, 930(1), 43-61.
37. Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code?. *Psychological bulletin*, 129(5), 770.
38. Juslin, P. N., & Västfjäll, D. (2008). Emotional responses to music: The need to consider underlying mechanisms. *Behavioral and brain sciences*, 31(05), 559-575.
39. Kim, Y. E., Schmidt, E. M., Migneco, R., Morton, B. G., Richardson, P., Scott, J., ... & Turnbull, D. (2010, August). Music emotion recognition: A state of the art review. In *Proc. ISMIR* (pp. 255-266).
40. Kimura, D. (1979). Neuromotor mechanisms in the evolution of human communication. *Neurobiology of social communication in primates: An evolutionary perspective*, 197-220.
41. Kirthika, P., & Chattamvelli, R. (2012, July). A review of raga based music classification and music information retrieval (MIR). In *Engineering Education: Innovative Practices and Future Trends (AICERA), 2012 IEEE International Conference on* (pp. 1-5). IEEE.
42. Kirby, S. (2011). Darwin's musical protolanguage: an increasingly compelling picture. *Language and Music as Cognitive Systems*, 96-102.
43. Kirchhoff, B. A., Anderson, B. A., Barch, D. M., & Jacoby, L. L. (2012). Cognitive and neural effects of semantic encoding strategy training in older adults. *Cerebral cortex*, 22(4), 788-799.
44. Knapp, M., Hall, J., & Horgan, T. (2013). *Nonverbal communication in human interaction*. Cengage Learning.
45. Kotz, S. A., & Schmidt-Kassow, M. (2015). Basal ganglia contribution to rule expectancy and temporal predictability in speech. *Cortex*, 68, 48-60.
46. Kung, S. J., Chen, J. L., Zatorre, R. J., & Penhune, V. B. (2013). Interacting cortical and basal ganglia networks underlying finding and tapping to the musical beat. *Journal of Cognitive Neuroscience*, 25(3), 401-420.

47. Leow, L. A., & Grahn, J. A. (2014). Neural mechanisms of rhythm perception: present findings and future directions. In *Neurobiology of Interval Timing* (pp. 325-338). Springer New York.
48. Libert, S., & Pletcher, S. D. (2007). Modulation of longevity by environmental sensing. *Cell*, 131(7), 1231-1234.
49. Masataka, N. (2009). The origins of language and the evolution of music: A comparative perspective. *Physics of Life Reviews*, 6(1), 11-22.
50. Matchin, W., Groulx, K., & Hickok, G. (2014). Audiovisual speech integration does not rely on the motor system: evidence from articulatory suppression, the McGurk effect, and fMRI. *Journal of cognitive neuroscience*, 26(3), 606-620.
51. Meyer, M., Elmer, S., & Jäncke, L. (2012). Musical expertise induces neuroplasticity of the planum temporale. *Annals of the New York Academy of Sciences*, 1252(1), 116-123.
52. Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological review*, 63(2), 81.
53. Morrison, M., Gan, S., Dubelaar, C., & Oppewal, H. (2011). In-store music and aroma influences on shopper behavior and satisfaction. *Journal of Business Research*, 64(6), 558-564.
54. Mountcastle, V. B. (1998). *Perceptual neuroscience: the cerebral cortex*. Harvard University Press.
55. Münzel, S. C., Seeberger, F., & Hein, W. (2002). The Geißenklösterle Flute—discovery, experiments, reconstruction. *Studien zur Musikarchäologie III*, 107-118.
56. Musacchia, G., Sams, M., Skoe, E., & Kraus, N. (2007). Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *Proceedings of the National Academy of Sciences*, 104(40), 15894-15898.
57. Nelson, A. B., & Kreitzer, A. C. (2014). Reassessing models of basal ganglia function and dysfunction. *Annual review of neuroscience*, 37, 117.
58. Norman-Haignere, S., Kanwisher, N., & McDermott, J. H. (2013). Cortical pitch regions in humans respond primarily to resolved harmonics and are located in specific tonotopic regions of anterior auditory cortex. *The Journal of Neuroscience*, 33(50), 19451-19469.
59. Nozaradan, S., Peretz, I., & Mouraux, A. (2012). Selective neuronal entrainment to the beat and meter embedded in a musical rhythm. *The Journal of Neuroscience*, 32(49), 17572-17581.
60. Nozaradan, S., Zerouali, Y., Peretz, I., & Mouraux, A. (2013). Capturing with EEG the neural entrainment and coupling underlying sensorimotor synchronization to the beat. *Cerebral Cortex*, bht261.
61. Patel, A. D. (2003). Language, music, syntax and the brain. *Nature neuroscience*, 6(7), 674-681.

62. Patel, A. D. (2010). Music, biological evolution, and the brain. *Emerging disciplines*, 91-144.
63. Patel, A. D. (2012). Advancing the comparative study of linguistic and musical syntactic processing. *Language and music as cognitive systems*, 248-253.
64. Patel, A. D., & Iversen, J. R. (2007). The linguistic benefits of musical abilities. *Trends in cognitive sciences*, 11(9), 369-372.
65. Pichon, S., & Kell, C. A. (2013). Affective and sensorimotor components of emotional prosody generation. *The Journal of Neuroscience*, 33(4), 1640-1650.
66. Pinker, S. (1999). How the mind works. *Annals of the New York Academy of Sciences*, 882(1), 119-127.
67. Puts, D. A., Gaulin, S. J., & Verdolini, K. (2006). Dominance and the evolution of sexual dimorphism in human voice pitch. *Evolution and Human Behavior*, 27(4), 283-296.
68. Quinto, L., FordeThompson, W., & Keating, F. L. (2013). Emotional communication in speech and music: the role of melodic and rhythmic contrasts.
69. Rickard, N. S. (2004). Intense emotional responses to music: a test of the physiological arousal hypothesis. *Psychology of music*, 32(4), 371-388.
70. Rogalsky, C., Rong, F., Saberi, K., & Hickok, G. (2011). Functional anatomy of language and music perception: temporal and structural factors investigated using functional magnetic resonance imaging. *The Journal of Neuroscience*, 31(10), 3843-3852.
71. Rogalsky, C., & Hickok, G. (2011). The role of Broca's area in sentence comprehension. *Journal of Cognitive Neuroscience*, 23(7), 1664-1680.
72. Romani, G. L., Williamson, S. J., & Kaufman, L. (1982). Tonotopic organization of the human auditory cortex. *Science*, 216(4552), 1339-1340.
73. Rose, S. (1992). *The Making of Memory: From Molecules to Mind*. Anchor Books/Doubleday (NY)
74. Salimpoor, V. N., Zald, D. H., Zatorre, R. J., Dagher, A., & McIntosh, A. R. (2015). Predictions and the brain: how musical sounds become rewarding. *Trends in cognitive sciences*, 19(2), 86-91.
75. Sammler, D. (2009). *The neuroanatomical overlap of syntax processing in music and language: Evidence from lesion and intracranial ERP studies* (Doctoral dissertation, Max Planck Institute for Human Cognitive and Brain Sciences Leipzig).
76. Sampson, G. (2015). *The earliest writing*. Equinox Publishing.
77. Schön, D., Magne, C., & Besson, M. (2004). The music of speech: Music training facilitates pitch processing in both music and language. *Psychophysiology*, 41(3), 341-349.
78. Schön, D., & Besson, M. (2005). Visually induced auditory expectancy in music reading: a behavioral and electrophysiological study. *Journal of cognitive neuroscience*, 17(4), 694-705.

79. Schön, D., Anton, J. L., Roth, M., & Besson, M. (2002). An fMRI study of music sight-reading. *Neuroreport*, 13(17), 2285-2289.
80. Smith, J. C., Marsh, J. T., Greenberg, S., & Brown, W. S. (1978). Human auditory frequency-following responses to a missing fundamental. *Science*, 201(4356), 639-641.
81. Smith, M. (2014). Emotional intensity in prosody: Acoustic correlates of emotion as intended by a male and female actor (Doctoral dissertation, MGH INSTITUTE OF HEALTH PROFESSIONS).
82. Spagnol, S., Geronazzo, M., & Avanzini, F. (2013). On the relation between pinna reflection patterns and head-related transfer function features. *Audio, Speech, and Language Processing, IEEE Transactions on*, 21(3), 508-519.
83. Spangenberg, E. R., Grohmann, B., & Sprott, D. E. (2005). It's beginning to smell (and sound) a lot like Christmas: the interactive effects of ambient scent and music in a retail setting. *Journal of business research*, 58(11), 1583-1589.
84. Strait, D. L., Parbery-Clark, A., Hittner, E., & Kraus, N. (2012). Musical training during early childhood enhances the neural encoding of speech in noise. *Brain and language*, 123(3), 191-201.
85. Thelen, E. (1983). Learning to walk is still an "old" problem: a reply to Zelazo (1983). *Journal of Motor Behavior*, 15(2), 139-161.
86. Thompson, W. F., Marin, M. M., & Stewart, L. (2012). Reduced sensitivity to emotional prosody in congenital amusia rekindles the musical protolanguage hypothesis. *Proceedings of the National Academy of Sciences*, 109(46), 19027-19032.
87. Todd, N. P. M., & Cody, F. W. (2000). Vestibular responses to loud dance music: A physiological basis of the "rock and roll threshold"? *The Journal of the Acoustical Society of America*, 107(1), 496-500.
88. Todd, N. P., Paillard, A. C., Kluk, K., Whittle, E., & Colebatch, J. G. (2014). Vestibular receptors contribute to cortical auditory evoked potentials. *Hearing research*, 309, 63-74.
89. Trainor, L. J., Shahin, A. J., & Roberts, L. E. (2009). Understanding the benefits of musical training. *Annals of the New York Academy of Sciences*, 1169(1), 133-142.
90. Van Dyck, E., Moelants, D., Demey, M., Deweppe, A., Coussement, P., & Leman, M. (2013). The impact of the bass drum on human dance movement. *Music Perception: An Interdisciplinary Journal*, 30(4), 349-359.
91. Whorf, B. L., Carroll, J. B., Levinson, S. C., & Lee, P. (2012). *Language, thought, and reality: Selected writings of Benjamin Lee Whorf*. Mit Press.
92. Worrall, D. (2009). The use of sonic articulation in identifying correlation in capital market trading data.
93. Wright, A. A., Rivera, J. J., Hulse, S. H., Shyan, M., & Neiworth, J. J. (2000). Music perception and octave generalization in rhesus monkeys. *Journal of Experimental Psychology: General*, 129(3), 291.

CHAPTER 2

SUBJECTIVE RESPONSES TO MUSIC

INTRODUCTION

Musical experiences are deeply linked to emotions (Blood 1999, 2001; Juslin 2003; Quinto 2013). Musical stimuli are used in many different situations to manipulate listener's emotional states, from shopping malls, to dance clubs, to churches (Areni 1993; Spanegenberg 2005). Despite their presumed effect, people have a range of responses to music (Rickard 2004; Juslin 2008; Caldwell 2007). However, music that successfully engages listeners' emotions is consistently rated as better (Salimpoor 2009). There is substantial evidence from a variety of fields, from machine learning, to neuroimaging that shows that music has some attributes that serve as universal cues to the emotional content of the piece (Fritz 2009; Kim 2010). These attributes exist in the auditory domain at a variety of time scales.

Describing emotions is a complex scientific problem as they are subjective states. However, it is possible to describe emotions as part of a two-dimensional space: this is the commonly used valence-arousal model (Chanel 2007; Kensinger 2004; Barrett 1998). Although this tool is not without its flaws (Sun 2009), it provides a simple, easy-to-understand metric for classifying and rating human emotions.

Neuroimaging studies using this tool have shown that the two axes of the valence-arousal space are processed orthogonally in the brain (Dolcos 2004). In that study, the left dorsolateral prefrontal cortex was more active for arousal, ventromedial prefrontal cortex was sensitive to positive valence (Dolcos 2004). The PFC is at the

core of the somatic marker hypothesis of Bechara (2005). Music engages a wide swath of limbic and paralimbic regions (Blood 1998, Rogalsky 2011, Salimpoor 2015). Engagement of the PFC by musical stimuli could explain some degree of why music itself is so reinforcing - it is a sort of “puzzle” to figure out as there are multiple conflicting cues about the intention of the piece. Music that lacks a level of complexity appropriate for the listener will not generally be appreciated by that listener.

This experiment sought to clarify the nature of listener responses to music. By exposing subjects to a wide variety of stimuli and having them provide subjective feedback, several regularities were found in their response data. These regularities in their responses were then used to investigate auditory features of music in order to characterize what features convey different emotions to listeners.

METHOD

Participants in this experiment were placed in a quiet room with a computer running Matlab 2013b. They were given set of instructions on the task they were to perform, with particular emphasis on the fact that they were to categorize the music they heard based on what they perceived to be the emotional content or message of the piece, and not based on how the music made them feel personally. This was done to characterize the intentionality of the music and limit the effect of genre preference on listener responses. Subjects were then instructed to rate the music on a two-dimensional axis - the valence-arousal space. Valence was defined as a continuum of pleasant / unpleasant emotions. Subjects were asked to describe each of the four corners of the grid in their own terms to ensure that they correctly understood the nature of the space

and instructed to employ the entire space in order to ensure that a full range of responses would be used.

The music stimuli were 400 excerpts (40 of each genre) from the GTZAN database, converted in to WAV format (16-bit, 44,100 KHz). The genres present in the database are 'blues', 'classical', 'country', 'disco', 'hiphop', 'jazz', 'metal', 'pop', 'reggae', and 'rock.'

Subjects

Subjects were recruited through the UCI Sona Systems program. Subjects were all native English speakers and enrolled as undergraduates at UCI for the summer of 2015. 25 subjects were used (14 female, 11 male).

RESULTS

There was a large variance in responses to many of the stimuli, although for certain cases, there was a very tight grouping in the responses. These stimuli were labeled as exemplars and saved for later analyses. Most responses to the stimuli varied across participants, but there was much agreement in the basic content of most the excerpts (i.e. subjects placed their responses in the same half or quadrant of the graph).

There is a bias towards staying on cardinal and diagonal axes, and when viewed as a whole, there is a bias towards rating music and moderately high arousal and positive valence. This latter effect was expected to some degree, as music listening is generally rated as "pleasant" no matter the context.

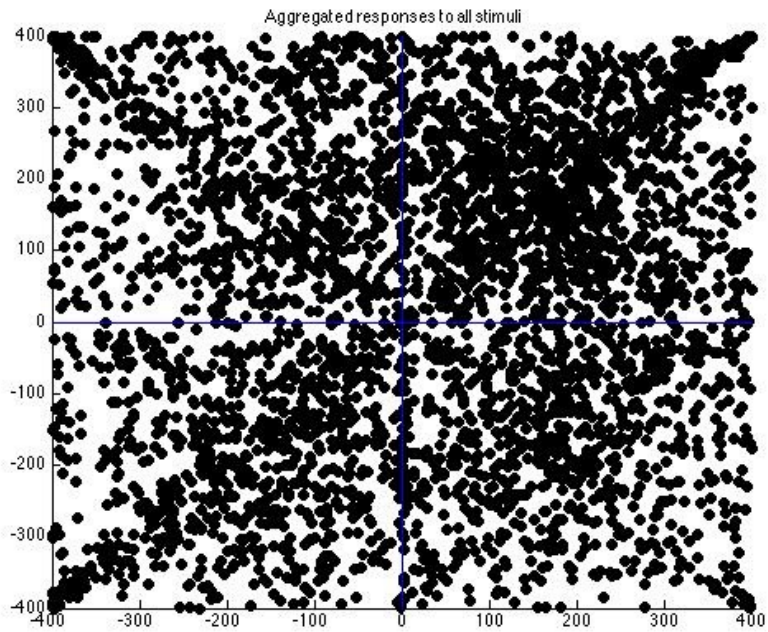


Figure 1. Aggregated responses to all stimuli.

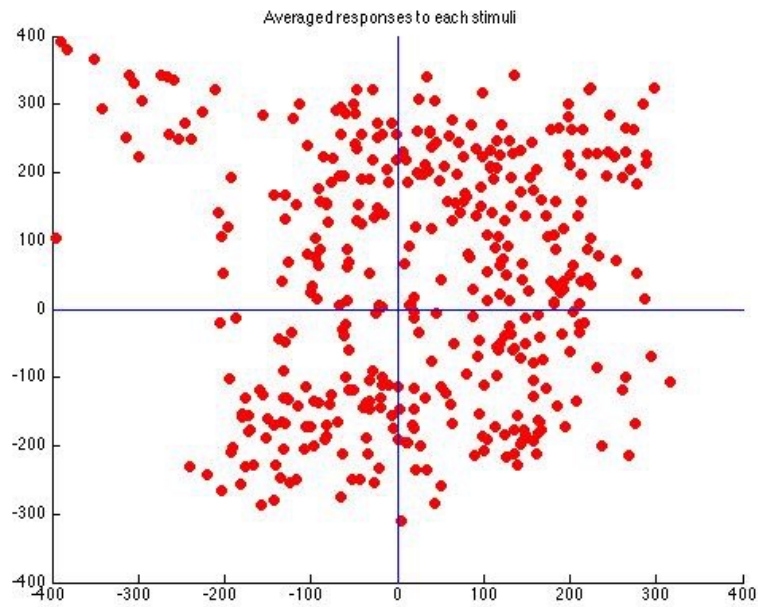


Figure 2. Averaged responses for each stimuli.

Root Mean Square

Intensity of sound is an important statistic to measure when dealing with auditory information. Although the tracks were peak-normalized to a set standard, RMS values of the stimuli still varied. These values can vary because of recording techniques, genres, or simply artistic choices in the music.

Correlation coefficients were calculated for total RMS of the stimulus and both arousal / valence. Additionally, correlation coefficients were calculated for the variance in the RMS and arousal / valence. The reasoning was that variations in the energy of the track could induce higher ratings of emotional content.

<i>N</i> = 400	Stimulus RMS	Stimulus RMS variance
Valence	-.0882 (<i>ns</i>)	-.0541 (<i>ns</i>)
Arousal	.4873 ($p < .001$)	.2671 ($p < .001$)

Table 2.1 RMS Correlations

The RMS is not correlated with ratings for the valence of the music stimuli, but an extremely strong correlation is found between ratings of arousal and the RMS of a musical stimulus. This is an excellent lead for a future study - if tracks that were rated as low arousal were brought up to a higher RMS rating via compression and limiting, it is possible that their ratings of arousal will rise in accordance.

Tempo

Faster tempos were hypothesized to correlate strongly with higher ratings of arousal. Correlation analysis revealed that songs with faster tempos were rated higher on the arousal axis. Tempo did not provide a strong correlation with valence, suggesting orthogonality on this axis.

The Presence of Lyrics

A t-test was performed on the entirety of the tested database with the presence of lyrics as the grouping variable. It was found to have an effect on both arousal and valence. However, interpreting this result is difficult given that the preponderance of music without lyrics was from either the jazz or classical genres.

df=398	t	p	95% CI
Valence	-3.0	0.0026	4.4 , 20.8
Arousal	6.2	<.0001	-20.9 , -40.3

Table 2.2 The Presence of Lyrics

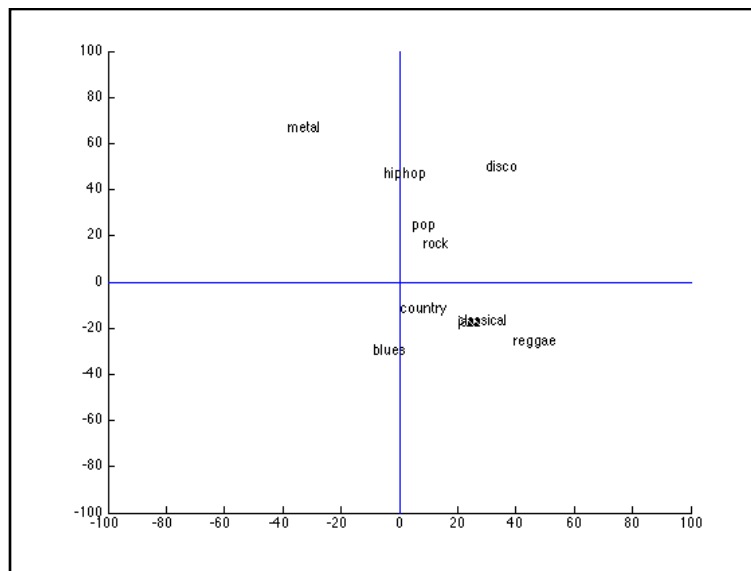


Figure 3. Average responses for each genre.

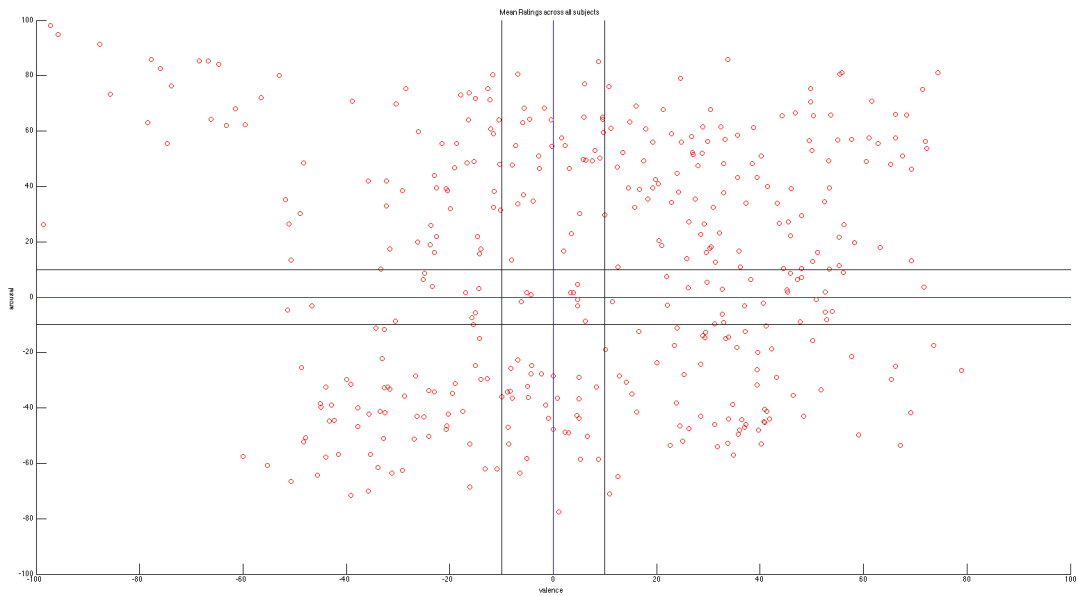


Figure 4. Divisions of stimuli for emotional content analysis

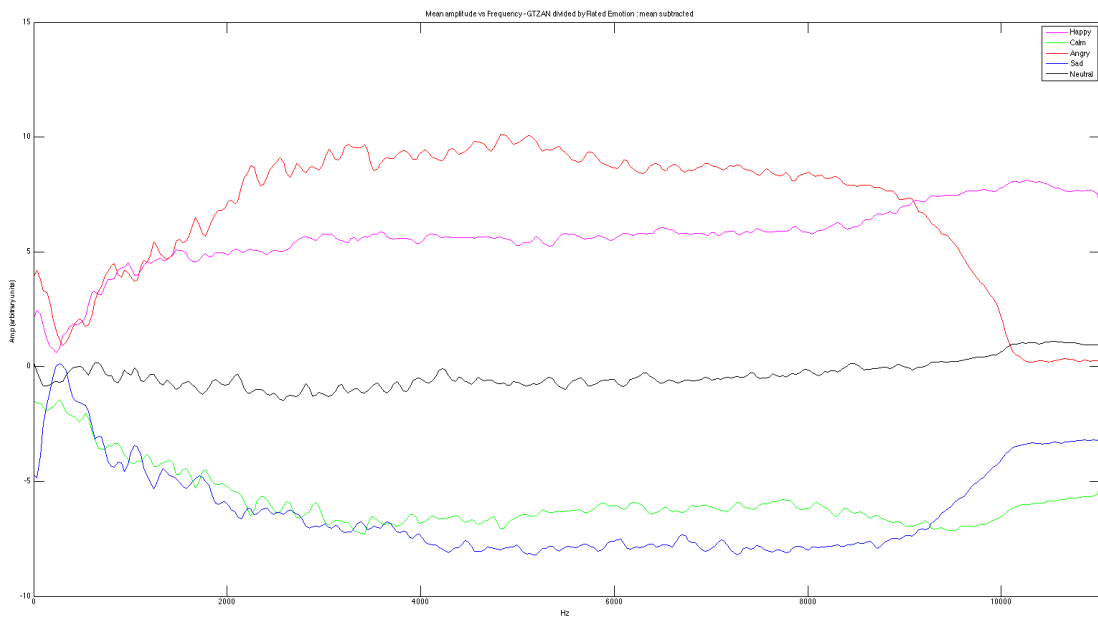


Figure 5. Spectral power for each emotional group, mean subtracted. Red: Angry; Purple: Happy; Black: Neutral; Green: Calm; Blue: Sad.

Dividing up the stimuli based on their average emotional classification reveals a marked overrepresentation of mid-band energy (i.e. 2KHz to 8KHz) for stimuli rated as angry, and a similar over-representation of the same region for happy music. Both calm and sad music have relatively weak mid-bands, leading them to sound “smoother” than their counterparts.

Relationship between variance in responses and musical attributes

Some stimuli, particularly a few exemplars from the metal genre, had extremely low variance in subject’s responses. In order to quantify what attributes are rated consistently, a series of correlational analyses of musical attributes and response variance was performed. The MIR toolbox (Lartillot 2007) was used to compute the spectral centroid, mode, pulse clarity, tempo and roughness. These features were then correlated with variance in valence, arousal as well as the means in valence and arousal. Results are shown in the table below.

Feature	r (Valence variance)	p (Valence variance)	r (Arousal variance)	p (Arousal variance)
Centroid	-0.0241	0.6313	-0.1649	9.28E-04**
Mode	-0.0021	0.9669	0.0482	0.3366
Pulse Clarity	-0.1126	0.0244	-0.0157	0.7550
Tempo	0.0305	0.5430	0.0341	0.4965
Roughness	-0.1328	0.0078*	-0.1559	0.0018*

Table 2.3 Musical features and variance correlations (*=p<.01, **=p<.001, ***=p<.0001).

Feature	r (Valence mean)	p (Valence mean)	r (Arousal mean)	p (Arousal mean)
Centroid	-0.0919	0.0663	0.6592	3.169E-51***
Mode	0.1800	2.969E-04**	-0.0283	0.5720
Pulse Clarity	0.3771	5.8E-15***	0.3447	1.343E-12***
Tempo	-0.0486	0.3323	0.0535	0.6308
Roughness	-0.2289	3.737E-06***	0.6308	8.909E-46***

Table 2.4 Musical features and rating correlations (* = $p < .01$, ** = $p < .001$, *** = $p < .0001$).

Higher degrees of roughness was correlated with low variance in valence ratings ($r = -.1328, p < .01$). A higher spectral centroid and a higher rating of roughness were also correlated with low variance in arousal ratings, with $p < .001$ and $p < .01$, respectively.

When investigating the same features and the actual ratings, several of the features showed medium to strong correlations. For the valence domain, mode and pulse clarity were positively correlated with ratings, while roughness was negatively correlated. In the arousal domain, centroid, pulse clarity, and roughness were all positively correlated with ratings, at extremely high significance levels. In all analyses, tempo showed no correlation with listener responses. Although mode strength was correlated with ratings of valence, it showed no relationship to arousal or variance in responses.

DISCUSSION

There is a marked rightward tendency in the data visualizations, that is, subjects rated the bulk of music as fundamentally pleasant. This tendency could be primarily explained in two ways.

- I. Music listening is an enjoyable activity.
- II. This dataset was skewed towards pleasant music.

It's possible that the true explanation is a combination of some of these factors. The bulk of the evidence in the literature supports the first explanation. Even sad music is pleasurable to listen to (Huron 2011). Although the dataset may contain some bias, it was selected because it is fairly well balanced across a number of genres and moods. Still, without a larger dataset to compare to, it's not easily discoverable if the stimuli selected here are indeed a perfectly representative sample of the entirety of music.

Although this study engaged listeners there were some limitations. Taking more information about genre preference may have given some information about why listeners rated the music the way that they did. Moment by moment judgments were recorded, but most subjects stopped moving their pointer after they made a judgment, so only endpoints were used in the analysis.

Some of these attributes have direct correlations to biological states. Tempo particularly maps almost directly onto the range of human heartbeat speeds. Musical tempos generally range from 50 to 220 beats per minute, and heart rates also range from about 50 to 220 beats per minute. The tempo of a song accounts for a great deal of variance in the arousal axis, (positive correlation figure). States of higher arousal involve higher heart rates as well, whether they be positively valenced states such as excitement or negatively valenced states such as anxiety and rage. The fact that tempo seems to carry this emotional weight can be easily explained by this direct correlation. People in states of high physical arousal almost always have their hearts beating quickly.

Across all Western music, there is a preference for music with a tempo around 120bpm, particularly for dance music (Moelants 2002, 2003). This also closely aligns

with the gait speed of 2Hz (Leman 2013). The relationship between music and dance is undeniable but the relationship with walking is more abstract. It is most likely that there are underlying biological factors that explain the presence of consistent speeds across these categories. Indeed, it has been found that for other animals, prefer music designed with species specific attributes (Snowdon 2013, 2015).

When investigating variance in listener responses, several features were correlated with lower variance in either the valence or arousal domain. Spectral centroid, which here serves as a proxy for the overall spectral energy, or “brightness” of the recording, was negatively correlated with lower variance in ratings of arousal. Songs with more high frequency content are less ambiguous in the arousal domain. Roughness, as computed by the MIR toolbox, is a measure of “sensory dissonance,” essentially, the amount of dissonant intervals present in the waveform. Songs with higher ratings of roughness have more dissonant sounds present in them. This feature was negatively correlated with variance in both domains; large amounts of dissonant sounds send a clear message to listeners.

Looking at correlations between these five features and overall listener responses reveal several key cues as well. Spectral centroid, pulse clarity, and roughness were all highly positively correlated with arousal ratings at extremely high significant ratings. Brighter songs, with more dissonant sounds / small pitch intervals and a strong beat will generally be rated as having more “energy” or “arousal.” This appears to hold true across all groups. Conversely, darker songs with a more ambiguous beat and a smoother overall spectral character are rated as lower in arousal. In the valence domain, mode, pulse clarity were positively correlated and roughness was negatively cor-

related. That is to say, songs with a major key and a stronger beat are rated as having a more positive valence, whereas songs with a rougher texture are rated as having a more negative valence.

Tempo was not expected to be correlated to any attributes of ratings, and this held true in the analyses. Although it is tempting to assume that songs with a faster tempo will be more arousing, this is simply not the case. Interestingly, mode was not correlated with lower response variance. It was hypothesized that songs with a stronger modal footprint (i.e. more consistent major/minor key structure) would be less ambiguous to listeners, but this was not found in the present experiment.

Although there is a large amount of variance in responses to musical stimuli, the presence of these key regularities hints at a biologically-rooted response to certain characteristics of sounds. The hypothesis that music is a form of emotional communication is supported by the fact that there is a great deal of agreement in general. Communication systems are only effective if they can consistently convey meaning to receivers. In this regard, music appears to be an effective system.

REFERENCES

1. Barrett, L. F. (1998). Discrete emotions or dimensions? The role of valence focus and arousal focus. *Cognition & Emotion*, 12(4), 579-599.
2. Bechara, A., & Damasio, A. R. (2005). The somatic marker hypothesis: A neural theory of economic decision. *Games and economic behavior*, 52(2), 336-372.
3. Chanel, G., Ansari-Asl, K., & Pun, T. (2007, October). Valence-arousal evaluation using physiological signals in an emotion recall paradigm. In 2007 IEEE International Conference on Systems, Man and Cybernetics (pp. 2662-2667). IEEE.
4. Dolcos, F., LaBar, K. S., & Cabeza, R. (2004). Dissociable effects of arousal and valence on prefrontal activity indexing emotional evaluation and subsequent memory: an event-related fMRI study. *Neuroimage*, 23(1), 64-74.

5. Huron, D. (2011). Why is sad music pleasurable? A possible role for prolactin. *Musicae Scientiae*, 15(2), 146-158.
6. Kensinger, E. A., & Corkin, S. (2004). Two routes to emotional memory: Distinct neural processes for valence and arousal. *Proceedings of the National Academy of Sciences of the United States of America*, 101(9), 3310-3315.
7. Lartillot, O. & Toivainen, P. (2007). A Matlab toolbox for musical feature extraction from audio. *International Conference on Digital Audio Effects*, pp 237-244.
8. Leman, M., Moelants, D., Varewyck, M., Styns, F., van Noorden, L., & Martens, J. P. (2013). Activating and relaxing music entrains the speed of beat synchronized walking. *PloS one*, 8(7), e67932.
9. Moelants, D. (2002, July). Preferred tempo reconsidered. In *Proceedings of the 7th international conference on music perception and cognition* (Vol. 2002, pp. 1-4).
10. Moelants, D. (2003). Dance music, movement and tempo preferences. In *Proceedings of the 5th Triennial ESCOM Conference* (pp. 649-652). Hanover University of Music and Drama.
11. Salimpoor, V. N., Benovoy, M., Longo, G., Cooperstock, J. R., & Zatorre, R. J. (2009). The rewarding aspects of music listening are related to degree of emotional arousal. *PloS one*, 4(10), e7487.
12. Salimpoor, V. N., Zald, D. H., Zatorre, R. J., Dagher, A., & McIntosh, A. R. (2015). Predictions and the brain: how musical sounds become rewarding. *Trends in cognitive sciences*, 19(2), 86-91.
13. Snowdon, C. T., & Teie, D. (2013). Emotional communication in monkeys: music to their ears?. *The Evolution of Emotional Communication: From Sounds in Nonhuman Mammals to Speech and Music in Man*, 133-151.
14. Snowdon, C. T., Teie, D., & Savage, M. (2015). Cats prefer species-appropriate music. *Applied Animal Behaviour Science*, 166, 106-111.
15. Sun, K., Yu, J., Huang, Y., & Hu, X. (2009, June). An improved valence-arousal emotion space for video affective content representation and recognition. In *2009 IEEE International Conference on Multimedia and Expo* (pp. 566-569). IEEE.

CHAPTER 3

Prosodic Information is Conveyed by Spectral and Temporal Modulations in Speech

Language comprehension is an extremely complex task. Even in optimal conditions, the amount of information that has to be parsed to fully comprehend an utterance is staggering. Adding to the complexity of the problem, most listening environments are suboptimal; background noise or music, the presence of other speakers, accents, inarticulate speakers and distance from the speaker are among the factors complicating the task. After locking into an auditory stream, meaning can be built up from the many building blocks of speech (Hickok 2007). Despite all of these complexities, humans are quite good at understanding degraded speech (Ahissar 2001; Obleser 2011).

The Hickok model (2007) presents a unique combinatorial view of what parts of the brain are serving the many sub functions necessary for speech comprehension. This model divides the task into two processing streams, modeled in part after the dorsal/ventral model of visual processing. There is accumulating evidence for a third stream that focuses on processing emotional content in auditory signals (Blood 1999, 2001; Rogalski 2011; Sridharan 2007). Connections between auditory, motor and limbic regions suggest that the role of music is deeply linked to sound, movement and emotion (Zatorre 2007).

Not all parts of the speech waveform have equal informational value (Rosen 1992; Cassell 1999). Much work has been done in the past to uncover the necessary components of speech for comprehension. Speech-in-noise is a common paradigm

that has much biological relevance (Kalikow 1977; Obelser 2007). Investigations into audiovisual speech comprehensions, namely into the McGurk effect have found that the visual system can interact with the auditory system to alter comprehension but that the motor system does not effect comprehension (Matchin 2007).

The modulation power spectrum is a novel technique used to investigate the relative contributions of modulations in the spectral and temporal domains. This method was developed by Elliot & Theunissen (2009). A technique developed by Venezia et al (2016, in press) combined a visual filtering paradigm, “bubbles,” along with using speech filtered along the two axes of the the modulation power spectrum. The “bubbles” paradigm was a technique to investigate which parts of a face were most important for making judgments in recognition tasks.

Merging these two techniques, thus creating “bubbles” in the modulation power spectrum, allowed for an investigation into necessary regions of the MPS for speech intelligibility. Modulations in the spectral domain at around 1.5 cyc/Khz were found to be most important for intelligibility, which modulations in the temporal domain at less than 4-8 Hz were most important for intelligibility. Interestingly, the 4-8 Hz range is similar to the average musical note duration.

In the present experiment, the technique developed by Venezia et al (2016) was again applied to speech, but rather than investigating intelligibility, the ability of participants to differentiate utterances based on their emotional content was investigated. Jabberwocky sentences, where all nouns and verbs are replaced with phonemically appropriate nonsense words have been used in a number of investigations into speech comprehension (Rogalski 2011; Federenko 2015; Bonhage 2015). They are well suited

to the task at hand because they do not carry the emotional weight or associations that real words do. A behavioral paradigm was used to elucidate the necessary modulations for effective conveyance of emotional content via speech.

METHOD

Stimuli

The stimuli used in the first experiment were two sets of “jabberwocky” sentences, spoken by a female drama student recruited from UCI. There were 40 sentences recorded, each with five different intended emotions: happy, angry, sad, calm and neutral. The student was instructed to convey the emotions as clearly as possible while emphasizing a realistic portrayal of speech. Using speech from naturally occurring situations provides a more realistic impression of the speech used, but becomes nearly impossible to get all the categories necessary to convert even the limited range of emotions used here. Using a student well versed in acting is a good compromise between naturalism and experimental control.

The wav files were then normalized to an arbitrary RMS standard, truncated to 3 seconds each and bounced to 16-bit, 22.05Khz sampling rate, mono .wav file. Subsequently, the filtered stimuli were generated with a varying amount of “bubbles” in the temporal and spectral modulation space. Regions that fall in the “bubbles” were unchanged by the filtering procedure. Using multiple tokens of each filtered stimuli allowed for the entirety of the MPS to be mapped out for each subject.

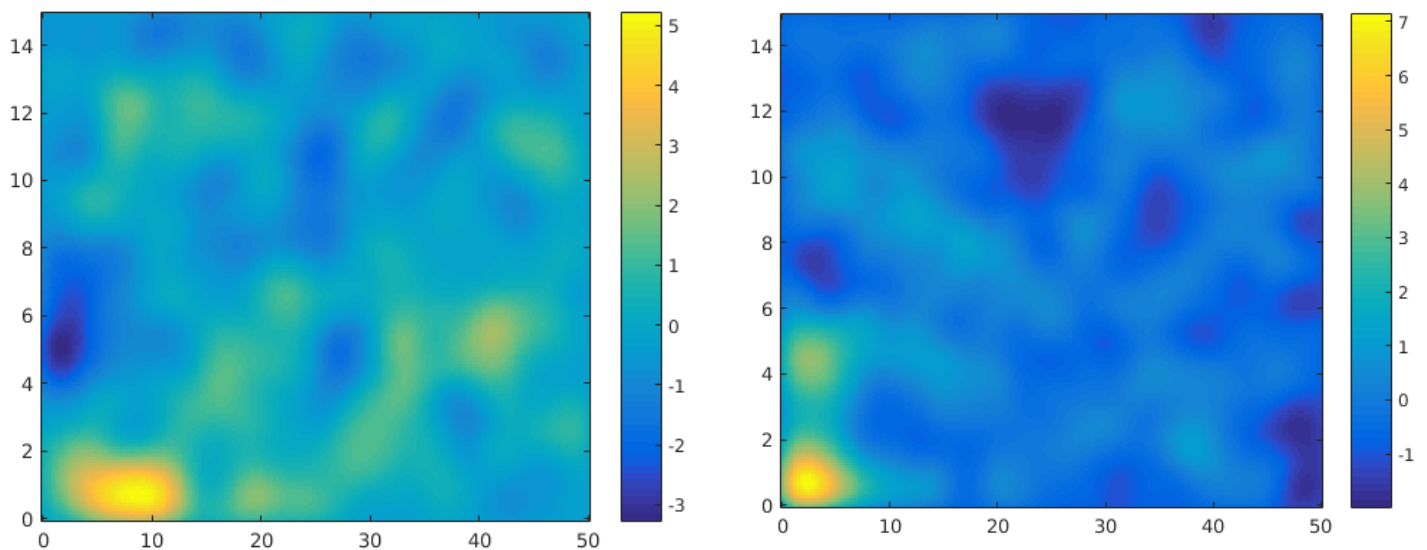
Subjects were instructed to state if the emotion contained in the stimuli was happy or angry in nature, and to provide a confidence rating from 1 (complete guess) to 9

(absolute certainty) for each stimulus presentation. An adaptive staircase procedure was used to control for the differing ability levels of the subjects. Only performance was used to change the difficulty during the experiment - confidence ratings were only used in the data analysis.

Participants

Participants were recruited from the UCI undergraduate community via the Sona Systems program. 20 subjects were used in total (12 female).

RESULTS



Figures 1 and 2: Angry map (L) and Happy map (R)

In order to analyze the results, an analysis paradigm wherein confidence ratings were used to weight the responses. Maps from the incorrect responses were subtracted from the maps for correct responses in order to characterize what regions of the

MPS were most important for correct transference of the emotional content of the stimulus.

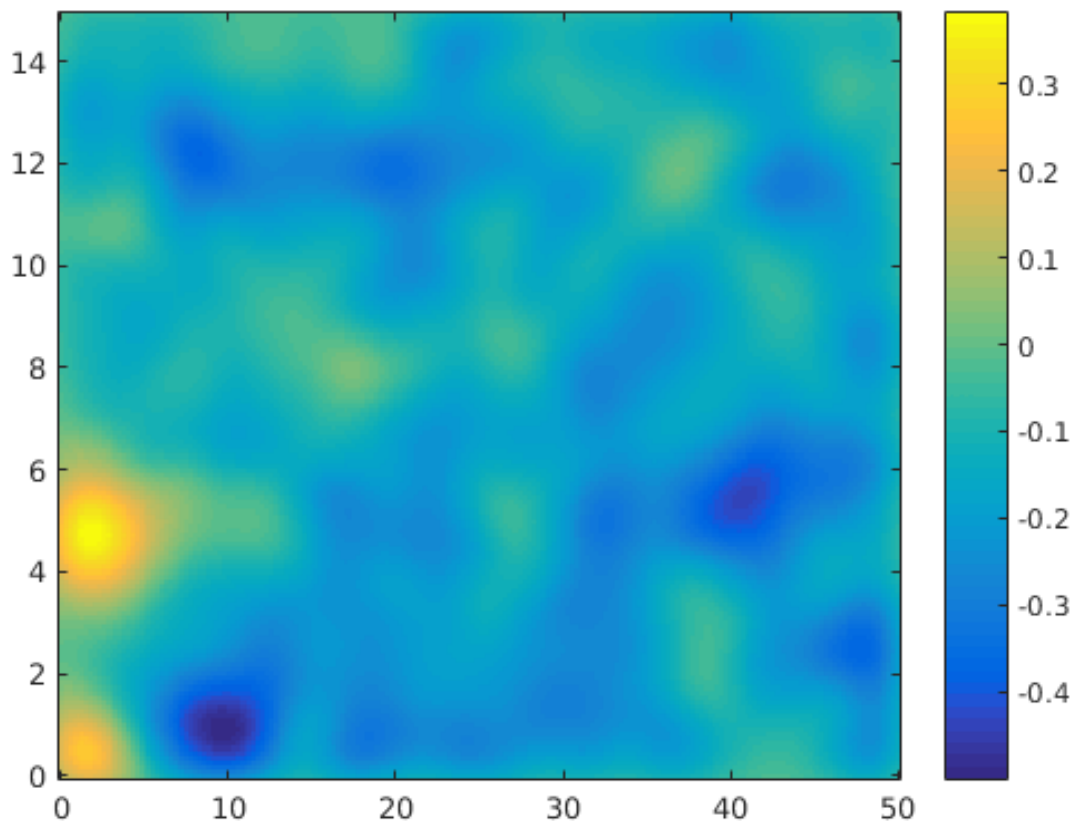


Figure 3. Difference map for the two conditions

DISCUSSION

Intriguingly, the two images produced via the analysis paradigm largely mirrored each other. This could have been a result of only having two types of stimuli. In order for a stimulus to be perceived as “happy,” the region of the MPS associated with pitch need to be intact. Happy speech tends to have a high degree of pitch movement, and without that pitch movement present, it appears that the primary cue is lost. Conversely, angry

speech has less pitch variation and a peak in the MPS regions associated with roughness.

It was hypothesized that the maps could be applied to speech with a neutral prosody to modulate the perception of the emotional content of those utterances, but in pilot experiments, this failed to show any effect.

Further investigations

In a short pilot follow-up to this experiment, an attempt to use 5 different emotions (happy, angry, sad, calm and neutral) as stimuli was made, but subjects were unable to perform at an adequate level with the paradigm remaining relatively similar to the present experiment. The rationale for using happy and angry as the primary types of emotions was that they would be closely matched for energy, even before RMS normalization. The differences between calm, neutral, and sad speech were thought to be too subtle to fully investigate without an inappropriately large time commitment from the participants.

Different speakers vary in the ways they display emotions in their speech. Different contexts will change the way people emote as well - different behavioral standards apply to the office, home, performances and other environments. Different speakers will have different baseline levels of expression as well. Tracking these differences across environments is a complex task for a listener, and yet, even when stimuli are severely degraded, subjects were able to correctly interpret the intended emotional meaning these utterances.

Musical training

Detecting signals from others is a core ability of social animals. In humans, these cues occur in the language domain, and in body language. There is substantial evidence that musical training aids in the auditory portions of this task. Trained musicians have stronger frequency-following responses in the brainstem (Musacchia 2007).

Conclusion

The Modulation Power Spectrum is a novel tool for visualizing the content of speech and other sounds. In this experiment, some of the fundamental parts of prosody were revealed by the technique. Although a huge space is covered by the visualization technique, only small parts of the area are relevant for interpreting sounds as happy or angry.

REFERENCES

1. Ahissar, E., Nagarajan, S., Ahissar, M., Protopapas, A., Mahncke, H., & Merzenich, M. M. (2001). Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. *Proceedings of the National Academy of Sciences*, 98(23), 13367-13372.
2. Blood, A. J., Zatorre, R. J., Bermudez, P., & Evans, A. C. (1999). Emotional responses to pleasant and unpleasant music correlate with activity in paralimbic brain regions. *Nature neuroscience*, 2(4), 382-387.
3. Blood, A. J., & Zatorre, R. J. (2001). Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion. *Proceedings of the National Academy of Sciences*, 98(20), 11818-11823.
4. Bonhage, C. E., Mueller, J. L., Friederici, A. D., & Fiebach, C. J. (2015). Combined eye tracking and fMRI reveals neural basis of linguistic predictions during sentence comprehension. *Cortex*, 68, 33-47.
5. Elliott, T. M., & Theunissen, F. E. (2009). The modulation transfer function for speech intelligibility. *PLoS comput biol*, 5(3), e1000302.
6. Fedorenko, E., Hsieh, P. J., & Balewski, Z. (2015). A possible functional localiser for identifying brain regions sensitive to sentence-level prosody. *Language, cognition and neuroscience*, 30(1-2), 120-148.
7. Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, 8(5), 393-402.
8. Kalikow, D. N., Stevens, K. N., & Elliott, L. L. (1977). Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *The Journal of the Acoustical Society of America*, 61(5), 1337-1351.
9. Matchin, W., Groulx, K., & Hickok, G. (2014). Audiovisual speech integration does not rely on the motor system: evidence from articulatory suppression, the McGurk effect, and fMRI. *Journal of cognitive neuroscience*, 26(3), 606-620.
10. Musacchia, G., Sams, M., Skoe, E., & Kraus, N. (2007). Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *Proceedings of the National Academy of Sciences*, 104(40), 15894-15898.
11. Obleser, J., Wise, R. J., Dresner, M. A., & Scott, S. K. (2007). Functional integration across brain regions improves speech perception under adverse listening conditions. *The Journal of Neuroscience*, 27(9), 2283-2289.
12. Obleser, J., & Kotz, S. A. (2011). Multiple brain signatures of integration in the comprehension of degraded speech. *Neuroimage*, 55(2), 713-723.
13. Rogalsky, C., Rong, F., Saberi, K., & Hickok, G. (2011). Functional anatomy of language and music perception: temporal and structural factors investigated using functional magnetic resonance imaging. *The Journal of Neuroscience*, 31(10), 3843-3852.

14. Salimpoor, V. N., Zald, D. H., Zatorre, R. J., Dagher, A., & McIntosh, A. R. (2015). Predictions and the brain: how musical sounds become rewarding. *Trends in cognitive sciences*, 19(2), 86-91.
15. Sridharan, D., Levitin, D. J., Chafe, C. H., Berger, J., & Menon, V. (2007). Neural dynamics of event segmentation in music: converging evidence for dissociable ventral and dorsal networks. *Neuron*, 55(3), 521-532.
16. Venezia, J. H., Thurman, S. M., Matchin, W., George, S. E., & Hickok, G. (2016). Timing in audiovisual speech perception: A mini review and new psychophysical data. *Attention, Perception, & Psychophysics*, 78(2), 583-601.
17. Zatorre, R. J., Chen, J. L., & Penhune, V. B. (2007). When the brain plays music: auditory–motor interactions in music perception and production. *Nature reviews neuroscience*, 8(7), 547-558.

CHAPTER 4

Functional neuroanatomical differences in Music and Language Processing

Introduction

Music and language share several surface-level characteristics; both convey information between individuals and although they primarily exist in the auditory domain, they also exist in other modalities. They employ overlapping brain regions and have a similar series of necessary computations for comprehension, including acoustic analysis, rhythm tracking, and long-range dependencies [1-3]. Music has a set of grammar-like properties, and those properties can vary among different musical cultures in a similar way to surface variations in linguistic grammars across cultures.

The presence of dissonant musical pitches can be viewed as analogous to grammatical violations in language. The presence of dissonant pitches generates mismatch negativities in similar brain regions as linguistic grammatical violations (Patel 1997). The key signature of a piece can be thought of as a probability map for the relative occurrence rate for musical pitches. Dissonant notes are those that have a lower probability of occurrence when compared to consonant notes, and thus their occurrence generates the MMN. There are several candidate brain regions for this computation—Broca's area is often mentioned as the prime candidate for processing these implicit "rules" of music.

Indeed, music and language are commonly described as sharing a large degree of neural resources. There is a particularly large degree of overlap at low levels of auditory processing (i.e. prior to A1). However, in a recent study by Rogalsky et al (2011), several regions of non-overlap were seen for music and language. That study also re-

vealed a marked lack of activation in Broca's area for sentential processing or phrase processing in music and argued that Broca's area serves as a hub for a more general short-term memory function. In that study, single note melodies were used to drive activation in the brain and it has been argued that those musical stimuli used were insufficiently complex to effectively drive activation in Broca's area. In order to further examine that potential explanation, the stimuli used here were piano chord progressions adapted from an experiment by Patel 1998 [36], where dissonant chords in musical stimuli drove activation in left frontal regions as measured by EEG. Both the present experiment and Rogalsky (2011) used passive listening as the task as the goal of these experiments was to investigate natural processing. Rather than processing syntactical variations in the stimuli, we posit that Broca's area is involved in an error detection task.

Method

Subjects

20 right-handed subjects (as measured by the Edinburgh inventory) native English speakers (X female, Y male, average age) were used in this study. Subjects had minimal formal musical training on an instrument (average of <3 years) and had never studied music theory. Subjects were recruited from the UCI population and surrounding area as per IRB guidelines. All subjects were free of neurological disease and gave informed consent under the University of California, Irvine's (UCI) Institutional Review Board.

Stimuli

The stimuli used were organized into 20 second blocks, with one type of stimuli per block. In all, 4 types of stimuli were used: jabberwocky sentences, scrambled jabberwocky sentences, piano melodies and piano melodies with intermittent out-of-key chords. Rest periods (jittered around 12s) were interspersed with the trial blocks to collect baseline activation.

The same jabberwocky sentences were used as in Rogalsky (2011). These sentences had grammatical structure but were devoid of semantic meaning (e.g. “it was the glandar in my nederop”). Scrambled sentences were created by randomly rearranging the words in the original recordings.

The musical stimuli were piano chord progressions used in Patel et al (1997). Melodies within the same key signature were used to form blocks so that there was no tonic center change within each block. Within normal musical blocks, all the chords present fit with rules of tonality. Within musical blocks with violations, there were several out-of-key chords that would appear at various points in the chord progression after the key was established. We only used out-of-key chords that were as far as possible from the tonic center (i.e. the tritone/ augmented 5th) to ensure that subjects would be able to notice them.

Stimulus Presentation

Stimuli were presented with the Cogent toolbox for MATLAB on an IBM ThinkPad laptop running Windows XP. Subjects were given sound-isolating, in-ear, non-ferro-

magnetic headphones (Sensimetrics Model S14) and subsequently fitted with over-the-ear covers (Pro Ears ultra 26) to further isolate them from scanner noise. With these measures in place, the estimated total scanner noise reduction was approximately -45dB. The volume of the stimuli was also calibrated for each subject in the scanner to ensure that stimuli were easily audible over the scanner noise without being unpleasantly loud.

Stimulus presentation order was randomized on the level of blocks. Subjects were instructed to attend to the stimuli and maintain their focus on the fixation cross. As there was no active task associated with the experiment, subjects were checked in on between each fMRI run to confirm that they were able to complete the task appropriately (approximately every 3 min).

Two subjects had to be excluded from the analysis due to severe inconsistencies in image alignment.

fMRI Data Acquisition

Data were collected on the 3T Phillips Achieva MR scanner at the UCI Research Imaging Center. T2-weighted functional scans (Echo-Planar Imaging, 43 slices, slice thickness of 2.5mm .5mm gap between slices, TR = 2s, TE = 25ms, FOV of 180x220x128, voxel size = 2.5mm³, SENSE acceleration factor = 2.4). 10 functional runs of 96 scans per run were completed for a total of 960 scans per subject. A continuous scanner paradigm was used rather than a sparse sampling one, in order to increase the power of the study and fit with the block design.

After the functional scans, a T1-weighted structural image was acquired (140 axial slices; slice thickness = 1 mm; field of view = 240 mm; matrix 240 × 240; repetition time = 11 ms, echo time = 3.55 ms; flip angle = 18°; SENSE factor reduction 1.5 × 1.5).

Data Processing and Analysis

Data was analyzed with AFNI on a PowerMac G4 running Mac OS X (v. 10.6.8). Preprocessing, alignment and analysis were run on UNIX with the AFNI command library. Preprocessing in AFNI consisted of aligning the images to individual anatomicals, blurring with 3dmerge at 6.0mm and masking with 3dAutoMask. Images were then aligned to Talarach space, using the TT_N27 dataset, for group analysis using AFNI's @auto_tlrc program.

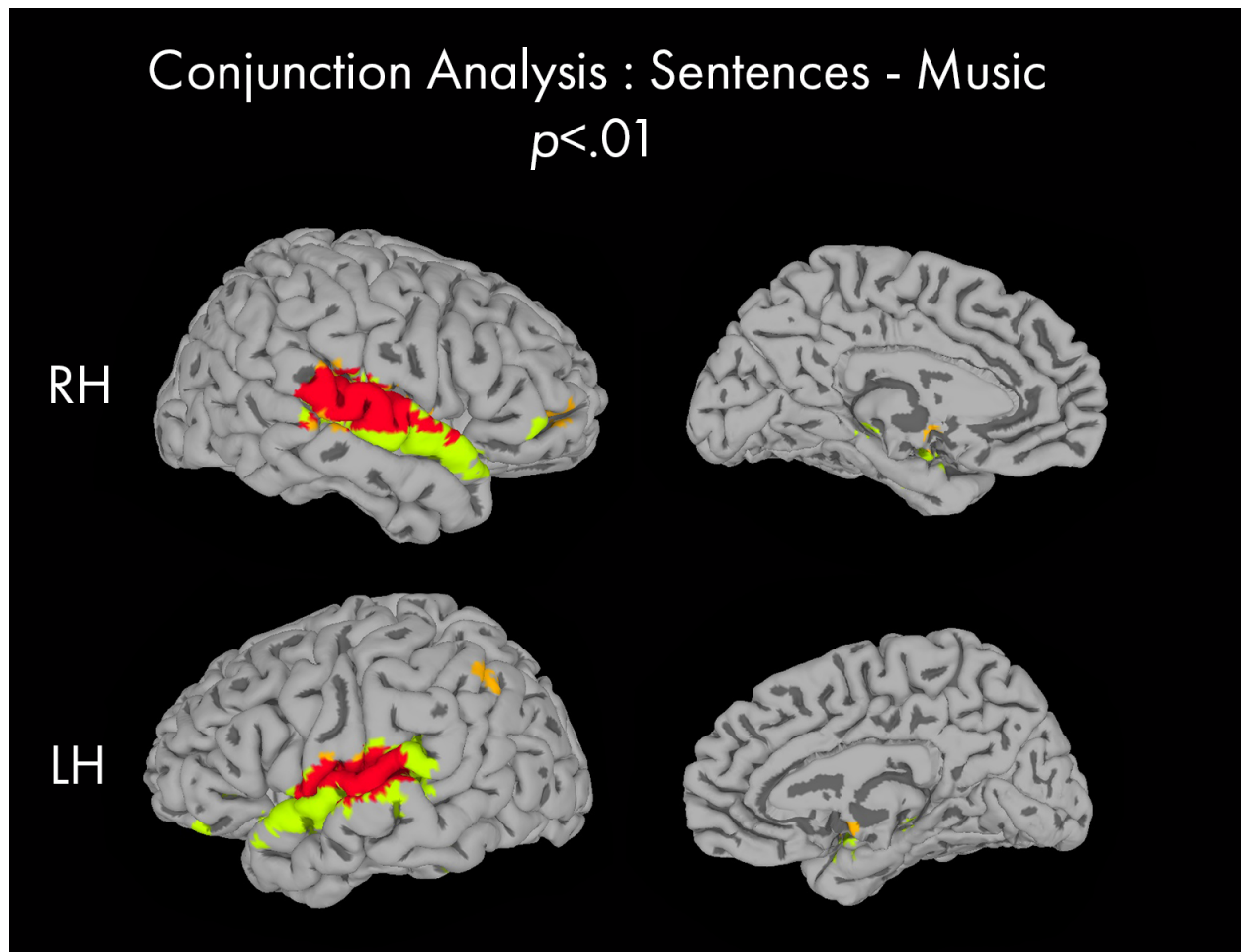
Analyses consisted of an omnibus ANOVA, several paired contrasts (using AFNI's 3dttest++ function) and a series of conjunction analyses. The planned paired contrasts were music vs musical violations, both at the block level and at the level of individual TRs, sentences vs scrambled sentences, and language vs music generally. Contrasts were performed at $p < .005$, uncorrected, except for sentences vs music, which was studied at $p < .001$, uncorrected. Data was limited to clusters of at least 20 voxels to limit false positives. Images were then produced with AFNI and SUMA.

Results

A series of paired contrasts designed to highlight functionally distinct areas active during the different stimuli types were used as the primary analyses. These tests re-

vealed differences in activation in auditory regions, several frontal areas as well as parts of the limbic system.

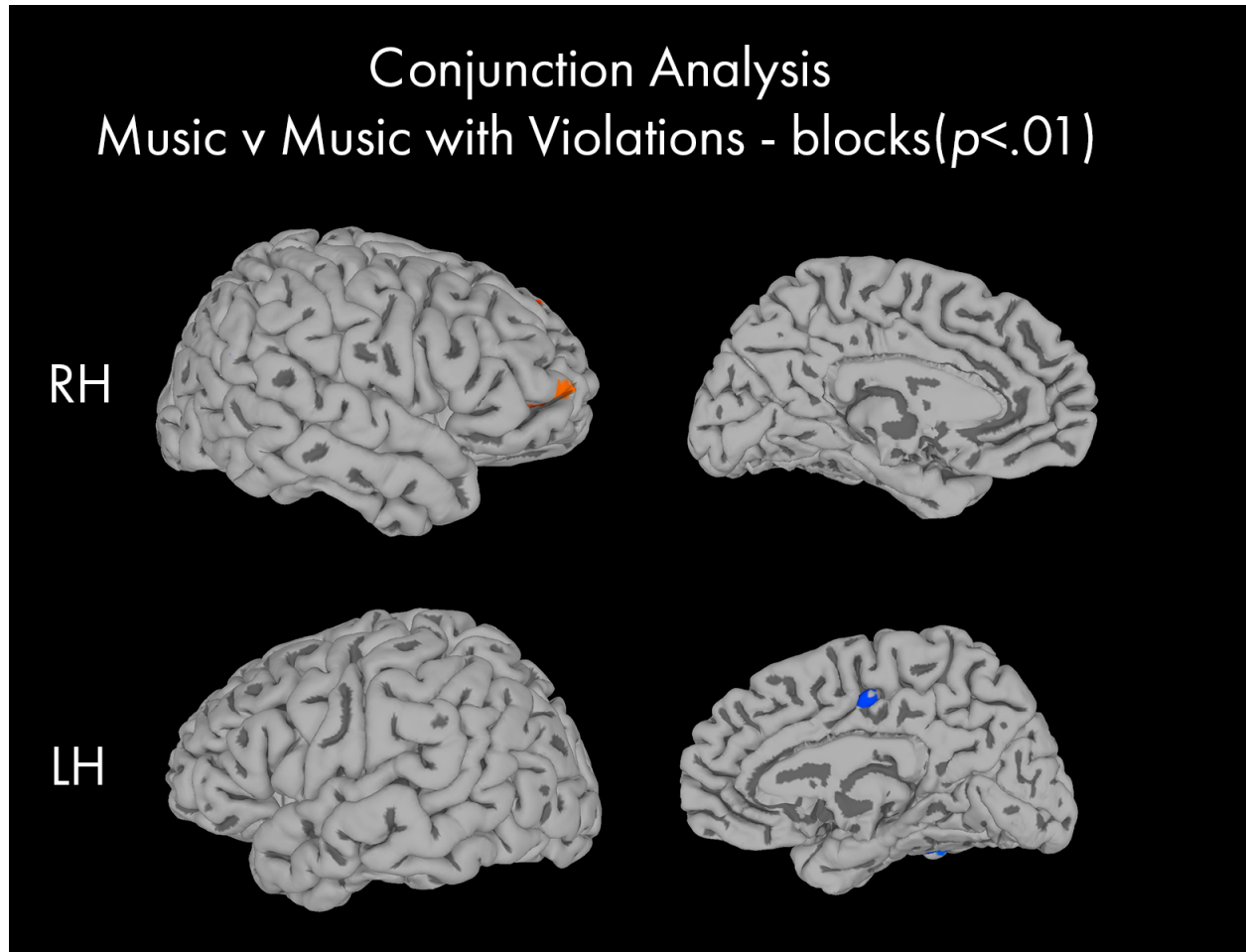
Sentences vs Music



Generally, language stimuli elicits more activity in a large swath of primary auditory regions (-61,-14,6) as well as BA 22 bilaterally (61, -21, 2). Music elicits more activation in a range of locations, including right MFG (28,15,50), right SFG (4,47,37), left

SFG (-14,39,39), left precuneus (-38,-75,33), and a large central cluster comprised of anterior cingulate, BA24, BA32, and rBA10 with peak activation at (4,33,-1).

Music vs Music with Violations



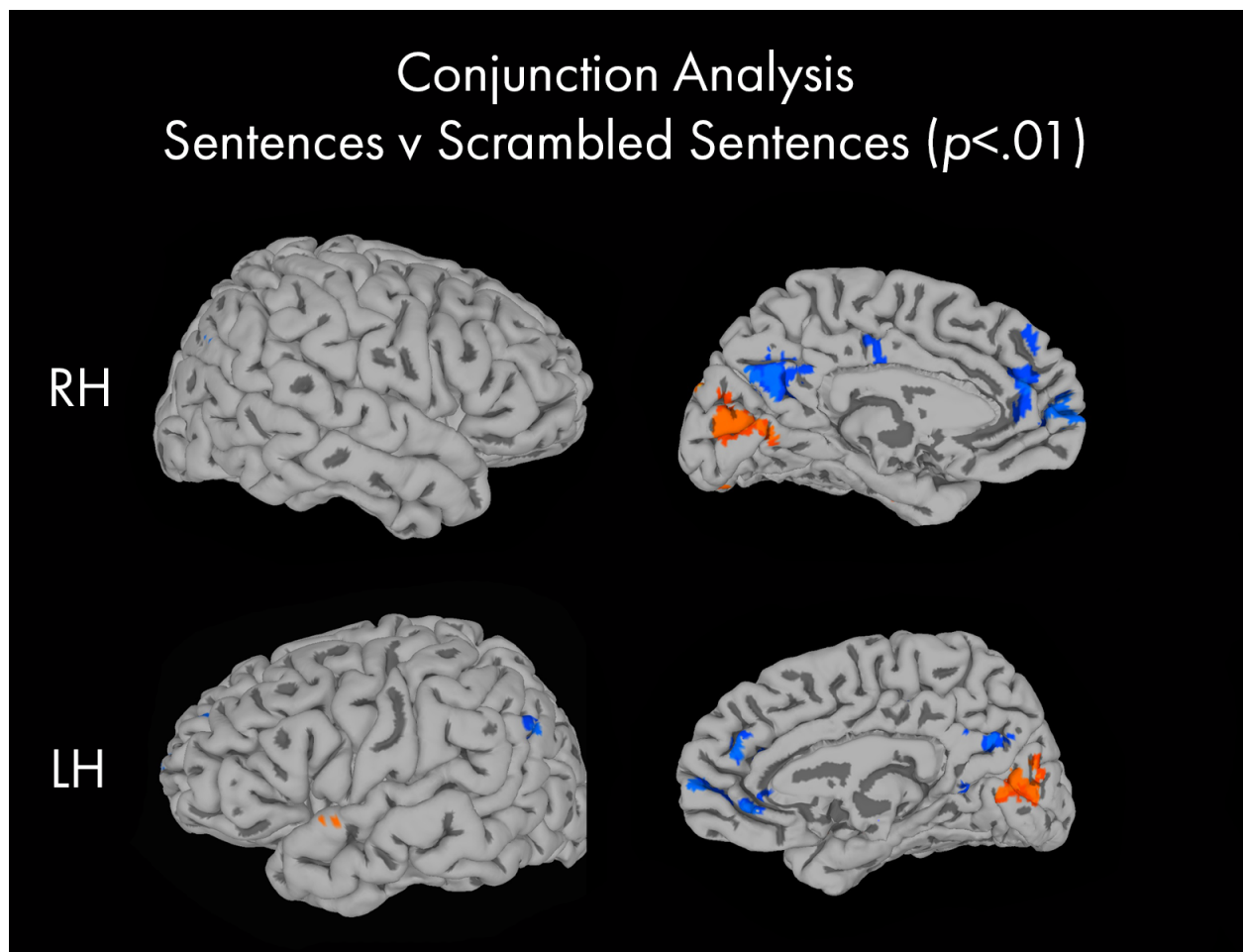
Musical violation blocks elicited greater activation in right BA20 (51,-43,-22), left cuneus (-9,-86,31), left precuneus (-48,-56,41), right cingulate gyrus (6,-16,41). Right

BA8 (15,47,40) and BA10 (39,53,8) responded more strongly to consonant music blocks than musical violation blocks.

Music vs Music with Violations (Event-related)

When analysis was run at the level of TRs rather than blocks, several cerebellar clusters, left BA13 (-39,-10,16), and a cluster in the right MFG (28,48,11) activated more for consonant music, and two cerebellar clusters activated more for dissonant chords, centered at (24,-72,-29) and (51,-47,-23).

Sentences vs scrambled sentences



The contrast of sentences vs scrambled sentences revealed a large number of differences at the $p < .005$ level. Grammatically intact sentences elicited more activity in several clusters in L BA37 (-41,-56,-19), L Declive (-17,-82,-20), R Cuneus (2,-78,7), and R BA 34 (13,1,-13). Several clusters activated more strongly for scrambled sentences, including bilateral MFG, including Brodmann Areas 9, 10, 31, 39, L posterior Cingulate (-16,-50,16), and bilateral Cingulate Gyrus (-4,-55,21).

Conjunction Analyses

Conjunction analysis ($p < .01$) shows that in earlier parts of the auditory pathway, activation is similar for language (green) and music (orange) (shared activation is shown in red). However, outside of primary auditory regions, there is minimal overlap in the two types of conditions. In the right anterior temporal gyri, there is adjacent, non-overlapping activation for language and music. There is activation for sentences in the left anterior pole, replicating a result for grammatically intact sentences. Sentences activated both L BA45 and BA47, but markedly, no activation in these areas was seen for music. There is greater activation in R BA10 (41,44,3) and L BA13 (-40,-13,16) for music. Both of these regions are associated with pitch processing in other studies.

Discussion

Similarities and Differences in Processing

Differential brain activation for music and language show that different networks are employed for the two systems. Overlap certainly exists in the early stages of the process stream, e.g. Heschl's gyrus, but a number of non-overlapping higher-level brain regions also exist, such as the IFG and temporal pole. In general, music activated frontal and medial regions more than language [18]. The increased BOLD signal in frontal regions suggests an a heavier cognitive load in domain-general regions.

Specialization effects

Although most humans are expert users of language, not every human is an expert musician. Receptive musical rules are acquired automatically, in a similar fashion to language [27, 28]. Most individuals are able to parse the various necessary parts of music (i.e. harmony, rhythm and melody). Although they may not be able to explicitly describe aspects of the music, they can appreciate the differences in the gestalt of the piece. The fact that non-expert musicians were selected could have influenced the results of this study. Much of the activation seen during musical stimuli is found in medial and frontal regions, which are generally believed to be domain-general (i.e. MFG/SFG, cingulate gyri and precuneus regions).

The default mode network is a possible alternate explanation for this midline activity. To control for this, we performed conjunction analyses while controlling for resting activation (i.e. resting activation was subtracted from activation during stimulus presen-

tation). Even in these cases, activity is seen in both anterior cingulate and angular gyrus, both areas associated with several forms of domain-general processing.

Consonance and the Limbic System

Humans generally rate refer frequency relationships with small, whole-number fractions (i.e. the perfect fifth, which has a ratio of 3/2) as more pleasant and analogously, match dissonant intervals with ratings of unpleasantness (Koelsch 2006). There are two primary systems at work here: pitch pattern recognition in the frontal cortex (i.e. several loci in BA10 and BA13) which activate more strongly to consistent pitch relationships and certain regions of the limbic system, particularly the parahippocampal gyrus and cingulate gyri which activate more strongly to the presence of dissonant chords. In non-musicians, pitch relationships are built up in frontal regions and domain-general error detection mechanisms handle the processing of mismatched pitches. A further inquiry would be needed to determine if this relationship holds in trained musicians, or if the regions involved in processing will change with sufficient training.

Broca's Area

Broca's Area has traditionally been viewed as a computational hub for sentential processing, however Rogalsky (2011) and the present study call this into question [5]. When investigating the contrast of sentences and scrambled sentences, no difference in Broca's area activation is seen, suggesting that the cognitive load on that region is iden-

tical for the two stimuli types, despite the fact that the intact sentences employ natural grammar and the scrambled sentences do not. Musical stimuli did not activate Broca's Area either, despite the presence of relatively complex musical cues. Some regions anterior to Broca's area did activate for musical phrase structure building, suggesting that perhaps activation seen in past studies. Taken together, these negative results call into question claims that Broca's Area is a computational hub for syntactical processing.

REFERENCES

1. Patel, A.D., *Language, Music, Syntax and the Brain*. Nature Neuroscience, 2003. **6**(7): p. 674-681.
2. Patel, A.D., *Music, Language and the Brain* 2010: Oxford University Press. 520.
3. Patel, A.D., et al., *Processing Prosodic and Musical Patterns: A Neuropsychological Investigation*. Brain and Language, 1998. **61**: p. 123-144.
4. Rogalsky, C., et al., *Functional Anatomy of Language and Music Perception: Temporal and Structural Factors Investigated Using Functional Magnetic Resonance Imaging*. Journal of Neuroscience, 2011. **31**(10): p. 3843-3852.
5. Novick, J.M., J.C. Trueswell, and S.L. Thompson-Schill, *Cognitive Control and parsing: Reexamining the role of Broca's area in sentence comprehension*. Cognitive, Affective & Behavioral Neuroscience, 2005. **5**(3): p. 263-281.
6. Kryszak, J.K., *Critical Review: Do individuals with musical training have enhanced brainstem encoding of linguistic pitch compared to those not musically trained?*, in *Clinical Science* 2008, UWO. p. 4.
7. Wong, P.C.M., et al., *Musical experience shapes human brainstem encoding of linguistic pitch patterns*. Nature Neuroscience, 2007.
8. Bidelman, G.M., J.T. Gandour, and A. Krishnan, *Cross-domain Effects of Music and Language Experience on the Representation of Pitch in the Human Auditory Brainstem*. Journal of Cognitive Neuroscience, 2010. **23**(2): p. 425-434.
9. Schmithorst, V., *The effect of musical training on music processing: a functional magnetic resonance imaging study in humans*. Neuroscience Letters, 2003. **348**(2): p. 65-68.
10. Bidelman, G.M. and A. Krishnan, *Neural Correlates of Consonance, Dissonance, and the Hierarchy of Musical Pitch in the Human Brainstem*. Journal of Neuroscience, 2009. **29**(42): p. 13165-13171.
11. Flanagan, J.R., et al., *Prediction Precedes Control in Motor Learning*. Current Biology, 2003. **13**: p. 146-150.
12. Janata, P. and S.T. Grafton, *Swinging in the brain: shared neural substrates for behaviors related to sequencing and music*. Nature Neuroscience, 2003. **6**(7): p. 682-687.
13. Schack, T. and F. Mechsner, *Representation of motor skills in human long-term memory*. Neuroscience Letters, 2006. **391**(3): p. 77-81.
14. Eck, D. and S.K. Scott, *New Research in Rhythm Perception and Production*. Music Perception: An Interdisciplinary Journal, 2005. **22**(3): p. 365-369.
15. Parsons, L.M., *Exploring the Functional Neuroanatomy of Music Performance, Perception and Comprehension*. Annals of the New York Academy of Sciences, 2006: p. 211-231.

16. Povel, D.J., *A Theoretical Framework for Rhythm Perception*. Psychological Research, 1984. **45**: p. 315-337.
17. Krohn, K.I., et al., *Neural Representations Of The Hierarchical Scale Pitch Structure*. Music Perception, 2007. **24**(3): p. 281-296.
18. Baeck, E., *The neural networks of music*. European Journal of Neurology, 2002. **9**: p. 449-456.
19. Jacome, D.E., *Aphasia with elation, hypermusia, musicophilia and compulsive whistling*. Journal of Neurology, Neurosurgery, and Psychiatry, 1984. **47**: p. 308-310.
20. Kotz, S.A., et al., *On the lateralization of emotional prosody: An event-related functional MR investigation*. Brain and Language, 2003. **86**(3): p. 366-376.
21. Kotz, S.A. and S. Paulmann, *When emotional prosody and semantics dance cheek to cheek: ERP evidence*. Brain Research, 2007. **1151**: p. 107-118.
22. Wartenburger, I., et al., *The processing of prosody: Evidence of interhemispheric specialization at the age of four*. NeuroImage, 2007. **34**(1): p. 416-425.
23. Rauschecker, J.P., *Cortical processing of complex sounds*. Current Opinion in Neurobiology, 1998. **8**: p. 516-521.
24. Schön, D., C. Magne, and M. Besson, *The music of speech: Music training facilitates pitch processing in both music and language*. Psychophysiology, 2004. **41**(3): p. 341-349.
25. Thompson, W.F., E.G. Schellenberg, and G. Husain, *Decoding Speech Prosody: Do Music Lessons Help?* Emotion, 2004. **4**(1): p. 46-64.
26. Balkwill, L., W.F. Thompson, and R. Matsunaga, *Recognition of emotion in Japanese, Western and Hindustani music by Japanese listeners*. Japanese Psychological Research, 2004. **46**(4): p. 337-349.
27. Brattico, E., et al., *Neural Discrimination of Nonprototypical Chords in Music Experts and Laymen: An MEG Study*. Journal of Cognitive Neuroscience, 2008. **21**(11): p. 2231-2244.
28. Levitin, D.J. and A.K. Tirovolas, *Current Advances in the Cognitive Neuroscience of Music*. Annals of the New York Academy of Sciences, 2009. **1156**(1): p. 211-231.
29. Harwood, D.L., *Universals in Music: A Perspective from Cognitive Psychology*. Ethnomusicology, 1976. **20**(3): p. 521-533.
30. Scherer, K.R., R. Banse, and H.G. Wallbott, *Emotion Inferences from Vocal Expression Correlate Across Languages and Cultures*. Journal of Cross-Cultural Psychology, 2001. **32**(1): p. 76-92.
31. Margulis, E.H., et al., *Selective neurophysiologic responses to music in instrumentalists with different listening biographies*. Human Brain Mapping, 2009. **30**(1): p. 267-275.

32. Warren, J.E., R.J.S. Wise, and J.D. Warren, *Sounds do-able: auditory–motor transformations and the posterior temporal plane*. Trends in Neurosciences, 2005. **28**(12): p. 636-643.
33. Hickok, G., *The functional neuroanatomy of language*. Physics of Life Reviews, 2009. **6**(3): p. 121-143.
34. Hickok, G., et al., *Auditory-Motor Interaction Revealed by fMRI: Speech, Music and Working Memory in Area Spt*. Journal of Cognitive Neuroscience, 2003. **15**(5): p. 673-682.
35. Pa, J. and G. Hickok, *A parietal–temporal sensory–motor integration area for the human vocal tract: Evidence from an fMRI study of skilled musicians*. Neuropsychologia, 2008. **46**(1): p. 362-368.
36. Patel, A.D., et al., *Processing Syntactic Relation in Language and Music: An Event-Related Potential Study*. Journal of Cognitive Neuroscience, 1998. **10**(6): p. 717-733.
37. Blood, A.J., et al., *Emotional responses to pleasant and unpleasant music correlate with activity in paralimbic brain regions*. Nature Neuroscience, 1999. **2**(4): p. 382-387.

CHAPTER 5

CONCLUSION

The Evolution of Music

In the past, evolutionary psychologists have explained music's existence via the argument of group selection- this argument is fundamentally incomplete. Although groups that can operate cohesively can have advantages over a scattered group of individuals, group selection on its own does a poor job of explaining the presence of so many of the necessary underlying abilities for music. It can perhaps explain some of the acceleration of the prevalence of music but that presupposes that all the underlying abilities are already there. Group selection is more of a form of cultural evolution than biological evolution. Music, being as multifaceted as it is, has a range of impacts on the activity and development of the brain, as well as some more general implications. In this final section, some of the reasons for music's persistence and near omni-presence in human society are considered.

Mate Selection

A strong argument can be made that music has an effect on mate selection. Organisms expend an huge amount of energy making themselves attractive to mates (Workman 2014). Reproduction is second only to survival in terms of biological drives; it is arguable that survival is only important for reproduction, but this may be a chicken

and egg scenario (Workman 2014). Anything that increases reproductive success can be said to a significant factor in evolution (Workman 2014)

Although music is one form of mating display, it is but one strategy among many. The fact that it is explainable as a mating display alone provides value to its existence (Fitch 2006). Human mating behaviors are fiendishly complex, and attempting to provide a complete explanation of the many factors at play is far beyond the scope of this paper. For some excellent reviews on the topic, see Miller (2000), as well as Sefcek (2007).

Displays of Aggression

Displays of aggression in the animal kingdom share a common trait; they make the animal appear larger. Whether it be a bear standing on its hind legs, the fur on a cat standing up or the plumage of a bird ruffling out in response to a threat, the goal appears to be the same. Sonic cues have a similar consistency- they are rich in low-frequency content and loud; this idea dates back to Plato, if not before (Kawasaki 1980, Ohala 1997).

Although most examples of this in the animal kingdom are limited to individuals, humans can work in groups to further increase the effectiveness of displays. Groups of humans marching in unison can create the perceptual experience of a single larger organism. This effect is used primarily in displays of military strength. The *haka* of the Maori is a perfect example. The display contains a number of movements and sounds that have the effect of making the performer(s) appear larger. A group of people per-

forming the display is quite intimidating. Other tribal war dances and military displays are largely similar - they generally consist of coordinated movements across members of a group. The effectiveness of this technique is achieved via perceptual grouping heuristics - sounds that occur at the same point in time and space are perceived as a single unit (Mountcastle 1998). Groups that could exploit this perceptual effect could be more effective in conflicts by weakening the resolve of enemy group members. The ability to coordinate movements across group members is crucial for this display to work, so it can be argued that rhythmic ability and sensorimotor synchronization generally has some evolutionary value based on this occurrence.

Birdsong and advances in musical technology

There is a huge body of work devoted to studying birdsong, its connections with evolution and the analogy of birdsong, human language and human music. Birdsong has long been hypothesized to be connected to mating. Birds with a more complex song can mate with more females and control a wider territory (CITATION). A recent meta-analysis of birdsong literature revealed a significant, but weak effect of song complexity on bird reproductive success (Soma 2011). Courtship behaviors are not one directional either, with female birdsong having demonstrable impacts on male behavior (Reichard 2011). Similarly, the songs themselves are subject to an “arms race” of sorts, with sexual selection acting to promote songs with certain attributes (Price 2004)

Birdsong evolution is constrained by physical abilities of the birds’ vocal tracts (Podos 2004). Human music has been decoupled from our physical capabilities. The

use of instruments first allowed for the creation of sounds that the human body itself cannot create. However, drums, flutes and other acoustic instruments were still limited by the capabilities of human movement. Although the existence of musical prodigies pushed the limits of what people thought possible from a performance, the advent of computer-generated music has provided the opportunity to create sounds completely divorced from the natural world, modulate acoustic parameters in ways that were previously unthinkable, and perform pieces that are physically impossible for a human to physically play. The factors that drive human music are now predominantly memetic. This is not to say that technological advances in music have stopped, rather the music technology industry is a huge and growing field, however, it is less driven by limits in physical reproduction and more limited by human ingenuity, or perhaps the development of musical AI.

Genetic Roots

There are a family of genes correlated to musical aptitude and listening (Levitin 2012; Ukkola-Vuoti 2011). One haplotype of the Arginine Vasopressin Receptor, AVPR1A, has been found to affect several factors of auditory processing and musical ability generally, as well as several other social tendencies (Ukkola 2009). AVPR1A appears to have a relationship with dancing as well, another highly structured activity deeply related to music (Bachner-Melman 2005). Another set of genes, centered at 3q21.3, is associated with the development of the inferior colliculus; certain single nucleotide polymorphisms were correlated with demonstrable differences in behavioral

abilities in a pitch detection and sound analysis (Oikkonen 2015). The serotonin transporter SLC6A4 is another candidate gene for musical behaviors, although more work has been done on the AVRP1A gene (Ukkla 2009).

Correlating genetic information with behavioral data is a field still in its infancy. Although the presence of these genes does appear to have significant impact on behavioral outcomes, it remains to be seen how much of a role epigenetic factors play in the development of musical skill and other human factors generally. Much of this work is done at the level of families- musical behaviors tend to run in families, but is this relationship strictly genetic or is it buttressed by social expectations as well?

The interaction between genes and environment is an ongoing conversation between the two systems. Most humans have the capacity to learn music, just as language is a fundamental human skill. Although there are variations in the quality of music and/or language production, the basic faculties necessary for the systems are present in everyone.

Benefits of musical training

Musical training engages a huge swath of brain systems. Several types of auditory processing are improved by music, including brainstem level pitch tracking (Wong 2007; Schön 2004), speech-in-noise comprehension (Pabery-Clark 2009; Strait 2012), motor skills (Schlaug 2005; Jabusch 2009), attention (Kraus 2010), recovery from stroke (Schlaug 2008, 2010; Särkämö 2008; Scheider 2007,2010; Johansson 2011), memory (George 2011; Roden 2014), plasticity (Ragert 2004; White-Schwoch 2013; Wan 2010;

Patel 2011) and vocal emotion detection (Strait 2007) are some of the neural benefits of musical training.

The ability to predict future events based on past occurrences is a foundational aspect of human learning. One theory about why music is rewarding involves predictions about what is coming next in the music (Salimpoor 2015). In this way, music can serve as a training ground for developing predictive skills.

Creative thought involves connecting disparate ideas. The foundations for this process were evolved a long time ago. Classical conditioning exists in a wide swath of the animal kingdom because it is very adaptive. Organisms that learn have better prospects for survival and organisms that can pass their learning down to their offspring increase the survival prospects for their progeny. It stands to reason that if an organism can evolve the ability to improve their learning skills, then the mechanism for that meta-learning would be widespread in that organism. The improvements in neural plasticity induced by musical training suggests that music is one such form of meta-learning.

Emotional Modulation

The core hypothesis of this paper is that music modulates emotional states via connections between auditory, motor and limbic regions. The motor system plays a role as a primary hub of dopaminergic activity. These connections have been shown to exist in a huge swath of the scientific literature (Zatorre 2007). Consonance and dissonance have been shown to strongly correlate with activity in auditory regions of the brain in both humans and monkeys (Bidelman 2009; Fishman 2001; McKinney 2001). One

study went further and tried to find neural loci of whole-number ratios in music and found a wider network involved for musicians, (the IFG, STG, IPL, and anterior cingulate) compared to nonmusicians, for whom only the IFG was activated (Foss 2007).

Music's ability to modulate emotions, whether it be calming, invigorating, or seductive effect is extremely useful. The ability of a mother to instill positive associations in her infant will increase the likelihood that the infant stays near her nurturing presence and survives. The ability of a singer to inspire romantic feelings in a potential partner increases the chances of successful mating. People suffering from Alzheimer's and other forms of dementia have been found to "come alive" in response to music from their past in an excellent documentary, "Alive Inside" (Sneed 2014). Alzheimer's takes a devastating toll on the brains of its victims, so the fact that these musical memories are so fully retained when activated hints at the depth and resilience of their encoding. Music can also be used to ameliorate depressive and aggressive behaviors in sufferers of dementia (Clark 1998; Goodall 2005; Svansdottir 2006; Ashida 2000). Ameliorative or palliative care are still among the best things medicine can do for terminal dementia. The fact that music works to regulate emotional states even in people with moderate to severe mental decline demonstrates again how deeply rooted in our brains it is. Several organizations have recently emerged to promote music therapy for people suffering from dementia (Morris 2016; "Music and Memory" 2016).

REFERENCES

1. Ashida, S. (2000). The effect of reminiscence music therapy sessions on changes in depressive symptoms in elderly persons with dementia. *Journal of Music Therapy*, 37(3), 170-182.
2. Bachner-Melman, R., Dina, C., Zohar, A. H., Constantini, N., Lerer, E., Hoch, S., ... & Granot, R. (2005). AVPR1a and SLC6A4 gene polymorphisms are associated with creative dance performance. *PLoS Genet*, 1(3), e42.
3. Bidelman, G. M., & Krishnan, A. (2009). Neural correlates of consonance, dissonance, and the hierarchy of musical pitch in the human brainstem. *The Journal of Neuroscience*, 29(42), 13165-13171.
4. Cassell, J., McNeill, D., & McCullough, K. E. (1999). Speech-gesture mismatches: Evidence for one underlying representation of linguistic and nonlinguistic information. *Pragmatics & cognition*, 7(1), 1-34.
5. Clark, M. E., Lipe, A. W., & Bilbrey, M. (1998). Use of music to decrease aggressive behaviors in people with dementia. *Journal of Gerontological nursing*, 24(7), 10-17.
6. Fishman, Y. I., Volkov, I. O., Noh, M. D., Garell, P. C., Bakken, H., Arezzo, J. C., ... & Steinschneider, M. (2001). Consonance and dissonance of musical chords: neural correlates in auditory cortex of monkeys and humans. *Journal of Neurophysiology*, 86(6), 2761-2788.
7. Fitch, W. T. (2006). The biology and evolution of music: A comparative perspective. *Cognition*, 100(1), 173-215.
8. Foss, A. H., Altschuler, E. L., & James, K. H. (2007). Neural correlates of the Pythagorean ratio rules. *Neuroreport*, 18(15), 1521-1525.
9. George, E. M., & Coch, D. (2011). Music training and working memory: an ERP study. *Neuropsychologia*, 49(5), 1083-1094.
10. Goodall, D., & Etters, L. (2005). The therapeutic use of music on agitated behavior in those with dementia. *Holistic nursing practice*, 19(6), 258-262.
11. Jabusch, H. C., Alpers, H., Kopiez, R., Vauth, H., & Altenmüller, E. (2009). The influence of practice on the development of motor skills in pianists: a longitudinal study in a selected motor task. *Human movement science*, 28(1), 74-84.
12. Johansson, B. B. (2011). Current trends in stroke rehabilitation. A review with focus on brain plasticity. *Acta Neurologica Scandinavica*, 123(3), 147-159.
13. Kawasaki, H., & Ohala, J. J. (1980). Acoustic basis for universal constraints on sound sequences. *The Journal of the Acoustical Society of America*, 68(S1), S33-S33.
14. Kraus, N., & Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nature Reviews Neuroscience*, 11(8), 599-605.
15. Levitin, D. J. (2012). What does it mean to be musical?. *Neuron*, 73(4), 633-637.

16. McKinney, M. F., Tramo, M. J., & Delgutte, B. (2001). Neural correlates of musical dissonance in the inferior colliculus. *Physiological and psychophysical bases of auditory function* (Breebaart DJ, Houtsma AJM, Kohlrausch A, Prijs VF, Schoonhoven R, eds), 83-89.
17. Miller, G. (2000). Evolution of human music through sexual selection (pp. 329-360). na.
18. Morris, H., Jones, L., Drake, A., & Navolynski, A. (2016). Music therapy.
19. Mountcastle, V. B. (1998). *Perceptual neuroscience: the cerebral cortex*. Harvard University Press.
20. Music and Memory. Website, retrieved 2016. <http://musicandmemory.org/>
21. Ohala, J. J., Hinton, L., & Nichols, J. (1997, August). Sound symbolism. In *Proc. 4th Seoul International Conference on Linguistics [SICOL]* (pp. 98-103).
22. Oikkonen, J., Huang, Y., Onkamo, P., Ukkola-Vuoti, L., Raijas, P., Karma, K., ... & Järvelä, I. (2015). A genome-wide linkage and association study of musical aptitude identifies loci containing genes related to inner ear development and neurocognitive functions. *Molecular psychiatry*, 20(2), 275-282.
23. Parbery-Clark, A., Skoe, E., Lam, C., & Kraus, N. (2009). Musician enhancement for speech-in-noise. *Ear and hearing*, 30(6), 653-661.
24. Patel, A. D. (2011). Why would musical training benefit the neural encoding of speech? The OPERA hypothesis. *The relationship between music and language*, 195.
25. Podos, J., Huber, S. K., & Taft, B. (2004). Bird song: the interface of evolution and mechanism. *Annual Review of Ecology, Evolution, and Systematics*, 55-87.
26. Price, J. J., & Lanyon, S. M. (2004). Patterns of song evolution and sexual selection in the oropendolas and caciques. *Behavioral Ecology*, 15(3), 485-497.
27. Pulli, K., Karma, K., Norio, R., Sistonen, P., Göring, H. H. H., & Järvelä, I. (2008). Genome-wide linkage scan for loci of musical aptitude in Finnish families: evidence for a major locus at 4q22. *Journal of medical genetics*, 45(7), 451-456.
28. Ragert, P., Schmidt, A., Altenmüller, E., & Dinse, H. R. (2004). Superior tactile performance and learning in professional pianists: evidence for meta-plasticity in musicians. *European Journal of Neuroscience*, 19(2), 473-478.
29. Reichard, D. G., Rice, R. J., Vanderbilt, C. C., & Ketterson, E. D. (2011). Deciphering information encoded in birdsong: male songbirds with fertile mates respond most strongly to complex, low-amplitude songs used in courtship. *The American Naturalist*, 178(4), 478-487.
30. Roden, I., Grube, D., Bongard, S., & Kreutz, G. (2014). Does music training enhance working memory performance? Findings from a quasi-experimental longitudinal study. *Psychology of Music*, 42(2), 284-298.

31. Rosen, S. (1992). Temporal information in speech: acoustic, auditory and linguistic aspects. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 336(1278), 367-373.
32. Sefcek, J. A., Brumbach, B. H., Vasquez, G., & Miller, G. F. (2007). The evolutionary psychology of human mate choice: How ecology, genes, fertility, and fashion influence mating strategies. *Journal of Psychology & Human Sexuality*, 18(2-3), 125-182.
33. Salimpoor, V. N., Zald, D. H., Zatorre, R. J., Dagher, A., & McIntosh, A. R. (2015). Predictions and the brain: how musical sounds become rewarding. *Trends in cognitive sciences*, 19(2), 86-91.
34. Särkämö, T., Tervaniemi, M., Laitinen, S., Forsblom, A., Soinila, S., Mikkonen, M., ... & Peretz, I. (2008). Music listening enhances cognitive recovery and mood after middle cerebral artery stroke. *Brain*, 131(3), 866-876.
35. Schlaug, G., Norton, A., Overy, K., & Winner, E. (2005). Effects of music training on the child's brain and cognitive development. *Annals of the New York Academy of Sciences*, 1060(1), 219-230.
36. Schlaug, G., Marchina, S., & Norton, A. (2008). From singing to speaking: why singing may lead to recovery of expressive language function in patients with Broca's aphasia. *Music perception: An interdisciplinary journal*, 25(4), 315-323.
37. Schlaug, G., Norton, A., Marchina, S., Zipse, L., & Wan, C. Y. (2010). From singing to speaking: facilitating recovery from nonfluent aphasia. *Future neurology*, 5(5), 657-665.
38. Schön, D., Magne, C., & Besson, M. (2004). The music of speech: Music training facilitates pitch processing in both music and language. *Psychophysiology*, 41(3), 341-349.
39. Schneider, S., Schönle, P. W., Altenmüller, E., & Münte, T. F. (2007). Using musical instruments to improve motor skill recovery following a stroke. *Journal of neurology*, 254(10), 1339-1346.
40. Schneider, S., Münte, T., Rodriguez-Fornells, A., Sailer, M., & Altenmüller, E. (2010). Music-supported training is more efficient than functional motor training for recovery of fine motor skills in stroke patients. *Music Perception: An Interdisciplinary Journal*, 27(4), 271-280.
41. Sneed, A. (2014). *Alive Inside: A Story of Music & Memory*. *Scientific American*, 311(4), 96-96.
42. Soma, M., & Garamszegi, L. Z. (2011). Rethinking birdsong evolution: meta-analysis of the relationship between song complexity and reproductive success. *Behavioral Ecology*, arq219.
43. Strait, D. L., Parbery-Clark, A., Hittner, E., & Kraus, N. (2012). Musical training during early childhood enhances the neural encoding of speech in noise. *Brain and language*, 123(3), 191-201.

44. Strait, D. L., Kraus, N., Skoe, E., & Ashley, R. (2009). Musical experience and neural efficiency—effects of training on subcortical processing of vocal expressions of emotion. *European Journal of Neuroscience*, 29(3), 661-668.
45. Svansdottir, H. B., & Snaedal, J. (2006). Music therapy in moderate and severe dementia of Alzheimer's type: a case–control study. *International psychogeriatrics*, 18(04), 613-621.
46. Ukkola, L. T., Onkamo, P., Raijas, P., Karma, K., & Järvelä, I. (2009). Musical aptitude is associated with AVPR1A-haplotypes. *PLoS One*, 4(5), e5534.
47. Ukkola-Vuoti, L., Oikkonen, J., Onkamo, P., Karma, K., Raijas, P., & Järvelä, I. (2011). Association of the arginine vasopressin receptor 1A (AVPR1A) haplotypes with listening to music. *Journal of human genetics*, 56(4), 324-329.
48. Wan, C. Y., & Schlaug, G. (2010). Music making as a tool for promoting brain plasticity across the life span. *The Neuroscientist*, 16(5), 566-577.
49. White-Schwoch, T., Carr, K. W., Anderson, S., Strait, D. L., & Kraus, N. (2013). Older adults benefit from music training early in life: biological evidence for long-term training-driven plasticity. *The Journal of Neuroscience*, 33(45), 17667-17674.
50. Wong, P. C., Skoe, E., Russo, N. M., Dees, T., & Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nature neuroscience*, 10(4), 420-422.
51. Workman, L., & Reader, W. (2014). *Evolutionary psychology*. Cambridge University Press.
52. Zatorre, R. J., Chen, J. L., & Penhune, V. B. (2007). When the brain plays music: auditory–motor interactions in music perception and production. *Nature reviews neuroscience*, 8(7), 547-558.