12-1-2013

The Role of Music-Specific Representations When Processing Speech: Using a Musical Illusion to Elucidate Domain-Specific and -General Processes

Christina M. Vanden Bosch Der Nederlanden
University of Nevada, Las Vegas, christinavb@gmail.com

THE ROLE OF MUSIC-SPECIFIC REPRESENTATIONS WHEN PROCESSING

SPEECH: USING A MUSICAL ILLUSION TO ELUCIDATE

DOMAIN-SPECIFIC AND –GENERAL PROCESSES.

By

Christina M. Vanden Bosch der Nederlanden

Bachelor of Arts in Psychology

Calvin College

2008

A thesis submitted in partial fulfillment

of the requirements for the

Master of Arts - Psychology

Department of Psychology

College of Liberal Arts

The Graduate College

University of Nevada, Las Vegas

December 2013

UNLV
UNIVERSITY OF NEVADA LAS VEGAS

THE GRADUATE COLLEGE

We recommend the thesis prepared under our supervision by

Christina M. Vanden Bosch der Nederlanden

entitled

The Role of Music-Specific Representations When Processing Speech: Using a Musical Illusion to Elucidate Domain-Specific and –General Processes

is approved in partial fulfillment of the requirements for the degree of

Master of Arts - Psychology
Department of Psychology

Eric E. Hannon, Ph.D., Committee Chair

Joel S. Snyder, Ph.D., Committee Member

Laurel Pritchard, Ph.D., Committee Member

Diego Vega, Ph.D., Graduate College Representative

Kathryn Hausbeck Korgan, Ph.D., Interim Dean of the Graduate College

December 2013

ABSTRACT

The role of music-specific representations when processing speech: Using a musical illusion to elucidate domain-specific and -general processes.

by

Christina M. Vanden Bosch der Nederlanden

Erin E. Hannon, Examination Committee Chair

Professor of Psychology

University of Nevada, Las Vegas

When listening to music and language sounds, it is unclear whether adults recruit domain-specific or domain-general mechanisms to make sense of incoming sounds. Unique acoustic characteristics such as a greater reliance on rapid temporal transitions in speech relative to song may introduce misleading interpretations concerning shared and overlapping processes in the brain. By using a stimulus that is both ecologically valid and can be perceived as speech or song depending on context, the contribution of low- and high-level mechanisms may be teased apart. The stimuli employed in all experiments are auditory illusions from speech to song reported by Deutsch et al. (2003, 2011) and Tierney et al. (2012). The current experiments found that 1) non-musicians also perceive the speech-to-song illusion and experience a similar disruption of the transformation as a result of pitch transpositions. 2) The contribution of rhythmic regularity to the perceptual transformation from speech to song is unclear using several different examples of the auditory illusion, and clear order effects occur because of the within-subjects design. And

finally, 3) when comparing pitch change sensitivity in a speech mode of listening and, after several repetitions, a song mode of listening, only a song mode indicated the recruitment of music-specific representations. Together these studies indicate the potential for using the auditory illusion from speech to song in future research. Also, the final experiment tentatively demonstrates a behavioral dissociation between the recruitment of mechanisms unique to musical knowledge and mechanisms unique to the processing acoustic characteristics predominant in speech or song because acoustic characteristics were held constant.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Music and language allow humans to effectively transfer ideas, emotions, and intentions. All known cultures have linguistic and musical traditions that share similarities such as the use of discrete notes or syllables, the organized structure of phrases, and the creation of expectation (Brown, Werker, & Wallin, 2000). Yet music and language have unique features and rules that enable adults to differentiate speech from song such as their relative reliance on fine-grained pitch differences, regular rhythmic groupings, or referential meaning. While music and language seem to be easily discriminable and require domain-specific representations in adulthood, there is less known about whether human infants are able to differentiate between speech and song. Human infants prefer to listen to and more easily learn from the musical, infant-directed speech of their caregivers commonly called "motherese" (Schachner & Hannon, 2011; Thiessen, Hill & Saffran, 2005). However, it is not known whether infants can distinguish between adult song and speech even though such discriminations may be advantageous for the acquisition of domain-specific knowledge. Speech and song may initially be processed using the same resources or representations in the brain. Indeed, even in adulthood detecting specific violations in speech and song may recruit similar neural processes (Patel, 2003). In the past few decades, a considerable amount of research has been focused on the intersection of language and music. Adults may rely on domain-specific regularities in speech and song to parse relevant information such as the mood or intent of a phrase, yet the extent that processes analyzing such regularities overlap or remain domain-specific has been difficult to characterize.

Adults and infants are able to gain knowledge about regular visual and auditory input through incidental experience with their environment, illustrating that adults and infants are able to use a general mechanism to extract statistical regularities and form expectations (Saffran, Johnson, Aslin, & Newport, 1999; Kirkham, Slemmer, & Johnson, 2002). However, speech and music require different representations of knowledge, with speech requiring the activation of phonetic, articulatory, and semantic networks (Hickok & Poeppel, 2007), and music requiring knowledge of key membership, implied harmony, and fine pitch discriminations (Jackendoff & Lerdahl, 2006; Trainor & Trehub, 1994; Zatorre & Baum, 2012). The existing literature comparing speech and music have identified key areas of the brain involved in processing similarities and differences in musical and linguistic stimuli (Koelsch, 2005; Peretz & Hyde, 2003), but conclusions are limited by the physical acoustic differences in language and music stimuli (Zatorre & Gandour, 2007). Rapid transitions in speech and pitch stability in music may cause observable differences in brain activation because of specializations in processing fast and slow changes occurring over time in the left and right hemispheres, respectively (Joanisse & Gati, 2003). Hemispheric specializations for different acoustic features may lead to the misinterpretation of findings that demonstrate general left asymmetries for speech and right lateralization for music. Rather than observing activation that is functionally segregated, greater left hemisphere activation may arise from the predominance of certain features, specifically rapid temporal transitions, that are characteristic of the speech sounds. Further studies, explained in detail below, have attempted to hold some of the physical qualities of the stimulus constant, but have lacked ecological validity (i.e., compare artificial sounds or isolated syllables) and fail to compare music and speech directly.

The current experiments aim to determine whether music-specific knowledge is differentially recruited while processing an ambiguous stimulus that can be perceived as either speech or music depending on the context. Additionally, acoustic and perceptual features that inform listeners' speech-to-song judgments are considered, including repetition, rhythmicity, and pitch stability. The following study will use an auditory illusion consisting of a single speech stimulus subjectively rated by musicians as sounding like speech and, following several exact repetitions, like song (Deutsch, Henthorn & Lapidis, 2011). This approach improves on the traditional study of music and language processing mechanisms by allowing only high-level contextual knowledge of the stimulus to drive the perceptual transformation from speech to song instead of acoustic characteristics unique to either domain. The following experiments examined 1) whether or not formal music training is necessary to perceive an auditory illusion, 2) whether rhythmic manipulations of the repeating stimulus can interfere with the illusory transition from speech to song, and 3) whether or not the reported mode of listening (i.e., speech or song) can influence listeners' accuracy and sensitivity to specific pitch manipulations.

CHAPTER 2

BACKGROUND

Why music?

While it is apparent that language is necessary for everyday communication, the role and functional significance of music in human interaction is unclear. Music has been present in civilizations throughout history and Charles Darwin wrote of the mysterious importance of music in daily human life (Darwin, 1871). There is considerable debate about the origins of human musical behavior; some argue that music is a byproduct of other adaptations, existing as nothing more than “auditory cheesecake,” while others cast it as a linguistic prototype that predated the evolution of language (Pinker, 1997; Brown, 2000). It has also been suggested that music is a selected trait as evidenced by its presence in cultures dating back tens of thousands of years (Peretz, 2006). Evolutionary survival value may arise out of musical behaviors that have been hypothesized to promote advantageous mate selection and the facilitation of social group cohesion (Brown, Werker & Wallin, 2000).

Today music remains a multifunctional tool aiding in caregiver-infant attachment and emotional regulation (Trainor & Trehub, 1998; Milligan, Atkinson, Trehub, Benoit, Poulton, 2003), rhythmic coordination through beat entrainment (Merker, Madison, Eckerdal, 2009), the remediation of daily tasks in disease states (Thompson, Moulin, Hayre, Jones, 2005; Ford, Malone, Nyikos, Yelisetty, & Bickel, 2010), and prosocial behavior at an early age (Kirschner & Tomasello, 2010). The clearly impactful functional role of music has encouraged the empirical study of musical knowledge and perceptual development to understand the origins and organization of music in its own right. Yet, music shares some specific functional roles and organizational structure with language. Whether or not language and music evolved independently as a result of specific selection pressures, these domains may rely on separate or overlapping networks to efficiently and effectively process speech and song.

Similarities and differences between music and language

At the structural level, music and language share similarities while remaining wholly distinct. Both language and music are systems of communication that unfold over time as an ordered series of discrete elements such as notes or syllables. This type of system is often referred to as a particulate system, in which a specific set of building blocks can be arranged into words, ideas, or phrases to create meaning (Patel, 2008). A hierarchical organizational structure can be applied to both language and music despite fundamental differences in the elements that make up such a hierarchy (Jackendoff, 2009). Music consists of groupings of notes varying in pitch and duration to create musical phrases. Lerdahl and Jackendoff's Generative Theory of Tonal Music (1983) has illustrated a prolongational hierarchy based off of the perceived stability of notes in a tonal scale around a tonal pitch center. In language, syntactic relationships among words create the same type of organizational hierarchy according to the grammatical ordering of different parts of speech. In both syntactic and tonal hierarchies, there is a headed structure that consists of a single organizing principle, or head, such as an entire musical or linguistic phrase. The headed structure also identifies the subcomponents that feed into the musical or linguistic phrase, such as noun phrases in speech, or patterns of tension and release in song (Jackendoff, 2009). Importantly, both hierarchies create expectations for the proceeding elements within a phrase, whether it is the expectation of who is doing what to whom in language, or the expectation of finishing a musical phrase on the root of the diatonic scale.

Music and language are also related through the concept of recursion, in which elements from an existing set of words or phrases can be strung together in an infinite number of meaningful ways. It has been debated whether one of the defining characteristics of human communication systems is the embedding of simple clauses within other clauses to create increasingly complex, yet intelligible phrases (e.g., The dog, who bit the girl, was taken to the pound). This feature occurs not only in language, but also in music (Hauser, Chomsky, & Fitch, 2002). While the presence of recursive structure in meaningful speech and song is an important similarity, the assignment of meaning is also one of the characteristic differences between music and language (Jackendoff, 2009). Speech can access specific semantic associations through the presentation of a single word such that the word "pomegranate" conjures up specific references to a red seeded fruit. However, for typical listeners, music does not have a one-to-one correspondence between a given note and its meaning. Instead, musical meaning arises out of a combination of pitches and durations conveying or evoking emotional states through relative tonal and rhythmic patterns of expectation. In Western musical traditions, positive and negative emotions can be evoked through major and minor modes. This sort of meaning in music can be most directly related to the combinations of stress, intonation, and rhythm that make up the prosodic features of speech. However, music has been shown to evoke extramusical associations, too, even conveying semantic descriptors such as "hero" or "river" (Koelsch et al., 2004). While music and language may employ similar hierarchical structures to bring about meaning, the content of such meaning is fundamentally different.

Finally, music and speech differ significantly with respect to characteristic physical acoustic attributes. In comparison to speech, music is comprised of a series of discrete pitch events with stable fundamental frequency that, together, ascribe to a tonal pitch center, the organizing principal in a musical scale as described above (Zatorre & Baum, 2012). In contrast, speech is characterized by continuous pitch glides over the course of a voiced speech sound. Adult-directed speech also makes use of a smaller pitch range than infant-directed speech or music (Fernald & Kuhl, 1987). These differences in the stability of the fundamental frequency also result in different spectral characteristics that are easily recognized in a spectrogram of spoken and sung speech such as harmonicity. The relative durations among notes in a musical phrase usually occur with less variability than spoken utterances as characterized by the normalized Pairwise Variability Index (nPVI), comparing the durations of successive notes or consonants and vowels in musical or spoken phrases, respectively (Patel & Danielle, 2003). However, at a phrasal level, linguistic stress can create a regular stressed-unstressed rhythmic pattern (Jackendoff, 2009), creating the perception of some rhythmic structure. Similarities and differences in the structure and physical characteristics of music and language have led researchers hypothesize similar relationships concerning the perceptual organization of speech and song in the human brain.

Functional organization

In recent years, a considerable amount of research on the organization of language and music in the brain has stemmed from questions about innateness, knowledge acquisition, and cross-domain transfer. Domain-specific and domain-general processing are the two prevailing theories of organization, with strong versions of either theory maintaining either wholly separate or wholly integrated mechanisms, respectively.

Debates concerning domain-specificity are typically informed through a Chomskyan and, more broadly, Fodorian view of processing that assigns distinct functions from a single domain to specific, isolable areas of the brain. These areas are also referred to as modules and are thought to be dedicated to unique modes of processing. This stance was taken in part as a response to the Skinnerian behaviorist notion that human behavior could be simply defined as learned reactions to external stimuli (Chomsky, 1959). Domain-specific approaches asserted that the impoverished nature of linguistic input necessitated mechanisms that could account for the difficult task of language learning without explicit instruction, *a priori* knowledge, and with the presence of such great variability in speech input (Marcus, 1993; Laurence & Margolis, 2001). Chomsky also put forth the possibility of a universal grammar binding all languages together at a deep syntactic level in addition to the existence of a language acquisition device, that together would hypothetically provide the frame and structure for language learners to efficiently and predictably acquire language. The poverty of the stimulus argument is also important from a purely evolutionary perspective of domain-specific mechanisms.

Cosmides and Tooby (1994) proposed a set of domain-specific modules shaped through natural selection to respond preferentially to biologically significant stimuli. For example, face processing may rely on a specific mechanism that creates a perceptual bias toward faces over other visual objects (Kanwisher, 2000). The specialization of a language or music mechanism would allow stimuli to be quickly and efficiently processed in either a language- or music-specific module accustomed to analyzing only one type of input. By extension, if the status (i.e., linguistic, musical, etcetera) of the stimulus should change based on high-level knowledge, a different module would be activated instead (Zatorre & Gandour, 2007). The complete encapsulation of information within a module does not allow it to receive input from modules dedicated to separate domains (Fodor, 1983). Theoretically, because domain-specific modules are designed to solve problems originating much earlier in phylogeny by virtue of the long course of natural selection (Cosmides & Tooby, 1994), they may be ill-equipped to handle complex or novel signals. It has also been offered that, alongside the existence of domain-specific modules, more general faculties exist for memory or planning recruited to carry out domain-specific tasks. These more general faculties may be separate from domain-specific faculties at least conceptually by contrasting "modules" and "faculties." Modules can be distinguished as mental organs that innately specify propositional thoughts or concepts, consistent with a Neo-Cartesian philosophy of the mind (Fodor, 1983). Modules would be separate from faculties that exist as physical mental structures, designed to carry out representation, retrieval, and elaboration of the module's content. In this sense, then, domain-specific modules may coexist with non-modular mechanisms.

In contrast, a combination of solely domain-general mechanisms would allow music and language to share processing resources along with other non-speech and non-music sounds. Domain-general approaches argue that the same stable statistical relationships were not consistently available throughout the environments of hunter-gatherers of long ago (Chiappe & MacDonald, 2005), requiring the regularities of their environment to be rapidly extracted during the course of a single lifespan, often quite early in development (Greenough & Wallace, 1987). Indeed, a major characteristic of Homo sapiens in the Pleistocene era was increased brain size, which has been attributed to selection pressure for organisms with greater adaptive flexibility (Potts, 1998). One prominent argument against such processes points out that a purely domain-general mechanism does not provide any of the motivations or biases necessary to preferentially process speech or music sounds over others. An inability to preferentially attend to speech sounds may not allow an infant to extract the relevant features of his or her language.

Yet, domain-general accounts of learning have been put forth that do not require domain-specific modules. When humans are faced with extracting statistical environmental regularities in the visual and auditory domains, several researchers have illustrated how perceptual constraints and preferences aid in the process of building accurate representations (Saffran, 2002; Newport, 2011). For example, children's short-term working memory may provide positive limitations when faced with processing long-distance dependencies in language, allowing children to extract information from phrases embedded within the larger sentence (Newport, 1990; Elman, 1993). Similarly, early musical preferences for processing tonal, perceptually stable scales over scales with no tonic (Trehub, Schellenberg, Kamenetsky, 1999) or complex (3:2 ratio) and simple meters (2:1 ratio) over highly complex meters (7:4 ratio) may arise out of perceptual or cognitive constraints (Hannon, Soley, Levine, 2011). With experience, performance on these tonal or metrical tasks becomes increasingly culture-specific. The youngest infants show precocious change detection abilities, yet, for instance, still fail to notice meters that are highly complex.

However, as will be outlined below, it is not likely that music and language are processed in completely general, overlapping regions in the brain. Knowledge representations for unique types of input such as music or language may arise out of an interaction of perceptual constraints and domain-general mechanisms (Trehub & Hannon, 2006; Newport, 2011). In this case, acoustic cues unique to music or language stimuli would dictate which aspects of the general mechanisms are employed to efficiently process auditory information. For example, an incoming linguistic or musical sound with rapid temporal transitions would be processed using general mechanisms that are more efficient at extracting information from fast changes over time (Zatorre & Gandour, 2008). The benefit of a processor that is built to preferentially analyze the salient characteristics of many sound types is the hypothesized greater flexibility for such mechanisms when processing novel sounds (Chiappe & MacDonald, 2005) and perhaps sounds that do not fit categorically into either spoken or sung utterances. Similarly, the encapsulation of information is not a necessary characteristic of general processors, allowing information from other faculties to be incorporated into ongoing analyses or operations. Domain-general systems offer a simple, yet powerful alternative to specialized and isolated approaches to both language and music perception and cognition.

An integral part of these discussions is the question of innateness. While innateness and domain-specificity are separable concepts, some researchers propose that a domain-specific organization of music and language in the brain would indicate that music and language were adaptive tools in evolution rather than byproducts of other selected traits (Peretz & Coltheart, 2003). That is, a brain that is specially suited to process particularly relevant types of information is proposed to have been created through thousands of years of selective pressures, honing a processing module that provides the organism specific survival advantages over those organisms without or with less specialized processing mechanisms. Cosmides and Tooby (1994) use domain-specificity to differentiate between what the human brain can successfully analyze and what it has been designed to process. Researchers have posited that the language faculty alone is "special" and suggest that music exists only as part of a parasitic relationship with language (Sperber, 1996).

Proponents of a general set of mechanisms for music and language processing offer the theory of a small number of primitive, innate motivations and limitations that interact with evolutionarily advantageous and powerful general mechanisms to produce specialized systems (Chiappe & MacDonald, 2005; MacDonald & Hershberger, 2005). Important implications of domain-general models include the facilitation of cross-domain transfer that would take advantage of the overlap for processing shared components of music and language or other domains. For example, improvements in reading abilities and pitch discrimination in spoken utterances have been demonstrated following music training but not painting in childhood (Moreno et al., 2011). Studies comparing domain-general and domain-specific organization have yielded only mixed support, leaving both theories, as well as a combination of these approaches, as viable possibilities.

Investigating music and language

One argument for domain-specific processing of music and language comes from evidence of double dissociations, in which impaired speech or music perception is present while the opposite faculty remains spared (Ayotte, Peretz, & Hyde, 2002; Peretz et al., 2002). Case studies of either acquired or congenital amusia illustrate participants' inability to differentiate consonance from dissonance or to detect melodic violations (in- or out-of-key), even though performance on linguistic intonation tasks remains statistically similar to matched controls (Peretz et al., 2002; Peretz & Hyde, 2003; Peretz, Brattico, Jarvenpaa, & Tervaniemi, 2009). Amusics have been shown to possess increased cortical thickness in the right inferior frontal gyrus and right secondary auditory area that correlated negatively with performance on the Montreal Battery of Evaluation of Amusia (MBEA) possibly indicating a malformation affecting normal melodic perception but not prosody (Hyde et al., 2007). Similarly, when speech difficulties are present after left frontal lobe infarctions, the ability to sing remains intact and has even been used as a therapy technique with non-fluent aphasics (Jacome, 1984).

Such studies are taken as evidence for domain specificity because a component that is important for comprehension in both domains, such as pitch perception, is unaffected in one domain and impaired in the other. This dissociation can be taken as evidence for information encapsulation within the given module. Despite these findings, studies indicate that participants with amusia do have some difficulty with normal and emotional speech prosody (Patel, 2008; Thompson, Marin, Stewart, 2012), leaving open the possibility for some level of shared processing resources. Lesion studies are often difficult to interpret because the size of the affected brain area and the extent of the damage extending to other mechanisms is not identical across participants. Moreover, drawing inferences by comparing non-neurotypical subjects may bring additional factors into play such as aberrant connections formed in response to a neurological insult. It has also been argued that double dissociations cannot be interpreted as evidence for domain-specific mechanisms because the use of multiple domain-general processes remains a possibility (Dunn & Kirsner, 2003). For instance, fine- and coarse-grained pitch discrimination may recruit separate mechanisms for accurate pitch analysis, both part of a general mechanism for detecting differences between frequencies. Yet, there are specific testable predictions that arise out of domain-general and domain-specific explanations of music and language.

Early behavioral studies tapped into isolable knowledge representations for speech and non-speech to demonstrate domain-specificity. Whalen and Liberman (1987) demonstrated the ability for "duplex perception" in which a single stimulus can be perceived as both speech and non-speech at the same time by varying the relative intensity of the 3^{rd} formant transition. The rapid 3^{rd} formant transition is what distinguishes a "d" stop consonant from a "g," identified by a descending or ascending sweep, respectively. When each formant transition was heard in syllabic context it was accurately discriminated, but participants categorized stimuli at chance when the 3^{rd} formant transition was heard in isolation. These findings were described as evidence for a special route diverting input to a module for phonetic encoding unique to processing sounds in a speech "mode." This study also highlights that different acoustic cues may not be enough to drive accurate categorization across domains, leaving this type of discrimination unique to speech.

Speech perception has long drawn on the example of the categorical perception of phonemes as evidence for special speech mechanism recruited by human listeners. Categorical perception is separate from one's ability to categorize stimuli in that, for instance, there is an abrupt change in perception at a perceptual boundary between stop consonants. Additionally, discrimination within each category is worse than between categories, presumably because listeners have adapted to the noise inherent to talker variability and multiple pronunciations of the same phoneme or syllable in different contexts. It is crucial to perceive several versions of the same phoneme as belonging to the same phonetic category so that variable pronunciations can be efficiently processed as representing a single phoneme. While many stimuli can be perceived categorically, it was hypothesized that speech is unique because language users do not perceive vowel and consonant pairs according to a continuous gradient of responses. In contrast, in the absence of musical training, discrimination thresholds when respondents were asked to label a continuum of musical intervals by identifying the wider of two sequentially presented intervals increasing by increments of 12.5 cents (i.e., between a major second and a tri-tone) does not indicate the presence of categorical perception (Burns & Ward, 1978). However, this study suggests categorical perception can be gained through extensive training, not a phenomenon unique to the speech domain.

Several recent studies have demonstrated categorical perception of non-speech stimuli using varying tone-onset times (analogous to voice-onset times in speech) given experience with the distributional regularities of these categories (Holt, Lotto, & Diehl, 2004). More recent research describes categorical perception as a general sensitivity to the distributional probabilities of one's linguistic or even musical environment that is developed through experience (Werker, Yeung, Yoshida, 2012; Trehub & Hannon, 2006). In infancy perceptual mechanisms are characterized by flexibility and generality, but through implicit experience infants build knowledge representations and perceptual mechanisms that are specially tuned to process culture-specific input (Karmiloff-Smith, 1992).

A prominent trend in studies comparing the processing of speech and music is a left-hemisphere lateralization for speech sounds and generally a right-hemisphere lateralization for musical sounds (Zatorre, Evans, Meyer, & Gjedde, 1992; Auzou et al., 1995; Tervaniemi, et al., 2000). However, many of these early studies compared sentences or syllables to random note sequences or isolated chords, ignoring the hierarchical structural components music and language have in common. By taking advantage of the formation of expectations in musical or spoken phrases, Patel (1998) reported similar patterns of left lateralized activation for both incongruent musical and incongruent linguistic syntax, suggesting a shared mechanism for processing syntax in these two domains. Levitin & Menon (2003) demonstrated that the temporal structure of music activated the left inferior frontal cortex, an area that had previously been associated only with semantic linguistic perception. Similarly, Koelsch et al. (2005) found evidence of mutual interference during processing of music and speech, indicated by reduced left anterior negativity when processing incongruent sentences and chords or tones simultaneously. However, the data concerning the organization of music and language-specific elements is not entirely consistent with a model for shared syntactic processing. More recent studies directly comparing music and language within subjects have found no such overlap (Rogalsky, Rong, Saberi, & Hickok, 2011) and have speculated that implications drawn from "violation" studies may be task-specific. Instead of identifying low-level processes specific to language or music, violation studies may be eliciting general responses unique to higher-level processes such as cognitive control or working memory.

Many of the differences related to music and speech processing may be related to the temporal and spectral attributes unique to each domain. Computer algorithms set with the task of differentiating music and language focus on specific features that characterize music and language to make accurate discriminations, namely spectral stability and temporal rhythmicity of music (Scheirer & Slaney, 1997; Schlueter & Sonnleitner, 2012). Remez, Rubin, Pisoni, & Carrell (1981) designed several experiments using an analogue to authentic speech by constructing sine-wave tracings of human speech. Sine-wave speech is constructed by extracting the formant frequencies of a recorded speech sound and replacing those frequencies with typically three or four sine-waves in order to preserve some of the temporal and spectral characteristics of speech without the full complexity of natural human speech sounds. When a sine-wave analogue is presented in isolation, knowledge of the original stimulus can be recruited to decode sine-wave speech and increase intelligibility.

fMRI studies have compared the perception of sine-wave speech when it is reported as sounding like non-speech and speech. After participants are exposed to the original speech tokens that the sine-wave analogues were mimicking (i.e., they were presumably hearing the stimulus in “speech mode”), greater activity was observed in the left superior temporal sulcus (Dehaene-Lambertz et al., 2005; Mottonen et al., 2006). However, Dehaene-Lambertz and colleagues report greater left hemisphere activation than the right hemisphere for processing sine-wave syllable analogues even when perceived as non-speech. Vouloumanos et al. (2001) demonstrated greater activation in left posterior superior temporal gyrus, the classical Wernicke's area for speech compared to sine-wave speech. These authors discuss the potential role attention may play in modulating enhancements in these brain regions as a result of greater attentional priorities for speech compared to nonspeech. While this study argues for the presence of domain-specific processing of speech sounds, the role of attention in processing speech and nonspeech sounds may introduce confounding factors when studying brain organization, eliciting greater activation due to attention and not speech compared to nonspeech. Finally, Joanisse and Gati (2003) reported greater left hemisphere activation for speech and non-speech sounds that had either rapid temporal or spectral transitions using an fMRI paradigm. This raises the possibility that the left hemisphere is not specially suited for speech processing, but for rapid temporal or spectral transitions (Zatorre & Gandour, 2007). Zatorre, Belin, & Penhune (2002) and Poeppel (2003) proposed hemispheric lateralization might instead be drawing on left hemispheric abilities to extract information from stimuli unfolding rapidly over time, whereas the right hemisphere may be more suited for slowly unfolding temporal or spectral information.

Creative studies using auditory illusions can provide a way to examine the recruitment of speech- or song-specific knowledge while keeping the low-level properties of the stimulus constant. Sine-wave speech studies have provided insight into how top-down processes influence perception, but their findings can only be extended to the processing of speech and noise. Other auditory illusions that engage structural and functional similarities unique to speech and music may shed light on the specificity or generality of these structures. There are many examples in popular culture that test the bounds of what can easily be classified as speech or song. For instance, Minimalist composer Steve Reich uses the repetition of spoken motifs to rhythmically weave speech into a piece of music (Reich, 1966). Other artists have sought to blur the lines between language and music by vocally imitating the sound of a stringed instrument (e.g., Yo Yo Ma & Bobby McFerrin: Hush) or vice versa (Charles Spearin's The Happiness Project). Similarly, Arnold Schoenberg developed a method of singing designed to sound more like speech called Sprechstimme (literally "speech voice") and much of popular rap music plays with the rhythmic cadences of speech. An every day occurrence of song-like speech can also be found in infant-directed speech of adults across cultures (Falk, 2004). Indeed, illusory verbal transformations brought about by the repetition of normal words have been used to study attention, awareness, and the formation of auditory percepts (Kondo & Kashino, 2007).

The auditory illusion of particular interest to the current study is the illusory transformation from speech to song demonstrated by Deutsch (2003) and Deutsch, Henthorn, & Lapidis (2011) whereby a single spoken utterance looped several times is rated as sounding more like song by musically trained individuals. Researchers reported the apparent recruitment of domain-specific knowledge when participants were asked to listen to one spoken repetition, one sung repetition, or ten repetitions of the spoken phrase and reproduce it exactly as they heard it. Participants in the single sung repetition and ten repetition groups reproduced the spoken phrase at pitches closer to the original phrase than the single spoken repetition condition. Additionally, these two groups' reproductions assimilated the original intervals to more closely match the Western diatonic musical scale. The salience of the pitch contour in this particular type of auditory illusion may be pronounced because of relatively stable fundamental frequency information over the voiced segment of each speech syllable (Tierney, Dick, Deutsch, & Sereno, 2012).

In order to test the recruitment of music-specific knowledge in a purely perceptual task, previous studies have compared participants' ability to detect changes that do or do not violate musical key membership. In music, the perception of key membership and harmony progresses with age for listeners of Western tonal music. Seven-year-olds and adults are able to detect in-key, but out-of-implied-harmony violations in a simple melody, whereas 5-year-olds only detect out-of-key violations well (Trainor & Trehub, 1994). The involvement of culture-specific representations in seven-year-olds and adults and more culture-general representations in 5-year olds suggests that such representations are built over time with musical experience, but are fully defined by late childhood. Since the musicians in the Deutsch study outlined above reproduced Western tonal scale interpretations of the auditory illusion, such domain-specific representations could be used in adulthood to objectively measure each participant's current mode of listening.

One other study has employed the speech-to-song illusion using fMRI to determine whether other music-related brain areas were recruited in the processing of sentences that transform readily from speech to song with exact repetition. Tierney and colleagues (2012) used two sets of stimuli, one set of sentences perceived as song upon several repetitions and one set of stimuli perceived as speech after the same number of repetitions. When the fMRI activation for these conditions was compared, results indicated several areas more activated in response to the transforming, song-like stimuli than the non-transforming stimuli including the anterior and posterior regions of the superior temporal gyrus, the supramarginal gyrus, and the medial temporal gyrus.

However, no areas were more activated for the non-transforming stimuli compared to song. The areas showing greater activation for song-like stimuli were divided into two functional groupings of pitch processing sensitive areas and areas implicated in vocalization and auditory-motor integration. While this study provides important information about the role of increased pitch salience in this type of perceptual illusion, it does not take advantage of comparing the same stimulus when it is heard as speech and when it is subsequently heard as music. Indeed, the two groups of stimuli differed acoustically in terms of the stability of the fundamental frequency and differed marginally with respect to isochronous rhythmicity that was measured by the degree of regular spacing between the onsets of stressed syllables. The current study will use the auditory illusions presented by Deutsch et al. (2011) and Tierney et al. (2012) to determine the role of domain-specific and domain-general processes in language and music perception. This study will also characterize the acoustic and perceptual features associated with the classification of speech and song.

CHAPTER 3

CURRENT STUDY

In Experiment 1, non-musicians provided subjective ratings when listening to the illusory transformation from speech to song demonstrated by Deutsch, Henthorn, & Lapidis (2011). Listeners were in one of two conditions, either the *untransposed* condition in which each of 10 repetitions of the stimulus were not altered in any way or a *transposed* condition in which the middle eight repetitions (leaving the first and tenth repetitions acoustically identical) were altered by shifting the fundamental frequencies of the entire spoken phrase up or down by two-thirds or four-thirds of a semitone. These two manipulations were used in Deutsch et al. (2011) to determine whether the transformation from speech to song relied on the exact repetition of the stimulus' pitch information. Deutsch and colleagues found that altering the pitch of each repetition by slightly more or less than a semitone disrupted the perception of song for their participants, all of whom had formal music training. This result suggests that transposed features of the original stimulus alone (i.e., the pitch contour and intervals) are not sufficient to result in the subjective transformation to song, but rather it is necessary to repeat the stimulus at the same pitch level to elicit a change in perception. It is possible that the transformation from speech to song is unique to musicians who likely have greater experience listening to and labeling musical sounds. Similarly, it is possible that slight pitch transpositions during the intervening presentation of repetitions will have no effect on naïve, non-musician participants. In order to use this type of ambiguous stimulus to better understand the recruitment of music- and language-specific networks in adulthood, findings must be generalizable to more than just the musician population. Experiment 1 determines whether this auditory illusion is perceived as such by non-musicians and whether the transformation from speech-to-song is equally disrupted by slight transpositions in fundamental frequency. The type of transposition (i.e., an increase or decrease in pitch) will also be compared to shed light on the characteristics that contribute to participants rating a stimulus as sounding more like song or speech.

Experiment 2 further investigates the nature of the perceptual transformation by determining how rhythmic features of the stimulus play a role in transformation of participants' perception. Although past studies have indicated a slight trend for the presence of rhythmic regularity in the other transforming speech-to-song illusions (Tierney et al., 2012), the contribution of rhythmic regularity in hearing this type of auditory illusion has not been experimentally addressed. Specifically, Experiment 2 will investigate whether disrupting the rhythmic regularity of several illusions will lower participants' subjective ratings compared to several unaltered speech-to-song illusions, in a similar manner as the pitch transpositions from Experiment 1. Disruptions were created in a similar manner as Hannon & Trehub's (2005) structure-disrupting condition by adding, and in the current case, deleting an eighth note equivalent (100 ms duration) that was based on the tempo of each utterance. An investigation of song-like characteristics will also be considered here by comparing participants' ratings for different types of rhythmic alteration (i.e., shortened versus lengthened syllables).

Finally, Experiment 3 will examine the potential role of music-specific knowledge through listeners' detection of pitch changes within spoken stimuli. This study will compare each participant's accuracy for detecting pitch manipulations that do or do not conform to musical representations of each stimulus, both when the stimulus is perceived as speech and, after several repetitions, when it is presumably perceived as song. Pitch perception abilities that are unique to the music domain should produce increased sensitivity for nonconforming pitch change types and stable or decreased sensitivity for conforming pitch changes provided participants perceive the stimuli more like song at the final repetition compared to the initial repetition. Frances (1988) demonstrated the predicted effect of interference as a result of context-specific expectations in music. Frances systematically flatted the same two critical notes and recorded them in differing musical contexts. One context contained descending arpeggiations and the other ascending. Mistunings were noticed less when they conformed to the expected structure in the first context and were more easily noticed in the second context when they did not conform to context-specific expectations. As such, the current experiment should elicit interfering and facilitating effects depending on the recruitment of musical knowledge.

CHAPTER 4

EXPERIMENT 1:

Musician and non-musician ratings of Deutsch's auditory illusion

Method

Participants

Ninety-six undergraduates and members of the surrounding community (48 musicians, 48 non-musicians; 24 subjects in each condition) were recruited through the University's Psychology subject pool or through flyers and word-of-mouth communication. Participants received course credit for participation and musicians not eligible for course credit were given the option to be enrolled into a drawing for a $40 gift card. Musicians were defined as having five or more years of formal music training and non-musicians had fewer than five. Musicians (27 females) were an average age of 21.7 years (range: 18 to 56 years) and had an average of 9.6 years of formal music training. Non-musicians (24 females) were 20.7 years old on average (range: 18 – 47 years) and reported an average of 1.2 years of formal music training. All subjects provided informed consent before participation. Upon completion of the study, participants provided hearing, music, and language background information by filling out a confidential demographic questionnaire. All participants reported normal hearing and provided informed consent prior to participation and all forms and protocols were reviewed and approved by the University's Institutional Review Board.

Apparatus

Stimuli were presented and participants' responses were recorded using a custom script written with Psyscope X, Build 57 (Cohen, MacWhinney, Flatt, Provost, 1993) on an Apple Mac mini Intel core duo with OS X. Stimuli were presented through headphones at about 60 dB SPL and subjects indicated their responses using a computer keyboard.

Stimuli

The spoken utterance "…sometimes behave so strangely…" was used to partially replicate and extend the study by Deutsch et al. (2011), which demonstrated the auditory illusion from speech to song in musicians. In the *untransposed* condition, the stimulus was first presented in full context as follows: "The sounds as they appear to you are not only different from those that are really present, but they sometimes behave so strangely as to seem quite impossible." At 2300 ms after the presentation of the full context, 10 repetitions of the speech segment "…sometimes behave so strangely…" was presented with a 2300 ms interstimulus interval (ISI). At 2300 ms following the final segment repetition, the final context was again presented. In the *transposed* condition, the same speech segment was used; however, each of the eight intervening repetitions were moved up or down in pitch by two-thirds or four-thirds of a semitone relative to the first repetition in Praat (Boersma and Weenink, 1992) using the PSOLA overlap-add method in order to preserve formant frequency characteristics. All stimuli were exported from Praat as AIFF sound files. The order of presentation for the transposed stimuli was identical to Deutsch et al. (2011), keeping the first and last of the 10 repetitions at the original untransformed pitch frequency and the middle eight repetitions transposed relative to the original stimulus in the following order: up 2/3 semitone, down 4/3 semitone, up 4/3 semitone, down 2/3 semitone, up 4/3 semitone, down 4/3 semitone, up 2/3 semitone, and down 2/3 semitone.

Procedure

Participants were randomly assigned to one of two conditions: *untransposed* or *transposed*. Participants were instructed that they would first hear a full sentence, and that one piece of the full sentence would be repeated 10 times. Participants were told that they were to rate how each repetition sounded to them and were reminded that their responses were subjective. The rating scale was from 1 through 5, where a 1 indicated that the segment sounded "exactly like speech" and 5 indicated that the segment sounded "exactly like singing." Participants were reminded that the study was not about *what* the sound was, but rather *how* it sounded to them in order to prevent participants from overriding their perception of the stimulus with their knowledge that the stimulus was a speech sample. Finally, participants were made aware of the short ISI in which they needed to respond (about two seconds). Participants were encouraged to keep their fingers hovering over the number pad on the keyboard in order to respond within the given time frame. During testing, stimuli were presented through headphones and the experimenter sat in the same room as the participants. Participants were asked to rate each repetition of the stimulus as well as the initial and final full context sound clips.

Results

Participants' responses for repetition 1 (initial) and repetition 10 (final) were analyzed to quantify perception of the auditory illusion before and after repetition. Responses were entered into a 2 x 2 x 2 (Condition [untransposed, transposed] x Repetition [initial, final] x Group [musician, non-musician]) mixed design Analysis of Variance (ANOVA). The main effect of group was statistically significant, $F(1, 92) = 12.694$, $p < .01$, $\eta_p^2 = .121$, with musicians providing consistently higher ratings than non-musicians (musician's mean rating = 2.61, SD = .86; non-musicians mean rating = 2.02, SD = .84). No significant interactions with group were found ($p > .1$), allowing for all subsequent analyses to combine musicians' and non-musicians' ratings.

For all subjects there was a highly significant main effect for repetition, $F(1, 92) = 126.490$, $p < .001$, $\eta_p^2 = .579$, with higher (i.e., sounds like singing) ratings given for final than for initial repetitions (mean initial rating = 1.58, SD = .84; mean final rating = 3.05, SD = 1.33). A main effect for condition was also found, $F(1, 92) = 9.380$, $p < .01$, $\eta_p^2 = .093$, with higher overall ratings for participants in the untransposed condition (mean = 2.575, SD = .89) compared to the transposed condition (mean = 2.06, SD = .83). Most importantly, an interaction effect between repetition and condition was found $F(1, 92) = 8.710$, $p < .01$, $\eta_p^2 = .086$, as illustrated in Figure **1**. Post-hoc t-tests confirmed that the effect of condition for initial repetitions was not significant, $t(94) = .766$, $p = .470$, but the effect of condition on final rating was significantly different, $t(94) = 3.480$, $p < .01$, indicating that the interaction was driven by differences in participants perception at the final repetition. The current results show that final ratings in the untransposed condition were perceived as sounding more like song than those in the transposed condition and that transposing the pitch of each repetition lowers the final subjective ratings (i.e., sounds more like speech) for both musicians and non-musicians. Importantly, these findings replicate the results from Deutsch et al. (2011) and extend this pattern of results to non-musicians, as well.

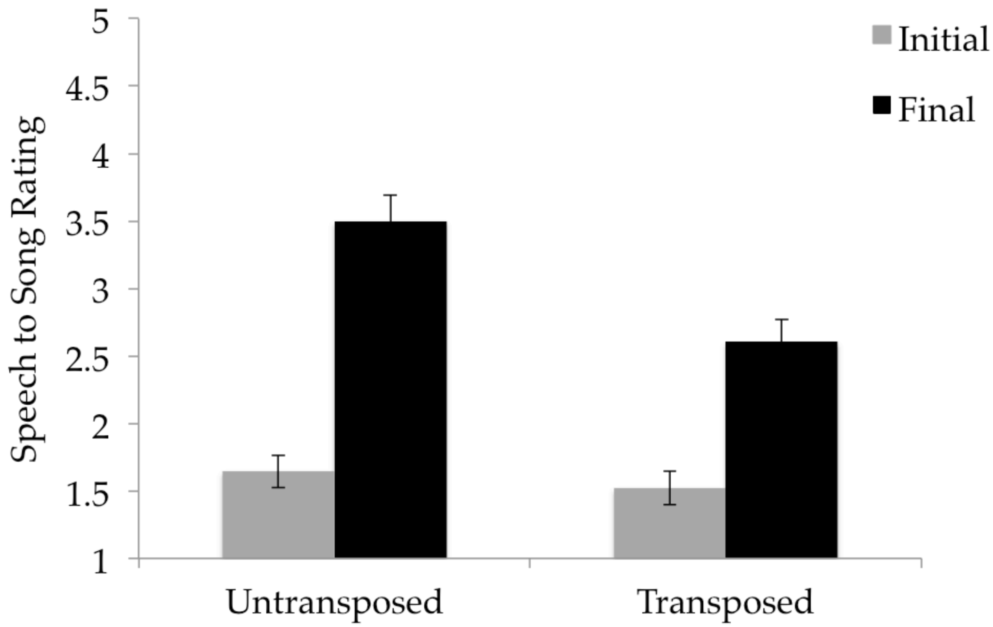

Figure 1: Grand averages across all subjects for the untransposed and transposed conditions for the initial and final repetitions. A disruption of the transformation is evident by the difference in subjective ratings by condition for the final repetition.

Additional analyses compared the subjective ratings for the transposition direction on the transposed condition. We averaged separately repetitions that raised the pitch (repetitions 2, 4, 6, and 8) and those that lowered the pitch (repetitions 3, 5, 7, and 9). These two averages were submitted to a paired-samples t-test comparing each transposition direction (higher versus lower) for participants in the transposed condition only. A significant difference was found between higher and lower transpositions, $t(47) = 3.013$, $p < .01$. As illustrated in Figure **2**, repetitions going higher in pitch were rated significantly higher than repetitions going lower in pitch (mean higher = 2.6, SD = 1.0; mean lower = 2.2, SD = .9).

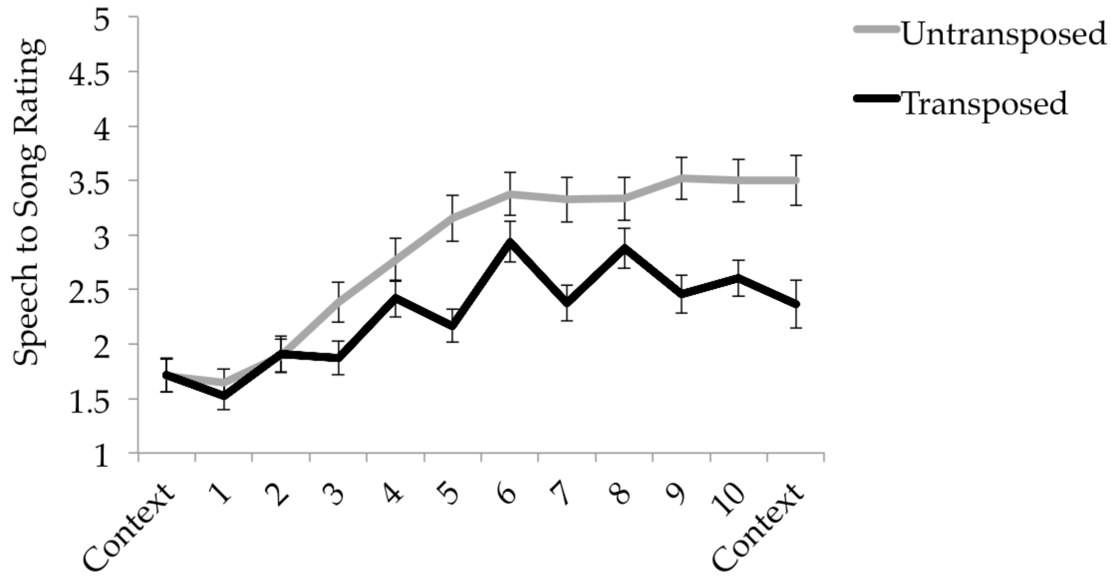

Figure 2: All repetitions for the untransposed and transposed conditions, showing that transpositions going higher in pitch (repetitions 2, 4, 6, and 8) were perceived as more musical than those going lower in pitch (repetitions 3, 5, 7, and 9).

Because it is possible that this effect is confounded with participants' use of the 1 to 5 rating scale, tending to move higher on the rating scale when the overall pitch was raised. A small pilot analysis of 4 subjects was run with the rating scale reversed for the transposed condition (i.e., 5 = "sounds like speech" and 1 = "sounds like song"). Participants' responses suggest they do not conflate responses with the scale, but similarly report transpositions raising the overall pitch as more musical than lower transpositions (mean lower = 3.4, mean higher = 2.8, with higher numbers indicating more like speech). The current experiment demonstrated that non-musicians are capable of reliably rating auditory illusions in the same manner as musicians. Although non-musicians provided lower ratings overall, musicians and non-musicians nevertheless showed the same pattern of responses for the current auditory illusion under exact and transposed repetition. Also, in the transposed condition, all participants rated changes in pitch going upward as sounding more like song than transpositions moving downward in pitch. Non-musicians, then, also perceive the pitch information from this auditory illusion to be as equally salient as musicians, as evidenced by the same disruption of the transformation from speech to song in the transposed condition, demonstrated previously in Deutsch et al. (2011). Specific musical training is not required for the perception of this illusion, indicating that further investigations using similarly ambiguous sounds have the potential to generalize to all adults and may perhaps be extended to the study of functional organization during development.

While this experiment investigated the role pitch plays in this type of auditory illusion, the role of rhythmic regularity remains experimentally untested. Previous studies concerning the role of rhythm in the transformation of speech to song have only been correlational in nature, allowing for the possibility that rhythmic or metrical features of the stimulus play a unique role in causing this perceptual change. Experiment 2 compared non-musicians' subjective ratings during the manipulation of regular rhythmic qualities of several auditory illusions and determined whether rhythmic alterations have as strong of an effect on perception during final ratings as similarly subtle alterations in pitch.

CHAPTER 5

EXPERIMENT 2:

The role of perceived rhythmicity in auditory illusion transformation

Method

Participants

Twenty-four undergraduate participants (10 females) participated for Psychology course research credit and were recruited through the University's undergraduate Psychology subject pool. Participants provided verbal and written informed consent before participating and filled out a background and demographic questionnaire upon completion of the experiment. Participants were non-musicians reporting less than 5 years of formal musical training, with an average age of 20.7 years.

Apparatus

Apparatus was identical to Experiment 1.

Stimuli

Six stimuli from Experiments 1 and 3 were used to create two conditions using the same type of alterations as those outlined in Experiment 1. The six chosen stimuli had the highest average final ratings (near an average of 4.0) and also had the lowest average initial ratings (near an average of 1.5) in preliminary analyses of Experiment 3, ensuring that, on average, these stimuli transformed from speech to song. The average rhythmic regularity of these six stimuli was measured using nPVI (Ramus et al., 1999) that captures the degree of variability between successive vocalic intervals in an utterance, taking into account language-specific effects of vowel reduction. The original set of transforming stimuli from Tierney et al. (2012) had an nPVI of 49.58, while their non-transforming stimuli had an nPVI of 61.22. Using an independent samples t-test, these sets of stimuli are significantly different from each other ($p < .05$). The calculated nPVI for the six stimuli used in the current experiment was an average nPVI of 51.44, statistically indistinguishable from the transforming stimuli ($p = .77$) and marginally different from the non-transforming stimuli ($p = .19$). All nPVI are similar to reported estimates of British English from previous studies (Patel & Daniele, 2003; Grabe & Low, 2002).

The two conditions in the current experiment were similar to the conditions from Experiment 1 and were called *unaltered* and *altered* rhythm conditions (see Figure **3**). The unaltered rhythm condition contained no manipulations of the speech segment, while the altered rhythm condition disrupted the rhythmic structure by shortening or lengthening two syllables within the speech segment. Using Praat, four different duration tiers were created using custom scripts to shorten and lengthen each of the two designated speech syllables. These new duration tiers were then used to replace the original duration tier of each spoken segment in Praat, creating four new spoken segments (one for each of the two syllables that were both shortened and lengthened by 100 ms). Rhythmic patterns and metrical structures were determined by having a musician provide rhythmic transcriptions of each speech sample. The two shortened or lengthened syllables from each of the speech samples were chosen based on their presence on a strong metrical beat (i.e., 1st or 3rd beat) within the utterance. All trials were presented in the following order, with S1 and S2 representing the two syllables that differed depending on the sentence: No alteration, +100ms S1, -100ms S2, +100ms S1, -100ms S1, +100ms S2, -100ms S2, +100ms S2, -100ms S1, No alteration.

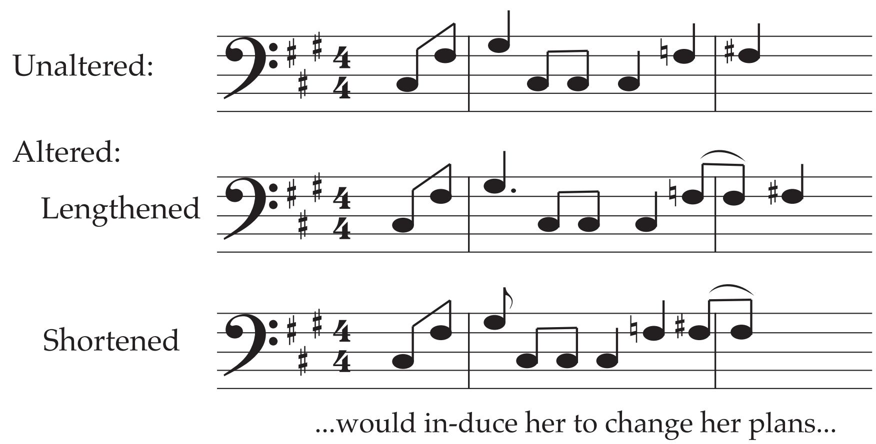

Figure 3: Example of an altered trial for one of the stressed/metrically strong syllable alterations. The syllable "-duce" from the word "induce" is shortened and lengthened by 100 ms (equivalent to an 8^{th} note change with reference to the typical syllable rate or tempo), creating a disruption in the metrical and rhythmic structure. The other syllable changed in this example was "change" and those alterations are not illustrated in the current example.

Procedure

In a quiet room, participants sat at a computer and provided responses individually using the keyboard, while listening to stimuli through headphones. Verbal instructions identical to those in Experiment 1 were given to participants by the experimenter. Participants provided responses for six trials, where each trial began with the full context and was followed by 10 repetitions of one short segment taken from an original, contextualized phrase. Each speech segment repetition was separated by a 2300 ms interstimulus interval during which participants were instructed to respond with their rating of 1 to 5 where "1" indicated that the previous repetition sounded "exactly like speech" and "5" indicated "exactly like song." Participants listened to three sentences in the unaltered rhythm condition and three sentences in the altered rhythm condition. A within-subjects approach was attempted here in order to determine whether auditory illusions could be perceived as transforming and non-transforming stimuli within a single study and for a single participant, which may be of use for subsequent study designs. Trials were presented in one of four semi-random orders, counterbalanced so that each sentence was heard in the unaltered and altered conditions across all participants. Presentation orders were also counterbalanced for hearing an unaltered or altered condition stimulus first.

Results

Average ratings were examined with a 2 x 2 (Repetition [initial, final] x Condition [unaltered, altered]) repeated measures ANOVA. Analysis revealed a main effect for condition, $F(1, 23) = 4.325$, $p < .05$, $\eta_p^2 = .158$, with unaltered trials higher than altered trials (unaltered mean = 2.7, SD = .72; altered mean = 2.4, SD = .43). A main effect was also found for repetition, $F(1, 22) = 102.131$, $p < .001$, $\eta_p^2 = .816$, with the final repetition rated as much more song-like than the initial repetition for both conditions (unaltered: initial mean = 1.96, SD = .76, final mean = 3.53, SD = .90; altered: initial mean = 1.57, SD = .61, final mean = 3.27, SD = .78) as illustrated in Figure **4**. No significant interaction was found for repetition by condition ($p = .586$), however, indicating that the transformation from speech to song occurred regardless of the presence or absence of rhythmic alterations. These results suggest that subtle alterations in the metrical structure of the stimuli do not play a role similar to subtle transpositions in pitch in Experiment 1.

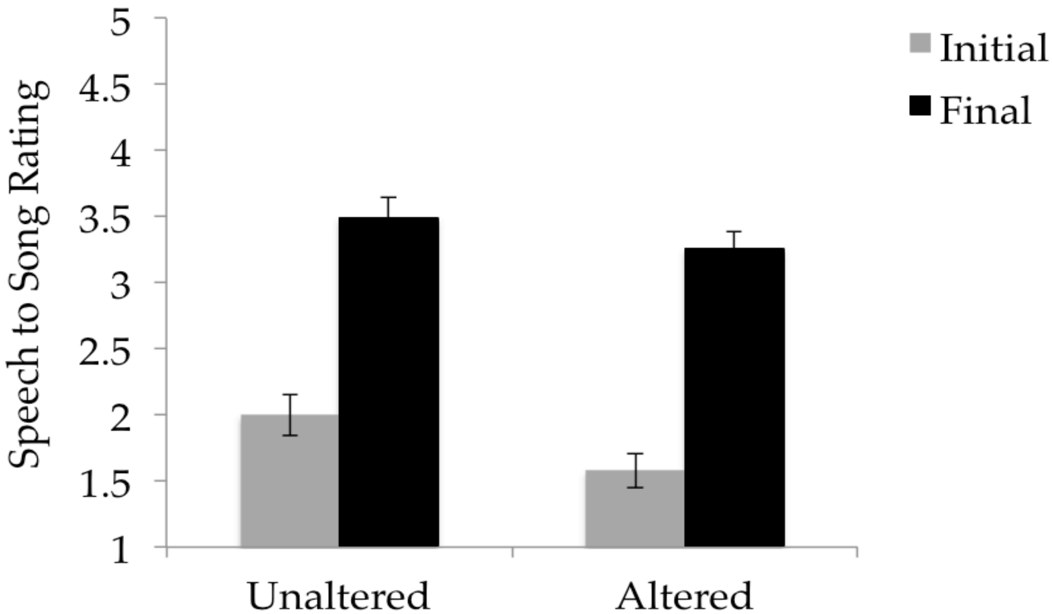

Figure 4: Initial and Final repetitions for the unaltered and altered rhythm conditions. There is no interaction effect between condition and repetition, only two significant main effects for condition and repetition.

Because there was a main effect for condition in the absence of an interaction effect with repetition, the current results indicate that both the final and the initial repetitions were rated higher in the unaltered but not the altered rhythm condition. Order effects may be able to account for this finding such that hearing one condition for the first trial may have primed participants to base subsequent ratings off of the first trial type. For this reason, the between-subjects factor of order was added to the previous analysis. Ratings were evaluated with a 2 x 2 x 2 (Repetition [initial, final] x Condition [unaltered, altered] x Order [unaltered first, altered first]) mixed design ANOVA. In addition to significant main effects for condition, $F(1, 22) = 6.721$, $p < .05$, $\eta_p^2 = .234$, and repetition, $F(1, 22) = 108.654$, $p < .001$, $\eta_p^2 = .832$, a significant interaction effect was found for condition by order, $F(1, 22) = 13.746$, $p < .01$, $\eta_p^2 = .385$. A potentially marginally significant interaction for repetition by order was found, $F(1, 22) = 2.469$, $p = .130$, $\eta_p^2 = .101$, inspection of the data reveals that this trend may be driven by higher average initial ratings for unaltered versus altered trials. The lack of a robust interaction effect for repetition by order indicates that order effects primarily affect responses on a condition level basis. To that end, when participants heard an unaltered trial first, no difference was found between conditions, whereas when an altered trial was heard first, a difference was observed with higher average ratings for unaltered compared to the average of altered rhythm ratings (unaltered first: unaltered mean = 2.3, SD = .72, altered mean = 2.5, SD = .5; altered first: unaltered mean = 3.2, SD = .42, altered mean = 2.4, SD = .37). Results are depicted in Figure **5**.

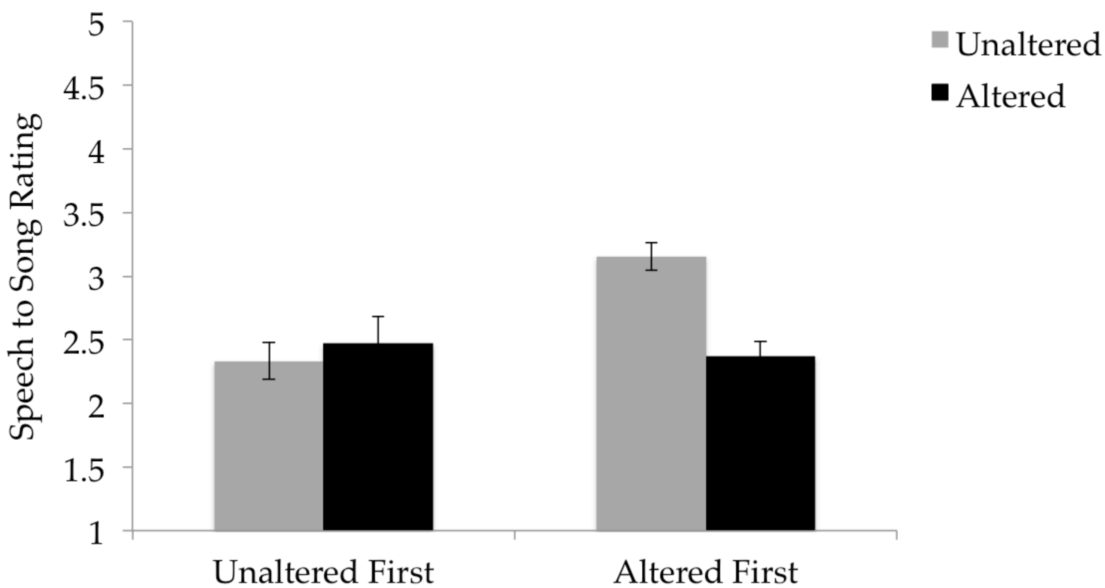

Figure 5: A comparison of ratings (average of initial and final repetitions) for unaltered and altered trials, grouped by whether the first trial presented was in the unaltered or altered condition. When participants heard an altered trial first, unaltered trials were rated higher than altered overall and hearing the unaltered trials first showed no difference.

Even when taking into account the order effects dependent on the type of trial heard first, no disruption of the transformation from speech to song was found by altering the regular rhythmic structure of auditory illusions. By disrupting the regular meter of the original stimulus, the current experiment failed to clearly characterize the contribution of a consistent, regular rhythmic structure to the transformation of speech to song. Clear order effects were found, suggesting that participants based their ratings off of the first trial type that was encountered. Because of this effect, subsequent studies should guard for order effects by keeping with a between-subjects design as was employed in Experiment 1 or a blocked design to minimize effects from prior context.

Finally, the effect of alteration type was investigated for lengthened versus shortened syllable alterations by submitting average ratings to a paired-samples t-test for the altered condition only. No significant difference was found, $t(23) = .606$, $p = .551$ and no difference was found even after accounting for the order effects explained above (all p values > .1). Indeed, a comparison of the average ratings for the same repetitions in unaltered conditions revealed no significant difference between the unaltered or altered repetitions (p's > .4, with ratings for altered condition: long [2, 4, 6, 8] = 2.98, SD = .58, short [3, 5, 7, 9] = 2.91, SD = .70; unaltered condition: long = 2.97, SD = .74, short = 3.0, SD = .74). The results indicate that the effect of lengthening or shortening the syllables was equivalent and did not systematically move ratings toward a more song-like or speech-like rating.

Given the lack of an interaction between condition and repetition, these results imply that altering the rhythm of metrically strong beats within auditory illusions does not affect the transformation of speech to song. It is apparent, then, that when an auditory illusion is repeated the tendency to rate it as transforming from speech the song is quite strong. However, it is possible that even though participants report that the stimulus sounds like song after several repetitions they are recruiting speech-specific representations to process these sounds. Instead of simply using subjective responses to indicate the recruitment of music-specific representations when processing ambiguous speech a more objective measure should be used. Experiment 3 employs subjective ratings in conjunction with participants' pitch change sensitivity for two types of changes anticipated to differ depending on the recruitment of music-specific representations.

CHAPTER 6

EXPERIMENT 3:

Pitch manipulation detection in speech and song

Method

Participants

Forty-eight participants (24 musicians and 24 non-musicians) were recruited in the same manner as Experiments 1 and 2. Participants gave verbal and written informed consent prior to the start of the experiments and completed demographic questionnaires at the experiment's end. Musicians (12 females) were 23.56 years of age on average (range: 18 – 56 years) and reported an average of 10.56 years of formal music training. Non-musicians (14 females) were 19.26 years old on average (range: 18 – 22) and reported .77 years of formal music training on average.

Apparatus

Apparatus was identical to Experiment 1.

Stimuli

Twenty-four auditory illusion stimuli were selected for this experiment. All of the stimuli were previously demonstrated to reliably transform from speech to song (Tierney et al., 2012) (see appendix 1 for repeated segments and contexts). The stimuli were taken from online, open-access audiobook recording websites and consisted of speech segments from three male speakers. Relative to a matched corpus of non-transforming stimuli, these stimuli were reported to have significantly greater fundamental frequency stability, but showed marginally greater rhythmicity as measured by the regularity of the intervals between stressed syllables (Tierney et al., 2012).

Two tasks were used within a single trial. The first task was identical to Experiment 1, in which each participant provided his or her subjective ratings (speech to song) for each repetition (*rating task*). All of the rating task stimuli were unaltered and were presented at the same frequency as the original audio recordings. The second task required a “same” or “different” judgment to indicate the discrimination of two repetitions of a single speech segment (*discrimination task)*. For “different” trials, the second stimulus contained a change in the pitch of a single syllable. Pitch manipulations were produced using Praat and the PSOLA overlap-add method as in Experiment 1.

Pitch manipulations either did or did not conform to Western musical scale relationships. For conforming manipulations, pitches were altered in a manner that was consistent with the melody that would have been produced after the stimulus transformed to song and are referred to as *perceived contours*. The perceived contour for each segment was estimated in a similar fashion as Deutsch et al. (2011) by reproducing the utterances in song as well as with a musical instrument and constraining these contours to fit a diatonic scale. Conforming pitch manipulations moved the average pitch, calculated over the voiced duration of a given syllable, toward the perceived contour; nonconforming pitch manipulations, by contrast, moved away from the perceived contour as depicted in Figure **6**. Average pitches for each of the actual contour syllables were calculated using Praat’s autocorrelation method. All pitch manipulations consisted of one-semitone movements up or down in pitch from the original, average pitch of the syllable. Even though using a fixed pitch manipulation size results in only approximating the conforming pitch center, using identical changes does not bring confounding issues of the detecting changes of different magnitude. Seven subjects gave informed consent to participate in pilot analysis and provided similarity judgments (1 = not similar to 5 = very similar) comparing the perceived contour (played as a piano melody) and the contour they perceived after listening to 10 repetitions of each spoken stimulus. All perceived contours were rated as very similar (average rating of 3 or above) or had at least two "very similar" ratings, suggesting that the majority of the contours were perceived as conforming to other's melodic perception of the stimulus as well. Thus, discrimination tasks had three types: conforming, nonconforming, or no manipulation (same). For each of the 24 stimuli, pitch manipulations only affected one syllable in the second speech segment, but the location of the manipulation (i.e., which syllable was altered) varied between stimuli.

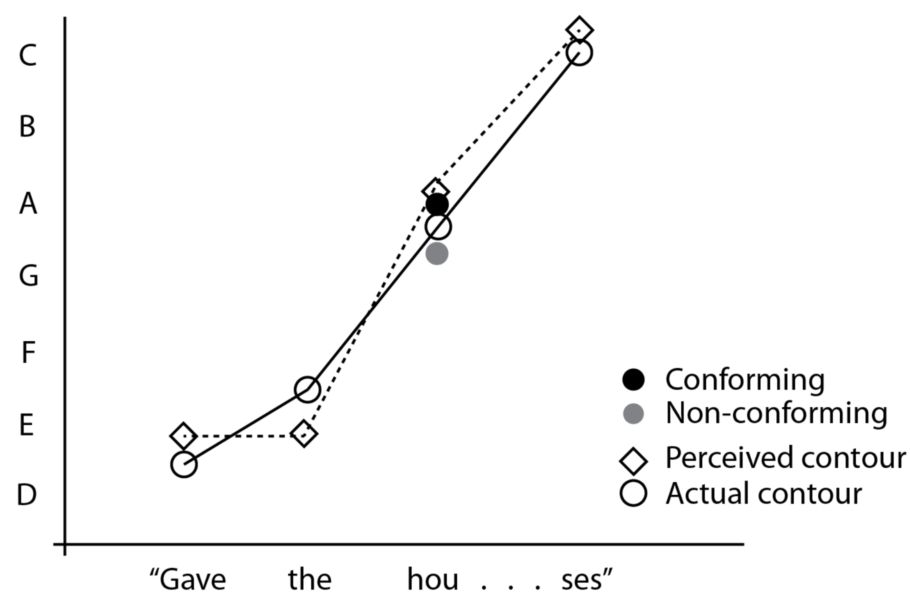

Figure 6: Illustration of conforming and nonconforming pitch manipulations where the green dot illustrates a conforming pitch manipulation that moves from the average actual pitch (solid line) toward the perceived melodic interpretation (dotted line) of the spoken segment. A nonconforming pitch manipulation (grey) moves away from the melodic interpretation.

Procedure

Subjects participated individually in a quiet room wearing headphones. After providing informed consent, participants were given verbal and written instructions by the experimenter. Instructions highlighted the presentation order of the rating and discrimination tasks within a single trial. Each trial began by presenting the full context of the embedded sentence segment (average length = 8604 ms, range 5483 – 15612 ms). At 2300 ms following the presentation of the full context, participants heard one repetition of the speech segment (average length = 1333 ms, range 841 – 1799 ms) and were asked to give their subjective rating from 1 to 5 on a continuum of speech to song using the computer keyboard (see Experiment 1 for details). The next two repetitions were part of the discrimination task, in which the participant was instructed to provide a response of "S" for "same" or "K" for "different." Following the participant's key press (response in the discrimination task), ratings were obtained for each of the subsequent five repetitions (occurring with 2300 ms ISI). Finally, two more repetitions were compared as part of the final discrimination task. The custom Psyscope script prompted participants to press the spacebar to continue on to the next trial. For each trial, a participant heard 10 repetitions of the stimulus with a total of six iterations of the rating task and two iterations (four repetitions total) of the discrimination task (see Figure **7** for a depiction of a single trial).

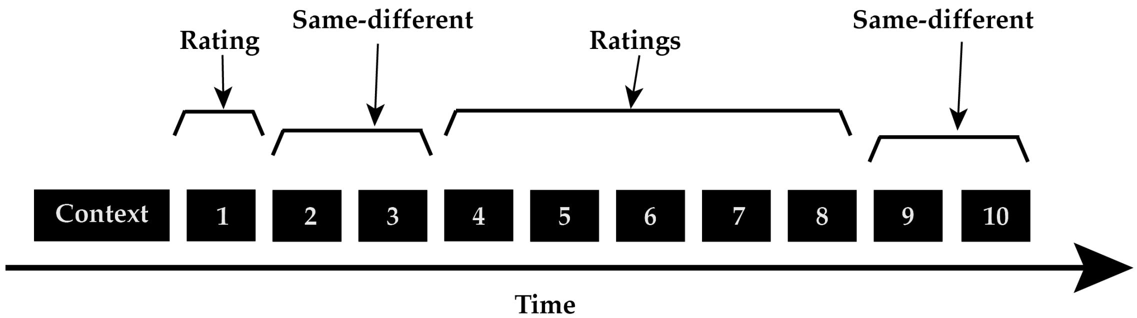

Figure 7: The order of rating and discrimination tasks within a trial. Each trial consists of the full context and the repetition of a single speech segment from the full context of the speech sample.

The discrimination task always began with a standard, unaltered stimulus and was followed by a comparison stimulus that was or was not altered. Each participant was presented with eight trials with no pitch manipulation ("same" trials) and 16 trials with a pitch manipulation ("different" trials). Among the "different" trials, there were eight conforming and eight nonconforming pitch manipulations. Additionally, the order of discrimination task pairs (i.e., initial and final discrimination tasks within a trial) was semi-random such that, over the course of the experiment, participants were presented with 12 trials containing the same manipulation type pair (e.g., same-same, nonconforming-nonconforming, etc.) while the other 12 trials contained contrasting pairs (e.g., same-conforming, nonconforming-conforming, etc.). Participants provided responses for 24 trials administered within a single test block. Three semi-random presentation orders were created using a Latin square design, allowing each sentence to be presented with all three types of discrimination task pitch manipulations in a between-subjects fashion.

Results

Results are reported using the dependent measure d' (sensitivity) for conforming and nonconforming pitch manipulations during different trials for the initial and final discrimination tasks. d' was calculated to take participant bias into account as raw data comparing non-musicians and musicians indicated different rates of responding "same" for all different trials (musicians = 43.9%, non-musicians = 65.5%). Each participant's hit rate was calculated as the percentage of different trials correctly identified as different and false alarms were considered same trials mistakenly identified as different. Because participants theoretically used an independent-observation strategy, estimating differences by making comparisons to a standard stimulus instead of estimating the difference between the pair of discrimination task stimuli (as is theorized for participants using a differencing strategy), the independent-observation strategy calculation for d' in a same-different task was used (Macmillan & Creelman, 2005).[1] The sensitivity for detecting conforming and nonconforming pitch manipulations will determine whether pitch detection abilities vary according to perceived context (i.e., perceived as speech versus song). Sensitivity scores (d') were entered into a 2 x 2 x 2 (Group [musicians, non-musicians] x Position [initial, final] x Manipulation [conforming, nonconforming]) repeated-measures ANOVA. Results are shown in Table 1.

---

[1] Because some participants had zero false alarms or a 100 percent hit rate, one-half (.5) was added to the number of hits and false alarms for each participant and divided by n+1 in order to prevent infinite values for d', where n is the number of trials being averaged (Hautus, 1995). This correction is called the log-linear correction and is less biased than other corrections.

d' for all trials	*df*	*F*	*p*	*Partial eta²*
Group	1, 46	6.374	.015*	.122
Position	1, 46	15.409	.000***	.251
Manipulation	1, 46	.099	.755	.002
Position x group	1, 46	3.355	.073+	.068
Manipulation x Group	1, 46	1.912	.173+	.040
Manipulation x Position	1, 46	1.576	.216	.033
Manipulation x Position x Group	1, 46	.612	.438	.013

d' for all trials, non-musicians only	*df*	*F*	*p*	*Partial eta²*
Position	1, 23	2.085	.162+	.083
Manipulation	1, 23	.497	.488	.021
Manipulation x Position	1, 23	.105	.748	.005

d' for all trials, musicians only	*df*	*F*	*p*	*Partial eta²*
Position	1, 23	17.472	.000***	.432
Manipulation	1, 23	1.692	.206	.069
Manipulation x Position	1, 23	2.213	.150+	.088

Table 1: d' F test results for all trials together (* < .05, ** = < .01, *** = < .001, + = marginal significance).

A main effect was found for position, where sensitivity at the final position was significantly higher than sensitivity at the initial discrimination task position (initial d' = .99, final d' = 1.62). There was no main effect for the type of manipulation and, against predicted findings, no interaction between manipulation and position (see Table 1). A main effect for group was found, such that musicians' sensitivity was greater initial and final repetitions (musicians: initial d' = 1.18, final d' = 2.11; non-musicians: initial d' = .80, final d' = 1.13). Different patterns of sensitivity for non-musicians and musicians were indicated by a marginally significant interaction between position and group.

Because of and interaction between position and group, two separate 2 x 2 repeated-measures ANOVAs with position and manipulation were run for each group. Surprisingly, no significant main effects or interactions were found for non-musicians (all p's > .1, see Table 1), but sensitivity was significantly greater than chance $F(1, 23) = 36.201$, $p < .001$, $\eta_p^2 = .611$. The results for all participants above seem to be primarily driven by musicians who had a significant main effect for position, but no main effect for manipulation type and similarly no interaction effect for position by manipulation (see Figure **8**). Musicians, but not non-musicians, benefited from a greater number of repetitions and exhibited greater sensitivity for the final discrimination task relative to the initial discrimination task.

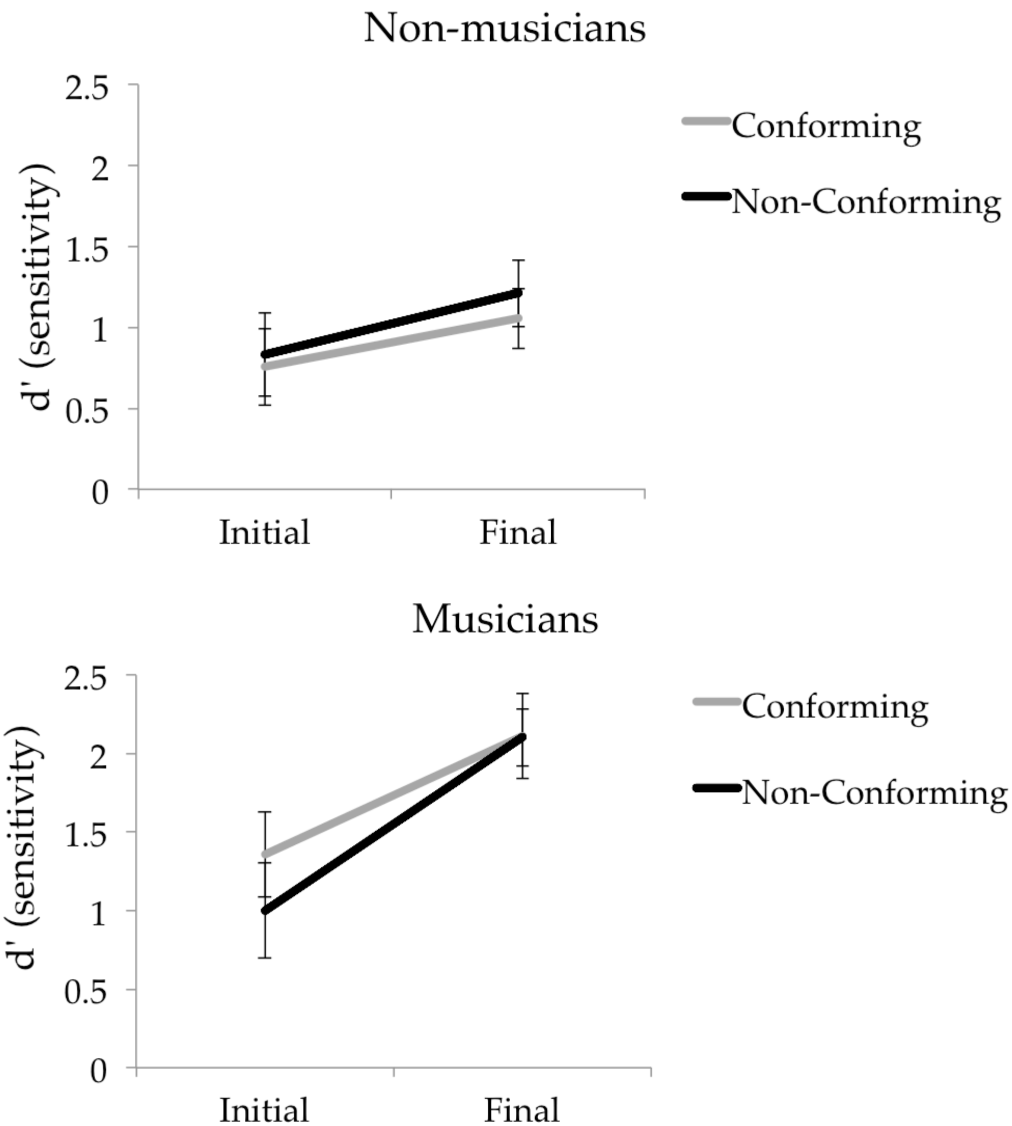

Figure 8: Sensitivity (d') computed separately for non-musicians (top) and musicians (bottom) comparing conforming and nonconforming pitch manipulation detection for the initial and final discrimination task positions.

The previous analysis indicated that musicians and non-musicians did not demonstrate different patterns of sensitivity for conforming compared to nonconforming pitch manipulation types. That is, the detection of conforming pitch manipulation types did not remain stable from initial to final discrimination task position and the detection of nonconforming pitch manipulation increased over time as predicted. These findings were not consistent with the anticipated results, suggesting that the pitch change detection paradigm used in this experiment did not indicate the recruitment of music-specific representations and, similarly, perceived melodic structure does not differentially mask or facilitate pitch manipulation detection. However, the previous analyses did not take into consideration participants' subjective ratings of speech or song for the initial or final discrimination task position.

In order to address the original question of whether hearing the stimulus in a speech mode compared to a music mode of listening would bring about different sensitivity for conforming and nonconforming pitch manipulation types, the adjacent rating trials occurring before the initial and final discrimination tasks were used to estimate participants' perception during each discrimination task. That is, the first and the eighth repetitions (see Figure **7**) were used to estimate mode of perception for the discrimination of repetitions 2 to 3 and 9 to 10. These rating responses were used to categorize each of the participants' trials into two possible percepts: stable perception over time (e.g., speech percept to speech percept or song to song) or transforming (e.g., speech to song). Ratings of a 1 or 2 were taken to indicate "speech" and 3-5 were taken to indicate "song" according to the results of Experiment 1. These two outcomes were used to calculate new d' scores using the same calculations reported above. d' scores were entered into a 2 x 2 x 2 x 2 (Group [musicians, non-musicians] x Position [initial, final] x Manipulation [conforming, nonconforming] x Percept [stable, transformed]) repeated-measures ANOVA. Results are shown in Table 2. This procedure resulted in list-wise deletion of several subjects who had 0 trials in a given cell of the factorial design.

Analyses were run on 16 non-musicians and 18 musicians, for a total of 34 participants.

d' for all trials, categorized by percept	*df*	*F*	*p*	*Partial eta²*
Group	1, 32	3.067	.089⁺	.087
Position	1, 32	5.229	.029*	.140
Manipulation	1, 32	.002	.965	.000
Percept	1, 32	.406	.529	.013
Position x Group	1, 32	.688	.413	.021
Manipulation x Group	1, 32	1.121	.298	.034
Percept x Group	1, 32	.281	.600	.009
Manipulation x Percept	1, 32	1.352	.253	.041
Position x Manipulation	1, 32	4.704	.038*	.128
Position x Percept	1, 32	.399	.532	.012
Position x Manipulation x Percept	1, 32	1.982	.169⁺	.058
Position x Manipulation x Group	1, 32	.172	.681	.005
Position x Percept x Group	1, 32	.311	.581	.010
Manipulation x Percept x Group	1, 32	.181	.673	.006
Position x Manipulation x Percept x Group	1, 32	.553	.463	.017

Table 2: Sensitivity (d') for trials categorized by each participant's percept at the beginning and end of a given trial. Stable perception indicates no change (i.e., speech-speech or song-song) while transforming perception indicates a change from speech to song (* < .05, ⁺ = marginal significance).

A main effect for position was found, again showing that d' scores were higher at the final position than initial position. The main effect of group was marginally significant such that musicians demonstrated higher overall discrimination. No main effects for manipulation or percept were found, yet a statistically significant interaction effect was found for position by manipulation. Most importantly, a marginally significant 3-way interaction for position by manipulation by percept, as illustrated in Figure **9**, is consistent with the predicted effect. This finding indicates a trend toward the recruitment of music-specific representations only when stimuli are perceived as transforming from speech to song. All other main effects and interactions did not reach significance. Post hoc analyses comparing stable and transforming percept trials in separate ANOVAs reveal a significant interaction of position by manipulation for transforming trials with $F(1, 37) = 5.637$, $p < .05$, $\eta_p^2 = .132$, and a marginally significant main effect for position with $F(1, 43) = 3.871$, $p = .056$, $\eta_p^2 = .083$ for stable trials. For both analyses all other effects did not reach statistical significance.

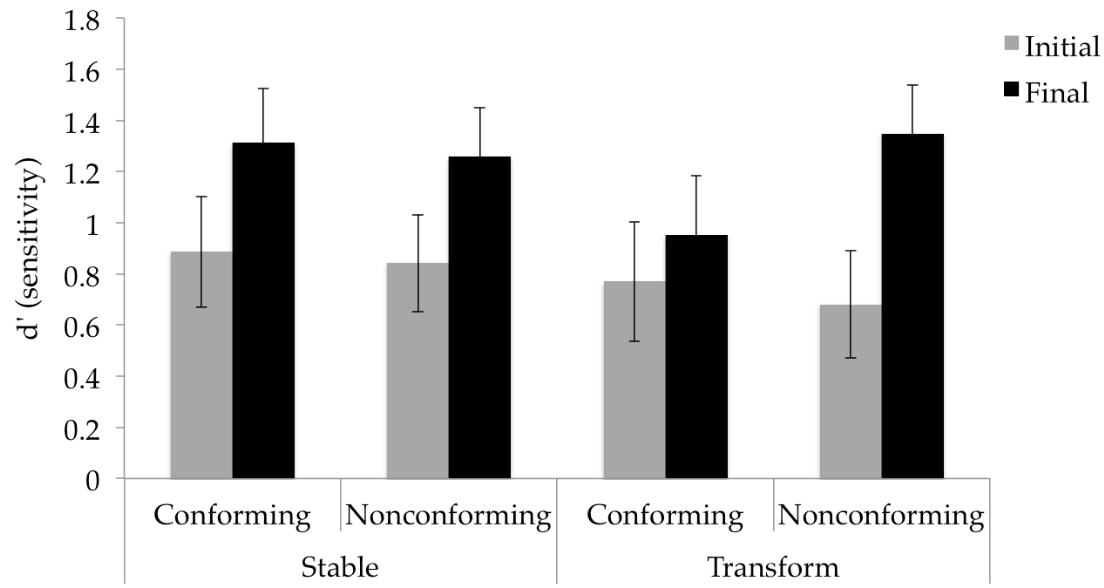

Figure 9: d' categorized by participants' speech to song ratings for all subjects. Stable perceptual experience shows an effect of position, possibly due to better encoding after hearing several repetitions of the stimulus. Trials in which participants perceived a transformation from speech to song show a significant change in sensitivity from initial to final position on nonconforming trials only.

These results must be taken with caution however, given that the number of trials in each condition for each participant was sometimes small and may have resulted in a biased measure of d'. These results indicate a trend toward the anticipated findings, and either a greater number of trials or participants are needed to fully understand the role of perception for the detection of conforming or nonconforming pitch chances.

Beyond the subjective appraisal of these ambiguous stimuli as either speech or song, the current experiment tentatively demonstrates an objective measure for the recruitment of domain-specific knowledge for processing music after several repetitions of a single stimulus. By keeping low-level acoustic characteristics constant, the presence of a perceptual shift indicates that high-level, top-down processes are involved in the conscious perception of music in this auditory illusion. Most importantly, when no perceptual shift from speech to song is reported, only a marginally significant effect of repetition is found with greater sensitivity for final discrimination tasks compared to initial regardless of pitch change type. As anticipated, when participants reported the transformation of sentences from speech to song, Western musical knowledge interfered with the detection of conforming and facilitated the detection of nonconforming pitch changes.

CHAPTER 7

DISCUSSION

This collection of experiments helped to clarify the nature and recruitment of higher-level knowledge elicited by auditory illusions from speech to song. In Experiment 1, non-musicians' subjective ratings for the initial and final repetitions of the stimulus were lower overall than musicians, but showed the same pattern for subjective ratings during *transposed* and *untransposed* conditions. That is, non-musicians also reported a disruption in the transformation from speech to song when the middle eight repetitions were transposed slightly in pitch by about a half step on the Western musical scale, indicating the importance of exact pitch information. Deutsch et al. (2011) also found that mixing the presentation order of the syllables within the sentence similarly disrupted the transformation, demonstrating the importance of pitch relationships and not simply the presentation of syllables with a more stable fundamental frequency (Tierney et al., 2012). The generalizability of these findings to non-musicians suggests that music training is not necessary to form a melodic percept of this ambiguous sentence, implying a reliance on implicit knowledge for tonal or rhythmic structure and musical syntax through experience (Bigand & Poulin-Charronnet, 2006; Hannon & Trainor, 2007). These results replicate and extend previous studies by including subjective ratings of non-musicians for Deutsch and colleagues' (2011) "sometimes behave so strangely" stimulus.

A curious effect in the transposed condition was found, whereby any change moving higher in pitch was rated as sounding more song-like and any change going lower in pitch was rated more speech-like. However, it is possible that if the mean fundamental frequency of one's singing voice is typically higher in pitch than one's speaking voice, except in the case of infant-directed speech and singing (Bergeson & Trehub, 2002), participants may tap into past experiences with spoken and sung sounds and rate the current transpositions according to this dichotomy.

As the mere presence of repetition is not enough to bring about a perceptual transformation from speech to song, the exact pitches are critical to creating this perceptual transformation. Yet other characteristics may be necessary such as interval relationships and metrical or rhythmic regularity. Experiment 2 demonstrated that altering the temporal regularity of these stimuli does not reliably disrupt the perception of song after several repetitions. Instead, order effects, depending on the first trial type presented, indicated that hearing a trial from the altered condition first resulted in greater ratings overall for unaltered trials, but hearing a trial from the unaltered condition first resulted in no difference between conditions. While these results don't seem to have bearing on the role of regularity during the perception of the current auditory illusions, they point out that participants used past experience as a referent for future speech-to-song judgments.

Interestingly, participants did not systematically respond differently to increases or decreases in the duration of a single syllable within an illusion. Shortened and lengthened stressed syllable durations from the altered condition were statistically indistinguishable from each other and indistinguishable from the analogous repetitions from the unaltered condition trials. This raises the question of whether the alterations were perceptually large enough to bring about a disruption. It is possible that the shortened and lengthened syllables are being regularized toward their original metrical position and duration. Previous studies have demonstrated that slight perturbations are perceived as regular because cognitive mechanisms involved in beat perception are biased toward integer ratios (Motz, Erickson, & Hetrick, 2013). Also, previous entrainment studies have shown that musicians entrain to higher levels of a beat hierarchy than non-musicians (Drake, Jones & Baruch, 2000). Because changes were meant to alter the metrically strong beats of the sentences, perhaps non-musicians paid more attention to lower levels of the hierarchy and did not perceive these disruptions. Although 100 ms changes are rather large in the speech domain, larger alterations in duration may be necessary to completely disrupt the perceived regularity of auditory illusions. Finally, individual differences in experience with speech or music (e.g., percussionists) may interact with the perception of these stimuli in the light of recent work suggesting that listeners can entrain to speech rhythms (Lidji et al., 2013) and do so differently based on linguistic background. This same study also suggested that the English language, with greater nPVI than French, might actually be more rhythmic due to the greater alternation of strong and weak beats. If this is the case then the current manipulations may have inadvertently increased variability and subsequently rhythmicity causing no change in the altered and unaltered conditions. Further studies should take into consideration other measures of rhythmicity as well as participants' individual biases toward beat- or interval-based strategies for rhythm processing (Snyder, Pasinski & McAuley, 2011). Individual differences may produce differing perception of the stimuli in terms of perceiving alterations as speeding up or slowing down which may have an effect on subjective ratings.

While the results of Experiment 2 do not rule out the contribution of rhythmic or metrical regularity to the transformation from speech to song, they may suggest a lesser role than pitch information. The hypothesis from Experiment 2 should also be tested using a between-subjects design as was done in Experiment 1 to determine whether the potential effects of rhythmic regularity were obscured by order. It may be that two conditions within a single block produce order effects, while a block of auditory illusions presented within one condition (e.g., either all unaltered or all altered) allows participants to properly assess repetitions as spoken or sung as was the case in the final experiment's subjective ratings.

Experiment 3 demonstrated a trend toward the recruitment of music-specific knowledge for both musicians and non-musicians when perceiving ambiguous speech. When participants reported the final rating task of a trial to be more like song, they recruited music-specific representations to process these stimuli as exhibited by better discrimination for the final than initial nonconforming change trials compared to conforming change trials. Importantly, the unique nature of these stimuli allowed the same physical stimulus to be perceived as either speech or song depending on context and in some cases individual differences. When the stimulus was perceived as speech during the final rating task of a given trial, sensitivity marginally increased for both conforming and nonconforming pitch change manipulations. Sensitivity was generally better for musicians than non-musicians, fitting with previous research suggesting musical training increases pitch change detection abilities (Schon, Magne & Besson, 2004). High-level musical knowledge may allow all listeners to form melodic expectations for the upcoming repetition of a sentence's contour. When a pitch change moves toward a musically expected pitch height (i.e., a conforming pitch change), listeners may perceive the changed pitch to fit well within their melodic representation of the sentence, whereas nonconforming pitch changes may create a pop-out effect against the expected contour. These results are consistent with hypotheses indicating greater enculturation for native musical scale intervals, with a more internalized framework for the Western major and minor scales compared to the Javanese Pelog scale for Western adults (Lynch et al., 1990, 1991). They are also consistent with studies showing context-dependent auditory object identification (Krishnan et al., 2013), where changes that are incongruent with the auditory context are more readily detected than those that are congruent.

A domain-specific theory of music and language organization is consistent with the current results because, as articulated by Zatorre & Gandour (2007), despite identical acoustic characteristics, higher-level knowledge altered the status of the stimulus. Moreover, a purely domain-general view of organization is inconsistent given that perceiving the stimulus as speech or song differentially affected pitch change detection. However, it seems likely that the higher-level status of the stimulus was altered because of changes in the salience of specific acoustic features (e.g., stable pitch information) as a result of repetition, a hypothesis that is incompatible with strong forms of domain-specific processing. Rather, a combination of these theories appears most plausible with higher-level knowledge being recruited depending on the relative weighting of elements within the sound stimulus. Contrary to previous work suggesting that detection of pitch changes is a general mechanism within the spoken and sung domains (Schon, Magne, & Besson, 2004), it is possible that top-down knowledge may inhibit lower-level frequency change detection mechanisms depending on the expectations instantiated within a given domain. The current results tentatively provide evidence for domain-specific mechanisms in adulthood through the preferential recruitment of knowledge particular to the music domain. Yet the processes that bring about the perception of an identical stimulus as either speech or song may involve domain-general mechanisms. Further studies using manipulations based on domain-specific knowledge such as the greater importance of pitch relationships in music compared to speech (Zatorre & Baum, 2012) are warranted to further clarify the processes involved in the recruitment of music- or language-specific mechanisms and what types of manipulations remain tied to domain-general processes.

Why specific speech excerpts can be perceived as song upon several repetitions is still not fully understood. As was indicated by the experiments above, specific physical characteristics that are perhaps more song-like in nature contribute to the transformation. A more stable fundamental frequency than typical speech might allow listeners to extract and process pitch information more readily. Repetition of the utterance is also essential to bringing about the speech-to-song illusion. Margulis (2013) hypothesized that repetition is one of the characteristic differences between language and music and suggested that the mere presence of repetition may delineate sung versus spoken utterances. While the current results from the transposed condition in Experiment 1 and the existence of a non-transforming set of stimuli in the Tierney et al. study do not support this claim, repetition may serve an important role in directing the locus of attention. Margulis points out that repetition may act to alert the listener to another aspect of the auditory object, such as the prosodic features of a speech stream. Upon the first repetition, listeners may recruit the highest level of the auditory hierarchy required to perform the listening task and only upon greater task demands or upon subsequent repetitions are other, less salient aspects attended to. This interpretation is consistent with the process of top-down guided perception in Reverse Hierarchy Theory (Ahissar & Hochstein, 2004; Nahum, Nelken, &

Ahissar, 2008). Applying this theory to auditory illusions would imply that high-level phonological representations are accessed initially because of contextual task demands, yet, with repetition, more low-level, suprasegmental information becomes more salient. If auditory illusions also make use of a greater pitch range than non-transforming sentences, then musical knowledge could be more easily mapped onto each sentence's contour.

The mechanisms involved in domain-specific processing and implicated by Reverse Hierarchy Theory lead to testable hypotheses concerning the neural organization and processing of speech-to-song utterances. Deutsch et al. (2011) hypothesized the presence of inhibitory control over pitch processing areas in the brain during the processing of speech. With repetition, pitch information areas are no longer inhibited and sentences can be processed as sung. Tierney and colleagues' (2012) findings may be consistent with this hypothesis, as greater activation was found in working memory for pitch and auditory-vocal motor integration areas of the brain for transforming compared to non-transforming speech samples.

Margulis (2013) likened the auditory illusion to semantic satiation in which repeating or viewing a single word in succession forces the word to briefly lose semantic meaning. Recent work suggests that increased processing time for repeated words is not caused by meaning satiation (i.e., adaptation at the level of semantic neural representations) or lexical satiation (i.e., adaptation at the level of orthographic or phonological representation), but by a temporary disruption between lexical and semantic nodes for the repeated word, called associative satiation (Tian & Huber, 2010). Specifically, during a speeded category-matching task, reaction times were not affected when words or meanings were repeated in separate experiments, but reaction times slowed when both the word and its associated meaning was repeated. The authors implicated synaptic depression as the mechanism for the apparent dissociation caused by repetition (Tian & Huber, 2013). Synaptic depression is described as separate from other mechanisms of adaptation in that the sending and receiving neurons or neuronal populations remain active (as evidenced by no change in performance for word- or meaning-only repetition experiments described above), but synaptic resources become depleted causing weakened transmission of the signal (Tian and Huber, 2013). Other forms of adaptation such as suppression involve the active inhibition of irrelevant features or streams of stimuli represented by distinct neural populations (Bidet-Caulet et al., 2009; Zanto & Gazzaley, 2009). Suppression in such ERP studies was evidenced by a pattern of distinct activation for faciliatory and inhibitory effects part of the mechanism for selective attention.

In light of Tierney et al.'s findings, semantic satiation may be unrelated to the speech-to-song illusion because, instead of a reduction in synaptic connectivity, a change in the locus of attention may have suppressed activation for speech processing regions because greater activity was seen in areas of the brain related to pitch processing and production. These results may implicate a reallocation of attention instead of a simple reduction in connectivity. Yet, because task demands in the current study and previous work do not require reactivation of speech-specific representations, it is difficult to determine whether synaptic depression, attentional shifts, or a combination of both leads to the transformation from speech to song.

Further research should address the neural mechanisms involved in processing the speech-to-song illusion to determine which areas of the brain are involved in music- or language-specific processing when physical acoustic characteristics are held constant. The unique characteristics of auditory illusions can further research efforts concerning information processing theories such as Reverse Hierarchy Theory or mechanisms of adaptation and their combined effects on the perception ambiguous sounds. Comparisons in infancy and childhood will provide insight into the development of domain-specific and domain-general mechanisms. Using music and language to characterize the nature of the organization and processing of sounds in the brain will help to inform learning interventions by providing insight into linguistic and musical overlap or separation. Research may also suggest specific types of interventions that are most effective during the development of music- or language-specific knowledge.

APPENDICES

Appendix A: Tierney stimuli sentence segments and contexts

1. And sir **here is no less** joynge (i.e., joyful) and rejoicing in these parties for the birth of our prince…
2. The beams were painted red, or blue, or black according to the owners taste and this **gave the houses** a very picturesque look
3. Time wore on pleasantly and likewise smoothly on the whole, **snags and sandbars** grew less and less frequent
4. But Tom, his grandmother, **and his two sisters**, Bett and Nan, were not restricted they had all the floor to themselves and might sleep where they chose.
5. It seems absurdly simple and yet **somehow I can get** nothing to go upon there's plenty of thread no doubt but I can't get the end of it into my hand
6. We will be shown into the sitting room to wait for the lady, but it is probable that **when she comes she may find** neither us nor the photograph
7. **Some little distance down** Fred Needle street upon the left-hand side there as you may have remarked a small little angle in the wall
8. For it had never been his good luck to own and eat one , **there was a cold drizzle** of rain, the atmosphere was murky, it was a melancholy day
9. And it was not thought advisable by the council to begin a new reign by the death of the greatest nobleman in the kingdom, who had been condemned by a sentence so unjust and tyrannical, **Hume's history of England**, volume 3, page 307, end of footnote.

10. The landlady informed me that he had left the house **shortly after eight o'clock** in the morning, I sat down beside the fire however with the intention of awaiting him however long he might be.

11. To say nothing of half a dozen other **people in the neighborhood** in whom I was not in the least interested, but whose biographies I was compelled to listen to.

12. In truth, being a king is not all dreariness, **it hath its compensations** and conveniences.

13. The night lights by the beds of the three children continued to burn clearly they were awfully nice little night lights and **one cannot help wishing** that they could have stayed awake to see Peter.

14. I am sorry that Ms. Sutherland has troubled you about this this little matter for I think it is far better not to wash **linen of this sort in public**

15. It is not yet nine, the streets will be crowded so I trust that you may be in safety and yet you **cannot guard yourself** too closely.

16. "It had escaped my memory, **I have had nothing since breakfast**." "Nothing?" "Not a bite, I had no time to think of it."

17. But she could hardly have allowed anyone to acquire so deep an influence over her that **would induce her to change her plans** so completely.

18. It is the fairy language, you ordinary children can never hear it, but if you were to hear it, **you would know that you had heard it** once before.

19. For, presently, we were aware of a **sudden commotion on the deck** of the stranger who immediately afterward ran up a British flag and hauling her wind bore up directly upon us

20. But the next moment he was himself disturbed by it and showed discomposure for **this was the only service** he had been permitted to do with his own hands during the meal.
21. But bethought him of the nuts he had brought away from dinner **and joy it would be** to eat them with no crowd to eye him.
22. So **he restored the pretty things** to their several places and soon was cracking nuts and feeling almost naturally happy.
23. He was hardly even heard so great the turmoil, **the prince continued to struggle** for freedom and to rage against the treatment he was suffering
24. The attendants flew to his assistance but he put them aside and said “Trouble me not, it is **nothing but a scurvy faintness**.”

Appendix B: IRB Approval

Social/Behavioral IRB – Expedited Review

Revisions Request

DATE: March 12, 2013

TO: **Dr. Erin Hannon,** Psychology

FROM: Office of Research Integrity – Human Subjects

RE: Notification of IRB Action
Protocol Title: **Recruitment of musical and linguistic knowledge while processing ambiguous sound patterns**
Protocol #: 1303-4388

This memorandum is notification that the project referenced above has been reviewed by the UNLV Social/Behavioral Institutional Review Board (IRB) as indicated in Federal regulatory statutes 45CFR46.

Please complete the following revisions and resubmit to ORI-HS for review.

1. Protocol Proposal Form:
   a. Section 17.3.1 indicates that participants will receive $5 cash award. This should be deleted.
2. Informed Consent:
   a. Participants - Verbiage should be clarified. As worded, it is not clear whether a person could participate if she/he considers self as a musician, wants to participate for subject pool credit, but did not have three or more years of formal training before age 12.
   b. Separate consent sections are needed for audio **and** for video recording.
   c. The term 'understand' should be removed from the audio and for video consent sections (implies that PI can assess understanding).
3. Questionnaire:
   a. Questionnaire requests social security number. SS# cannot be used.

To continue the review of the protocol named above, please make the revisions requested above and submit all revised documents.

If you have questions or require any assistance, please contact the Office of Research Integrity – Human Subjects at IRB@unlv.edu or call 895-2794.

REFERENCES

Ahissar, M. & Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning. *TRENDS in Cognitive Sciences, 8*(10), 457-464.

Auzou, P., Eustache, F., Etevenon, P., Platel, H., Rioux, P., Lambert, J., Lechevalier, B., et al. (1995). Topographic EEG activations during timbre and pitch discrimination tasks using musical sounds. *Neuropsychologia*, *33*(1), 25–37.

Ayotte, J., Peretz, I., & Hyde, K. (2002). Congenital amusia: A group study of adults afflicted with a music-specific disorder. *Brain*, *125*(2), 238–251.

Bergeson, T. R. & Trehub, S. E. (1999). Mothers' singing to infants and preschool children. *Infant Behavior & Development, 22*(1), 51-64.

Bidet-Caulet, A., Mikyska, C., & Knight, R. T. (2010). Load effects in auditory selective attention: Evidence for distinct facilitation and inhibition mechanisms. *NeuroImage, 50*, 277-284.

Bigand, E., & Poulin-Charronnat, B. (2006). Are we "experience listeners"? A review of the musical capacities that do not depend on formal music training. *Cognition, 100*, 100-130.

Brown, S. (2000). The "musilanguage" model of music evolution. In N. L. Wallin, B. Merker, & S. Brown (Eds), *The origins of music* (pp 271-300). Cambridge, MA: MIT Press

Brown, S., Merker, B., & Wallin, N. L. (2000). An introduction to evolutionary musicology. In N. L. Wallin, B. Merker, & S. Brown (Eds), *The Origins of Music* (pp 3-24). Cambridge, MA: MIT Press

Boersma, P. & Weenink, D. (2001). Praat, a system for doing phonetic by computer. *Glot International 5*(9/10), 341-345.

Burns, E. M., & Ward, W. D. (1978). Categorical perception – phenomenon or epiphenomenon: Evidence from experiments in the perception of melodic musical intervals. *Journal of the Acoustical Society of America, 63*(2), 456-468.

Chiappe, D., & MacDonald, K. (2005). The Evolution of domain-general mechanisms in intelligence and learning. *The Journal of General Psychology*, *132*(1), 5–40.

Chomsky, N. (1959). A review of B. F. Skinner's verbal behavior. *Language*, *35*(1), 26–58.

Cohen J.D., MacWhinney B., Flatt M., and Provost J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments, and Computers, 25*(2), 257-271.

Cosmides, L., & Tooby, J. (1994). Origins of domain specificity: The evolution of functional organization. *Mapping the mind: domain specificity in cognition and culture* (pp. 85–116). Cambridge, UK: Cambridge University Press.

Darwin, C. (1871). *The Descent of Man*. London: John Murray

Dehaene-Lambertz, G., Pallier, C., Serniclaes, W., Sprenger-Charolles, L., Jobert, A., & Dehaene, S. (2005). Neural correlates of switching from auditory to speech perception. *NeuroImage*, *24*(1), 21–33.

Deutsch, D. (2003). *Phantom Words, and Other Curiosities*. (Philomel Records, La Jolla) Compact Disc; Track 22.

Deutsch, D., Henthorn, T., &Lapidis, R. (2011). The illusory transformation of speech to song. *Journal of the Acoustical Society of America, 129* (4), 2245-52.

Drake, C., Jones, M. R., Baruch, C. (2000). The development of rhythmic attending in auditory sequences: attunement, referent period, focal attending. *Cognition, 77*, 251-288.

Dunn, J. C., & Kirsner, K. (2003). What can we infer from double dissociations? *Cortex*, *39*(1), 1–7.

Elman, J. L. (1993). Learning and development in neural networks: the importance of starting small. *Cognition*, *48*(1), 71–99.

Falk, D. (2004). Prelinguistic evolution in early hominins: Whence mothers? *Behavioral and Brain Sciences, 27,* 491-541.

Fernald, A. & Kuhl, P. (1987). Acoustic determinants of infant preference for motherese speech. *Infant Behavior and Development, 10,* 279-293.

Fodor, J. A. (1983). *The modularity of mind: An essay on faculty psychology*. MIT Press.

Ford, M. P., Malone, L. A., Nyikos, I., Yelisetty, R., & Bickel, C. S. (2010). Gait training with progressive external auditory cueing in persons with Parkinson's disease. *Archives of Physical Medicine and Rehabilitation*, *91*(8), 1255–1261.

Frances (1988). *The perception of music*. Tranlsated by W. J. Dowling. Hillsadale, NJ: Erlbaum.

Greenough, W. T., Black, J. E., & Wallace, C. S. (1987). *Child Development, 58*(3), 539-559.

Hannon, E. E., Soley, G., & Levine, R. S. (2011). Constraints on infants' musical rhythm perception: Effects of interval ratio complexity and enculturation. *Developmental Science*, *14*(4), 865–872.

Hannon, E. E., & Trainor, L. J. (2012). Musical development. In D. Deutsch (Ed.), *The Psychology of Music, 3rd edition* (pp.423-498). San Diego, CA: Academic Press.

Hannon, E. E., & Trehub, S. E. (2005). Metrical categories in infancy and adulthood. *Psychological Science, 16*, 48-55.

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, *8*(5), 393–402.

Holt, L. L., Lotto, A. J., & Diehl, R. L. (2004). Auditory discontinuities interact with categorization: Implications for speech perception. *The Journal of the Acoustical Society of America*, *116*(3), 1763–1773.

Hyde, K. L., Lerch, J. P., Zatorre, R. J., Griffiths, T. D., Evans, A. C., & Peretz, I. (2007). Cortical thickness in congenital amusia: When less is better than more. *The Journal of Neuroscience, 27*(47), 13028-13032.

Jacome, D. E. (1984). Aphasia with elation, hypermusia, musicophilia and compulsive whistling. *Journal of Neurology, Neurosurgery, and Psychiatry, 47,* 308-310.

Jackendoff, R., & Lerdahl, F. (2006). The capacity for music: What is it, and what's special about it? *Cognition*, *100*(1), 33–72.

Joanisse, M. F. & Gati, J. S. (2003). Overlapping neural regions for processing rapid temporal cues in speech and nonspeech. *NeuroImage, 19*, 64-79.

Kanwisher, N. (2000). Domain specificity in face perception. *Nature Neuroscience*, *3*(8), 759–763.

Karmiloff-Smith, A. (1992). *Beyond modularity: A developmental perspective on cognitive science*. MIT Press.

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: evidence for a domain general learning mechanism. *Cognition*, *83*(2), B35–B42.

Kirschner, S., & Tomasello, M. (2010). Joint music making promotes prosocial behavior in 4-year-old children. *Evolution and Human Behavior*, *31*(5), 354–364.

Koelsch, S., Kasper, E., Sammler, D., Schulze, K., Gunter, T., Friederici, A. D. (2004). Music, language, and meaning: brain signatures of semantic processing. *Nature Neuroscience 7*(3), 302-307.

Koelsch, S. (2005). Neural substrates of processing syntax and semantics in music. *Current Opinion in Neurobiology*, *15*(2), 207–212.

Krishnan, S., Leech, R., Aydelott, J., & Dick, F. (2013). School-age children's environmental object identification in natural auditory scenes: Effects of masking and contextual congruence. *Hearing Research, 300*, 46-55.

Kondo, H. M., & Kashino, M. (2007). Neural mechanisms of auditory awareness underlying verbal transformations. *NeuroImage*, *36*(1), 123–130.

Laurence, S. & Margolis, E. (2001). The poverty of the stimulus argument. *British Society for the Philosophy of Science, 52*, 217-276.

Levitin, D. J., & Menon, V. (2003). Musical structure is processed in "language" areas of the brain: a possible role for Brodmann Area 47 in temporal coherence. *NeuroImage*, *20*(4), 2142–2152.

Lidji, P., Palmer, C., Peretz, I., & Morningstar, M. (2011). Listeners feel the beat: Entrainment to English and French speech rhythms. *Psychonomic Bulletin Review, 18*, 1035-1041.

Lynch, M. P., Eilers, R. E. , Oller, K. D., & Urbano, R. C. (1990). Innateness, experience, and music perception. *Psychological Science, 1*(4), 272-276.

Lynch, M. P., Eilers, R. E. , Oller, K. D., & Urbano, R. C. (1991). Influences of acculturation ad musical sophistication on perception of musical interval patterns. *Journal of Experimental Psychology: Human Perception and Performance, 17*(4), 967-975.

Ma, Y. Y. & McFerrin, B. (1992). *Hush* [CD]. Sony.

Marcus, G. (1993). Negative evidence in language acquisition. *Cognition, 46*(1), 53-85.

Margulis, E. H. (2013). Repetition and emotive communication in music versus speech. *Frontiers in Psychology, 4*, 1-4.

Merker, B. H., Madison, G. S., & Eckerdal, P. (2009). On the role and origin of isochrony in human rhythmic entrainment. *Cortex*, *45*(1), 4–17.

Milligan, K., Atkinson, L., Trehub, S. E., Benoit, D., & Poulton, L. (2003). Maternal attachment and the communication of emotion through song. *Infant Behavior and Development*, *26*(1), 1–13.

Moreno, S., Bialystok, E., Barac, R., Schellenberg, E. G., Cepeda, N. J., & Chau, T. (2011). Short-term music training enhances verbal intelligence and executive function. *Psychological Science*, *22*(11), 1425–1433.

Möttönen, R., Calvert, G. A., Jääskeläinen, I. P., Matthews, P. M., Thesen, T., Tuomainen, J., & Sams, M. (2006). Perceiving identical sounds as speech or non-speech modulates activity in the left posterior superior temporal sulcus. *NeuroImage*, *30*(2), 563–569.

Motz, B. A., Erickson, M. A., & Hetrick, W. P. (2013). To the beat of your own drum: Cortical regularization of non-integer ratio rhythms toward metrical patterns. *Brain and Cognition, 81*, 329-336.

Nahum, M., Nelken, I., Ahissar, M. (2008). Low-level information and high-level perception: The case of speech in noise. *PLoS Biology, 6*(5), 0978-0991.

Newport, E. L. (1990). Maturational constraints on language learning. *Cognitive Science*, *14*(1), 11–28.

Newport, E. L. (2011). The modularity issue in language acquisition: A rapprochement? Comments on Gallistel and Chomsky. *Language Learning and Development*, *7*(4), 279–286.

Patel, A. D. (1998). Syntactic processing in language and music: Different cognitive operations, similar neural resources? *Music Perception: An Interdisciplinary Journal*, *16*(1), 27–42.

Patel, A. D. & Daniele, J. R. (2003). An empirical comparison of rhythm in language and music. *Cognition, 87,* B35-B45.

Patel, A. D. (2003). Language, music, syntax and the brain. *Nature Neuroscience*, *6*(7), 674–681.

Patel, A. D. (2010). *Music, language, and the brain*. Oxford University Press.

Peretz, I., Ayotte, J., Zatorre, R. J., Mehler, J., Ahad, P., Penhune, V. B., & Jutras, B. (2002). Congenital amusia: A disorder of fine-grained pitch discrimination. *Neuron*, *33*(2), 185–191.

Peretz, I. (2006). The nature of music from a biological perspective. *Cognition*, *100*(1), 1–32.

Peretz, I., Brattico, E., Järvenpää, M., & Tervaniemi, M. (2009). The amusic brain: in tune, out of key, and unaware. *Brain*, *132*(5), 1277–1286.

Peretz, I., & Coltheart, M. (2003). Modularity of music processing. *Nature Neuroscience*, *6*(7), 688–691.

Peretz, I., & Hyde, K. L. (2003). What is specific to music processing? Insights from congenital amusia. *Trends in Cognitive Sciences*, *7*(8), 362–367.

Pinker, S. (1997). *How the mind works*. New York, NY: W. W. Norton & Company.

Poeppel, D. (2003). The analysis of speech in different temporal integration windows: cerebral lateralization as "asymmetric sampling in time." *Speech Communication*, *41*(1), 245–255.

Potts, R. (1998). Variability selection in hominid evolution. *Evolutionary Anthropology, 7,* 81-96.

Reich, S. (1966). *Come out*. New York, NY: Nonesuch.

Remez, R., Rubin, P., Pisoni, D., & Carrell, T. (1981). Speech perception without traditional speech cues. *Science*, *212*(4497), 947–949.

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.

Saffran, J. R. (2002). Constraints on Statistical Language Learning. *Journal of Memory and Language*, *47*(1), 172–196.

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, *70*(1), 27–52.

Schachner, A., & Hannon, E. E. (2011). Infant-directed speech drives social preferences in 5-month-old infants. *Developmental psychology*, *47*(1), 19–25.

Scheirer, E., & Slaney, M. (1997). Construction and evaluation of a robust multi-feature speech/music discriminator. *IEEE International Conference on Acoustics, Speech, and Signal Processing, 2,* 1331 –1334.

Schlüter, J., & Sonnleitner, R. (2012). Unsupervised feature learning for speech and music detection in radio broadcasts. In *Proceedings of the 15th International Conference on Digital Audio Effects*.

Schoen, D., Magne, C., Besson, M. (2004). The music of speech: Music training facilitates pitch processing in both music and language. *Psychophysiology, 41*, 341-349.

Snyder, J. S., Pasinski, A. C., McAuley, J. D. (2011). Listening strategy for auditory rhythms modulates neural correlates of expectancy and cognitive processing. *Psychophysiology, 48*, 198-207.

Spearin, C. (2009). *The happiness project*. Toronto, ON: Arts & Crafts.

Sperber (1996). Explaining culture: A naturalistic approach. Blackwell: Cambridge, MA

Tervaniemi, M., Medvedev, S. V., Alho, K., Pakhomov, S. V., Roudas, M. S., van Zuijen, T. L., & Näätänen, R. (2000). Lateralized automatic auditory processing of phonetic versus musical information: A PET study. *Human Brain Mapping*, *10*(2), 74–79.

Thiessen, E. D., Hill, E. A., & Saffran, J. R. (2005). Infant-directed speech facilitates word segmentation. *Infancy*, *7*(1), 53–71.

Thompson, R. G., Moulin, C. J. A., Hayre, S., & Jones, R. W. (2005). Music enhances category fluency in healthy older adults and alzheimer's disease patients. *Experimental Aging Research*, *31*(1), 91–99.

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.

Tian, X. & Huber, D. E. (2010). Testing an associate account of semantic satiation. *Cognitive Psychology, 60*, 267-290.

Tian, X. & Huber, D. E. (2013). Playing "Duck Duck Goose" with neurons: Change detection through connectivity reduction. *Psychological Science, 24*(6), 819-827.

Tierney, A., Dick, F., Deutsch, D., & Sereno, M. (2012). Speech versus song: Multiple pitch-sensitive areas revealed by a naturally occurring musical illusion. *Cerebral Cortex*.

Trainor, L. J., & Trehub, S. E. (1994). Key membership and implied harmony in western tonal music: Developmental perspectives. *Perception & Psychophysics*, *56*(2), 125–132.

Trehub, S. E., Schellenberg, E. G., & Kamenetsky, S. B. (1999). Infants' and adults' perception of scale structure. *Journal of Experimental Psychology-Human Perception and Performance*, *25*(4), 965–975.

Trehub, Sandra E., & Hannon, E. E. (2006). Infant music perception: Domain-general or domain-specific mechanisms? *Cognition*, *100*(1), 73–99.

Vouloumanos, A., Kiehl, K. A., Werker, J., F., Liddle, P. F. (2001). Detection of sounds in the auditory stream: Event-related fMRU evidence for differential activation to speech and nonspeech. *Journal of Cognitive Neuroscience, 13*(7), 994-1005.

Werker, J. F., Yeung, H. H., & Yoshida, K. A. (2012). How do infants become experts at native-speech perception? *Current Directions in Psychological Science*, *21*(4), 221–226.

Whalen, D. H., & Liberman, A. M. (1987). Speech perception takes precedence over nonspeech perception. *Science (New York, N.Y.)*, *237*(4811), 169–171.

Zanto, T. P. & Gazzaley, A. (2009). Neural suppression of irrelevant information underlies optimal working memory performance. *The Journal of Neuroscience, 29*(10), 3059-3066.

Zatorre, R., Evans, A., Meyer, E., & Gjedde, A. (1992). Lateralization of phonetic and pitch discrimination in speech processing. *Science*, *256*(5058), 846–849.

Zatorre, R. J., Belin, P., & Penhune, V. B. (2002). Structure and function of auditory cortex: music and speech. *Trends in Cognitive Sciences*, *6*(1), 37–46.

Zatorre, R. J., & Gandour, J. T. (2008). Neural specializations for speech and pitch: moving beyond the dichotomies. *Philosophical Transactions of the Royal Society: Biological Sciences*, *363*(1493), 1087–1104.

Zatorre, R. J., & Baum, S. R. (2012). Musical melody and speech intonation: Singing a different tune? *PLoS Biology 10*(7).

VITA

Graduate College
University of Nevada, Las Vegas

Christina M. Vanden Bosch der Nederlanden

Degree:

Bachelor of Arts in Psychology; Minor in Music, 2008
Calvin College, Grand Rapids, Michigan

Special Honors and Awards:

Graduate and Professional Students Association (GPSA) Travel Grant, 2012
GPSA Research Forum – Honorable Mention, 2013
Best Graduate Poster at the Center for Research on Brain, Language, and Music (CRBLM) Symposium on Music and Language Development, 2013
Patricia Sastaunik UNLV Graduate College Scholarship, 2011, 2012

Publications:

Hannon, E.E., Vanden Bosch der Nederlanden, C.M., Tichko, P. (2012). Effects of perceptual experience on children's and adults' perception of unfamiliar rhythms. *Annals of the New York Academy of Sciences, 1252*, 92-99.

Ullal-Gupta S., Vanden Bosch der Nederlanden C.M., Tichko P., Lahav A., and Hannon E.E, (2013) Linking prenatal experience to the emerging musical mind. *Frontiers in Systems Neuroscience, 7*(48). doi:10.3389/fnsys.2013.00048

Thesis Title: The role of music-specific representations when processing speech: Using a musical illusion to elucidate domain-specific and –general processes.

Thesis Committee:

Chairperson, Erin E. Hannon, Ph.D.
Committee Member, Joel S. Snyder, Ph.D.
Committee Member, Laurel Pritchard, Ph.D.
Graduate Faculty Representative, Diego Vega, Ph.D.