8-2010

Inferring rules from sound: The role of domain-specific knowledge in speech and music perception

Aaronell Shaila Matta
University of Nevada, Las Vegas

INFERRING RULES FROM SOUND: THE ROLE OF DOMAIN-SPECIFIC

KNOWLEDGE IN SPEECH AND MUSIC PERCEPTION

by

Aaronell Shaila Matta

Bachelor of Science
The Pennsylvania State University
2005

A thesis submitted in partial fulfillment of
the requirements for the

Master of Arts in Psychology
Department of Psychology
College of Liberal Arts

Graduate College
University of Nevada, Las Vegas
August 2010

THE GRADUATE COLLEGE

We recommend the thesis prepared under our supervision by

Aaronell Shaila Matta

entitled

Inferring Rules from Sound: The Role of Domain-Specific Knowledge in Speech and Music Perception

be accepted in partial fulfillment of the requirements for the degree of

Master of Arts in Psychology

Erin E. Hannon, Committee Chair

Joel S. Snyder, Committee Member

Jennifer L. Rennels, Committee Member

Eugenie I. Burkett, Graduate Faculty Representative

Ronald Smith, Ph. D., Vice President for Research and Graduate Studies
and Dean of the Graduate College

August 2010

ABSTRACT

Inferring Rules from Sound:

The Role of Domain-Specific Knowledge in Speech and Music Perception

By

Aaronell S. Matta

Dr. Erin E. Hannon, Examination Committee Chair
Assistant Professor of Psychology
University of Nevada, Las Vegas

Speech and music are two forms of complex auditory structure that play fundamental roles in everyday human experience and require certain basic perceptual and cognitive abilities. Nevertheless, when attempting to infer patterns from sequential auditory input, human listeners may use the same information differently depending on whether a sound is heard in a linguistic vs. musical context. The goal of these studies was to examine the role of domain-specific knowledge in auditory pattern perception. Specifically, the study examined the inference of "rules" in novel sound sequences that contained patterns of spectral structure (speech or instrument timbre) and fundamental frequency (pitch). Across all experiments, participants were first familiarized to a sequence containing pitch or syllable patterns that followed a particular rule (e.g., ABA), and they were subsequently asked to rate the similarity of novel sequences that were consistent or inconsistent with that rule. In two experiments participants were familiarized to either a pitch or syllable rule, and in a third experiment they were familiarized to simultaneous conflicting rules (e.g. pitch following ABA but syllables following ABB). Although participants readily detected inconsistent stimuli after familiarization to a single rule, in the conflicting two-rule condition they gave high

similarity ratings to any stimulus that obeyed the syllable rule, regardless of whether or not the music dimension was consistent or inconsistent with familiarization. Three additional experiments took the same approach but tested rule-learning on the basis of pitch or timbre (instrument) sequences. In these experiments, participants gave the highest similarity ratings when the pitch rule was preserved, regardless of variation in the timbre dimension. These results support the notion that adults "filter" information according to domain-specific knowledge and expectations, presumably due to perceptual learning processes that take place during early development.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

Music is unique, ubiquitous, and profoundly important to human beings. There are no studied human cultures that do not produce music in some form, making music a human universal (Christensen-Dalsgaard, 2004). Music plays a fundamental role in social bonding between parents and infants (Trehub, 2001), providing a mode of communication before language is useful (Christensen-Dalsgaard, 2004). Music serves many other purposes as well. It provides a means of emotional self-regulation, with certain styles of music making an individual feel happier or more energized, or by allowing listeners to empathize with a certain mood or emotion (Schellenberg et al., 2008). There has also been evidence that music can be effective in soothing or relieving certain types of pain in clinical settings (Roy et al., 2008).

Music may also play a role in mate selection and attraction. Famous musicians, regardless of their attractiveness, seem to be able to attract a fervent, loyal fanbase of screaming females who profess their undying love and commitment, as well as beautiful women to date or marry. For example, singer Rod Stewart, whose physical appearance might be considered unremarkable, has been married three times, all to actresses or supermodels, and has fathered seven children between them. Music might therefore function in the same way as a peacock's tail feathers by allowing individuals with musical skills to attract mates. This idea dates back to Charles Darwin (1871), who first proposed that music might be involved in sexual selection.

Music may also function to enhance group bonds (i.e., performing in a marching band, orchestra, or choral group together), or it might allow individuals to socialize at events such as concerts, where they might share a common bond over a particular musical artist or group that is performing. Even hearing-impaired individuals are able to enjoy some forms of music, or in the case such as the famous composer Ludwig von Beethoven, are able to compose it, presumably by attending to the vibrations that can be felt on the skin and through the body (Green, 1988). Given that music can play such diverse and fundamental social and emotional roles, an obvious question is why music evolved and whether or not music capacities benefit human survival.

Many well-respected thinkers, including Darwin, have argued that music specifically evolved through natural selection, as a way to promote survival. By contrast, there are many scholars who argue that music is not an evolutionary adaptation in and of itself, but rather, a spandrel -- or by-product -- of language evolution. Steven Pinker (1997) referred to music as "auditory cheesecake", having no real survival advantage, but rather acting as "an exquisite confection crafted to tickle the sensitive spots of at least six of our mental faculties" with no innate or evolutionary importance independent of language. By this, he proposed that music is essentially a "parasite" on language that feeds upon what we already know. According to Pinker's view, music has no real adaptive function or circuitry independent from language or other innate abilities.

CHAPTER 2

REVIEW OF RELATED LITERATURE

Modularity of Language and Music

The concept of modularity has been applied to many psychological domains, including language, music, and face recognition (Conway & Christiansen, 2005). Modularity refers to the notion that the mind is divided into innate structures and processes that act independently of one another, each specifically developed or specialized to serve a certain functional purpose (Elsabbagh & Karmiloff-Smith, 2004). According to Fodor (1983), a module is domain specific, which means it is associated with distinct neurological structures and specialized to receive only certain types of input, and it is informationally encapsulated, meaning that it does not need to refer to other psychological systems to function properly (Fodor, 1983).

Cosmides & Tooby (1994) suggested that domain-specific mechanisms are a product of evolution dating back to the Pleistocene (ending approximately 10,000 years ago), and were likely favored by natural selection. They argued that without the specialization associated with domain-specific processes, humans would essentially have to undergo a process of trial and error until the necessary solution or mechanism was determined, which, from the perspective of natural selection, could be the difference between life and death. In other words, because domain-general adaptations are not designed to maintain specific behaviors or processes, it is not likely that such mechanisms alone could work efficiently enough to promote success or survival.

A great deal of controversy surrounds the question of whether or not modularity exists in the human mind and which systems, in particular, should be considered modular.

Bishop (1997) suggests that Fodor's notion of modularity is better suited for the way adult minds function, but such standards might be unrealistic to apply to a developing child. It may be normal for systems to interact during development, but in time, each system may assume relative autonomy (Hulme & Snowling, 1992).

To support the modularity of mind view, researchers examine the extent to which brain areas can be characterized as modular or specialized for specific functions. It is widely assumed that different parts of the brain are associated with different functions, but this does not necessarily support the idea of modularity. In the case of modularity, specific dedicated neural structures presumably developed for the sole purpose of serving a particular domain. Many scholars embrace the notion of a language module, arguing that humans are born with an innate capacity for language that develops in every human being, including individuals who may be otherwise impaired (Chomsky, 1965; Fodor, 1983; Pinker, 1994). Chomsky (1965, 1988) was one of the earliest proponents of the idea that language is both modular and innate. He suggested that all humans are pre-wired with what he called a "universal grammar," or innate knowledge of a limited set of rules for organizing all languages. By extension, nonhumans are assumed to lack such a language module.

Attempts at teaching language to nonhuman animals have generally been unsuccessful. For example, when researchers attempt to teach American Sign Language to apes, the apes do not actually learn real ASL signs, but rather a number of pantomimes and gestures instead of a full language (Terrance et al., 1979). This suggests that nonhuman animals are not pre-programmed to use language, making language learning distinctly unique to humans. The abilities of nonhuman animals stand in stark contrast to the abilities of human children not only to acquire language but to create language, as in the case of pidgins and creoles (Bickerton, 1981). Pidgins are defined as simplified languages that develop in order for two or more groups to communicate when they do not have a language in common. Bickerton found that when human children were exposed to ungrammatical "pidgin English" spoken by their parents, they had a tendency to incorporate increasingly complex grammatical structures where there were previously none present, creating what is referred to as a "creole language." This process suggests that human children have innate grammatical capacities that they impose while learning language.

In defiance of Pinker's characterizations that music is a parasitic by-product of language, many have argued for the existence of a music module (Peretz & Coltheart, 2003). Evidence supporting this view comes from music-specific deficits observed in individuals with certain types of brain injuries. In these cases, patients are unable to recognize melodies that were once familiar to them, but they readily recognize spoken lyrics and words, familiar voices, and other common sounds such as animal sounds and traffic noises. This condition is referred to as 'acquired amusia'. There are also individuals who suffer from 'congenital amusia' deficits in recognizing melodies exhibited from birth, even though they recognize spoken lyrics (Ayotte et al., 2002; Peretz & Hyde, 2003). Congenital amusics exhibit severe deficits in processing of pitch variations, but they readily process and recognize speech prosody, environmental sounds, and human voices (Ayotte et al., 2002). Because double dissociations have been used to support the notion that language is modular, double dissociations in the music domain may also support the notion that music is modular.

Others argue that neither music nor language are truly modular, but instead share many important and overlapping processes and structures in behavior and the brain (Trehub & Hannon, 2006). Williams Syndrome (WS) is a genetic condition that results in individuals who have deficits in multiple areas including psychomotor coordination, visual-spatial-organization, and adaptation to novelty (Brown et al., 2003; Karmiloff-Smith et al., 2003; Mervis et al., 1999) as well as difficulty with complex areas of language, including comprehension and pragmatics (Anderson & Rourke, 1995). Often these deficits are paired with a lack of social judgment skills (Donnai & Karmiloff-Smith, 2000) or difficulty with face processing (Karmiloff-Smith et al., 1995). However, individuals with WS are often highly verbal and able to perform well on simple verbal tasks. They also demonstrate good singing skills, can easily learn songs, and show an abnormally high rate of absolute (perfect) pitch as compared to normal individuals (Don et al., 1999). Although children with Williams Syndrome tend to perform better on verbal than on nonverbal tasks, they also excel at tasks within a musical domain, which undermines the claim for a distinct and separate language module. Instead, it would appear that there are shared aspects of auditory processing that allow individuals with Williams Syndrome to show relative strengths in both the language and the music domains.

Jentschke et al. (2008) provided further evidence for a lack of distinct language and music modules in the brain through their work with Specific Language Impairment (SLI). These children do not acquire language as rapidly or effortlessly as other children and often have syntactic difficulties, though their lexical and pragmatic skills are relatively intact, which is the opposite of children with Williams Syndrome (Jentschke et al., 2008; Karmiloff-Smith et al., 2003). These children also show comparable deficits in their processing of musical syntax, as evidenced through ERPs investigating Early Right Anterior Negativity (ERAN) and late negativity (N5). ERAN typically reflects early and fairly automatic processing of syntax, while N5 indicates processing of harmonic integration; neither was elicited in SLI children, but both were evoked in children who show typical language development (Jentschke et al., 2008). These findings provide even further evidence to suggest that many components of music and language are processed by shared neural systems, not separate modules.

Just as with language, musical double dissociations can be challenged. Closer examination of the language processing abilities of amusics does, in fact, reveal abnormalities in their sensitivity to vocal intonation (Patel et al., 2008). When presented with sentences uttered by a female native speaker of American English or continental French, 30% of amusic individuals had difficulty distinguishing sentence pairs on the basis of vocal intonation, such as statements and questions that have distinct rising and falling pitch patterns. Similarly, Foxton et al. (2004) found that British amusics had difficulty judging the direction of short pure-tone pitch glides that precisely mimicked speech intonation. These results suggest that amusia may cause impairments not only in music processing, but also in processing language prosody. In summary, such evidence challenges the notion that there are exclusively language-specific and music-specific disorders and thereby contradicts modular accounts of both language and music.

Additional challenges to the modular view come from evidence that music training can influence performance in non-musical domains. For example, adults who are musically trained are better than untrained adults in identifying emotions (i.e., happy, sad, fearful, angry) conveyed through speech prosody; these findings extend to 6-year-olds as well (Thompson et al., 2004). Magne et al. (2006) used both behavioral tasks (recognizing incongruent music or language samples) and ERP recordings to demonstrate that musician children are able to detect pitch violations in both music and language better than are non-musician children.

Evidence has also indicated that there is a positive effect for individuals with long-term music exposure in terms of encoding linguistic pitch at the brainstem as compared to non-musicians (Wong et al., 2007). These findings suggest that musicians may actually have enhanced sensitivity to language (in terms of pitch or prosody) due to their musical experience.

Examination of brain responses to music and language among both musical and nonmusical individuals further muddy the modular account of music. For example, Steinbeis & Koelsch (2008) used electroencephalography (EEG) recordings to measure various brain waves during a dual-task procedure in which participants engaged in a task involving chord identification in Western tonal music and a simultaneous semantic task involving sentence recognition. They documented an event-related potential (ERP) that responded both to violations of harmonic structure in music and semantic structure in language.

Other evidence suggests that Broca's area, a part of the brain once thought to be uniquely dedicated to language, is activated during both music and language listening (Maess et al., 2001). Participants were presented with numerous chords that were occasionally harmonically inappropriate. According to magnetoencephalography (MEG) recordings, brain responses to chords that violated harmonic expectancies were localized to Broca's area. This supports the notion that music may not be modular, in that it may not be distinctly separate from speech and language.

Domain-Specificity and Development

One of the best ways to examine the question of innate language and music modules is by studying infants, because any domain-specific biases or processes in young listeners are unlikely to have been learned. If music and speech were indeed separate modules, then we would expect infants to show specialized, domain-specific mechanisms in terms of how they process music and language information. However, if the music and language modules were more domain-general, we would expect that infants would show similar or shared processing mechanisms for both types of stimuli.

Domain-specific preferences have been observed in very young infants. Vouloumanos & Werker (2004) examined 2- to 7-month-old infants' preferences for human speech to determine if speech is a "privileged signal". In other words, they wanted to know if infants are already somewhat attuned to certain dimensions of human speech at birth (Jusczyk et al., 1990). Infants' listening preferences were measured during presentation of a speech stimulus (consisting of nonsense words and syllables) and a non-speech stimulus (consisting of sinusoidal waves that tracked the main regions of significant energy in natural speech). Infants as young as 2.5 months demonstrated a preference for the speech stimuli, suggesting that a bias for speech is present very early in development.

A bias for speech has even been observed as early as 1-4 hours after birth (Vouloumanos & Werker, 2007). Newborns were presented with isolated syllables of speech (nonsense words) and with non-speech stimuli consisting of sine-wave analogues.

They monitored the amplitude of the infant sucking on a pacifier while presenting speech and non-speech stimuli in alternation. The infants significantly increased their sucking on a pacifier to listen to speech but not non-speech analogues, suggesting that infants did show a bias for speech. They concluded that this bias arises from language-specific processes that facilitate infants' learning of the properties of the language in the surrounding environment.

While provocative, Vouloumanos & Werker's (2007a) study was met with some criticism. Prenatal experience has an influence on infants' speech perception, suggesting that there may be other factors contributing to any potential speech biases (Curtin & Werker, 2007). For example, at birth, infants prefer their mothers' voices to other female voices (DeCasper & Fifer, 1980) and they will also change their sucking patterns to hear stories that were recited by their mothers during the final six weeks of pregnancy versus novel passages (DeCasper & Spence, 1986). These findings suggest the womb is permeable to sounds that the fetus is able to perceive, encode, and remember (Vouloumanos et al., 2010). As a result, this prenatal exposure may influence neonates' preference for speech over other non-speech sounds that may not be as familiar at birth. Though Vouloumanos & Werker (2007a) argued that the speech preference for infants is innate, it may be the case that prenatal exposure to speech sounds instead plays a role.

Further objection from Rosen and Iverson (2007) pointed out important acoustic differences between the speech and non-speech stimuli, including differences in fundamental frequency (voice pitch), spectral shape, and amplitude (loudness). These acoustic features were probably salient to the developing infants, and could have driven preferences independent of a domain-specific bias. They also contended that the sine-wave analogues were insufficient as non-speech stimuli because while the speech had a strikingly salient voice pitch with exaggerated, child-directed melodic contours, the non-speech sine-wave analogues sounded very similar to one another, and lacked a sense of any type of melodic contour. Thus, infants' preference might have been driven by exaggerated voice melody in the speech condition (Rosen & Iverson, 2007).

Such differences in the acoustic properties of vocalizations are known to be important in driving infant preferences, as shown by research on "motherese" (now referred to as "infant-directed speech"), or speech typically directed toward young children and infants (Grieser & Kuhl, 1988). This speech has a unique acoustic quality, typically higher in pitch, slower in tempo, and with expanded intonation contours (Ferguson, 1964). These characteristics actually make the speech almost song-like in quality, leading some to refer to this kind of vocal input as "musical speech" (Fernald, 1989; Trehub & Trainor, 1998). Infants show strong preferences for the music-like qualities of infant-directed speech to adult-directed speech.

The literature on infant-directed vocalizations makes an interesting prediction: if voice melody drives infants' preferences for vocal stimuli, then musical vocalizations should be preferred over non-musical, infant-direct vocalizations. Indeed, 6-month-old infants showed a preference for maternal singing over maternal speech. Infants heard audio recordings of their own mother either speaking or singing and their behavior (specifically body movement and gaze) was video recorded. Cumulative looking times were greater and body movements were minimal for maternal singing vs. maternal speech. These findings suggest that infants prefer the exaggerated melody in music, particularly when sung by mothers to their infants. Thus, there might be a bias not for any particular domain, but rather for exaggerated voice melody (Nakata & Trehub, 2004).

Behavioral evidence thus provides only ambiguous support for the notion that speech is special for infants, which leads to the question of whether or not infants' brains respond to music and language in a domain-specific fashion. Because it is known that language functions are lateralized to the left cerebral hemisphere, one obvious question is whether infants' brains are lateralized. Dehaene-Lambertz (2000) examined brain activity in 4-month-old infants in response to an unexpected alteration to a speech or non-speech auditory stimulus. To study this, high-density ERPs were recorded from infants' scalps while they listened to speech stimuli with occasional oddball trials, which were expected to elicit an early brain response, or mismatch negativity. Mismatch responses to a voice change versus a phoneme change were compared. Four-month-old infants did not show the same activation patterns in the left hemisphere of the brain that adults did when presented with the speech stimuli, supporting the notion that the brain does become more specialized as it matures.

Similarly, Imada et al. (2006) examined brain responses to tones, harmonics, and syllables in the left hemispheres of neonates, 6- and 12-month-old infants using magnetoencephalography. Consistent with past research findings, even newborns showed activation in the superior temporal cortex, where the brain conducts auditory analysis. However, activation of the inferior frontal region, which involves speech motor analysis and is unique to speech stimuli, only emerged around 6 months of age, and increased with age. These results support the notion that speech perception becomes increasingly specific between the ages of 6 and 12 months and also acknowledge that there is an emerging left-hemisphere link that is not present at birth, but that may develop after 6 months of age. Thus, the left hemisphere advantage in auditory processing which adults typically exhibit may not be present, or may only be very weak in young infants, suggesting that specialized modules in the brain may need to develop.

Later, Dehaene-Lambertz, Dehaene, & Hertz-Pannier (2002) conducted work on three-month-old infants, using fMRI to measure brain activity evoked by normal and backwards speech. Results indicated that left-lateralized regions of the brain similar to those of adults were already active, suggesting that precursors of adult cortical language become active in infants before they begin speech production. This implies that the left auditory cortex presents a higher responsiveness to auditory stimuli in general, and is not necessarily specific to linguistic input, due to the response also evoked by backwards speech.

Further review of infants' neural processing by Dehaene-Lambertz & Gliga (2004) found that, in terms of phoneme processing, behaviors and ERP readings are similar between initial (infancy) and mature (adulthood) stages, suggesting continuity in processing and neural structure. In this case, it may be that infants do have access to phonemic representations at the beginning of life, but it isn't until around 5-6 months of age that the influences of their native languages become more pronounced and apparent. Research by Kotilahti et al. (2009) further confirms this idea, finding that newborns show typical speech lateralization and language-related brain activity to the left hemisphere, but it is not yet fully developed. They did not, however, find significant right hemisphere activation in response to music, suggesting that newborns may not have as strong or developed music-related brain activity as they do for speech and language.

However, contrary to the notion that the brain becomes more domain-specific over time, Schön et al. (2010) conducted an fMRI experiment in which adult participants listened to pairs of spoken words, 3-note melodies sung to the same syllable (*Vocalise*), or words sung to melody and were asked to distinguish whether the pairs were the same or different. Similar networks of brain regions (specifically, the middle and temporal gyri) were more activated while listening to spoken words, *Vocalise*, and sung words, as compared to a control condition containing noise that was neither speech- nor music-related. Furthermore, both brain hemispheres appeared to be involved in both speech and music processing, though the degree of activation varied such that the left hemisphere was more involved in speech processing, and the right hemisphere was more involved in music processing. Overall, because these brain regions do not respond exclusively to language nor music stimuli, this evidence supports a more nuanced view in which both shared and non-shared processes are involved in language and music processing.

Though infants may not necessarily begin life with domain-specific knowledge, they might possess speech-specific learning mechanisms, such as the ability to infer certain types of patterns or "rules" from language input (Marcus et al., 1999; Marcus 2000; Marcus et al., 2007). Numerous studies have highlighted infants' ability to detect patterns in various types of input, whether linguistic, musical, or even visual (Altmann, 1999; Christiansen & Curtin, 1999; Eimas, 1999; Elman, 1999; Kirkham et al., 2002; McClellend & Plaut, 1999; Saffran et al., 1999; Seidenberg & Elman, 1999). These studies report that both adults and infants segment sequences of elements (such as syllables, tones, or shapes) according to their sequential statistics (i.e. how likely one element is to be preceded or followed by other elements), and that they do this similarly for all types of input (Saffran et al., 1999). For example, 8-month-old infants have demonstrated the ability to apply statistical learning to segment words from fluent speech (Saffran et al., 1996), as well as the learning of tone sequences. Such findings are important because they show that simple but powerful domain-general learning mechanisms might be used to acquire domain-specific knowledge that was once assumed to rely on innate modules.

By contrast, the work of Marcus et al. (1999, 2007) suggests that although infants do possess simple statistical learning abilities, they also possess a rule-learning ability (i.e., the capacity to abstract algebraic rules) that applies exclusively to linguistic input. These "algebraic" rules are described as "open-ended abstract relationships for which [the listener] can substitute arbitrary items" (Marcus et al., 1999, pp. 77).

Several studies suggest that infants can abstract algebraic rules over linguistic input. These studies used the preferential-looking paradigm (Jusczyk & Aslin, 1995), in which auditory stimuli are paired with a flashing light and infants' gaze duration (in s) is used as an indicator of listening preference. Infants are first familiarized with a series of synthesized speech samples containing sixteen 3-syllable sentences that followed either an ABA (i.e., "la ti la") or ABB rule (i.e., "la ti ti"). According to Marcus, if infants use the simple statistical learning mechanisms described above, they should only be able to recognize familiarized patterns containing previously heard syllables. If, however, infants can infer the underlying algebraic rules that govern such sequences, they should be able to apply the learned rule to entirely novel syllables. Indeed, in a subsequent test phase consisting of entirely novel syllables, whose sequential arrangement was either consistent or inconsistent with the rule established during familiarization, 7-month-old infants showed a novelty preference for the inconsistent sentences. Novelty preferences were obtained across all comparison conditions (e.g., ABA contrasted with ABB and ABB vs. AAB). Because novelty preferences were obtained based solely on the rule structure and not on the familiarity of specific syllables these findings bolstered the claim that infants can extract algebra-like rules from linguistic input.

This initial finding was met with a great deal of opposition from those who argued that abstract algebra-like rules were tantamount to general pattern-learning based on statistical properties. Seidenberg & Elman (1999) argued that Marcus et al.'s findings resulted from infants' use of statistics, not rules. For example, in the ABB condition, infants could form a "same-different-different" representation of the syllable sequence, even if the specific syllables were distinct between the familiarization and test phases of the study.

While many researchers did not question the basic findings of Marcus at al. (1999), it was widely agreed that additional work was needed to validate the claims that were made. One proposed solution was for Marcus et al. to conduct a control study that familiarized infants with a sentence structure of AAB and then used BAB and AAB rule structures in the test phase (Eimas, 1999). This way, if infants showed a preference for the novel structure, despite the final two syllables remaining unchanged, Marcus et al.'s conclusion for the acquisition of an algebraic rule would be better supported.

Other critics created connectionist networks to demonstrate that built-in algebraic rules are not necessary to succeed at generalizing patterns to novel stimuli (Altmann, 1999). In a simulation of Marcus et al.'s study, simple recurrent networks were trained on Marcus' own stimuli using the ABA grammar or the ABB grammar. After training, each network was given ABA or ABB sentences in random order on input nodes that had not been used during training. Results showed that, like the infants in the Marcus et al. study, the networks successfully discriminated between the consistent and inconsistent test stimuli. Since simple recurrent networks can only use statistical learning (i.e., no sensitivity to rules is built in), Altmann argued that simple statistical learning mechanisms could explain the cognitive processes that infants used in the Marcus study.

In addition to disputing the idea that infants were using algebra-like rules rather than statistical learning, Elman (1999) pointed out that by 7 months of age infants have heard over 6 million words and have already formed representations of speech sounds in their native languages. It is therefore unreasonable to assume that the stimuli used in the Marcus study were completely novel.

If infants' success in the task was simply due to a type of statistical learning, then, Marcus argued, infants should be able to infer rules in both speech and non-speech domains, just as statistical learning has been demonstrated in speech and non-speech domains (Saffran et al., 1999; Fiser & Aslin, 2002; Kirkham et al., 2002). Thus, Marcus et al. (2007) incorporated additional non-speech stimuli to ask whether infants could infer a rule instantiated in speech and then generalize that rule to sequences of non-speech stimuli. In the first experiment, infants were familiarized with tone patterns or sung syllables (i.e., where pitch level and syllable identity varied simultaneously) that followed either an AAB or ABB rule. In the tones condition, infants showed no noticeable preference for a novel sequence that violated the previously established rule. However, in the sung syllables condition, infants showed a strong preference for the novel rule. Thus, infants were unable to abstract rules from tone patterns when they did not have speech information available, but readily did so when rules were instantiated by syllables.

In a second experiment, ABB vs. ABA rules were instantiated using timbre, consisting of synthesized instrument sounds or animal sounds instead of the tones and sung syllables. Thus, in this experiment, there was no speech-related information contained within the stimuli. Infants did not show a noticeable preference for the novel stimuli in either condition, leading the researchers to conclude that infants are unable to abstract rules over non-speech sequences. In a third experiment, infants were familiarized with spoken speech sequences that followed the ABB or ABA rule (as in Marcus et al., 1999) and then tested on their ability to discriminate the same rule, but in non-speech stimuli such as tones, musical instruments, or animal sounds. In this case, infants showed a preference for the stimuli that followed the novel rule, whether they were tones, timbres, or animal sounds. Marcus et al. (2007) argued that speech is a domain-specific catalyst for infants' ability to generalize rules, whereas statistical pattern learning is separate and domain-general.

Contrasting evidence suggests that four-month-old infants are fully capable of learning rules in non-speech domains. For example, 7-month-old infants were shown to detect and generalize patterns when the stimuli consist of pictures of dogs and cats, (Saffran et al., 2007), thus suggesting that rule learning is not even specific to auditory patterns. Within the auditory modality, 4-month-olds infants learned rules from chord- and tone-sequences without any prior familiarization to syllables (Dawson & Gerken, 2009). However, unlike 4-month-olds, 7-month-olds did not learn rules from tone and chord sequences, replicating Marcus et al (2007). The failure of rule-learning leads to important questions about why more experienced listeners would perform worse than less experienced listeners. One possibility is that acquired knowledge of music might interfere with successful rule-learning in the Marcus (1999; 2007) tasks. For example, familiarization sequences never followed a consistent contour pattern (i.e., there were conflicting patterns of "up-down-up" and "down-up-down"). The shape of a melody (e.g. up down), and not the sequence of absolute pitches, is typically the most salient aspect of a musical pattern, at least for adults and infants over the age of 5 months (Trehub et al., 1984). By varying contour within a particular rule structure (i.e., by instantiating ABB by both down-up-up and up-down-down) Marcus et al. (2007) may have made it confusing for experienced listeners, leaving them unlikely to be able to infer music-relevant pattern information, as it was unclear what rule he or she was supposed to follow. In this sense, the musical conditions in these studies were unfairly pitted against real music, since any listener bringing music-specific expectations or knowledge to the learning situation would already be biased against the to-be-learned rule. Thus, the findings of Dawson and Gerken (2009) might arise because older infants pay more attention to the rising and falling contours than to the actual pitches themselves, creating a potential bias in terms of where their attention was to be focused (i.e., pitch patterns vs. varying pitch contour).

Given the large range of potentially meaningful patterns in the world, individuals may attempt to constrain the available information by seeking the most relevant structures in a given context, which may vary according to domain. Thus, in a rule-learning paradigm, the to-be-learned rules or patterns might need to be domain-appropriate. Music-specific rules or biases may be acquired through music lessons, for example (Thompson et al., 2004), or through implicit perceptual learning processes that result from every-day exposure to music in the environment. Once acquired, domain-specific biases would be expected to influence how listeners infer patterns or rules from a given stimulus, depending on whether they perceive it to be music or speech.

Perceptual Learning & Development

Perceptual learning is the process through which perceptual and cognitive networks are modified in a relatively long-lasting manner in response to encounters with predominant structures, norms, beliefs, and values in the environment. Perceptual learning occurs as individuals adapt to their environments and learn how to respond to them appropriately (Goldstone, 1998).

Perceptual learning processes are fundamental to the acquisition of knowledge about both speech and music. One salient example of early perceptual learning comes from the literature on infant speech perception and learning, where perceptual abilities undergo a reorganization that is dependent upon their native language. Werker & Tees (1984) first compared English-speaking adults, Thompson-speaking adults, and 6-month-old infants born to English-speaking parents to determine how well they were able to perceive Thompson speech contrasts. English-speaking adults performed significantly worse than Thompson-speaking adults and English infants, while the latter two groups did not show a significant difference in performance. In further experiments incorporating Hindi and Salish speech contrasts, English infants aged 6-8 months performed slightly better than 8-10 month-old infants and significantly better than infants aged 10-12 months. Hindi and Salish infants' performances were similar to those of the 6-8 month-old English infants. These findings support the theory that younger infants are better able to discriminate phonetic distinctions that occur across many different languages, but this ability decreases with age, typically by the end of the first year of life, as perceptual reorganization takes place.

Similar trends have been reported for infants learning Hindi, Japanese, Spanish, and Mandarin (Kuhl, et al., 2006; Rivera-Gaxiola, et al., 2005; Tsao, et al., in press; Werker & Lalonde, 1988). Moreover, infants' discrimination shows a significant increase for native language contrasts and a significant decline for nonnative perception, indicating that through perceptual learning, the brain's neural circuitry becomes organized in a way such that is specialized to the properties of native-language speech.

Comparable developmental trends have been observed in the music domain. Whether adults are musically trained or not, music is processed via perceptual and cognitive networks that are sensitive to experience (Hannon & Trainor, 2007). In one such example, North American adults, Bulgarian and Macedonian adults, and North American infants were presented with musical stimuli containing either a simple meter (similar to that found predominately in Western music) or a complex meter (often found in Bulgarian and Macedonian folk music) (Hannon & Trehub, 2004). North American adults performed more accurately when rating simple meter patterns. Bulgarian and Macedonian adults performed equally as well in both meter patterns, and North American six-month-olds showed performance very similar to the Bulgarian and Macedonian adults, in that they were able to differentiate between both simple and complex meters quite well. These findings imply that the biases shown by North American adults, in particular, are likely a result of perceptual learning processes, due to their exposure to primarily (or exclusively, perhaps) native music meter.

Perceptual learning can also have an effect on individuals' perceptions of musical interval patterns (Lynch et al., 1991) and melodic changes based on key harmony and implied harmony (Trainor & Trehub, 1994). When American participants were presented with melodies based on interval patterns from both Western and Javanese musical scales, they had trouble detecting mistuning to Javanese but not Western scales, whereas infants' detection was comparable for both Western and Javanese scales (Lynch et al., 1991).

Acquired sensitivity to implied harmony and key also interferes with adult listeners' processing of wrong notes (Trainor & Trehub, 1994). To illustrate, if Western listeners have a strong grasp of key membership and implied harmony in Western music, they should find it easier to detect a wrong note that goes out of key than to detect a wrong note that conforms to key and implied harmony, primarily because such changes are common in Western music. Consistent with this hypothesis, adults and older children were better at detecting out-of-key and out-of-harmony changes than within-key, within-harmony changes, whereas infants detected both types of changes equally well. Thus, unlike adults and older children, younger listeners have not yet received enough musical exposure to ignore violations of melodic structure that, nevertheless, maintain key and harmony.

Adults of both the United States and Turkey were tested on recognition memory using three musical traditions: Chinese orchestral music, Western music, and Turkish music (Demorest et al., 2008). Participants were first played three 30-s excerpts of each music tradition, one at a time, and always in the same order. They subsequently heard additional excerpts and indicated whether or not they'd previously heard each clip of music. All of the participants were significantly better at recognizing novel excerpts from their own culture. Turkish participants were better at remembering Western music than Chinese music, though the Chinese music was equally unfamiliar for both cultural groups. These results suggest that an individual's detailed recognition and interpretation of musical information is influenced by their specific listening experiences.

The knowledge that allows listeners to meaningfully interpret and infer patterns in auditory input is thus acquired at least in great part through perceptual learning processes in response to environmental input, in both the music and speech domain. Typically, this learning starts affecting responses to auditory sequences during infancy, as shown by infants' preferences and processing and memory advantages for structures of their own culture. A perceptual tuning process may act both within and across domains, such that over time, individuals may acquire increasingly specific knowledge about the types of patterns that dominate in music as well as in speech. In principle, such domain-specific knowledge could determine the extent to which adult listeners attend to various types of information in a pattern- or rule-learning context, such as that used by Marcus et al. (1999; 2007).

Current Study

The present experiments examine the types of rules that adults infer when presented with novel auditory patterns (similar to Marcus et al., 2007) that contain information relevant to both speech and music. Based on evidence suggesting that individuals are able to follow rules instantiated through speech (e.g., Marcus et al., 2007) and music (e.g., Dawson & Gerken, 2009), and that such abilities may be influenced by domain-specific knowledge, the study attempted to determine whether or not adults rely on different acoustic features to infer rules in a context-dependent fashion.

Research by Kolinsky et al. (2009) has indicated that vowels are most similar in structure to musical tones in terms of shared processing. In a classification task using bisyllabic nonwords sung on two-tone intervals, musically naïve adults were asked to classify stimuli into two categories during two different tasks. The melodic task involved identifying ascending or descending intervals, and the phonological task required participants to attend only to syllables, identifying which syllable they heard. Irrelevant pitch interval variation disrupted the phonological task, and likewise, irrelevant vowel changes interfered with participants' ability to classify pitch intervals. These findings reveal that vowel and pitch changes appear to be processed in an interactive way. In a second study, Kolinsky et al. (2009) demonstrated that consonants are less integrated with musical intervals. In this experiment, the middles of the nonwords contained varying stop consonants (produced by stopping the airflow in the vocal tract, such as with the letters K or T). Less interval interference was found for the stop consonants than for the vowels. A third experiment showed that nasal consonants (where air may escape freely through the nose, but not the mouth, such as with the letters M or N) interfered less with phonological tasks than did vowels. Overall, these findings support the notion that, with the exception of vowels, most speech sounds do not appear to be vulnerable to interference from pitch (Mehler et al., 2006).

If individuals do, in fact, place more emphasis on syllabic information when they perceive the situation to be within a linguistic context, this raises the question of whether or not the same biases would be observed in a musical context. For example when pitch varies independent of instrument timbre, since timbre, like syllables, is also defined by spectral envelope and temporal characteristics of event onset. Every sound is made up of several different perceptual components, including pitch, loudness, length, timbre, and location. However, in a musical context, pitch typically receives the most focus (Patel, 2008). Moreover, relative and not absolute pitch information predominates in a music context, because listeners are still capable of recognizing melodies in different registers, so long as the pitch relations remain consistent. Pitch contrasts have orderly perceptual distances that can be measured through a system of intervals that allows higher-level relations to emerge (e.g., even though pitches are different, the distance between a set of two notes may be recognizable as the same distance between two different notes) (Patel, 2008). For example, if two individuals were asked to hum the melody to a commonly known song without any further instruction, it is unlikely that they would sing at the same volume or choose the same starting pitch, and even less likely that they would have matching vocal timbres. However, as long as the pitches were sung in the correct order and the appropriate intervals were maintained, listeners would still be able to identify the song.

Timbre contrasts in music, unlike pitch, are somewhat difficult to organize systematically (Patel, 2008). While many instruments are capable of producing a variety of timbral contrasts (e.g., a violin may be bowed, plucked, or struck to produce a sound), these contrasts rarely form the basis for structural organization the way that pitch forms the basis for scales, tonality, and key. Though there have been attempts to organize timbral intervals, the reoccurring issue is that timbre relations are often not perceived with enough uniformity to provide a solid basis for developing a system such as the system of intervals that exists for pitches (Krumhansl, 1989).

Auditory streaming refers to when two or more repeating sounds that differ by at least one acoustic element are perceived as two (or more) separate streams (Bregman & Pinker, 1978; Snyder & Alain, 2007). In the case of timbre, Gregory (1994) found that, in the absence of timbre variation, ascending and descending scales that crossed in the middle were perceived (streamed) as separate high and low half-scales. However, increases in timbre difference led to more complete perceptions of the scales, played by each instrument. Thus, these findings suggest that continuity is an important aspect of timbre perception in music.

While timbre is what allows us to differentiate between musical instruments, it seems to be more relevant to speech than music (Patel, 2008). For example, to the extent that instrument timbres and syllables can be considered similar, speech contains successions of timbres that are meaningful to listeners, but in a musical context, listeners tend to place more emphasis on the melody than on the changing instruments.

Experiment 1 examines how listeners infer rules in a speech context. Listeners are familiarized to sequences of sung syllables containing rule information through spoken syllables (spectral structure) and through pitch contour (fundamental frequency). During familiarization, listeners hear a sequence of triplets containing syllables, pitches, or both syllables and pitch played simultaneously that follow a specific rule or set of rules. During a subsequent test phase, listeners hear stimuli containing novel syllables and pitches that are consistent or inconsistent with the rule instantiated during familiarization, and they are asked to rate how similar each stimulus is to the initial rule. Accuracy is defined as the tendency to rate inconsistent items as less similar than consistent items. To make sure that the single-dimension rules are not too difficult for listeners to infer, two baseline conditions will be run. These conditions will focus on rules instantiated by spectral structure (syllables) while pitch is held constant, or fundamental frequency (pitch contour) rules when spectral structure is held constant. It is expected that listeners will extract rules from monotone sequences of syllables, in replication of Marcus et al. (1999, 2007). It is also expected that listeners will extract pitch contour-based rules in monosyllabic sequences. A third condition will pit spectral and fundamental frequency against each other (i.e., opposite rules are established by syllables vs. pitch contour). Based on the findings of Marcus et al. (2007) and the notion that rule-learning is language-specific, one prediction is that listeners will focus entirely on rules that are instantiated through speech, but will ignore rules instantiated in pitch. However, such an outcome would also be consistent with the notion that adult listeners have acquired domain-specific biases, and that due to the presence of syllabic information, listeners will selectively attend to the feature that they have learned is most relevant in the speech domain. It is thus essential to determine whether or not biases for spectral information would also be observed in the context of music.

Experiment 2 thus examines rule learning in a musical context. Specifically, variation in spectral structure is achieved through instrumental timbre rather than through syllable variation. Fundamental frequency will again be represented through pitch contours, as in Experiment 1. Two baseline conditions will again determine whether or not single-dimension rules are learnable. In these conditions, it is expected that listeners will be able to extract the rules without difficulty in both domains. However, in the third condition, timbre and pitch will be played simultaneously while following conflicting rules. If Marcus et al. (2007) is correct in suggesting that rule learning is language-specific, it suggests that adults will not be capable of learning any of the rules in Experiment 2, due to the lack of language-specific information. On the contrary, if listeners have a general bias to infer patterns on the basis of spectral structure (in this case, timbre) rather than fundamental frequency (pitch contour), they should learn the rules instantiated through timbre, but ignore those established through pitch contour. A third possibility is that listeners may alter their dimension of focus depending on the context (e.g., speech vs. music). Since sequential musical patterns are largely defined by pitch changes over time, and timbre has the tendency to be less meaningful than pitch in music (Patel, 2008), in this case, it would be predicted that adult listeners, who presumably possess domain-specific musical knowledge and expectations, will be biased to infer rules from sequential pitch information rather than from sequential timbre information.

CHAPTER 3

METHODOLOGY

EXPERIMENT 1: RULE LEARNING IN A SPEECH CONTEXT

Participants

Participants for this study were 40 students recruited from the subject pool at the University of Nevada, Las Vegas. Participation was voluntary and each individual received one research credit per hour in partial fulfillment of the requirement for Psychology courses. The sample consisted of 24 females and 16 males, aged 18 to 40 years (M = 23.45). Participants ranged in formal music training from 0 to 14 years (M = 2.9). All participants reported having normal hearing.

Apparatus

Participants were tested at individual computer stations over headphones. The computer program PsyScope X, Build 51 (Cohen, MacWhinney, Flatt & Provost, 1993) presented stimuli and recorded button-press responses from the computer keyboard.

Stimuli

In the Speech Only (spectral structure) condition, modeled after prior infant work by Marcus et al. (1999; 2007), participants were familiarized with 13 unique computer-synthesized phoneme triplets following either an ABA (i.e., “ga, ti, ga”) or an ABB rule (i.e., “ga, ti, ti”) and spoken in a monotone set at 261.6 Hz (Middle C). The familiarization triplets were comprised of a combination of 9 different possible syllables. Individual syllables were synthesized using the software program MBROLA (Dutoit et al., 1996). The program Sound Studio 3 for Macintosh was used to combine and order syllables into long AIFF files. Each syllable within a triplet was approximately 300 ms in duration, with a 250 ms gap between syllables and a 1 s gap between triplets. Each triplet was approximately 2.4 s in length. The familiarization stimulus consisted of two iterations of each unique triplet, ordered randomly, resulting in a sound file made up of 26 triplets and lasting approximately 1 min in duration. During the test phase, participants heard eight stimuli (comprised of 4 novel syllables) containing triplets of syllables that either followed or violated the rules established during familiarization (*Table 1*). For example, if a participant heard a familiarization stimulus following the ABA rule, that participant then heard novel triplets of both ABA ("wo, fe, wo") and ABB ("wo, fe, fe") structure during the test phase. Each test trial contained 2 repetitions of each of 4 novel triplets. Test trial order was randomized by the computer.

In the Pitch Only (fundamental frequency) condition, rules were instantiated in monosyllabic tone patterns, having no phoneme variation. Sequences in the Pitch Only condition therefore contained the syllable "da" presented at varying fundamental frequencies that were no lower than Middle C (261.6 Hz), followed an ABA or ABB rule, and a musical contour of low-high-low or low-high-high, using the notes C, C#, D, F#, G#, A, and A# for the familiarization phase and E, F, D# and G for the test phase. All pitches were within a perfect fifth of Middle C (*Table 1*). Maintaining musical contour was essential, as contour provides vital information for identifying whether auditory information constitutes music or some other type of sound (Trehub et al., 1984). As previously described, Dawson & Gerken found that inconsistent musical contour might cause listeners to divert their attention from the actual pitches themselves and place more emphasis on the rising and falling patterns within the stimuli. Consistent musical contour allowed the listeners to focus on the pitches, and made it more evident that the context was intended to be musical.

Just as in the Syllable Only condition, the individual syllables were synthesized using the software program MBROLA (Dutoit et al., 1996), and the program Sound Studio 3 for Macintosh was again used to combine and order syllables into long AIFF files. Each syllable within a group was approximately 300 ms in duration, with a 250 ms gap between syllables and a 1 s pause between triplets. Each triplet was approximately 1.4 s in length. Again, like the Speech Only condition, the familiarization phase contained 26 groups of stimuli and lasted approximately 1 min. Each test trial contained 2 repetitions of each of 4 novel groups. Test trial order was random.

In the Simultaneous Speech/Pitch (SSP) condition, the phonemes *and* pitches in familiarization stimuli varied and followed conflicting rules. Thus, familiarization stimuli followed either a Speech ABA/Pitch ABB rule (i.e., syllables follow an ABA rule while pitches follow an ABB rule) or a Speech ABB/Pitch ABA rule (*Table 1*). In the test phase, participants heard four types of novel stimuli: 1) Pitch Inconsistent/Speech Consistent where the pitch sequence violated the rule established during familiarization but the syllable sequence followed the original rule, 2) Pitch Consistent/Speech Inconsistent, where syllables but not pitches violated the original rule, 3) Fully Consistent, where both the pitches and syllables followed the rules established during familiarization sequence, or 4) Fully Inconsistent, where both pitches and syllables violated the rules. The phonemes and pitches used in the familiarization and test phases of Speech Only and Pitch Only conditions were also used in the SSP condition. Just as in the control conditions, the familiarization phase contained 26 groups of stimuli and each test trial contained 2 repetitions, but this time, with 8 novel groups. Test trials were random.

Procedure

All conditions of Experiment 1 contained a familiarization phase followed by a test phase. During familiarization, participants heard a sequence of sounds that all followed a general pattern or rule (i.e., units within a group followed an ABA or ABB order). During the test phase, participants then heard novel sound sequences that either obeyed or violated the rules established during familiarization. The participants' task was to decide, for each test sequence, the extent to which that sequence followed the rule established during familiarization. Each participant took part in each of three stimulus conditions: Speech Only, Pitch Only, and Simultaneous Speech/Pitch (SSP).

Initial instructions were given to each participant in verbal and written form (over a computer monitor). Each participant completed three experimental blocks, corresponding to the three conditions described above, with order of block (condition) and specific rule type (i.e., ABA or ABB) counterbalanced. Each block began with the familiarization phase, during which participants were instructed to listen to the sample of speech, pitch, or speech and pitch. After the familiarization phase, participants completed the test phase, during which were asked to rate how closely each test stimulus matched the rule established during the familiarization phase on a scale from 1 ("Did not follow rules at all") to 9 ("Followed rules perfectly"). In all conditions and blocks, test stimulus types were ordered randomly. After the participant finished all three blocks, he or she was asked to complete a demographic/music information questionnaire, obtaining information about any hearing problems, language background, and formal music training.

Rule: ABB	Familiarization Phase	Test Phase
Speech Condition (pitch level held constant at Middle C)	ga ti ti, ga na na, ga la la, li na na, li ti ti, li la la, ni ti ti, ni la la, ta la la, ni na na, ta na na, ta ti ti, ga gi gi	de ko ko, de ko de wo fe fe wo fe wo
Pitch Condition (syllable held constant at "Da")	C#DD, CC#C#, AA#A#, G#AA, CDD, F#G#G#, G#A#A#, F#AA, DF#F#, F#A#A#, C#F#F#, DAA, C#G#G#	D#GG, D#GD# EFF EFE
Simultaneous Speech/Pitch Condition (SSP)	ga(C#)ti(D)ti,(C#) ga(C)na(C#)na(C) ga(A)la(A#)la(A) ni(G#)ti(A)ti(G#) li(C)ti(D)ti(C) li(F#)la(G#)la(F#) ni(G#)ti(A#)ti(G#) ni(F#)la(A)la(F#) ta(D)la(F#)la(D) ni(F#)na(A#)na(F#) ta(C#)na(F#)na(C#) ta(D)ti(A)ti(D) ga(C#)gi(G#)gi(C#)	de(D#)ko(G)ko(G) de(D#)ko(G)ko(D#) wo(E)fe(F)fe(F) wo(E)fe(F)fe(E) de(D#)ko (G)de(D#) de(D#)ko(G)de(G) wo(E)fe(F)wo(E) wo(E)fe(F)wo(F)

Table 1: Stimuli Distribution for ABB rule: Experiment 1. Note. Half of the subjects received the opposite rule (ABA) during the familiarization phase (Speech ABA/Pitch ABB in the SSP Condition). Here, the SSP Condition follows the rule sequence of Speech ABB/Pitch ABA.

CHAPTER 4

RESULTS: EXPERIMENT 1

Control Condition: Syllable-Only

Similarity ratings were submitted to a 2 x 2 (Test Item [inconsistent, consistent] x Familiarization Grammar [ABA, ABB]) mixed design Analysis of Variance (ANOVA), with Test Item as a within-subjects factor and Familiarization Grammar as a between-subjects factor. This ANOVA yielded a main effect of Test Item, $F(1,38) = 100.177$, $p < .001$. *Figure 1* presents overall similarity ratings for consistent and inconsistent test items (averaged across the two grammars), and illustrates that participants gave lower similarity ratings to inconsistent test items ($M = 2.11$, $SD = 1.66$) than to consistent test items ($M = 7.03$, $SD = 2.26$). There were no main effects of Familiarization Grammar, $F(1,38) = .985$, $p < .327$, and no interaction between Test Item and Familiarization Grammar, $F(1,38) = 1.199$, $p < .280$, which indicates that rules were readily learned regardless of whether they were ABA or ABB. This experiment thus replicates prior work (Marcus et al., 1999), demonstrating that adults readily infer rules on the basis of syllable patterns when pitch is held constant, as reflected by their accurate differential ratings of consistent and inconsistent test items.

Control Condition: Pitch Contour-Only

Similarity ratings in the Pitch Contour-Only condition were submitted to a 2 x 2 (Test Item [inconsistent, consistent] x Familiarization Grammar [ABA, ABB]) mixed design ANOVA, with Test Item as a within-subjects factor and Familiarization Grammar as a between-subjects factor. This ANOVA yielded a main effect of Test Item, $F(1,38) = 105.130$, $p < .001$. As shown in *Figure 1*, participants gave lower similarity ratings to inconsistent test items (*M* = 2.86, *SD* = 2.19) than to consistent test items (*M* = 7.99, *SD* = 1.79). There were no main effects of Familiarization Grammar, $F(1,38) = 2.120$, $p < .154$, and no interaction between Test Item and Familiarization Grammar, $F(1,38) = .398$, $p < .532$. Thus, adults readily inferred rules on the basis of pitch contour, just as they did rules on the basis of syllable patterning.

Given claims made in prior work (Marcus et al., 2007) that speech is somehow privileged for rule-learning, it was important to ensure that no baseline differences existed in the difficulty of learning syllable and pitch contour rules. Ratings from both the Syllable-Only and Pitch Contour-Only conditions were submitted to a 2 x 2 (Test Item [inconsistent, consistent] x Domain [syllable, pitch contour]) repeated-measures ANOVA, yielding a main effect of Test Item, $F(1,39) = 257.975$, $p < .0001$, and a main effect of domain, $F(1,39) = 11.144$, $p < .002$, but no interaction between test item and domain, $F(1,39) = .074$, $p < .788$. Thus, participants' similarity judgments were based on test item, and not whether the stimuli were presented as syllables or pitch. The main effect of domain suggests that participants may have attributed greater importance to individual syllables than to specific pitches. As such, they rated the inconsistent syllable triplets lower than pitch triplets when stimuli were both inconsistent (M_s=2.11 vs M_p=2.86) and consistent (M_s=7.11 vs. M_p=7.99). However, it does not appear that these potentially perceived differences made the Syllable-Only condition more difficult than the Pitch Contour-Only condition. Instead, it may simply imply that most participants had more experience with identifying speech components than exact pitches.

Overall, these results clearly indicate that participants were able to accurately differentiate inconsistent from consistent test items, inferring rules regardless of the specific rule presented during familiarization (ABA vs. ABB) and whether the rule was instantiated in syllable or pitch contour patterns.

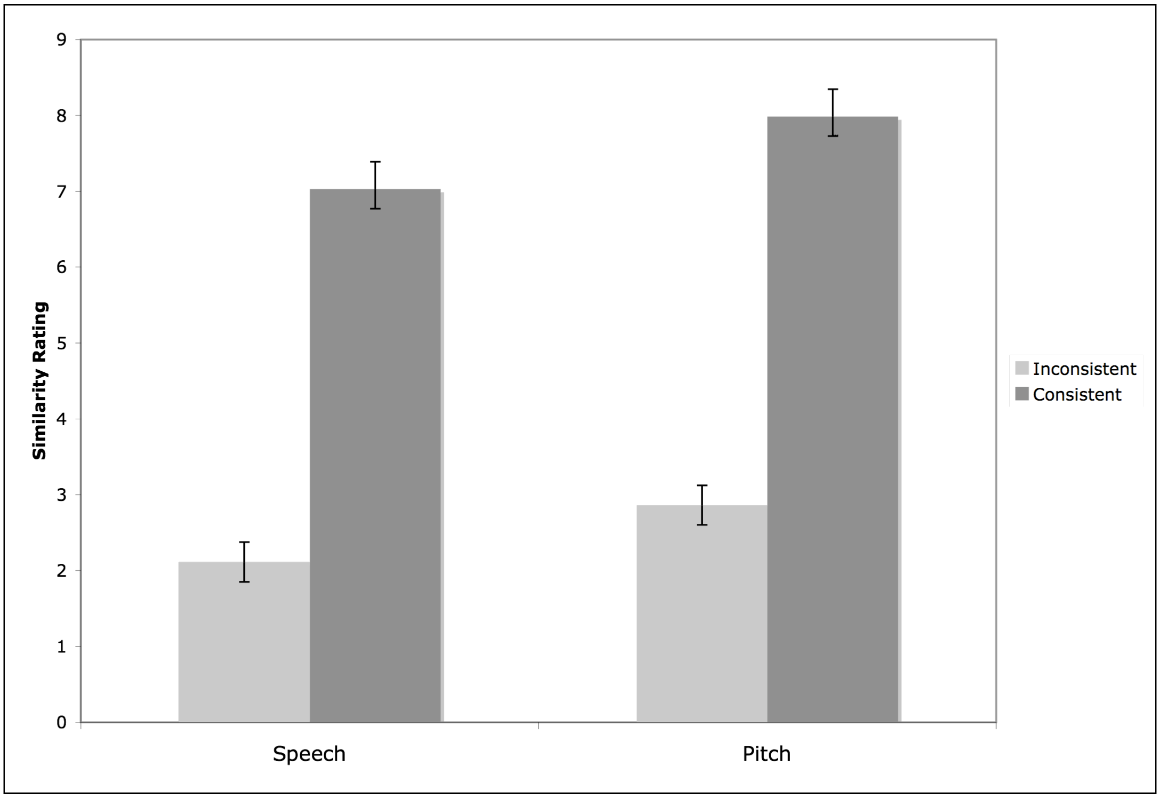

Figure 1: Similarity ratings for consistent and inconsistent test items in the Speech Only and Pitch Contour Only conditions of Experiment 1.

Simultaneous Syllable/Pitch Contour Condition (SSP)

In the SSP condition, subjects were presented with four test items instead of two, with each test item either conforming or not conforming to the syllable rule, the pitch contour rule, or both. For the statistical analysis, two factors were created to indicate whether or not the test item was consistent or inconsistent with the syllable rule (Test

Item Syllable), and to the pitch contour rule (Test Item Pitch Contour). Thus, similarity ratings in the SSP condition were submitted to a 2 x 2 x 2 (Test Item Syllable [consistent, inconsistent] x Test Item Pitch Contour [consistent, inconsistent] x Familiarization Grammar [Syllable ABA/Pitch Contour ABB, Syllable ABB/Pitch Contour ABA]) mixed design ANOVA, with Test Item Syllable and Test Item Pitch Contour as within-subjects factors and Familiarization Grammar as a between-subjects factor. This analysis yielded a main effect of Test Item Syllable, $F(1,38) = 86.28$, $p < .001$, a main effect of Test Item Pitch Contour, $F(1,38) = 29.54$, $p < .001$, and an interaction between Test Item Syllable and Test Item Pitch Contour, $F(1,38) = 61.28$, $p < .001$. The interaction occurred because participants' discrimination of inconsistent and consistent test items differed by domain, such that violations in Syllable and Pitch Test Items were weighed differently.

Figure 2 presents mean similarity ratings for each of the four test items. Using Bonferroni correction, planned post-hoc pairwise *t*-tests revealed that participants gave significantly higher similarity ratings to fully consistent items, $M = 7.37$, $SD = 1.85$, than to fully inconsistent items, $M = 2.68$, $SD = 2.16$, $t(39) = 9.82$, $p < .01$. Similarity ratings were also significantly higher for partially inconsistent items that violated the pitch contour rule, $M = 6.07$, $SD = 2.70$), than for partially inconsistent items that violated the syllable rule, $M = 3.08$, $SD = 2.67$), $t(39) = -3.87$, $p < .01$, suggesting that a change of syllable structure had a greater impact on perceived similarity than did a change of pitch contour. Fully consistent items were rated significantly higher than partially inconsistent items, whether partially inconsistent items violated the syllable rule, $t(39) = 7.48$, $p < .001$ or pitch contour rule, $t(39) = 3.22$, $p < .01$. However, fully inconsistent items were only rated significantly less similar than partially inconsistent items when the difference between the two test items involved a violation of the syllable rule, $t(39) = 5.24$, $p < .01$, and not when the difference involved the violation of a pitch contour rule, $t(39) = 1.07$, $p = .293$. This implies that test items were perceived as dissimilar whenever they violated the syllable rule, but that the additional violation of the pitch contour rule did not further distinguish fully inconsistent items.

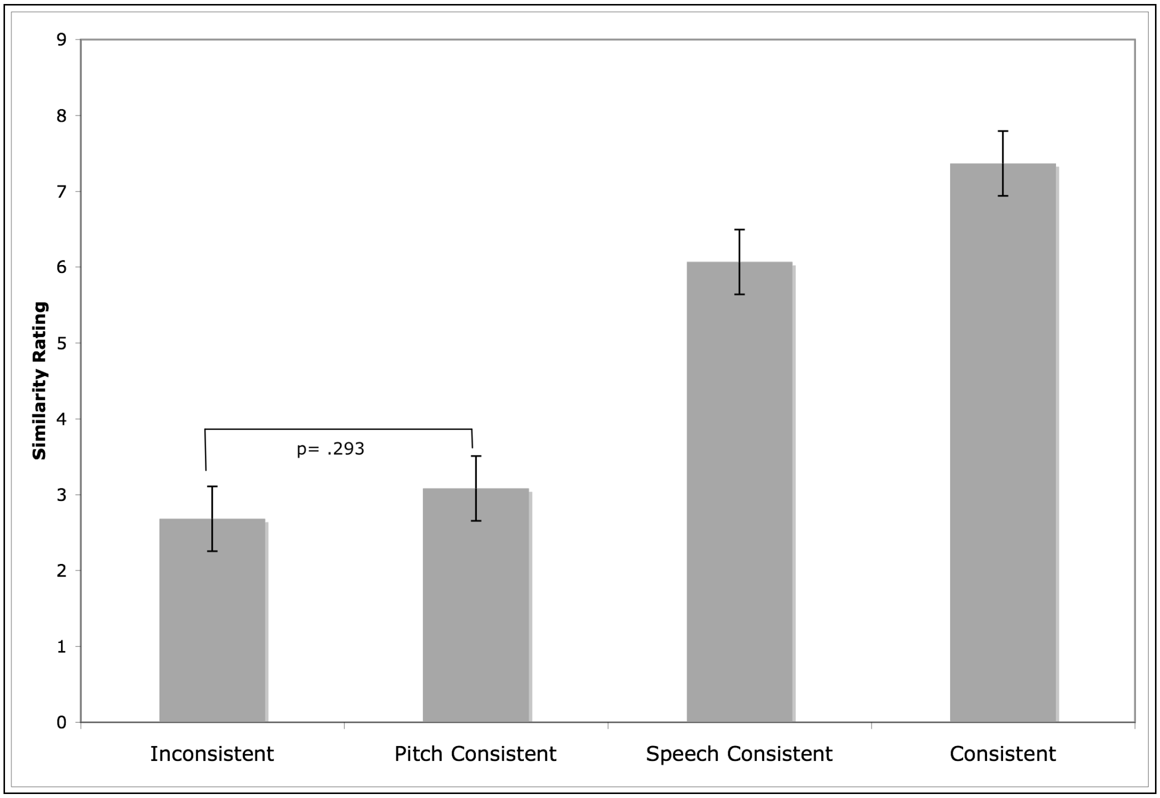

Figure 2: Mean Similarity Ratings for Fully Inconsistent, Syllable Violation, Pitch Contour Violation, and Fully Consistent Test Items in the Simultaneous Speech/Pitch condition of Experiment 1.

To summarize, results in the SSP condition imply that when familiarization stimuli contain conflicting syllable and pitch contour rules, syllables and pitch contour make asymmetrical contributions to post-familiarization similarity ratings. Specifically, participants were more likely to notice when the syllable rule was violated than when the pitch contour rule was violated.

CHAPTER 5

DISCUSSION: EXPERIMENT 1

Results from the present experiment show that like infants (Marcus et al., 1999, 2007) adults readily infer rules from monotone syllable sequences (Endress, Scholl, & Mehler, 2005). Our results indicated very low similarity ratings for inconsistent (novel) speech stimuli over the high-similarity-rated consistent (familiar) stimuli, suggesting that participants were successful in identifying the rule that was instantiated during the familiarization phase of the experiment.

According to Marcus et al. (2007), rule-learning is better when rules are instantiated in speech (i.e., syllable patterns). However, results from the present study indicated that when pitch contour was maintained, adults readily learned rules on the basis of pitch patterns.

One important contribution of the present experiments was to determine whether or not experienced listeners are able to infer rules in non-speech domains when those rules are domain-appropriate. Our results demonstrated that individuals were clearly able to learn rules when they were instantiated in a music-only context when the to-be-learned rule was consistent with predominant structures in music (e.g. melodic contour was maintained). These findings suggest that perhaps musical contour was a potentially confounding factor in Marcus et al.'s (2007) stimuli design because perhaps the infants were misled by the inconsistent contour, and were not able to discern which rule they were expected to learn. Individuals are certainly able to learn rules when they are presented in a speech context, and prior evidence suggests they are just as able to do so in a number of other non-speech domains (Fiser & Aslin, 2002; Kirkham et al., 2002;

Saffran et al., 1999), and our evidence suggesting they can learn contour rules over pitch sequences, lends further support to the notion that rule-learning is not exclusive to language and speech.

The final condition of Experiment 1, however, indicates the presence of domain-specific biases in adults, particularly in contexts when rules can be inferred from multiple and conflicting dimensions of sound. When listeners in the present study were familiarized to conflicting syllable and pitch contour rules, they subsequently gave lower similarity ratings to any test stimulus that presented inconsistent syllable patterns, regardless of pitch contour. This supports the contention that listeners weigh syllable and pitch contour differently when those dimensions conflict by giving more weight to the speech-relevant dimension (syllable patterning) than to the less speech-relevant dimension (pitch contour).

Together, findings across three conditions in Experiment 1 suggest that adults are readily able to infer rules from sequences on the basis of either speech (syllable), or melody (pitch), but that when speech and pitch dimensions are in conflict, adults focus almost exclusively on syllable patterns, and virtually ignore pitch contour. If the findings of Dawson & Gerken (2009) are indicative of not just older infants, but how all experienced listeners respond, this may explain why previous studies have failed to find rule-learning in a pitch context. This possibility would help provide an explanation as to why the infants in the Marcus et al. (2007) study responded so differently to the language and music stimuli.

In the present study, it was highly likely that the presence of syllables prompted listeners to interpret the context as linguistic, thus causing them to place more weight on the language-relevant information when both speech and pitch dimensions varied simultaneously. However, another possibility was that listeners simply found spectral variation more salient than pitch variation. If this is true, then individuals should also attend to other forms of spectral structure when paired with varying pitches.

To test this theory, Experiment 2 examines rule-learning in a music context, replacing speech syllables with musical instrument timbres. If rule-learning is context-dependent, participants in Experiment 2 should direct their attention to pitch, a music-relevant dimension of sound, and ignore or attend less to timbre, a dimension irrelevant to music. However, if it is true that listeners simply find spectral structure a more salient dimension regardless of domain or context, participants will base their similarity ratings on whether the timbral pattern is consistent or inconsistent with the familiarization sequence, rather than placing emphasis on what rules the pitch patterns are following.

Timbre can be a salient dimension in some musical instances, particularly when factoring in musical experience. Pitt & Crowder (1992) had both trained and untrained listeners judge whether two consecutively presented notes had the same pitch. The notes were played by one of two different synthesized instruments (one similar to a car horn and one similar to an organ) at three different pitch levels. Musicians with six or more years of musical training were better able to identify timbre differences than individuals with little or no musical training, who tended to focus on the pitch differences. Similarly, when attempting to recognize melodies, musicians have a harder time than nonmusicians ignoring variation in timbre (Poulin-Charronnat et al, 2004). After hearing nine musical excerpts all played by the same instrument/timbre (a piano or an orchestra), musicians had greater difficulty recognizing previously heard excerpts played by a novel instrument than played by the same instrument. By contrast, nonmusicians had poor recognition that was not affected by instrumentation. This evidence suggests that timbre is a salient feature of musical patterns, at least for musicians.

However, past research has suggested that timbre is perceived in absolute (exact) form, more often than in relative form. This is what makes it possible to identify different speakers simply by the sounds of their voices. In a musical context, it means that listeners are typically able to identify an instrument based solely on its timbre. Krumhansl & Iverson (1992) found that, in stimuli containing both pitch and timbre information, pitch perception was largely unaffected by instrument timbre, and participants did equally as well in pitch recognition tasks, regardless of whether the timbre remained constant or varied. Furthermore, as Patel (2008) pointed out, most normal music is not organized around timbral contrasts because many musical instruments are capable of producing a multitude of different timbres, and thus, this lack of uniformity often makes the intervals between the varying timbres difficult, if not impossible to identify. Another possibility may be that timbre is so multi-dimensional that determining an "interval" between two timbral sounds would be much less straight-forward than identifying an interval between two pitches, which have much clearer frequencies and pitch class dimensions.

In the following experiment, participants will be familiarized with stimuli that are Timbre-Only, Pitch-Only, or both Timbre and Pitch played simultaneously, similar to Experiment 1. If Krumhansl & Iverson (1992) are correct by suggesting that timbre variation does not affect pitch perception unless pitch remains constant, participants will be successful in providing high similarity ratings for consistent stimuli, as well as low similarity ratings for inconsistent stimuli in both the Timbre-Only and Pitch-Only conditions. These findings would be consistent with the results of Experiment 1 in demonstrating that, contrary to the claims of Marcus et al. (2007), rule-learning is possible when language-specific information is not present.

Moreover, if Patel (2008) is accurate in suggesting that pitch changes are typically more salient than timbral changes, in the Simultaneous Timbre/Pitch (STP) condition, participants will be more likely to base their similarity ratings on pitch-based rules, ignoring timbral rules in the process. This finding would support the argument that listeners use pitch cues (including contour) more readily than timbre when they perceive a music-based context. However, if the assumptions of Krumhansl & Iverson (1992) and Patel (2008) are incorrect, and listeners do not focus on pitch-based rules, but rather place a greater emphasis on timbre in the Simultaneous Timbre/Pitch condition, it will suggest that they are attending more to spectral structure when inferring patterns from auditory sequences.

As in Experiment 1, pitch contour will remain constant to determine whether it plays a different role in a music context vs. a speech context, because contour is a very important aspect of what allows an individual to perceive something as being musical. If context has no effect, then pitch contour should play the same role in both experiments. However, if context is relevant to the listener, then consistent pitch contour will be essential to the stimuli.

CHAPTER 6

METHODOLOGY

EXPERIMENT 2: RULE-LEARNING IN A MUSIC CONTEXT

Participants, Apparatus, and Procedure

Forty new participants with normal hearing between the ages of 18-40 years were recruited from the subject pool at the University of Nevada, Las Vegas and received course credit. The sample consisted of 22 females and 18 males, aged 18 to 40 years (*M* = 20.2). Participants ranged in formal music training from 0 to 18 years (*M* = 2). All other aspects of the apparatus and procedure were identical to Experiment 1.

Stimuli

Stimuli in Experiment 2 were identical to those used in Experiment 1, except that instead of syllables, sequential changes in spectral structure were created using instrumental timbre rather than speech syllables. Subjects therefore participated in each of three testing conditions: Timbre Only, Pitch Only and Simultaneous Timbre/Pitch (STP). For the Instrument Only condition, rules were presented as sequential changes in instrumental timbre, such as flute-violin-flute (ABA) or flute-violin-violin (ABB) (*Table 2*). Rules were therefore defined by changes in instrumental timbre, while pitch was held constant. Thirteen instrumental timbres, each set at 261.6 Hz (or middle C) were created using Propellerhead Reason 4.0 synthesizer/sampling software. Test stimuli (*Table 2*) contained entirely novel instrument timbres arranged in triplet patterns that were consistent or inconsistent with the rule that was established during familiarization. Timbres were based off of those used by Marcus et al. (2007). To ensure that timbres were sufficiently distinctive, all possible instrument timbre pairings were rated using a scale of 1 ("Did not sound alike at all) to 5 ("Sounded exactly alike"). Only timbre pairings with an average rating of 2.5 to 3.5 were selected as stimuli. Each instrument within a triplet was approximately 500 ms in duration, with a 250 ms gap between timbres and a 1 s gap between triplets. Each triplets lasted approximately 2 s. The familiarization phase contained 26 triplets of stimuli and lasted approximately 1 min 3 s. Each test trial contained 2 repetitions of each of 4 novel triplets. Test trial order was random.

In the Pitch Only condition, rules were presented through changes in fundamental frequency while instrument was held constant. This condition was identical to the Pitch Only condition of Experiment 1, except that instead of using the syllable "da" to represent each pitch, a flute timbre was used. The STP Condition was identical to the SSP condition of Experiment 1, except that instrumental timbres and pitches presented the conflicting rules.

Rule: ABB	Familiarization Phase	Test Phase
Instrument Condition (Pitch level held constant at Middle C)	bell-flute-flute bell-vib-vib flute-harp-harp flute-piano-piano guitar-harp-harp guitar-org-org harp-org-org harp-piano-piano piano-bell-bell piano-vib-vib sax-org-org sax-vib-vib vib-org-org	clar-frh-clar clar-frh-frh trum-viol-trum trum-viol-viol
Pitch Condition (Timbre held constant at Flute)	C#DD, CC#C#, AA#A#, G#AA, CDD, F#G#G#, G#A#A#, F#AA, DF#F#, F#A#A#, C#F#F#, DAA, C#G#G#	D#GG, D#GD# EFF EFE
Simultaneous Timbre/Pitch Condition (STP)	bell(C#)-flute(D)-flute(C#) bell(C)-vib(C#)-vib(C) flute(A)-harp(A#)-harp(A) flute(G#)-piano(A)-piano(G#) guitar(C)-harp(D)-harp(C) guitar(F#)-org(G#)-guitar(F#) harp(G#)-org(A#)-org(A#) harp(F#)-piano(A)-piano(F#) piano(D)-bell(F#)-bell(D) piano(F#)-vib(A#)-vib(F#) sax(C#)-org(F#)-org(C#) sax(D)-vib(A)-vib(D) vib(C#)-org(G#)-org(C#)	clar(D#)-frh(G)-clar(D#) clar(D#)-frh(G)-frh(G) clar(D#)-frh(G)-frh(D#) clar(D#)-frh(G)-clar(G) trum(E)-viol(F)-trum(E) trum(E)-viol(F)-viol(F) trum(E)-viol(F)-viol(E) trum(E)-viol(F)-trum(F)

Table 2: Stimuli Distribution for ABB rule in Experiment 2. Note: Half of the subjects received the opposite rule (ABA) during the familiarization phase (Instrument ABA/Pitch ABB in the STP Condition). Here, the STP Condition follows the rule sequence of Timbre ABB/Pitch ABA. Instrumental abbreviations are as follows: “viol”=violin, “clar”=clarinet,, “frh”=french horn, “org”=organ, “trum”=trumpet, “sax”=saxophone, “vib”=vibraphone.

CHAPTER 7

RESULTS: EXPERIMENT 2

Control Condition: Instrument Timbre-Only

Similarity ratings were submitted to a 2 x 2 (Test Item [inconsistent, consistent] x Familiarization Grammar [ABA, ABB]) mixed design ANOVA, with Test Item as a within-subjects factor and Familiarization Grammar as a between-subjects factor. This ANOVA yielded a main effect of test item, $F(1,38) = 56.24$, $p < .001$. *Figure 3* presents overall similarity ratings for consistent and inconsistent test items (averaged across the two grammars), and illustrates that participants gave lower similarity ratings to inconsistent test items ($M = 2.88$, $SD = 2.54$) than to consistent test items ($M = 7.77$, $SD = 2.17$). There were no main effects of Familiarization Grammar, $F(1,38) = 0.77$, $p = .386$, and no interaction between Test Item and Familiarization Grammar, $F(1,38) = 0.57$, $p = .457$.

Control Condition: Pitch Contour-Only

Similarity ratings in the Pitch Contour-Only condition were submitted to a 2 x 2 (Test Item [inconsistent, consistent] x Familiarization Grammar [ABA, ABB]) mixed design ANOVA with Test Item as a within-subjects factor and Familiarization Grammar as a between-subjects factor. This ANOVA yielded a main effect of Test Item, $F(1,38) = 40.21$, $p < .001$. As shown in *Figure 3*, participants gave lower similarity ratings to inconsistent test items ($M = 2.76$, $SD = 2.49$) than to consistent test items ($M = 7.51$, $SD = 2.47$). There were no main effects of Familiarization Grammar, $F(1,38) = 0.01$, $p = .927$, and no interaction between Test Item and Familiarization Grammar, $F(1,38) = 0.64$,

p = .430, indicating that rule-learning was unaffected by the form of the grammar (ABA vs. ABB).

Again, to ensure that no baseline differences existed in the difficulty of learning instrument timbre and pitch contour rules, participants' similarity judgments from both the Timbre-Only and Pitch Contour-Only conditions were combined and submitted to a 2 x 2 (Test Item [consistent, inconsistent] x Domain [Timbre, Pitch Contour]) mixed-design ANOVA, where Test Item was within-subjects and Domain was between-subjects. This ANOVA confirmed that performance varied by Test Item, $F(1,39) = 77.20$, $p < .001$, but, unlike Experiment 1, did not vary by Domain, $F(1,39) = 0.75$, $p < .391$, with no interaction between Test Item and Domain, $F(1,39) = 0.03$, $p < .875$. The lack of a main effect of domain may simply be due to participants perceiving both the instrument timbres and pitches as musical in nature, and thus, did not find one to have more distinct differences than the other. In other words, participants readily inferred rules in both baseline conditions and did so equally well whether learning instrument timbre rules or pitch contour rules.

Overall, these results provide clear evidence that participants were able to accurately differentiate inconsistent from consistent test items, inferring rules regardless of the specific rule presented during familiarization (ABA vs. ABB) and whether the rule was instantiated in timbre or pitch contour patterns.

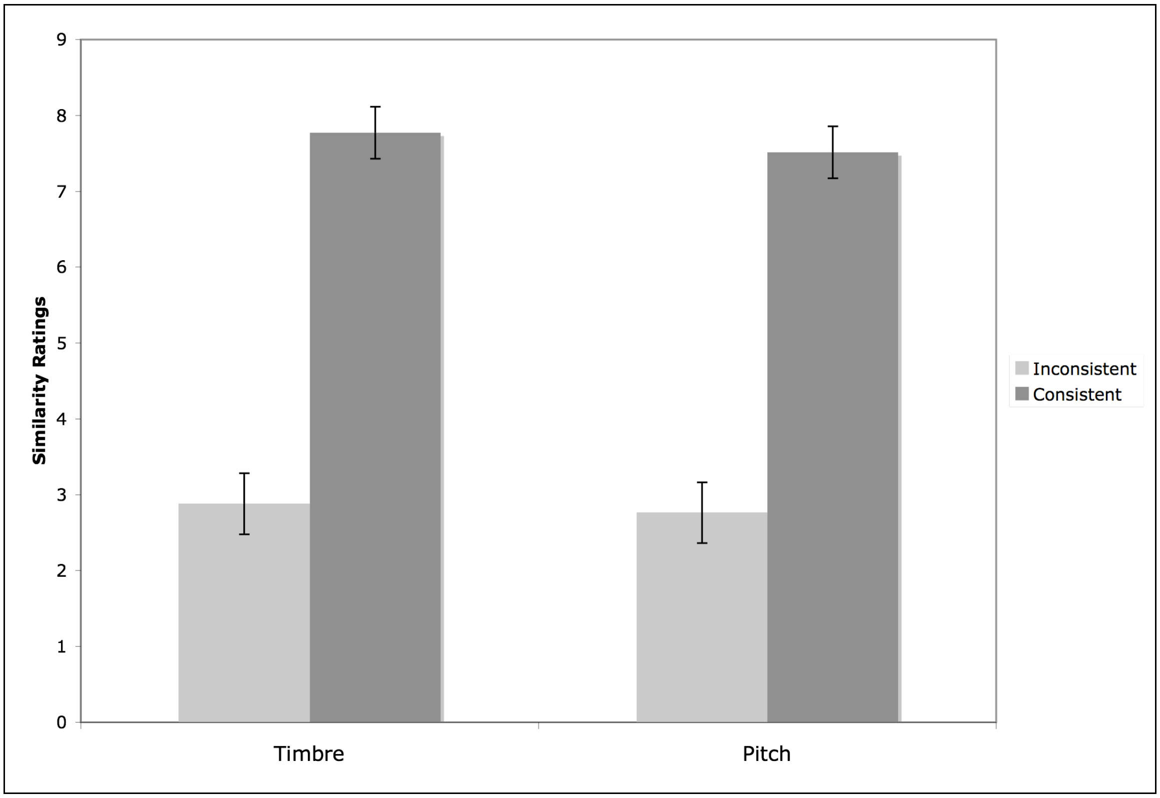

Figure 3: Similarity ratings for Consistent and Inconsistent test items in the Timbre Only and Pitch Contour Only conditions of Experiment 2.

Simultaneous Timbre/Pitch Contour Condition (STP)

As in Experiment 1, subjects were presented with four test items instead of two, with each test item conforming or not conforming to the timbre rule, the pitch contour rule, or both. Again, two factors were created to indicate whether or not the test item was consistent with the timbre rule (Test Item Timbre) and to the pitch contour rule (Test Item Pitch Contour). Similarity ratings in the STP condition were submitted to a 2 x 2 x 2 (Test Item Timbre [consistent, inconsistent] x Test Item Pitch Contour [consistent, inconsistent] x Familiarization Grammar [Timbre ABA/Pitch Contour ABB, Timbre ABB/Pitch Contour ABA]) mixed design ANOVA, with Test Item Timbre and Test Item

Pitch Contour as within-subjects factors and Familiarization Grammar as a between-subjects factor. This analysis yielded a main effect of Timbre Test Item, $F(1,38) = 4.818$, $p < .05$ a main effect of Pitch Test Item, $F(1,38) = 31.549$, $p < .001$, and an interaction between Timbre Test Item and Pitch Test Item, $F(1,38) = 6.959$, $p < .01$. This interaction indicates that participants' discrimination of inconsistent and consistent test items differed by domain, such that violations in Timbre and Pitch Test Items were weighed differently.

Figure 4 presents mean similarity ratings for each of the four test items. Using Bonferroni corrected ost-hoc pairwise t-tests revealed that participants gave significantly higher similarity ratings to fully consistent test items, $M = 6.66$, $SD = 2.43$, than to fully inconsistent test items, $M = 3.22$, $SD = 2.31$, $t(39) = 5.54$, $p < .01$. Similarity ratings were also significantly higher for partially consistent test items that maintained the pitch contour rule, $M = 4.93$, $SD = 3.03$, than for test items that maintained the timbre rule, $M = 3.37$, $SD = 2.51$, $t(39) = -2.33$, $p < .05$, suggesting that a change of pitch structure had a greater impact on perceived similarity than did a change in timbre. Fully consistent items were rated significantly higher than partially inconsistent items, whether partially inconsistent items violated the timbre rule, $t(39) = 2.93$, $p < .01$, or pitch contour rule, $t(39) = 5.76$, $p <.01$. However, fully inconsistent items were only rated significantly less similar than partially inconsistent items when the pitch rule was maintained (and the timbre rule violated), $t(39) = 3.06$, $p <.01$. Ratings did not differ between fully inconsistent and partially inconsistent items that maintained the timbre rule (and violated the pitch contour rule), $t(39) = .34$, $p =.739$. This implies that test items were perceived as dissimilar whenever they violated the pitch rule, but also that additional violation of the timbre rule did not further distinguish fully inconsistent items.

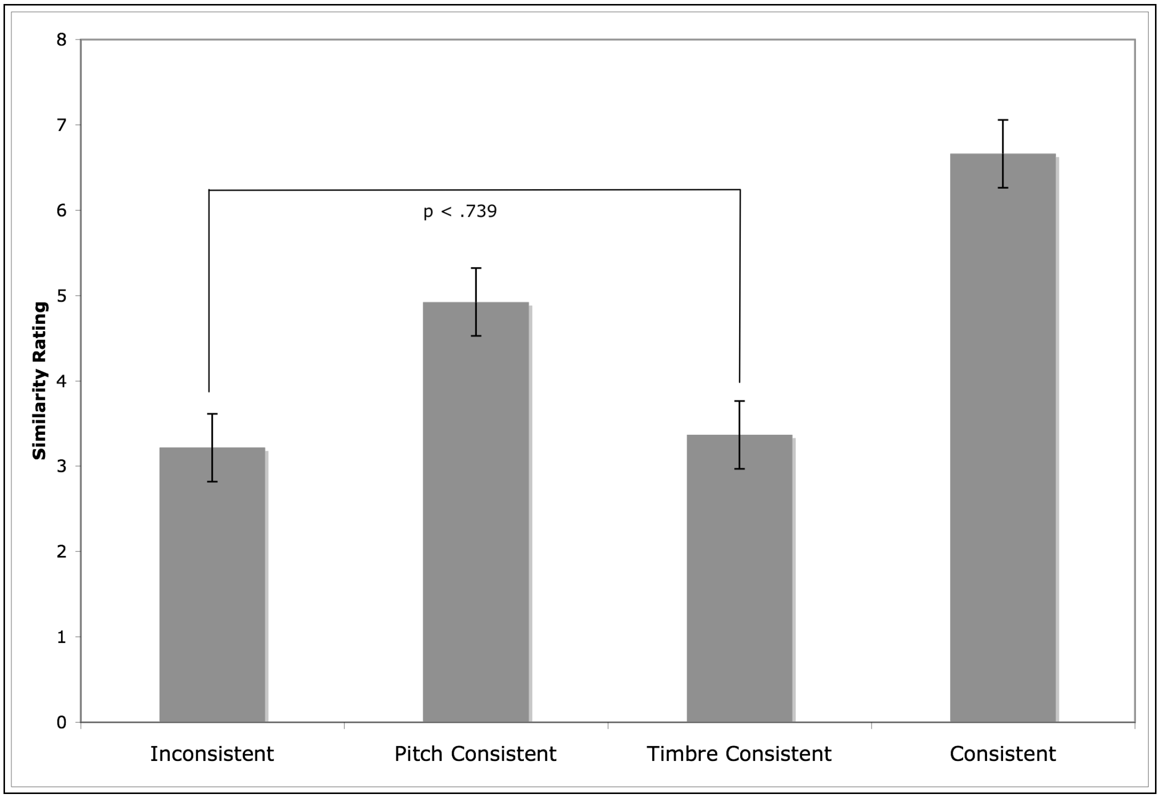

Figure 4: Mean Similarity Ratings for Fully Inconsistent, Timbre Pitch Violation, Pitch Contour Violation, and Fully Consistent Test Items in the Simultaneous Timbre/Pitch condition of Experiment 2

To summarize, results in the STP condition imply that when familiarization stimuli contain conflicting timbre and pitch contour rules, timbre and pitch contour make asymmetrical contributions to post-familiarization similarity ratings. Specifically, participants were more likely to notice when the pitch contour rule was violated than when the timbre rule was violated.

CHAPTER 8

DISCUSSION: EXPERIMENT 2

Individuals may develop speech and music biases that allow them to infer rules depending on what information is the most contextually relevant. In Experiment 1, by pairing syllable and pitch trios that followed conflicting rule patterns, individuals paid more attention to the syllabic information, ignoring the pitch information. Two possible explanations for this finding were that 1) individuals generally attend more to spectral structure than fundamental frequency when inferring rules from auditory sequences; or 2) individuals believed they were operating within a speech context, and thus found the syllabic information most salient as a domain-relevant cue. The purpose of Experiment 2, then, was to tease apart these two possibilities.

In this study, results showed that listeners based their similarity ratings on whether the pitch contour rules were similar with familiarization, and tended to ignore the timbral patterns, regardless of whether or not the information violated or conformed to the established timbre rules. This suggests that domain-specific biases lead listeners to focus more on the pitch contour, due to the perception of a more musically oriented context. These domain-specific biases appear to drive pattern learning in a way such that a more music-like context leads listeners to focus on the pitch contour, which is typically more music-relevant than instrument timbre. These findings confirm that listeners were not simply drawn to pay more attention to spectral structure overall, but rather, that perceived context plays a role in what auditory information is deemed most appropriate or relevant at a given time.

Results from Experiment 2 indicated that, while in the baseline conditions, listeners readily differentiate between consistent and inconsistent stimuli (just as in baseline conditions of Experiment 1). However, when spectral structure (now instrument timbre) and fundamental frequency (pitch) were pitted against one another, listeners gravitated toward following the rules instantiated in pitch contour and ignoring the timbre patterns. These findings are completely opposite to those of Experiment 1, where listeners placed greater emphasis on the rules presented through the spectral structure (speech syllables) than the fundamental frequency (pitch contour).

In Marcus et al.'s (2007) timbre study, seven-month-old infants were not able to infer rules from any nonlinguistic familiarization stimuli. Marcus et al. took these findings to imply that infants are not able to extract rules in non-speech contexts, and suggested that perhaps, from as early as birth, infants simply prefer listening to speech over other forms of auditory sounds including animal sounds, instrument timbres, tones, and sine-wave analogues, possibly due to speech being highly familiar or salient. Dawson & Gerken's (2009) findings confirmed that 7.5-month-old infants were not able to detect rules in stimuli containing chords and tones; however, they also found that 4-month-old infants succeeded with the same stimuli. This finding could imply that acquired domain-specific musical knowledge (.e.g of the primacy of pitch contour) was interfering more with the performance of older infants than with that of younger infants. If this type of domain-specificity continues to develop over time, then it should be evident in adults' inference of rules over pitch or timbre sequences in speech or music contexts. In the case of the present experiment, the adult listeners appeared to have such domain-specific music knowledge, and thus, were biased to infer rules from the pitch information more often than the timbral information, which is less likely to be meaningful in a musical context.

This bias is consistent with a "musical mode" of listening because over time and with exposure to music, individuals learn that sequences of timbre are not meaningful, whereas sequences of relative pitch are meaningful in a music context. When listening to an orchestra play, for example, a single instrument typically plays the key melody, and only rarely (such as in the 20th century genre of notoriously challenging "total serialism") do separate instruments play each note of a melody. Similarly, Patel (2008) points out that very few instruments are actually organized around their various timbral contrasts (one notable exception being the Australian didgeridoo). As mentioned previously, most musical instruments are capable of producing multiple timbral contrasts, and there is not necessarily a defined way to measure the perceptual distance from one timbre to another. This is unlike pitch, which does have a system of intervals that are fairly recognizable.

Conversely, when listening to someone speaking, individuals typically pay more attention to the words being spoken to obtain the most meaning. This is not to say that timbre and pitch, respectively, are completely irrelevant acoustic elements in speech. Instead, it merely suggests that depending on the perceived context, they may not be the most salient components of the auditory stimuli.

CHAPTER 9

GENERAL DISCUSSION

Overall, the present studies aimed to examine whether or not the capacity to infer patterns from sound sequences is unique to language, as claimed by Marcus et al. (2007), or instead varies by context and according to listeners' knowledge of the dimensions of sound that are relevant or irrelevant in that context. The results of the present experiments support the latter by documenting domain-specific biases that probably emerge over the course of development, and that lead listeners to weight information differently depending on context. Thus, through exposure to the auditory structure in music and language domains, listeners learn to direct their focus toward the information that is most likely to be important in each context and to ignore information that does not tend to be useful.

The results of the present study, as well as other findings, demonstrating rule-learning from sequences of tones (Dawson & Gerken, 2009), animal faces (Saffran et al., 2007), objects (Fiser & Aslin, 2002) and colored shapes (Kirkham et al., 2002) clearly challenge the notion that rule-like learning is specific to language, as suggested by Marcus et al. (2007). Rather, these studies and the present evidence lend support to the growing evidence that abstract rule-learning is possible in multiple domains.

Through exposure to structured sound in the world throughout infancy, childhood, and adulthood, listeners acquire and maintain perceptual and cognitive networks that structure their systems of norms, beliefs, and values (Demorest et al., 2008). Through this acquired knowledge, individuals may also gain speech and music biases that begin to develop as early as 10-12 months of age, and become increasingly domain-specific as they grow older. Thus, even though adults may be perfectly competent at learning rules in different domains, the domain-specific context may determine which rules they will choose to follow.

To better determine whether these domain-specific biases are learned or innate, a future direction for research would be to examine how infants differentially weigh spectral vs. pitch information using a preferential looking paradigm. If domain-specific biases are innate, then infants should show preferences for spectral or pitch information depending on the context in which it is presented, as with the present adult studies. However, if domain-specific biases are learned over time, then younger infants should not exhibit the adult-like tendency to weigh syllable patterning over pitch contour in the speech context, and pitch contour over timbre patterning in a music context. In other words, unlike adults, young infants would be expected to treat any inconsistent test item as novel, whether that item violates a rule based on spectral structure or pitch contour. By contrast, older infants or children might begin to show context-dependent biases, favoring spectral information in a speech context (Marcus et al., 2007), and pitch contour information in a music context.

APPENDIX

IRB APPROVALS

UNLV
UNIVERSITY OF NEVADA LAS VEGAS

Social/Behavioral IRB – Expedited Review
Approval Notice

NOTICE TO ALL RESEARCHERS:

Please be aware that a protocol violation (e.g., failure to submit a modification for <u>any</u> change) of an IRB approved protocol may result in mandatory remedial education, additional audits, re-consenting subjects, researcher probation suspension of any research protocol at issue, suspension of additional existing research protocols, invalidation of all research conducted under the research protocol at issue, and further appropriate consequences as determined by the IRB and the Institutional Officer.

DATE: May 23, 2009

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

FROM: Office for the Protection of Research Subjects

RE: Notification of IRB Action by Dr. Paul Jones, Co-Chair
Protocol Title: **Adult's Perception of Music vs. Speech**
Protocol #: 0904-3084M

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 45 CFR 46. The protocol has been reviewed and approved.

The protocol is approved for a period of one year from the date of IRB approval. The expiration date of this protocol is May 13, 2010. Work on the project may begin as soon as you receive written notification from the Office for the Protection of Research Subjects (OPRS).

PLEASE NOTE:
Attached to this approval notice is the **official Informed Consent/Assent (IC/IA) Form** for this study. The IC/IA contains an official approval stamp. Only copies of this official IC/IA form may be used when obtaining consent. Please keep the original for your records.

Should there be *any* change to the protocol, it will be necessary to submit a **Modification Form** through OPRS. No changes may be made to the existing protocol until modifications have been approved by the IRB.

Should the use of human subjects described in this protocol continue beyond May 13, 2010, it would be necessary to submit a **Continuing Review Request Form** *60 days* before the expiration date.

If you have questions or require any assistance, please contact the Office for the Protection of Research Subjects at OPRSHumanSubjects@unlv.edu or call 895-2794.

BIBLIOGRAPHY

Altmann, G.T.M. (1999). Rule learning by seven-month-old infants and neural networks. *Science, 284,* 875.

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

Bickerton, D. (1981). *Roots of language.* Ann Arbor, Mich: Karoma.

Bishop, D.V.M. (1997). Cognitive neuropsychology and developmental disorders: Uncomfortable bedfellows. *Qualitative Journal of Experimental Psychology, 50A,* 899-923

Bregman, A. & Pinker, S. (1978). Auditory streaming and the building of timbre. *Canadian Journal of Psychology, 32*, 19-31.

Brown, J.H., Johnson, M.H., Paterson, S.J., Gilmore, R., Longhi, E., & Karmiloff-Smith, A. (2003). Spatial representation and attention in toddlers with Williams Syndrome and Downs syndrome. *Neuropsychologia, 41,* 1037-1046.

Chomsky, N. (1965). *Aspects of the theory of syntax.* Cambridge, Mass: MIT Press.

Chomsky, N. (1988). *Language and problems of knowledge: the Managua lectures.* Cambridge, Mass.: MIT Press.

Christiansen, M.H. & Curtin, S.L. (1999). Transfer of learning: rule acquisition or statistical learning? *Trends in Cognitive Science, 3,* 289-290.

Christensen-Dalsgaard, J. (2004). Music and the origin of species. *The Journal of Music and Meaning, 2,* 1-16.

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

Conway, C.M. & Christensen, M.H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 24-39.

Cosmides, L. & Tooby, J. (1994). Origins of domain specificity: The evolution of Functional organization. In L. Hirschfeld & S. Gelman (Eds.), *Mapping the Mind: Domain-specificity in cognition and culture.* New York: Cambridge University Press.

Curtin, S. & Werker, J.F. (2007). The perceptual foundations of phonological development. In G. Gaskell (Ed.), *The Oxford Handbook of Psycholinguistics* (pp. 579 599). Oxford University Presss.

DeCasper, A.J. & Fifer, W.P. (1980). Of human bonding: newborns prefer their mothers' voices. *Science, 208*, 1174-1176.

DeCasper, A.J. & Spence, M.J. (1986). Prenatal maternal speech influences newborns' perception of speech sounds. *Infant Behavior and Development, 9*, 133-50.

Dehaene-Lambertz, G. (2000). Cerebral specialization for speech and non-speech stimuli in infants. *Journal of Cognitive Neuroscience, 12,* 449-460.

Dehaene-Lambertz, G., Dehaene, S., & Hertz-Pannier, L. (2002). Functional neuro-imaging of speech perception in infants. *Science, 298,* 2013-2015.

Dehaene-Lambertz, G. & Gliga, T. (2004). Common neural basis for phoneme processing in infants and adults. *Journal of Cognitive Neuroscience, 16,* 1375-1387.

Darwin, C. (1871). *The Descent of Man.* London: Folio Society.

Dawson, C. & Gerken, L. (2009). From domain-generality to domain-sensitivity: 4-month-olds learn an abstract repetition rule in music that 7-month-olds do not. *Cognition, 111,* 378-382.

Demorest, S.M., Morrison, S.J., Beken, M.N., & Jungbluth, D. (2008). Lost in translation: An enculturation effect in music memory performance. *Music Perception, 23,* 213-223.

Don, A.J., Schellenberg, E.G., & Rourke, B.P. (1999). Music and language skills of children with Williams Syndrome. *Child Neuropsychology, 5,* 154-170.

Donnai, D. & Karmiloff-Smith, A. (2000). Williams syndrome: from genotype through to the cognitive phenotype. *American Journal of Medical Genetics, 97,* 164-171.

Dutoit, T., Pagel, V., Pierret, N., Bataille, F., & Van Der Vreken, O. (1996). The MBROLA project: Towards a set of high-quality speech synthesizers free of use for non-commercial purposes. *Proc. ICSLP'96,* Philadelphia, *3*, 1393-1396.

Eimas, P.D. (1999). Do infants learn grammar with algebra or statistics? *Science, 16,* 433.

Elman, J. (1999). Generalization, rules, and neural networks: A simulation of Marcus et al. (1999).

Elsabbagh, M. & Karmiloff-Smith, A. (2004). Modularity of mind and language. In *The Encyclopedia of Language and Linguistics – 2nd Edition.*

Endress, A., Scholl, B. & Mehler, J. (2005). The role of salience in the extraction of algebraic rules. *Journal of Experimental Psychology: General, 134*, 406-419.

Ferguson, C.A. (1964). Baby talk in six languages. *American Anthropologist, 66,* 103-114.

Fernald, A. (1989). Intonation and communicative intent in mothers speech to infants: Is the melody the message? *Child Development, 60*, 1497–1510.

Fiser, J. & Aslin, R.N. (2002). Statistical learning of higher-order temporal structure from visual shape sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 458-467.

Fodor, J.A. (1983). *The modularity of mind.* MIT Press: Cambridge, MA.

Foxton, J.M., Dean, J.L., Gee, R., Peretz, I., & Griffiths, T.D. (2004). Characterization of deficits in pitch perception underlying 'tone deafness'. *Brain: A Journal of Neurology, 127,* 801-810.

Goldstone, R.L. (1998). Perceptual learning. *Annual Review of Psychology, 49*, 585-612.

Green, L. (1988). *Music on deaf ears: Musical meaning, ideology, education.* Manchester University Press: Michigan.

Gregory, A.H. (1994). Timbre and auditory streaming. *Music Perception, 12,* 161-174.

Grieser, D.L. & Kuhl, P.K. (1988). Maternal speech to infants in a tonal language: support for universal prosodic features in motherese. *Developmental Psychology, 24,* 14-20.

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

Hannon, E.E. & Trainor, L.J. (2007). Music acquisition: effects of enculturation and formal training on development. *Trends in Cognitive Science, 11,* 466-472.

Hulme, C. & Snowling, M. (1992). Deficits in output phonology: an explanation of Reading failure? *Cognitive Neuropsychology, 9*, 47-72.

Imada, T., Zhang, Y., Cheour, M., Taulu, S., Ahonen, A, & Kuhl, P. K. (2006). Infant speech perception activates Broca's area: a developmental magnetoencephalography study. *NeuroReport,*

Jusczyk, P.W., Bertoncini, J., Bijeljac-Babic, R., Kennedy, L.J., & Mehler, J. (1990). The role of attention in speech perception by young infants. *Cognitive Development, 5,* 265-286.

Jusczyk, P.W. & Aslin, R.N. (1995). Infants' detection of the sound patterns of words in fluent speech. *Cognitive Psychology, 29,* 1-23.

Karmiloff-Smith, A., Klima, E., Bellugi, U., & Grant, J. (1995). Is there a social module? Language, face-processing, and theory of mind in individuals with Williams syndrome. *Journal of Cognitive Neuroscience, 7,* 196-208.

Karmiloff-Smith, A., Brown, J.H., Grice, S., & Paterson, S. (2003). Dethroning the myth: cognitive dissociations and innate modularity in Williams syndrome. *Developmental Psychology, 23,* 227-242.

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

Kolinsky, P., Lidji, P., Peretz, I., Besson, M., & Morais, J. (2009). Processing interactions between phonology and melody: Vowels sing but consonants speak. *Cognition, 112,* 1-20.

Kotilahti, K., Nissilä, I., Näsi, T., Lipiänen, L., Noponen, T., Meriläinen, P., Huotilainen, M., & Feilman, V. (2009). Hemodynamic responses to speech and music in newborn infants. *Human Brain Mapping,* 1-9.

Krumhansl, C. (1989). Why is musical timbre so hard to understand? In: S. Nielzen & O. Olsson (Eds.), *Structure and Perception of Electroacoustic Sound and Music* (pp. 43-54). New York: Excerpta Medica.

Krumhansl, C. & Iverson, P. (1992). Perceptual interactions between musical pitch and timbre. *Journal of Experimental Psychology: Human Perception and Performance, 18,* 739-751.

Kuhl, P.K., Stevens, E., Hayaski, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science, 9,* F13-F21.

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

Maess, B., Koelsch, S., Gunter, T.C., & Friederici, A.D. (2001). Musical syntax is processed in broca's area: an MEG study. *Neuroscience,*

Magne, C., Schon, D., & Besson, M. (2006). Musician children detect pitch violations in both music and language better than nonmusician children: behavioral and electrophysiological approaches. *Journal of Cognitive Neuroscience, 18,*199-211

Marcus, G.F., Vijayan, S., Bandi Rao, S., & Vishton, P.M. (1999). Rule-learning by seven-month-old infants. *Science, 283,* 77-80.

Marcus, G.F. (2000). Pabiku and ga ti ga: Two mechanisms infants use to learn about the world. *Current Directions in Psychological Science, 9,* 145-147.

Marcus, G.F., Fernandes, K.J., & Johnson, S.P. (2007). Infant rule-learning facilitated by speech. *Psychological Science, 18,* 387-391.

Mehler, J., Pena, M., Nespor., M. & Bonatti, L.L. (2006). The "soul" of language does not use statistics: Reflections on vowels and consonants. *Cortex, 42,* 846-854.

McClelland, J.L. & Plaut, D.C. (1999). Does generalization in infant learning implicate abstract algebra-like rules? *Trends in Cognitive Sciences, 3,* 166-168.

Nakata, T. & Trehub, S.E. (2004). Infants' responsiveness to maternal speech and singing. *Infant Behavior & Development, 27,* 455-464.

Patel, A.D. (2008). *Music, Language, and the Brain.* Oxford University Press: New York, NY.

Patel, A.D., Wong, M., Foxton, J., Lochy, A., & Peretz, I. (2008). Speech intonation perception deficits in musical tone deafness (congenital amusia). *Music Perception, 25,* 357-368.

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

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

Pinker, S. (1994). *The language instinct.* New York, NY: HarperCollins

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

Pitt, M.A. & Crowder, R.G. (1992). The role of spectral and dynamic cues in imagery for musical timbre. *Journal of Experimental Psychology: Human Perception and Performance, 18,* 728-738.

Poulin-Charronnat, B., Bigand, E., Lalitte, P., Madurel, F., Vieillard, S. & McAdams, S. (2004). Effects of a change in instrumentation on the recognition of musical materials. *Music Perception, 22*, 239-263.

Rivera-Gaxiola, M., Silvia-Pererya, J., & Kuhl, P.K. (2005). Brain potentials to native and non-native speech contrasts in 7- and 11-month –old American infants. *Developmental Science, 8,* 162-172.

Rosen, S. & Iverson, P. (2007). Constructing adequate non-speech analogues: what is special about speech anyway? *Developmental Science, 10,* 165-169.

Roy, M., Peretz, I., & Rainville, P. (2008). Emotional valence contributes to music-induced analgesia. *Pain, 134,* 140-147.

Sacks, O., Schlaug, G., Jancke, L., Huang, Y. & Steinmetz, H. (1995). Musical ability. *Science, 258,* 621-622.

Saffran, J.R., Aslin, R.N., & Newport, E.L. (1996). Statistical learning by 8-month-old infants. *Science, 274,* 1926-1928.

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,* 27-52.

Saffran, J.R., Pollak, S.D., Seibel, R.L., & Shkolnik, A. (2007). Dog is a dog: infant rule learning is not specific to language. *Cognition, 105,* 669-680

Schellenberg, E.G., Peretz, I., & Vieillard, S. (2008). Liking for happy- and sad-sounding music: effects of exposure. *Cognition and Emotion, 22,* 218-237.

Seidenberg, M.S. & Elman, J.L. (1999). Networks are not 'hidden rules'. *Trends in Cognitive Psychology, 3,* 288-289.

Snyder, J. & Alain, C. (2007). Toward a neuropsychological theory of auditory stream segregation. *Psychological Bulletin, 133,* 780-799.

Steinbeis, N. & Koelsch, S. (2008). Shared neural resources between music and languate indicate semantic processing of musical tension-resolution patterns. *Cerebral Cortex, 18,* 1169-1178.

Terrace, H.S., Petitto, L.A., Sanders, R.J., & Bever, T.G. (1979) Can an ape create a sentence? *Science, 206,* 891-902.

Thompson, W.F., Schellenberg, E.G., & Husain, G. (2004). Decoding speech prosody: do music lessons help? *Emotion, 4,* 46-64.

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

Trehub, S.E., Bull, D., & Thorpe, L.A. (1984). Iinfants' perception of melodies: the role of melodic contour. *Child Development, 55,* 821-830.

Trehub, S.E. & Trainor, L.J. (1998). Singing to infants: lullabies and play songs. *Advances in Infancy Research, 12,* 43-77.

Trehub, S.E. (2001). Musical predispositions in infancy. In: *The biological foundations of music.* Zatorre, R.J.; Peretz, I; New York, NY: New York Academy of the Sciences, 1-16.

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

Tsao, F.M., Liu, H.M. & Kuhl, P.K. (in press). Perception of native and non-native affricate-fricative contrasts: cross language tests on adults and infants. *Journal of the Acoustical Society of America.*

Vouloumanos, A., Hauser, M., Werker, J., & Martin, A. (2010). The tuning of human neonates' preference for speech. *Child Development, 81,* 517-527.

Vouloumanos, A. & Werker, J. (2004). Tuned to the signal: the privileged status of speech for young infants. *Developmental Science, 7,* 270-276.

Vouloumanos, A. & Werker, J. (2007). Listening to language at birth: evidence for a bias for speech in neonates. *Developmental Science, 10,* 159-171.

Vouloumanos, A. & Werker, J. (2007a). Why voice melody alone cannot explain neonates' preference for speech. *Developmental Science, 10,* 169-170.

Werker, J.F. & Tees, R.C. (1984). Cross-language speech perception: evidence for percep tual reorganization during the first year of life. *Infant Behavior and Development, 7,* 49-63.

Werker, J.F. & Lalonde, C. E. (1988). Cross-language speech perception: initial capabilities and developmental change. *Developmental Psychology, 24,* 672-683.

Wong, P.C.M., Skoe, E., Russo, N.M., Dees, T., & Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nature Neuro-Science, 10,* 420-422.

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

VITA

Graduate College
University of Nevada, Las Vegas

Aaronell Shaila Matta

Degrees:

Bachelor of Science, Psychology, 2005
The Pennsylvania State University

Special Honors and Awards:

Graduate and Professional Students Association (GPSA) Travel Award, 2009, 2010

GPSA Research Forum (Social Sciences Poster Session) – Honorable Mention, 2010

Thesis Title: Inferring Rules From Sound: The Role of Domain-Specific Knowledge in Speech and Music Perception

Thesis Examination Committee:

Chairperson, Erin E. Hannon, Ph.D.
Committee Member, Joel S. Snyder, Ph.D.
Committee Member, Jennifer L. Rennels, Ph.D.
Graduate Faculty Representative, Eugenie I. Burkett, Ph.D.