A Study of Auditory Perceptual Capacity of Certain Musical Elements in a Non-human Primate Species, the Common Marmoset (*Callithix jacchus*)

by
Yueqi Guo

A thesis submitted to Johns Hopkins University in conformity with the requirements for the degree of Master of Science in Engineering

Baltimore, Maryland

July, 2015

Abstract

It has long been debated whether music is unique to human, as man is the only species that actively produces and listens to music for entertainment and social interaction purposes. In this thesis, I investigated the ability of perceiving and innate preference for music elements in the common marmosets (*callithrix jacchus*), a highly vocal New-World monkey, which has emerged in recent years as a promising animal model in auditory research.

We first characterized marmoset's fundamental auditory perceptual abilities of discriminating pitch, the most fundamental element in music. Marmosets were trained on a discrimination task using operant conditioning procedures. We then measured the minimum change in frequency (i.e., the frequency difference limen, or FDL) that marmosets could detect using pure tones at eight frequencies with an octave spacing, covering their entire hearing range. We also tested with harmonic complex tones, a common sound type in music, at four different fundamental frequencies to measure their fundamental frequency difference limen (F0DL). These results revealed that marmosets have relatively fine pitch discrimination capacities, with FDL ranging from ~2.6 semitones to ~0.4 semitones, and F0DL ranging from ~1.6 semitones to ~0.4 semitones, depending on the test frequency. These characterizations help guide further studies of auditory behaviors of this species.

Given marmoset's capacity of discriminating musical pitch, I further investigated their high-level cognition about innate preferences in music aesthetics in terms of musical consonance and dissonance. Spontaneous behaviors were measured using a V-shaped

maze with two branches, to test whether marmosets prefer consonant over dissonant music as we humans do. A customized program was used to detect and record an animal's location in the maze (left or right branch), then select a stimulus to play through a speaker. One sound was played when the test subject moved into one branch, and a different sound was played when the test subject moved into the other branch. Thus the proportion of time an animal spent at each side serves as an indicator of its preference for the associated sound. The results showed that marmosets did not prefer consonant over dissonant stimuli in this test setting, although they did show preference of silence over white noise in the same setting. Further studies are needed to confirm this finding. Nevertheless, these results provide more insights into the evolutionary origin of music.

Acknowledgement

First and foremost, I would like to express my sincere appreciation to my thesis primary advisor, Dr. Xiaoqin Wang. My passion in auditory neuroscience and my great interest in music perception were inspired by Dr. Wang, when I was an undergraduate student at Tsinghua University in China. I still remember our exciting discussions about the intriguing questions on how we perceive pitch and how the auditory system process music, which made me realize that I could combine my personal interest with scientific career. Then I joined the BME Master program at Johns Hopkins where I could put my engineering and mathematical trainings into good use, and Dr. Wang became my master program advisor. He is a true scientist and scholar, who always encourages me to pursue what really interests me and gives us freedom to try what we want to do. He always provides guidance and support so that we don't get discouraged by our failures. I am lucky to have Dr. Wang be my advisor at the starting point of my scientific career.

I would also like to thank Dr. Eric Young, who leads my way in the adventure of neuroscience research by explaining the complex computational models, describing auditory system structures and functions patiently and clearly in the courses he taught. In addition, the journal club he has been running gave me a much broader view in neuroscience, as well as many chances to practice scientific presentation. He is an insightful, rigorous scientist. He always offers generous guidance and useful advices to young investigators like me.

During the past two years, I have had many fruitful discussions and counsel from all members of Wang lab. Xindong Song has been a terrific colleague and mentor to me.

Thesis readers:

Dr. Xiaoqin Wang (advisor)

Dr. Eric D. Young

Dr. Cynthia F. Moss

Table of Contents

Chapter 3: Consonance vs. dissonance preference by marmoset monkeys

List of Tables

List of Figures

Chapter 1: Introduction

1 Pitch perception and music perception

Music is a manipulation of sound that aims at expressing emotions, and eliciting

emotional responses in listeners. There are various basic musical elements such as pitch,

timbre, tempo and meter. Among these elements, pitch is arguably the most fundamental

and essential one, at least in most western music: sequential manipulation of different

pitches generates melodies; multiple pitches stacked simultaneously gives rise to

consonance or dissonance of a chord. However, it has been one of the most mysterious

questions in auditory neuroscience that how these different configurations of pitch in

physical space map to our internal emotional space. To lay some foundation for this

question, I will briefly review from the most basic physical aspects of pitch, to higher-

level cognition for the pitch-related musical elements in this section.

1.1 Pitch perception

1.1.1 Acoustics and psychophysics of pitch

It is not easy to define pitch accurately. According to American Standards Association,

pitch is "attribute of auditory sensation in terms of which sounds may be ordered on a

musical scale" (ASA 1960). However, a more recent and broader definition says "the

attribute of auditory sensation in terms of which sounds may be ordered on a scale

extending from low to high. Pitch depends primarily on the frequency content of the

sound stimulus, but also depends on the sound pressure and the waveform of the

stimulus" (ANSI 1994). This definition brought pitch to a broader sense, not only referring to music. No matter how pitch is defined, the most important aspect is that pitch is a perceptual concept, rather than a physical one. There is no single physical parameter that could determine the pitch of a sound, while periodicity of the waveform is considered as the most relevant parameter.

Sound is represented by vibration of the air and can be transformed by the cochlea to mechanical vibration, then to neural activity patterns. The physical aspects of how the air vibrates determine what we hear. Many natural sounds have a regular waveform, that is, periodic waveform, and the periodicity is what ties to a perception of pitch. For example, some sounds in nature like water bubbling does not have a periodic vibration pattern, giving no perception of pitch—one cannot say one bubbling sound is higher than the other. In contrast, the sounds from music instruments and most animal vocalizations have a clear pitch. These periodic vibrations are generated by the resonating part of an instrument, or the vocal tract. By changing the resonating frequencies, different pitches could be produced.

Another way to describe the attribute of pitch is to transform the temporal waveform to the spectral domain by Fourier Transform. The simplest case is when the sound wave only has one frequency component (pure tone), the spectral power will concentrate on a single frequency value, which is usually the perceived pitch value. More commonly, periodic sound has a complex spectrum, due to the physical nature of vibrating objects such as musical instruments. The peaks in the complex spectrum are often harmonically related—the lowest frequency component normally equals the reciprocal of the period of the waveform (defined as fundamental frequency, F0), and other frequency components

are integer multiples of that frequency. These make up harmonic complex tones. Although a harmonic complex tone elicits a percept of pitch of F0, it is not because of the presence of the fundamental frequency component. Schouten and Licklider confirmed that even if the F0 is removed or masked, the same pitch is still perceived—the "missing fundamental" effect, which laid the cognitive foundation for telephone industry to transfer only high-frequency components in order to minimize bandwidth (Schouten 1938, Licklider 1956). Thus, pitch is not simply the frequency that is mapped to the tonotopic axis, but requires complicated and nonlinear computational process by the auditory system to extract from the sound wave.

There have been two major theories explaining how pitch may be extracted from a periodic sound. One is based on temporal patterns—by extracting the temporal envelope of the waveform and the corresponding reciprocal of period is the pitch value (Licklider 1951); The other is based on excitation pattern on the basilar membrane, or place theory. Different frequency components maximally excite different locations on the basilar membrane, thus the spectrum of a sound is mapped to the tonotopic axis. It then requires a template matching process that matches the excitation pattern with some internal harmonic templates, and the interval of the best-match template is the pitch value (Goldstein 1973; Terhardt 1974). Another set of theories is based on a computational process of autocorrelation. For example, in Loeb's traveling wave theory (Loeb, White, and Merzenich 1983) and Shamma's delay-line theory (Shamma 1985a; Shamma 1985b), the information of phase dispersion along the basilar membrane is utilized by an array of coincide detectors to perform an autocorrelation. There is a new model proposed by Laudanski et al. recently, which is also based on autocorrelation, whereas the delay lines

3

are generalized to higher-level that can perform both within-channel correlation as well as across-channel correlation (Laudanski, Zheng, and Brette 2014).

Currently, the most commonly accepted hypothesis is that the time theory and place theory works together to extract pitch, and are responsible for different frequency ranges: In a harmonic complex tone, the lower harmonics separated by individual auditory filters serves as spectral cues. The higher harmonics that are filtered through the same filter, are considered to be providing mainly temporal cues. This hypothesis is not only supported by many psychophysical studies, but also by behavioral evidence from marmosets (Bendor, Osmanski, and Wang 2012).

Although more complex models are continuously being proposed, and are capable of explaining a wider range of psychophysical observations, the complete underlying mechanisms are still unknown. Additionally, more physiological evidence supporting these theories is still needed.

1.1.2   Pitch discrimination

Generally speaking, pitch can be evoked by temporal cues or spectral cues. However, the strength of pitch elicited by these cues can be different. Pitch strength is commonly believed to be correlated with the ability to discriminate changes in the pitch value, which is the frequency of a pure tone or the F0 of a harmonic complex tone (Micheyl and Oxenham 2007). The smallest change (difference limen, or just-noticeable difference) that a subject can discriminate is thought to be inversely proportional to pitch strength (a measure of the salience of the perceived pitch).

Therefore, by manipulating the temporal or spectral content of a sound, and measuring the corresponding difference limen, we can test which part of information is more essential for evoking a salient pitch perception. In humans, it has been revealed that harmonic complex tones have a greater pitch strength than a pure tone at the same F0, implying that the fundamental component alone cannot dominate the pitch strength of complex tones. What's more, when the lowest harmonic number of a harmonic complex sound increases up to 5~10, F0DL increases significantly, suggesting that the lower harmonics, which presumably contribute to pitch perception as spectral cues, elicits more salient pitch perception. Further evidence supported the idea that the importance of spectral cues and temporal cues depends on the order of harmonics: pitch strength reduces (F0DL increases) when the lower harmonics are shifted or jittered so that the harmonic pattern is disrupted, and spectral cues are partially lost; pitch strength also decreases when the temporal envelope is flattened when there are minimum spectral cues available. Again, this suggested a "dual-mechanism" for pitch processing.

1.2   Music perception

The human is the only species known to actively produce and listen to music as entertainment. Currently music is even used for therapy in clinical treatments of psychological disorders. Birds also sing "songs" but exclusively for communication, such as establishing and maintaining territories and signaling reproductive quality to mates (Elemans 2014). Thus, it becomes a mystery how the human's musical faculties develop even if they do not serve any obvious functions that are essential for survival from the evolutionary standpoint.

1.2.1 Music perception and emotion

There are many dimensions to manipulate when composing a piece of music, i.e. pitch, tempo, timbre, meter and so on. People have been using music for treating some mental problems, that is, music therapy. Yet, how does the auditory system process these complex sounds, analyze and extract meanings, further modulate the emotional states is one of the most interesting myth in audition. Among all these dimensions of music, the most important one might be pitch, at least in western music. In music theory, one of the most important rule to follow when composing is to balance the relief and tension created by different pitch combinations. The happiness-related pitch combination is also termed consonance. On the contrary, unpleasant-related pitch combination is dissonance. In this thesis, I will mainly focus on the "consonance and dissonance" dimension of music appreciation, for that it serves as the bridge between pitch, one of the most fundamental and best-understood aspects of musical sound, and the highest level of human cognition (emotions).

1.2.2 Concept of consonance and dissonance

The concept of consonance and dissonance could date back to the 6th century B.C. when Pythagoras first discovered that the sound produced by plucking a string is related to the length of that string, and that when two strings with different length are plucked together, the combined two notes may sound pleasant or unpleasant. He further found that it sounds better when the ratio of the lengths of the two strings is a small number. Hundreds of years later, Helmholtz (1863) proposed that dissonance is mainly characterized by the "roughness" in the sound, and that roughness is due to the "beating" between overtones

of two simultaneously played sounds when they are so close to each other that fall into the same auditory filter.

Currently, there is no official definition for consonance and dissonance regarding physical characteristics. They are generally defined according to the perceptual aspects: consonance is a harmony, chord, or interval considered stable (sound pleasant to most people), as opposed to a dissonance, which is considered unstable (sound unpleasant to most people). In music theory, different degrees of consonance are classified into three categories (perfect consonance, imperfect consonance, and dissonance), based on the frequency ratios, as well as listeners' perceptual basis (Figure 2). By balancing the usage of different degrees of consonance, composers create emotional contrast and create a piece of dynamically flowing music. Noticeably, consonance and dissonance can be applied either melodically (one constituent tone of the interval heard after the other) or harmonically (both heard at the same time).

1.2.3 *Consonance and dissonance perceptual mechanisms*

In addition to Helmholtz's theory about beating effects that explains roughness, another prevailing theory concerns about "harmonicity" of the combined spectrum. Consonant chords sound more unified—they are more likely to be perceived as one single sound, rather than a rough sound without a clear pitch. For two harmonic complex tones with a simple ratio of F0, there will be more overlaps on the spectrum, thus the overall combined spectrum is more harmonic, i.e. fewer outliers when being matched to a harmonic template.

When asked to rate the degree of pleasantness of an isolated chord, normal listeners tend to rate the consonant chords much higher than dissonant chords (Plomp and Levelt 1965; Kameoka and Kuriyagawa 1969a; Kameoka and Kuriyagawa 1969b). There are also some evidence showing that human infants and some other species also prefer consonance to dissonance (for detailed discussion about preference, see chapter 3). However, the mechanisms underlying consonance perception is still under debate. Recent studies on patients with "amusia" support the idea that harmonicity is more important for perception of consonance, compared to beating theory (Cousineau, Oxenham, and Peretz 2015; Cousineau and McDermott 2012; McDermott, Lehr, and Oxenham 2010). These patients are called amusics—they usually have problems with pitch perception, and they do not show preference for consonance. When tested their preference using chords with temporal beating removed, or harmonicity disturbed, they prefer the non-beating chord to beating chord, but do not prefer harmonic chords over the inharmonic ones. The results suggest that these patients may lack the ability of processing harmonicity of a sound, thus have problems in pitch perception, as well as discriminating consonance with dissonance.

2   The marmoset monkey as an animal model

2.1   Advantages of marmosets in neuroscience research

The common marmosets (*Callithrix jacchus*), living in the northeastern coast of Brazil, is a type of New-World monkeys. Comparing to rodents that share a common ancestor with primates approximately 90 million years ago, the separation between New World and Old World monkeys occurred only about 40 million years ago. A closer evolutionary

relationship to human-beings is crucial, as neuroscience investigations are ultimately aimed at understanding human cognitive processes.

In addition to the marmoset's evolutionary advantage over non-primate species, they also have some advantages compared to other non-human primates. First of all, they are small and easy to handle (~400g, approximately the same size as rats); Secondly, their gestation period is about 5 months and postnatal development is about 18 months (Burkart and Finkenwirth 2015), which not only improves the rate of production, but also provides convenience for developmental studies. In addition, as new genetic techniques being developed, the marmoset is rising as an attractive animal model for applying more advanced and efficient techniques such as two-photon calcium imaging and optogenetics, to probe brain mechanisms underlying behaviors. Finally, but importantly, their cortical organization is very similar to that in humans. Besides, the cortical surface of the marmoset is flat, which makes it is relatively easy to target almost every cortical area with electrophysiological and optical methods. However, behavioral studies of the marmoset are lacking. Therefore, it is important to characterize this species' perceptual capacities to provide guides for further research using marmosets in neuroscience.

2.2   Characterization of marmosets' auditory features

As an important animal model in auditory research, several fundamental properties of the auditory system of the common marmoset have been characterized, including the structure of the cochlea (Johnson, Santina, and Wang 2012), hearing range and sensitivity (Osmanski and Wang 2011), the peripheral frequency resolution (measured as equivalent rectangular bandwidth of the auditory filters) (Osmanski, Song, and Wang 2013),

vocalization features and functions (DiMattina and Wang 2006), and their ability to perceive pitch (Bendor, Osmanski, and Wang 2012). Many of these aspects are similar to those found in humans. In addition, the organization of the auditory cortex has also been extensively investigated. Particularly, the discovery of pitch-encoding neurons (Bendor and Wang 2005) that provides neurophysiological evidence for pitch processing, promotes the motivation to study pitch perception, and other pitch-related cognitive functions in this species, such as music perception, vocalization, auditory object grouping and so on. Thus, it is necessary to further investigate their auditory functions and higher cognition capacities, to promote our understanding of the auditory system working principles.

3   Thesis outline and objectives

In this thesis, questions about musical pitch perception on different levels will be addressed.

The first part (Chapter 2) asked the question whether marmoset monkeys have the fundamental capacity of perceiving and discriminating musical pitch. The basic auditory functions may vary from species to species. The common marmoset, a highly-vocal New-World monkey with a well-developed auditory system and close evolutionary relationship to humans, has been demonstrated to have similar basic hearing faculties to humans, including hearing range and pitch perception abilities. However, to what degree can they distinguish different pitches, is an essential ability in music perception, has not been thoroughly characterized. Here I will start with simple pure tones to measure their

frequency resolution behaviorally across the entire hearing range. Next, I will investigate how well they can tell apart different pitches of harmonic complex tones, which is the most common type of sound in music, usually produced by musical instruments or human voice. Then I will compare these behavioral data with those on other species, also with some other auditory perceptual abilities of marmosets, with an endeavor to explore how these discrimination abilities help us better understand the evolution of auditory system, as well as the mechanisms underlying pitch perception.

The second part of my thesis (Chapter 3) explored a higher-level cognitive question : do marmosets prefer consonance over dissonance? By manipulating different pitch combinations, composers can generate consonant or dissonant harmonies or melodies for expressing various emotions. Humans prefer pleasant consonance over unpleasantness or tension-evoking qualities of dissonance, which may underlie our enjoyment of music, although the perceptual mechanisms for this preference are unknown. Assuming marmosets' auditory capacities are sufficient for discriminating pitch in music, do they share the same emotional responses to combinations of pitches with humans? This question is not only important for the search of origin of music, but also helps to answer whether our appreciation for music is innately hard-wired in the brain, or acquired through exposure to music in the environment. This question can be partially answered in humans by the universality of music, even in remote tribes isolated from the rest of the world.

Chapter 2: Pitch Discrimination by Marmoset Monkeys

1 Background

1.1 Pure-tone pitch discrimination

Pure tones are the building blocks of sounds. The Fourier Transform is based on the theory that any complex waveform can be produced by summing pure tones of different amplitudes, frequencies and phases. From the computational perspective, the basilar membrane's job is to execute a continuous Fourier transform of short segments of the sound wave, decomposing it into those "building blocks" by a bank of auditory filters. Thus, an investigation of the frequency resolution of the auditory system is essential for understanding other auditory perception mechanisms.

1.1.1 Models of pure-tone FDL

Perceiving and discriminating pure tones are not simple cases. It is believed that a pure tone generates a percept of pitch, the value of which equals the frequency of that tone (Plack, Oxenham, and Fay 2006). It is also demonstrated that pure tone discrimination varies with frequency, duration and sound level in a complex way (Moore 1973; Wier, Jesteadt, and Green 1977; Moore and Ernst 2012; Nelson 1983; Micheyl and Oxenham 2007). Researchers also proposed mathematical models and tried to find the relationship between FDL and these parameters.

Traditionally people plot ΔF or $log\Delta F$ against $logF$, where ΔF represents frequency

difference limen (FDL), and the plot usually looks nearly linear. Wier (1977) first

proposed a simple equation used to relate FDL with frequency:

$$\log \Delta F = a\sqrt{F} + b$$

It turns out that this simple equation fits the FDL data pretty well, with R^2 value range

0.88~0.97, and the slope of the straight line depends on sensation level of the sound. In

order to incorporate sensation level as another parameter in the model, Nelson (1983)

proposed another equation wherein the base-10 logarithm of the FDL (in Hz), is predicted

as a linear combination of the square-root of frequency, f (in Hz), and of the reciprocal of

the sensation level, s (in dB)

$$\log_{10}[d(f,s)] = a\sqrt{f} + bs^{-1} + c$$

The most quantified and most thorough study to date is conducted by Micheyl et al.

(2012). They analyzed data from 12 published studies of pure-tone frequency

discrimination, including 583 FDL measurements from 77 normal-hearing humans, then

derived a relationship between FDL and 3 major parameters—frequency, duration and

sensation level. Conceptually, this model confirmed an inverse-square-root relationship

between log(FDL) and duration, and an inverse relationship between log(FDL) and

sensation level. However, contradictory to previous models, a relationship between

log(FDL) and a power function of frequency with an exponent of about 0.8, rather than

the square-root of frequency best fits the results (Micheyl, Xiao, and Oxenham 2012).

1.1.2   Theories explaining pure-tone FDL

Although there is no single model that could explain all the observations in

psychophysics yet, these models may provide some insights for understanding the

mechanisms of how the auditory system process pure-tone pitch to some degree. As we

can see from the models mentioned above, and from the numerous psychophysical

studies, the most important parameters influencing FDL are considered as frequency,

sensation level and duration of the sound. The observations are summarized as following:

1) FDL and frequency: FDL in Hz (absolute value) usually increases with frequency; 2)

FDL and duration: FDLs shift down as duration increase, and this effect is larger for

frequencies that are below ~4kHz (B. C. J. Moore 1973); 3) FDL and sensation level:

FDL is smaller at moderate to high levels, and the sensitivity to level change is larger for

high frequencies (Wier, Jesteadt, and Green 1977).

Considering the fact that a pure tone elicits a pitch perception, and the two major theories

(place-based and time-based) on pitch perception mechanism, the explanations on pure-

tone discrimination are also based on either the "place theory" or "time theory".

Generally speaking, there are two ways that the pitch of a pure tone could be

discriminated. 1) Place mechanism: a pure tone can generate an excitation pattern on the

basilar membrane, normally a peak excitation at a single point. Thus the FDL could

depend on the change of excitation level on that location, or depend on detection of the

shift of peak excitation point. In either case, frequency discrimination is predicted to

depend on the slope of the excitation pattern and on the smallest detectable change in

excitation level. 2) Temporal mechanism: by detecting changes in the temporal pattern of

phase-locked firings in the auditory nerve, or higher level auditory stages, two different

14

pitch can be discriminated. It seems these two mechanisms work together so that most of the psychophysical phenomena can be explained.

Explaining frequency and FDL: To our current knowledge, the human's ability to discriminating pure-tone pitch reaches a maxima in the middle part of the hearing range. That is, as frequency increases, the relative FDL decrease first, then starts to increase once more. This turning point was estimated to be 4-5kHz (Moore 1973). If we only consider place theory, FDL should be correlated with the width of "auditory filters" on the basilar membrane, which is quantified by the equivalent rectangular bandwidth (ERB). ERB is estimated to have a linear relationship with frequency that can be represented as $ERB = 24.7 \times (4.37F + 1)$, where ERB has a unit of Hz, F has a unit of kHz (Glasberg and Moore 1990). That means the frequency resolution is linearly related to frequency, thus the relative FDL should keep relatively unchanged across the whole hearing range. However, Moore's data suggested temporal cues must play an important role in FDL. It is believed that the dramatically deteriorated performance after 4-5kHz is due to the loss of phase locking, thus a loss of temporal information. On the other hand, the fact that our perception of musical melody and our ability to recognize musical intervals breaks down above 4-5kHz Hz is further evidence supporting the notion that temporal cues are responsible for pitch perception in low frequency range. Latest work suggested this point is actually higher, around 8kHz, rather than 4-5kHz (Moore and Ernst 2012). This is supported by the data showing that relative FDL increases from testing frequency of 2kHz, and then reaches a plateau from 8kHz to 14kHz. This might be a more reasonable way to infer the turning point for encoding strategy from temporal-dominant to spectral-dominant.

15

Explaining duration and FDL: Assuming temporal cues to be the most deterministic factor on pitch discrimination, the relative FDL should be related to the number of periods of the pure-tone stimulus. In that case, as the duration of sound increase, the smaller FDL should be. This idea is supported by the data in Wier (1977), which showed that longer duration leads to lower FDL, and that the influence of duration is larger in low frequency range than high frequency range, with a turning point around 4-5kHz. It again suggests that after the turning point, sound duration does not have much effects on FDL because of the lost phase-locked temporal information.

Explaining sound level and FDL: Generally the FDL at low sensation level is higher than that in moderate to high sensation level. In addition, as the sensation level increase, the rate of FDL decrease becomes slower (Wier, Jesteadt, and Green 1977). On one hand, it is believed that phase locking is more salient as sound level increase, thus the less temporal information at low sensation level condition explains the relatively high FDL. On the other hand, considering the mechanical property of the basilar membrane, as sound intensity increases, the excited region by a pure tone also increases, which makes discriminating two different excitation patterns more difficult, leading to a higher FDL. Therefore, at high sensation level, the combined effects of increased excitation region and stronger phase-locking influence FDL in a contrary way. This may lead to a slower decreasing rate of FDL, even an increase of FDL. The idea also fits into the mathematical model where log(FDL) has a linear relationship with reciprocal of sensation level, so that FDL shows a rapid decline at low levels and a shallower decline at high levels (Nelson 1983).

1.2   Harmonic complex tone pitch discrimination

When presented with a harmonic complex tone, such as a note on the piano, that has a fundamental frequency component (F0) and several other components with frequencies that are multiple integers of F0, we generally perceive a single "global pitch" that equals F0 (Plack, Oxenham, and Fay 2006). From human psychophysics, we know that pitch discrimination ability is better when tested with harmonic complex tones, compared to pure tones, indicating that the global pitch is more salient than pure-tone pitch (Henning and Grosberg 1968; Zeitlin and Cramer 1957). Further evidence supported that the lower harmonics mainly contributes to the increased pitch salience, which is defined as "dominance region".

1.2.1   *Definition of resolved and unresolved harmonics.*

What is special about lower harmonics is that each frequency component falls into an individual auditory filter, that is, resolved harmonics. The harmonics are equally spaced on the acoustic spectrum, whereas the peripheral auditory filters have increasing bandwidth as the center frequency increase; the lower harmonics can be separated by different filters, while two or more higher harmonics may fall into the same filter and generate beating effects in the temporal pattern of that filter output.

Despite that the shape of an auditory filter can be modeled in different ways, the simplest method of defining an auditory filter bandwidth is by calculating the equivalent rectangular bandwidth (ERB), which can be measured behaviorally by masking effects, and is linearly related to the center frequency (Glasberg and Moore 1990). Although the boundary between resolved and unresolved harmonics is vague based on the excitation patterns, the rough relationship is that harmonics with spacing larger than 1-1.25 times of

17

ERB are resolved (Ohgushi 1993; Carlyon 1994). The most direct way to define "resolvability" is to consider the perceptual separation: individual resolved harmonics can be "heard out". The number of resolved harmonics is estimated to be around 5-10 (Bernstein and Oxenham 2003; Ohgushi 1993).

1.2.2 Dominance region

Human psychophysical evidences supported that the resolved harmonics dominant our perception of pitch, that is, the more resolved harmonics in the harmonic complex tone, the more salient the pitch is. This was demonstrated by increasing the lowest harmonic number in a harmonic complex tone, and measuring the relative F0DL in humans (Bernstein and Oxenham 2003). It was shown that when the lowest harmonic number increased above 10, the relative F0DL increased dramatically. Unpublished data in our lab also showed that when testing marmosets on discrimination task with harmonic complex sound containing only the unresolved harmonics, F0DL is much higher than that measured with full harmonics or resolved harmonics. It suggests that resolved harmonics not only dominant humans', but also marmosets' pitch perception.

2 Methods and experimental designs

2.1 Behavioral paradigm and subjects

The basic operant task and apparatus were described before in a previously published study (Remington, Osmanski, and Wang 2012). During testing, marmosets were seated in a Plexiglass restraint chair mounted in the center of a single-walled sound isolation chamber (Industrial Acoustic, Model 400A) lined with 2-inch acoustic absorption foam

(Pinta Acoustics, model PROSPEC). Sounds were played through a loudspeaker (Tannoy, model Arena) mounted 40cm away in front of animal's head. All sound signals were generated using a customized Matlab program (Mathworks) and delivered at a nominal sampling rate of 100kHz through a multi-processer DSP unit (Tucker-Davis Technologies, RX6), followed by a programmable attenuator (Tucker Davis Technologies, PA5), and an audio amplifier (Crown Audio, Model D-75A).

In the current study, two types of stimulus were presented: pure tones (PT) and harmonic complex tones (HC). Repeating "background" sounds were presented, and the animal's task was to detect the "target" tones with frequencies or fundamental frequencies higher than the background tones. Testing conditions were classified based on the frequency (F) of the background in pure-tone testing, or the fundamental frequency (F0) of the background in harmonic complex tone testing. For easy description, in the following section I use the format "PT-A4" to represent the testing condition of pure-tone discrimination at F = A4 (440Hz) and "HC-A4" to represent testing condition of harmonic complex tone discrimination at F0 = A4 (440Hz). Within one testing condition, the background sound was fixed.

In the pure-tone discrimination tasks, eight different frequencies of the background sounds were chosen based on the music scale, ranging from "A3" to "A10" in terms of note names (corresponding to 220Hz to 28160Hz on a logarithm scale). A4 is also known as A440, the A note above "middle C", is the standard music tuning pitch (ISO 16). The relationship between music note names and frequencies is shown in Figure 1 in the bottom table. The top figure shows a complete piano keyboard and the corresponding positions of note A2~A7 (A8~A10 does not exist on a piano keyboard). The spacing

between adjacent keys is defined as one semitone, or a half step. In a twelve-equal-temperament tuning, one octave is divided into 12 equal parts with widths of a semitone, thus the frequency ratio of the interval between two adjacent notes is the twelfth root of two: $\sqrt[12]{2} \approx 106\%$, and the increment is roughly 6% of the frequency.

Each testing trial had a waiting time from 3 to 10 seconds. During the waiting time the background sound was repeatedly presented to the animal. After this waiting period, a "target" sound, which was always higher in frequency than the background, began to alternate with the background sound. Both the background and target sounds had a duration of 200ms with a 10ms linear ramp (rise/fall time). The inter-stimulus interval was fixed at 300ms. Thus the tones were presented with a 2Hz temporal rate. Animals could respond any time during the period when the target and background were alternating (i.e., the response window), which lasted for 4.8 seconds in total. The subject had to detect the pitch change and respond by licking at a feeding tube placed in front of its mouth during the response window ("hit") to receive food reward. However, if the subject licked before the response window onset, the chamber light was extinguished for 2-5 seconds as a warning signal. If the subject did not respond during the trial at all, a "miss" was recorded and the system automatically started the next trial.

Each experimental session contained at least 100 but not more than 200 trials, in which 70% trials were measuring hit rates on real targets randomly chosen from 7 possible target choices, corresponding to seven different frequency distances from the background sound. These possible F0 changes were equally spaced in the semitone scale. The remaining 30% of trials were sham trials in which no target sound was presented. Sham trials were used to measure false alarm rate as an indicator of how much the subject relied

on guessing during the task. As the deviation of targets from the background decreases, the hit rate drops from nearly perfect 100% to around false alarm rate.

Three adult common marmosets (all male) were used in the current study, ranging from 4 to 6 years old during the testing. Each of them had at least one-year experience in discrimination tasks, either with previous F0DL measurements for different types of stimulus, or vocalization discrimination tasks. The experiments were conducted collaboratively with another two lab members: Xindong Song and Michael Osmanski (X.S. generated the sound files, trained subjects M13W and M4Y; M.O. collected data from M55Y; I collected data from M13W, M4Y). M55Y was not head-fixed during all testing sessions; M4Y was head fixed during all testing sessions; M13W was not head-fixed for the testing conditions PT-A3, PT-A5, PT-A6, PT-A7, PT-A8 and PT-A9, because of some head-cap stability considerations; M13W was head-fixed during all other sessions. The three subjects finished all eight testing conditions for pure tones and four testing conditions for harmonic complex tones, except that M13W has not finished PT-A10, and that 55Y failed to obtain qualified threshold for the condition of HC-A4. In order to eliminate the training effects that may reduce the thresholds, the order of testing was varied from subject to subject, which is listed in Table 2, Table 3, Table 4.

Subjects were housed in individual cages in a large colony at the Johns Hopkins University School of Medicine. All subjects were maintained at approximately 90% of their free-feeding weight on a diet consisting of monkey chow, fruit, and yogurt and had ad libitum access to water. Subjects were tested five or six days per week between the hours of 0900 and 1800. All experimental procedures were approved by the Johns

Hopkins University Animal Care and Use Committee and were in compliance with the guidelines of the National Institutes of Health.

2.2   Data analysis

The measured hit rates for 7 targets were corrected by the equation: *Corrected hit rate = (Raw hit rate – False alarm rate) / (1 - False alarm rate)*. Discrimination thresholds were defined as the frequency difference that the animals correctly identified 50% of the time. Statistics of behavioral sessions were calculated according to the methods used in (Talwar and Gerstein 1999). That is, the corrected hit rate curve was linearly interpolated and the x-value of the point intersecting with 50% line was taken as frequency difference limen (FDL) for PT conditions, or fundamental difference limen (F0DL) for HC conditions.

The criteria for qualified sessions are: 1) At least 100 trials were finished for each session; 2) The maximum difference among thresholds measured from these sessions was smaller than the spacing between targets; 3) The false alarm rate of each session was smaller than 30%; 4) The total number of trials for each testing condition from one subject was at least 520, which means at least 52 repetitions for each target, and at least 156 sham trials; 5) The corrected hit rate curve did not pass below 50% multiple times.

Figure 5 to Figure 10 show the psychometric curves of individual sessions of pure-tone discrimination and harmonic complex tone discrimination tasks, for the three subjects, respectively. Testing continued until at least three consecutive experimental sessions

meet the criteria described above. The detailed experiment records are listed in Table 2, Table 3 and Table 4.

2.3   Stimulus design

For PT discrimination, the background level was calibrated to be around 40dB SL (~70dB SPL). Targets were adjusted in level to match the sensation level of the background, based on the marmoset audiogram (Osmanski and Wang 2011), to eliminate level differences as a potential cue. For HC discrimination, the maximum level of harmonics was calibrated to be around 50dB SPL. A consideration for pitch discrimination using HC tones is that our subjects could use spectral edges as a cue to perform this task. To minimize this possibility, we implemented roll-offs on the upper spectral edges, a similar method used in a human psychophysics study (Moore, and Moore 2003). The amplitude of the spectral envelope at a given frequency F within the sloping region was determined by a frequency variable x defined as $x = 1 - |(F - F_e)/1.5F0|$, where F_e is the frequency at the edge of the flat region ($F_e = 28.16kHz$). The amplitude relative to that in the flat region was set to $(10^x - 1)/9$. The amplitude was set to zero when x was less than or equal to zero. Background sound level was randomly roved within ± 3dB. Target sound level was always fixed. All sound levels were calibrated using a 1/2" free-field microphone (Brüel and Kjaer, type 4191) positioned at the same location as the animal's head, with a customized program.

3 Results

3.1 Pure-tone pitch discrimination

Figure 3 summarizes the frequency difference limen (FDL) for pure-tone pitch discrimination. The data show that: 1) the absolute FDL increases almost monotonically from 36.49Hz to 1.50kHz as the testing frequency increases from 220Hz to 28kHz, although there is a decrease of FDL between A6 (1.76kHz) and A7 (3.52kHz); 2) the relative FDL decreases from ~2.65 semitones until it reaches minimum value of ~0.42 semitones at 7.04kHz, then slightly increases as the testing frequency goes higher. The error bars in Figure 3 denote the standard deviations of thresholds measured in different sessions. Noticeably, the variation is also smaller when FDL get smaller. In addition, the range of test frequency with lowest relative FDL overlaps with the most sensitive region in the marmoset's audiogram (Osmanski and Wang 2011), as well as with the fundamental frequency range of their typical vocalizations, which is around 7kHz (DiMattina and Wang 2006). Detailed information for each data point is shown in Table 1. Mean values and standard deviations of FDL are presented in terms of absolute value in Hz, relative value in semitones, and relative value in percentage. Relative FDL values were calculated by dividing the absolute value of FDL (in Hz) by the test frequencies. These results revealed auditory perceptual capacities of the marmoset and help guide further studies of auditory behaviors of this species.

3.2 Harmonic complex tone pitch discrimination

We also tested these marmosets' discrimination abilities using harmonic complex tones with F0 of 110 Hz, 220 Hz, 440 Hz and 880 Hz (music notes A2, A3, A4 and A5), in order to compare their ability of pitch discrimination using pure tones and harmonic complex tones, and to test if fundamental frequency difference limen (F0DL) change with F0 in the same fashion as FDL change with frequency. As shown in Figure 4, the absolute F0DL exhibits a roughly increasing pattern from ~11.08Hz to ~32.95Hz, with a minimum value of 10.84Hz at testing F0=440Hz. The relative F0DL first decreases and then increases, with a highest value of ~1.66 semitones at 110 Hz and a lowest value of ~0.42 semitones at 440 Hz, as shown in Figure 4 in which the error bars represents standard deviation. Again, the variation across different sessions is smaller as F0DL gets smaller, suggesting a more reliable performance. Statistical numbers are listed in Table 1.

Compare to FDL of pure tones, the trend is similar. However, the minimum value was obtained at different pitch value (here, pitch value is simply defined by frequency of a pure tone, or fundamental frequency of a harmonic complex tone). Relative FDL reaches a minimum at test frequency 7.04kHz, while relative F0DL reaches minimum at F0=440Hz. In addition, at the same test frequency, FDL is much higher than F0DL, as shown in Table 1.

4 Conclusion and Discussion

The results of FDL and F0DL characterized the marmoset's auditory capacity of pitch discrimination using pure tones and harmonic complex tones. However, it is not clear

what mechanisms underlying pitch perception could explain the trend that FDL and

F0DL change with frequency. In the following discussion, I will compare the marmoset's

data with other species, along with their own fundamental auditory characteristics, so as

to find possible explanations for these results.

4.1 Model for relationship between FDL and frequency

In psychophysics, Weber's law is often used to describe the phenomenon that resolution

of perception diminished as the magnitude of stimulus increase. It is said "the just-

noticeable difference between two stimuli is proportional to the magnitude of the

stimuli". The way that FDL changes with frequency seems do not obey the Weber's law

completely in the whole frequency range. The simple model Wier (1977) proposed was

demonstrated to fit the FDL data of humans pretty well. In addition, the model fits

humans' ERB data even better with a high R^2 value. Here I used the same equation

$log\Delta F = a\sqrt{F} + b$ fit the FDL data from marmosets, as well as the equivalent

rectangular bandwidth (ERB) measured on marmosets by Osmanski et al. (Osmanski,

Song, and Wang 2013) The data points and fitting results are shown in Figure 12. It turns

out this model provides a reasonable fit for marmosets' FDL data ($R^2 = 0.9241$),

although not so good as humans' FDL ($R^2 = 0.98$ when measured at SL=40dB on

humans). The ERB fits better with the model ($R^2 = 0.9725$).

The comparable goodness of fitting with human data suggest that marmoset's FDL vary

with frequency in a similar way as humans do. Although the underlying mechanism is

still unknown, the two species may have similar mechanisms of pure-tone pitch

perception. However, we did not test how does FDL change with duration and sensation level of the sound, it is not clear whether these trends are consistent with humans.

4.2 Comparison of FDL for different species

Here, I aligned our marmosets' data with a dataset of FDLs from different species, including gerbils, squirrel monkey, owl monkey, blue monkey (*Cercopithecus*), macaque monkey, chimpanzee and human. Some of the data are available in the original paper, listed as numbers in a table. For those that are not available as exact numbers, I extracted the data by digitizing the figures in the original paper, or the figure in Wienicke's study on squirrel monkey (Wienicke, Hausler, and Jurgens 2001), in which he also summarized some of the previous data. The source of these data points are listed in Table 5, and all the data are plotted in the same figure (Figure 11). FDL is expressed in both absolute value and relative value. Digitized figures were used to extract the numbers only when the raw data were not accessible, and the accuracy can reach up to the second decimal place.

From the absolute values of FDL, we can see the similar trend that FDL generally increase with test frequency. Comparing across species, humans have the lowest FDL, chimpanzee (blue) and old world monkey (green) come next. New world monkeys (warm colors) have larger FDL compared to other primates. Data from current study are shown in the thick red line. For test frequency below 4kHz, marmoset's FDLs are closer to rodent (gerbil, gray), while above 4kHz FDLs become relatively smaller and closer to old world monkeys. This trend is also shown in the relative FDL plots (bottom, Figure 11). In addition, several other species (human, macaque monkey, squirrel monkey) also show the

27

same trend in relative FDLs that decrease first, then slightly increase as the test frequency goes from low to high. Generally speaking, the frequency with lowest relative FDL is smallest for humans, then chimpanzee, old world monkey, new world monkey and then rodent, consistent with an evolutionary order. It may suggest that the frequency range with highest resolution is becoming lower and lower from the evolutionary perspective. The similar trend was observed by comparing the audiograms across species that the whole hearing range as well as the most sensitive frequency region is shifting down. The results in this current cross-species comparison could provide more insights into the evolution of hearing.

4.3 Relationship between FDL and other auditory features

4.3.1 FDL and audiogram

The results of relative FDLs follow the same trend as the audiogram of marmoset monkeys. Shown in Figure 13 is an overlaid plot of marmosets' FDL on a semitone scale and their audiogram interpolated at the test frequencies. The audiogram was characterized by Osmanski and Wang (2011). The consistency with audiogram indicates that the frequency resolution is best at the most sensitive frequency region. This region also overlapped with the range of center frequency of the fundamental component in their primary vocalizations, especially the long distance 'phee' call (believed to be a species-specific contact call) that has a power peak at ~7kHz (DiMattina and Wang 2006). Noticeably, the sound intensity was compensated based on the audiogram so that all

sounds were presented roughly at the same sensation level. Thus sound intensity is not an influencing factor for a lower FDL at the frequency range with highest sensitivity.

4.3.2 Phase locking

The ability of phase locking for the auditory nerve has been measured on many species including barn owl, pigeon, starling, chicken, cat, guinea pig, and so on (Köppl 1997). However, electrophysiology recordings from auditory nerves of marmosets are not available yet. Roughly speaking, the phase locking ability (quantified with vector strength) at auditory periphery begins to drop below 50% of maximum at around 2~4kHz, and this value does not differ very much across different species. Comparing to the FDL profile, 3.5kHz is approximately the point where relative FDL value started to stop decrease, and slightly increase. It is possible that this turning point is due to the loss of phase locking at the peripheral auditory system. In addition, moving from periphery to central stages, temporal contributions can only be reduced, and the auditory cortex was estimated to have a synchronized response boundary of ~40Hz. Therefore, temporal information stops to contribute to the discrimination of two pure tones above ~3.5kHz.

4.3.3 Auditory filter bandwidth

For human frequency discrimination tasks, people have stated that FDL is a constant fraction of the critical band (CB), to support the idea that pitch of a pure tone is closely related to the excitation pattern produced by that tone on the basilar membrane, i.e. the place theory. However, a study in which the intensity cues were minimized showed that the ratio was not a constant, but rather decrease as the frequency increased (Moore 1974). The results are listed in Table 6.

Here, I also aligned the critical band and FDL data from marmosets to compare their relationship (Figure 14). The equivalent rectangular bandwidth (ERB), which is a parameter that quantifies the auditory filter bandwidth, was measured on marmosets (Osmanski, Song, and Wang 2013). Figure 14 shows the overlaid plot of relative ERB and relative FDL (relative values are calculated by dividing the absolute values by the testing frequencies). ERB values were measured at frequencies 500Hz, 1kHz, 7kHz and 16kHz and interpolated at the test frequencies from 220Hz to 28kHz. Y-axis is the relative numbers, in units of percentage.

First of all, the ERB curve is always above the FDL curve in the whole frequency range, indicating that it is unlikely to perform a pitch discrimination task based only on place cues of the excitation pattern. What's more, unlike human CB/FDL ratio, which decreases with frequency, the ERB/FDL ratio for marmosets decreases first and then starts to increase between 1.8kHz to 3.5kHz. Notice the different patterns of critical bandwidth between humans and marmosets: relative CBs for humans are almost constant, while marmosets have a drop-and-rise pattern on ERB profile, with a relatively small value at 7kHz (Osmanski, Song, and Wang 2013). The similar drop-and-rise profile in ERB/FDL ratio could be partially explained by the ERB characteristics on marmosets. However, given that the ERB/FDL ratio is also not a constant, place theory alone cannot explain pure tone pitch perception. The larger ratio between ERB and FDL in the lower frequency region may be contributed by temporal cues, while the reasons for increased ratio in higher frequency regions are not clear. Lastly, even though the peripheral auditory filter bandwidths have similar values for humans and marmosets, humans have much better pure tone pitch discrimination abilities than marmosets, which suggests that

some other central mechanisms may be more essential for performing pure-tone pitch discrimination tasks.

4.3.4 Head-related transfer function

It is interesting that in the high frequency region, FDL still remains low, while there are presumably fewer temporal cues available. Another consideration for this phenomena is that head-related transfer function (HRTF) may play a role. HRTFs on marmosets were measured by Slee and Young (2010). In the high frequency range, there are much larger fluctuations at the perceived sound level (sensation level, SL) compared to the low frequency range. For example, when the testing frequency is 28.16kHz, deviating 1 semitone from that frequency also generates a deviation of ~2 dB SL. In the most extreme case, if the test frequency falls into the large notch of HRTF profile that is believed to be an important cue for sound localization, 1 semitone difference in frequency could generate a ~20dB difference on sensation level (Data was estimated by digitizing the figure in the original paper, and calculated from pixel numbers). This may have facilitated animal's performance, especially for those who were head-fixed. Although in the current experiments, the sound level has a random value within the range of $\pm$ 3dB, it may not be enough to completely eliminate the possibility that the subjects were using sound level as a cue to perform the discrimination task. Thus the low FDL in high frequency range could possibly be accounted for by sound level fluctuation, especially for test frequencies that are larger than 20kHz where HRTF shows the greatest frequency-dependence. It is also the case in humans, as Henning et al. (1968) reported that when performing the frequency discrimination tasks with random sound levels, the subjects did worse compared to constant sound level condition, and the growth in FDL is

31

larger in high frequency range (Henning and Grosberg 1968). In addition to HRTF, the speaker's frequency response also has a larger variation at high frequencies, which could also contribute to the sound-level cues as the animals were performing a discrimination task. Although increasing the roving range can reduce this kind of artifact, it enhances the difficulty of the task for the animals. Thus, the rove of ± 3dB is a relatively reasonable number, and a trade-off between eliminating sound level cues and reducing task difficulty.

4.3.5 Summary

In summary, it seems unlikely that pure-tone pitch discrimination abilities could be explained simply by a single auditory peripheral feature. Although the FDL pattern is correlated with the audiogram pattern, it does not explain the underlying mechanisms. Therefore, it is reasonable to speculate that even the simplest pure tone generates pitch perception in a complex way that involves central pitch perception mechanisms. Whereas this mechanisms of how pitch is processed along the auditory pathway are still not well-understood. In order to have better understanding pure-tone pitch perception, further investigation is required, such as measuring the phase locking limitation to pure tones on the auditory nerves of marmoset monkeys, and experiments with better controlled sound intensity in terms of sensation level.

4.4 Pitch discrimination ability of harmonic complex tones

4.4.1 Harmonic complex tones elicit more salient pitch sensation

Comparing FDL (Figure 3) with F0DL (Figure 4), it is obvious that F0DL is always lower than FDL when tested at the same pitch value (A3: FDL=2.7 semitone; F0DL=1.1semitone; A4: FDL=2.4 semitone, F0DL=0.4 semitone; A5: FDL=1.7semitone, F0DL=0.6 semitone). This is consistent with findings in psychophysics that harmonic complex tones have a greater pitch strength than a pure tone at the same F0 (Henning and Grosberg 1968; Zeitlin and Cramer 1957).This phenomenon provides further evidence for the possibility that marmoset monkeys may possess the same pitch perception mechanisms as humans.

4.4.2 Resolvability correlates with pitch strength

F0DL were tested at fundamental frequencies 110Hz, 220Hz, 440Hz and 880Hz. Because of the upper limit of hearing ~28kHz, we did not test higher F0s to ensure that there were enough number of harmonics in the harmonic complex tones. Figure 4 Shows that F0DL decrease from A2 to A4, then slightly raised at A5. The best discriminability is found at A4 (440Hz). Further analysis based on their ERB data reveals a correlation between F0DL and number of resolved harmonics (Figure 15). The resolved harmonics here were defined as the harmonics that have a spacing larger than 1 ERB, thus the numbers of resolved harmonics can be estimated based on ERB data, at different F0s. The number of resolved harmonics were estimated to be largest at F0=440Hz (around 14~15 harmonics), and smallest at F0=110Hz (about 3~4 harmonics). Figure 15 shows a negative correlation between F0DL and number of resolved harmonics, with R^2=0.9633. The result suggests

that resolved harmonics increase the pitch strength, thus serves as the dominant components for pitch perception in a harmonic complex sound. This is consistent with the findings in human psychophysics—the dominance region for pitch perception is the lowest 5~8 harmonics (resolved harmonics) (Bernstein and Oxenham 2003).

4.5   Summary

These results characterized marmoset's capacity of pitch discrimination using pure tones and harmonic complex tones. Despite the fact that both FDLs and F0DLs are much larger than those in humans, they are sufficient for human music melody perception and discrimination, as the intervals in music are at least one semitone (mostly bigger than 2 semitones). It was also demonstrated that FDL is dependent on the test frequencies, and that the pattern of dependency is similar to that in humans. These characterizations of basic auditory functions are important for guiding further auditory research in marmosets. In addition, F0DL data suggest that the resolved harmonics dominant their perception of pitch in a harmonic complex tone, which has also been shown in humans. Other studies in our lab (unpublished data) demonstrated that their perceived pitch salience is sensitive to the harmonicity of resolved harmonics, and temporal regularity of unresolved harmonics—another consistent observation with psychophysics. Combining the electrophysiology evidence of pitch neurons found in the primary auditory cortex of marmosets (Bendor and Wang 2005), we have reasons to believe that this species may have the same pitch processing mechanisms as we humans, and may serve as an ideal animal model for investigating the underlying neuronal basis for pitch perception.

Figures and tables for Chapter 2
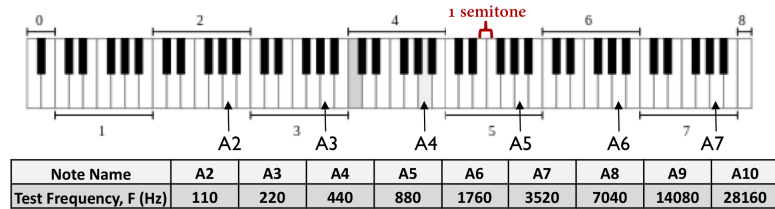
Figure 1. Musical note names and corresponding frequencies.
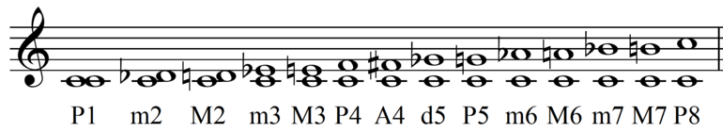
Interval (semitones)	Interval names	Interval Type	FO_2/FO_1
12	P8 (Octave)	Perfect consonance	2:1
11	M7	Dissonance	15:8
10	m7	Dissonance	16:9
9	M6	Imperfect consonance	5:3
8	m6	Imperfect consonance	8:5
7	P5 (Perfect fifth)	Perfect consonance	3:2
6	A4 (Tritone)	Dissonance	45:32
5	P4 (Perfect fourth)	Perfect consonance	4:3
4	M3	Imperfect consonance	5:4
3	m3	Imperfect consonance	6:5
2	M2	Dissonance	9:8
1	m2	Dissonance	16:15
0	P1 (Unison)	Perfect consonance	1:1

Figure 2. All possible music intervals within one octave, the corresponding interval types and frequency ratios.
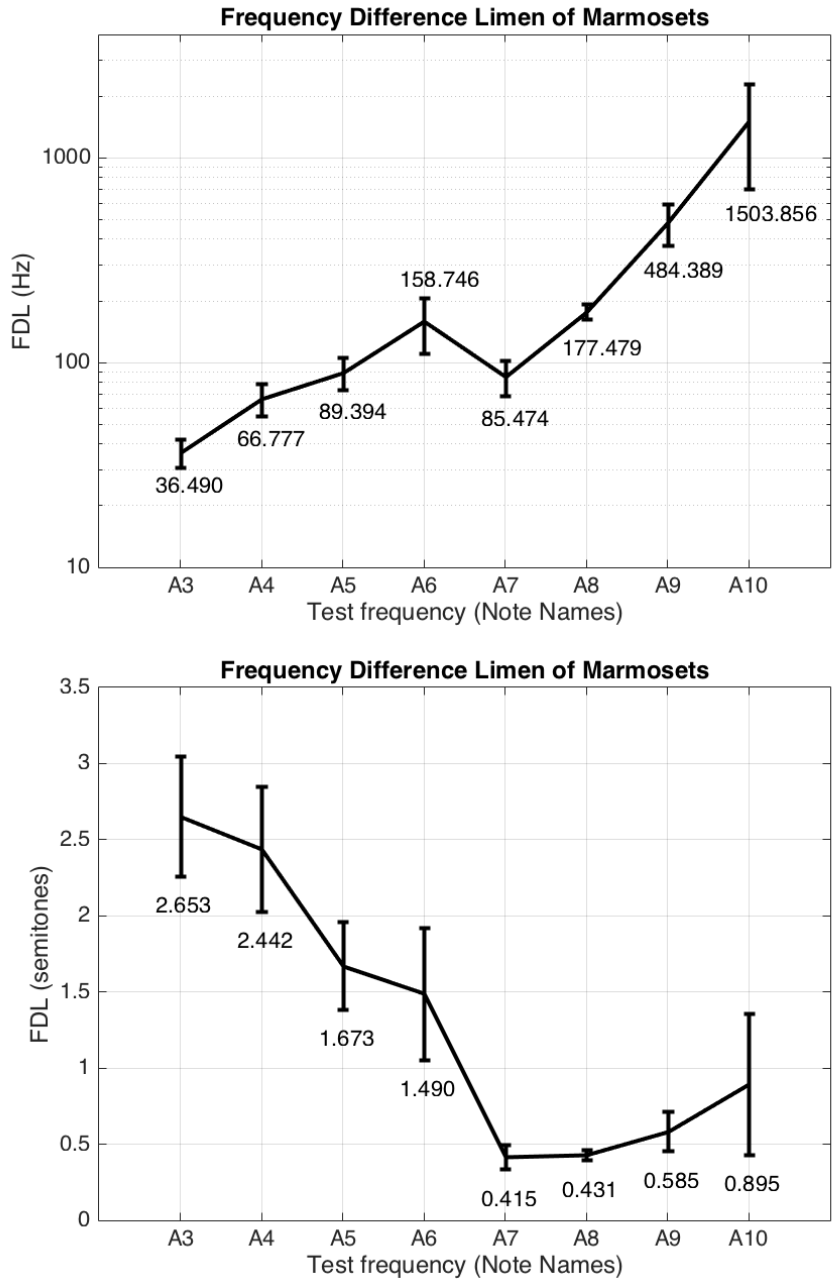
Figure 3. Frequency difference limen (absolute value and relative value) of marmosets.
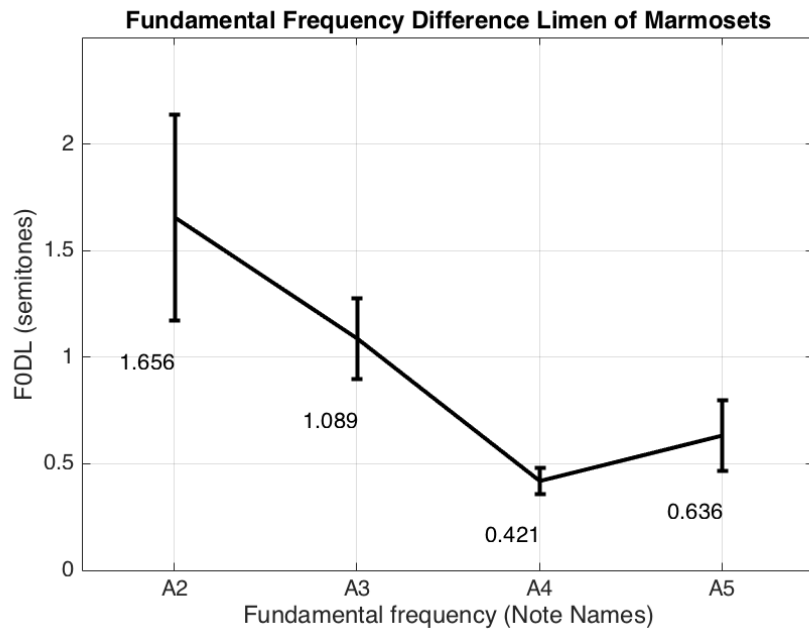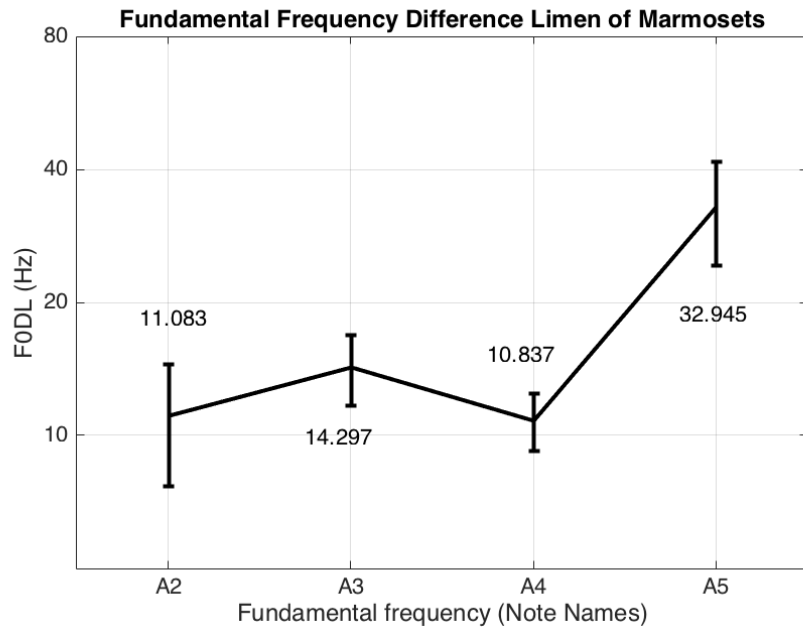
Figure 4. Fundamental frequency difference limen (absolute value and relative value) of marmosets.

Statistical data of discrimination task (pure tones)

Test Frequency, F (Hz)	220	440	880	1760	3520	7040	14080	28160
Note Name	A3	A4	A5	A6	A7	A8	A9	A10
# Subjects tested	3	3	3	3	3	3	3	2
# Trials	10	11	11	11	10	13	10	7
Mean FDL(Hz)	36.490	66.777	89.394	158.746	85.474	177.479	484.389	1503.856
Std. of FDL (Hz)	5.867	12.006	16.099	48.259	16.506	14.933	109.908	795.936
Mean relative FDL (semitone)	2.653	2.442	1.673	1.490	0.415	0.431	0.585	0.895
Std. relative FDL (semitone)	0.393	0.411	0.289	0.435	0.079	0.036	0.131	0.463
Mean relative FDL (%)	16.586	15.177	10.158	9.020	2.428	2.521	3.440	5.340
Std. relative FDL (%)	2.667	2.729	1.829	2.742	0.469	0.212	0.781	2.826

Statistical data of discrimination task (Harmonic complex tones)

Test Fundamental Frequency, F (Hz)	110	220	440	880
Note Name	A2	A3	A4	A5
# Subjects tested	3	3	2	3
# Trials	10	11	7	11
Mean FDL(Hz)	11.083	14.297	10.837	32.945
Std. of FDL (Hz)	3.411	2.595	1.606	8.696
Mean relative FDL (semitone)	1.656	1.089	0.421	0.636
Std. relative FDL (semitone)	0.483	0.192	0.062	0.165
Mean relative FDL (%)	10.075	6.499	2.463	3.744
Std. relative FDL (%)	3.101	1.179	0.365	0.988

Table 1. Statistics of the data from discrimination tasks.

Figure 5. Psychometric curves and response latencies for each session, pure-tone discrimination (subject: M4Y).

Figure 6. Psychometric curves and response latencies for each session, harmonic complex tone discrimination (subject: M4Y).

Figure 7. Psychometric curves and response latencies for each session, pure-tone discrimination (subject: M55Y).

Figure 8. Psychometric curves and response latencies for each session, harmonic complex tone discrimination (subject: M55Y).

Figure 9.Psychometric curves and response latencies for each session, pure-tone discrimination (subject: M13W).

Figure 10. Psychometric curves and response latencies for each session, harmonic complex tone discrimination (subject: M13W)

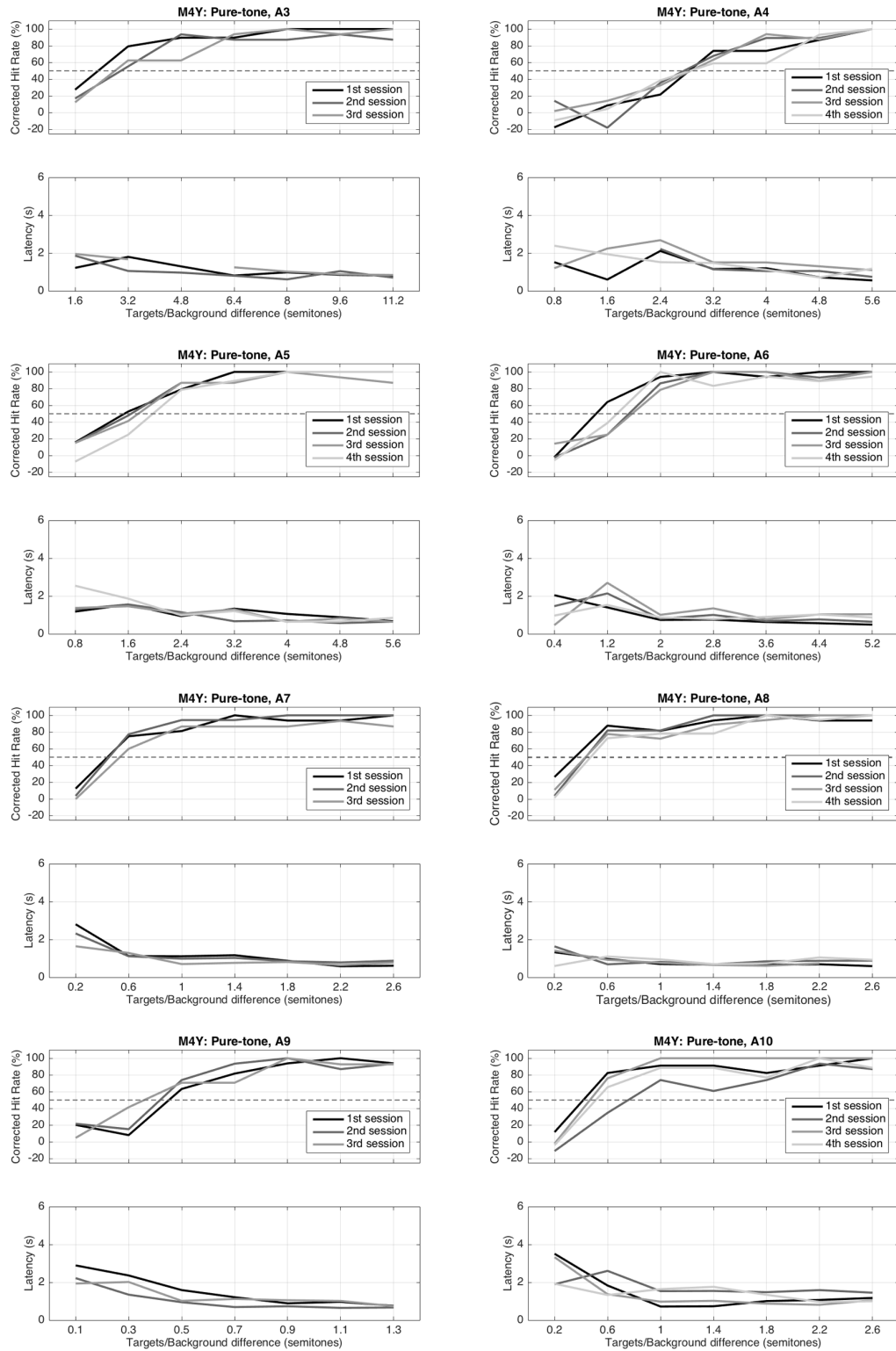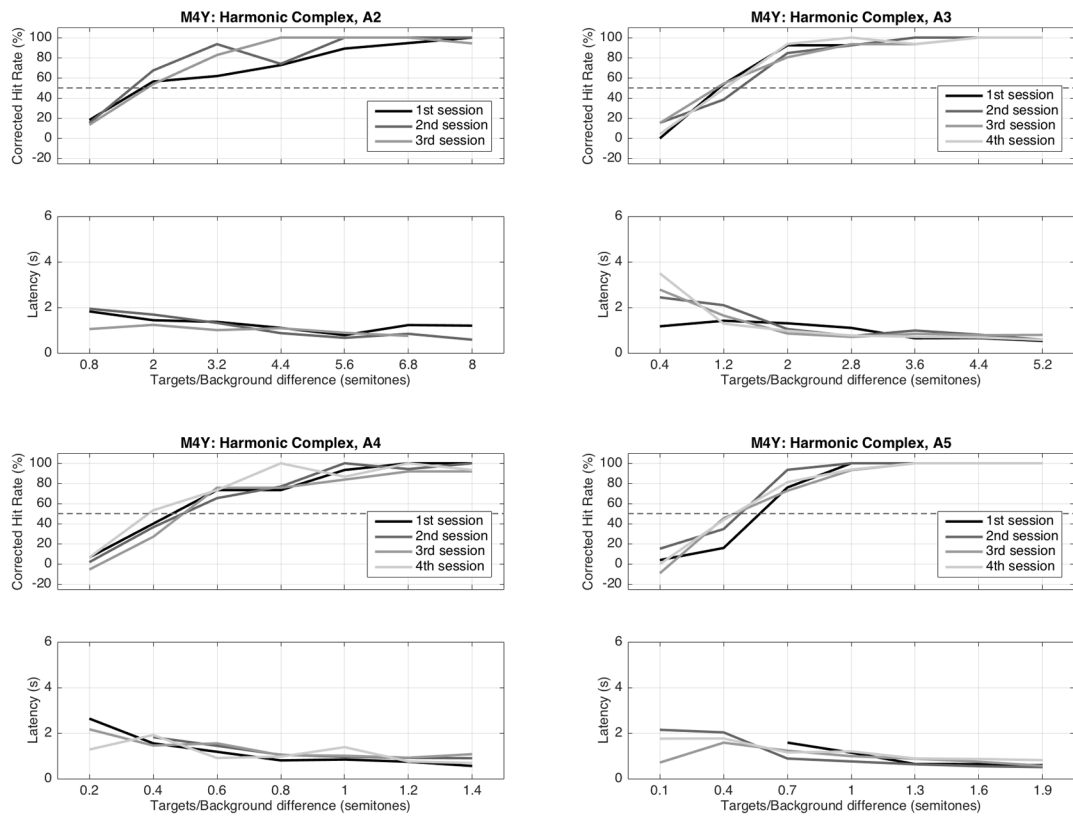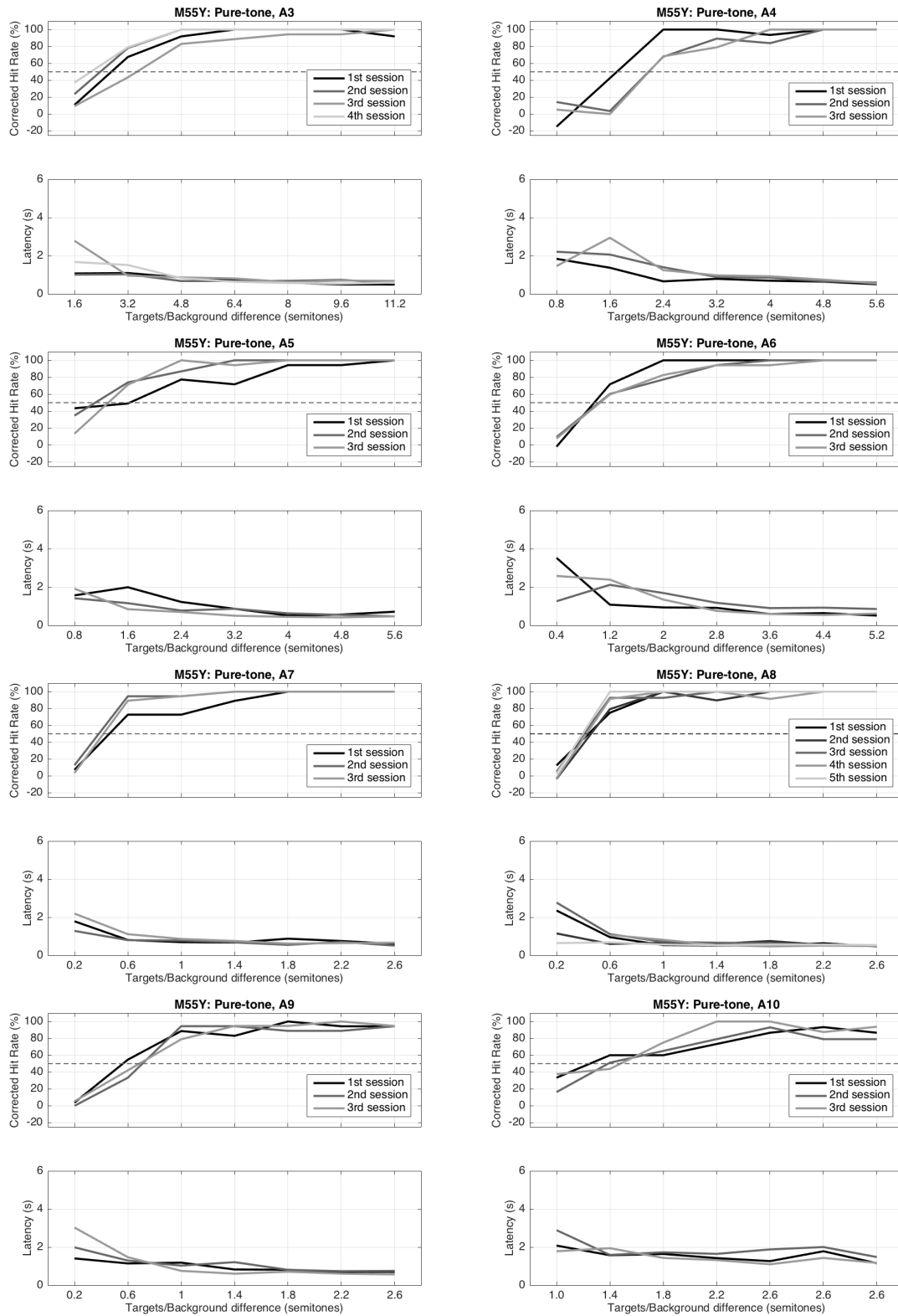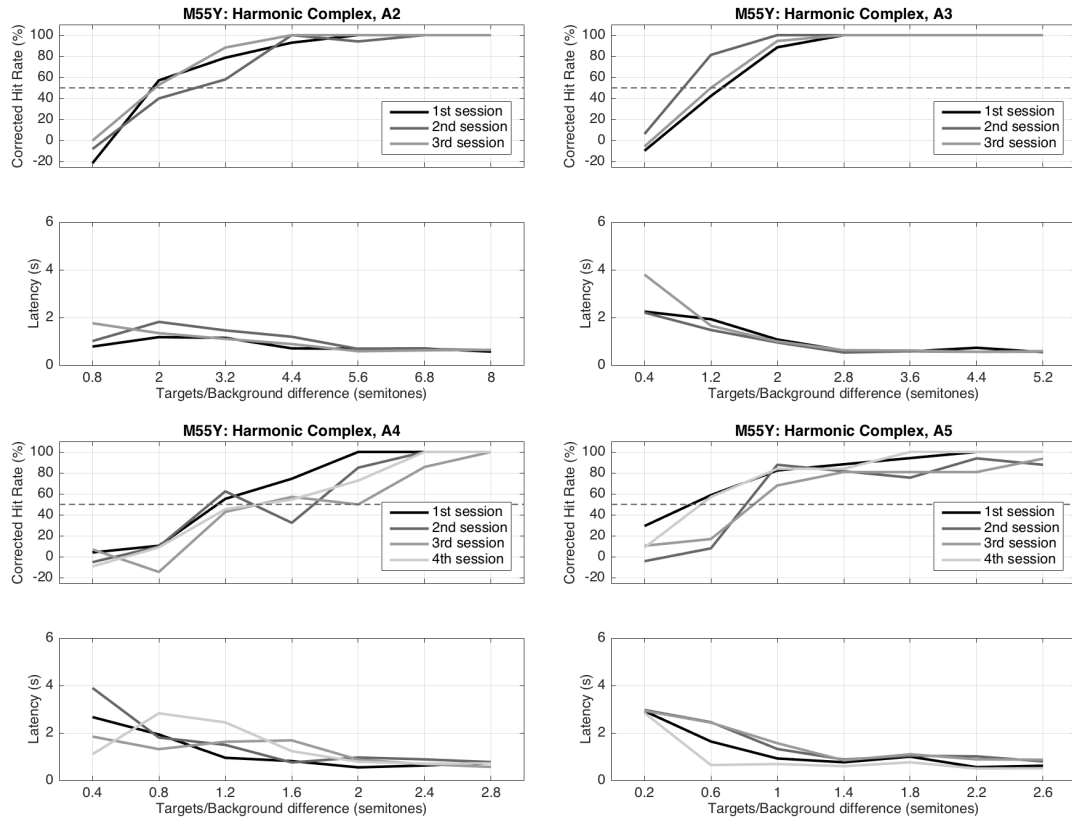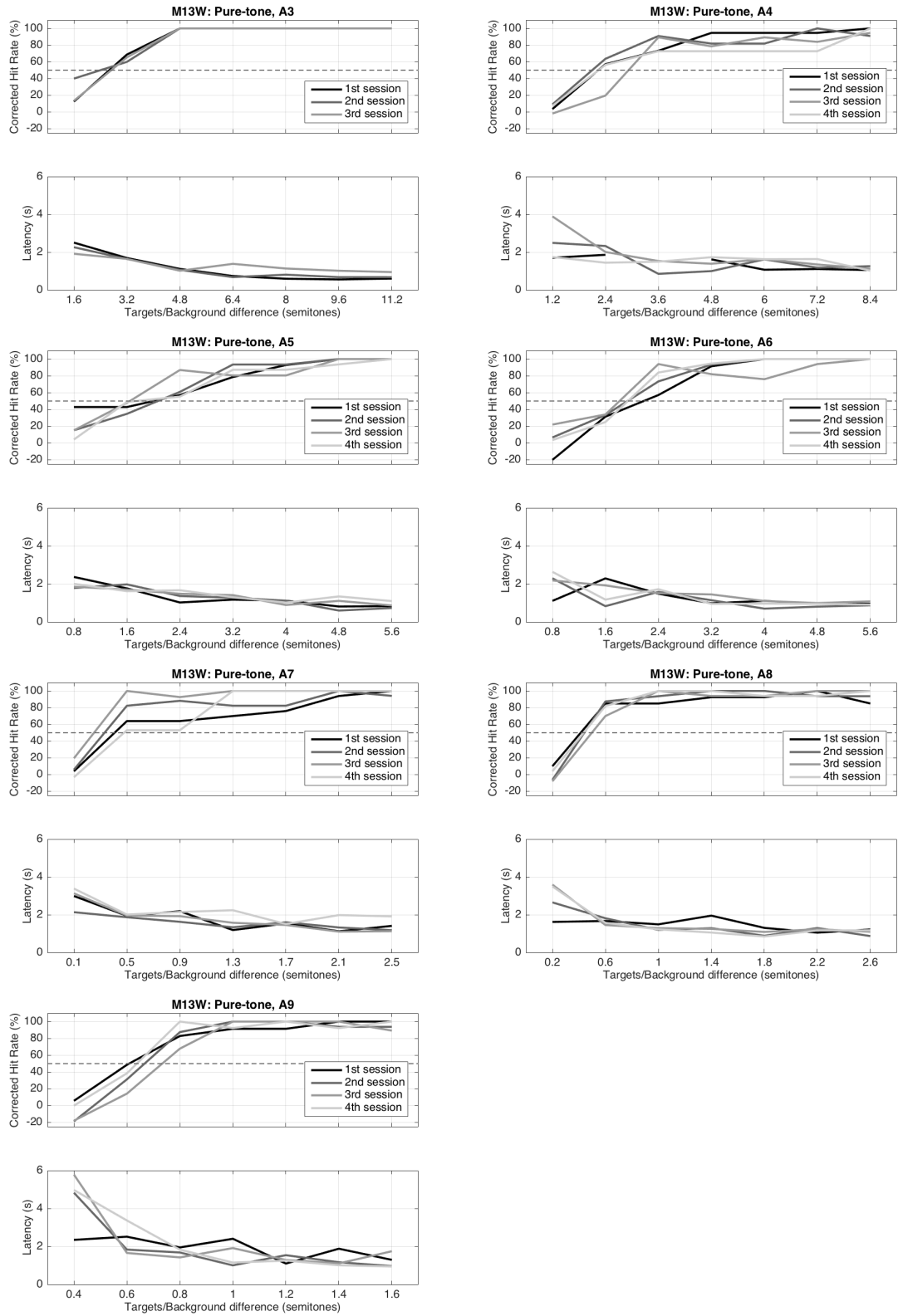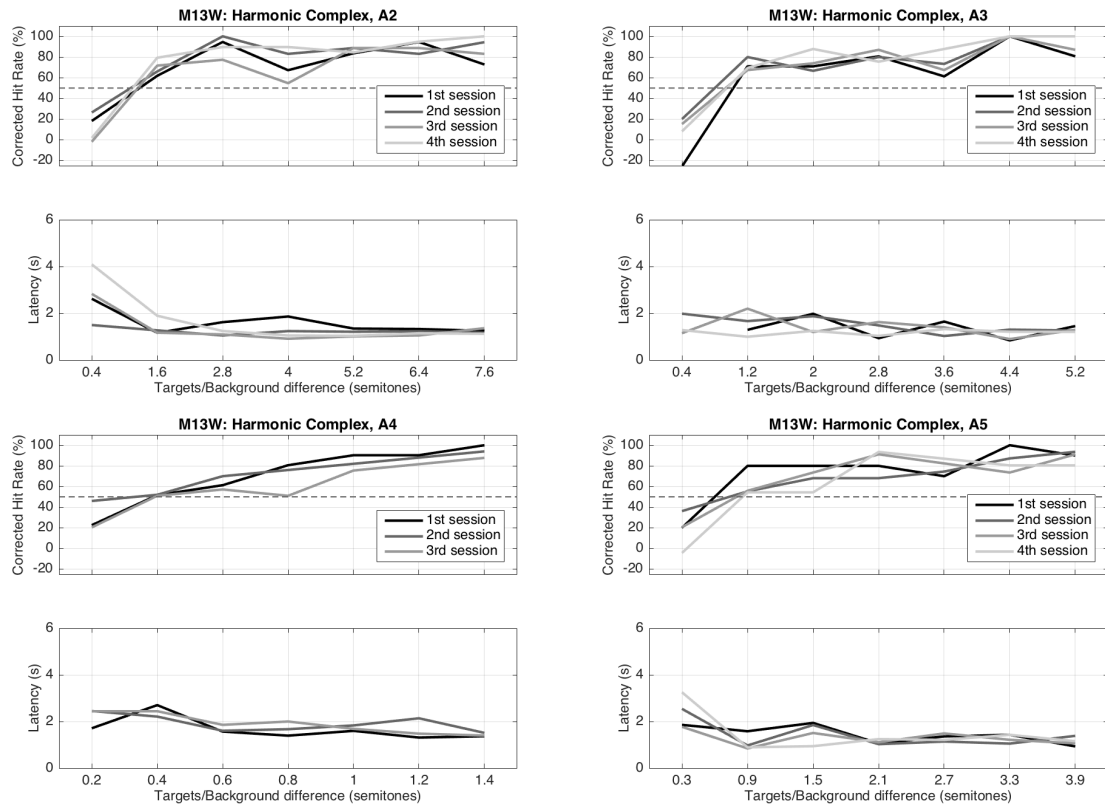| 4Y, Pure tone discrimination task								
Order	Targets	Threshold	#Trials	#Hit	HR	#FA	FAR	# Error
3	A3 F0 (16-112st, 16st interval)	2.293	120	73	86	7	19	27
		2.978	200	112	80	13	22	64
		2.800	200	112	80	12	20	48
1	A4 F0 (08-56st, 08st interval)	2.833	100	43	62	7	24	44
		2.756	110	47	59	5	15	60
		2.853	190	83	62	8	14	111
		2.844	180	74	59	10	19	56
2	A5 F0 (08-56st, 08st interval)	1.543	200	111	79	3	5	45
		1.644	200	115	82	14	23	45
		1.752	180	97	76	8	15	37
		1.973	120	64	77	8	23	40
4	A6 F0 (04-52st, 08st interval)	1.030	200	115	82	10	17	66
		1.526	190	104	78	13	22	36
		1.573	100	52	75	2	7	60
		1.346	200	103	73	6	10	11
5	A7 F0 (02-26st, 04st interval)	0.440	180	103	81	6	11	20
		0.451	200	117	83	7	12	15
		0.533	200	110	79	15	25	87
6	A8 F0 (02-26st, 04st interval)	0.353	200	120	86	11	19	77
		0.436	200	118	84	10	17	29
		0.433	200	112	80	6	10	38
		0.472	200	108	78	5	9	14
7	A9 F0 (01-13st,	0.452	200	101	73	11	19	42
		0.419	200	107	76	14	24	33
		0.358	160	81	72	7	14	35
8	A10 F0 (02-26st, 04st interval)	0.417	140	81	82	8	19	86
		0.756	190	89	66	11	20	53
		0.467	200	119	85	10	17	51
		0.511	110	60	77	7	22	24

4Y, Harmonic complex tone discrimination task								
Order	Targets	Threshold	#Trials	#Hit	HR	#FA	FAR	# Error
4	A2 HC (08-80st, 12st)	1.800	200	102	73	5	9	55
		1.600	200	117	83	14	24	42
		1.886	200	113	80	8	14	35
3	A3 HC (04-52st, 08st)	1.143	150	85	80	6	14	44
		1.400	170	97	82	12	24	50
		1.111	200	115	82	14	24	72
		1.219	200	115	83	13	21	40
1	A4 HC (02-14st, 02st)	0.460	200	108	77	15	25	108
		0.493	200	104	74	8	14	114
		0.494	150	73	70	8	18	66
		0.386	190	110	79	12	21	110
	A5 HC (01-19st, 03st)	0.570	200	106	75	10	16	65
		0.478	200	116	82	14	24	54
		0.450	190	104	79	13	23	67
2		0.450	200	111	79	12	20	60

Table 2. Experiment records for each qualified session (subject: M4Y).

55Y, Pure tone discrimination task								
Order	Targets	Threshold	#Trials	#Hit	HR	#FA	FAR	# Error
3	A3 (16-112st, 16st)	2.70	130	74	81	2	5.1	17
		2.37	200	122	87	5	8.3	19
		3.47	180	93	74	1	1.9	2
		2.09	150	93	89	2	4.4	16
2	A4 (08-56st, 08st)	1.70	200	112	80	13	22	106
		2.18	200	95	68	4	6.7	14
		2.18	200	93	66	3	5	44
1	A5 (08-56st, 08st)	1.63	200	110	79	7	12	47
		1.11	200	124	89	14	23	42
		1.31	200	119	85	8	13	35
7	A6 (04-52st, 08st)	0.96	200	117	84	7	12	37
		1.04	200	112	80	7	12	57
		1.05	200	112	80	8	13	58
6	A7 (02-26st, 04st)	0.46	200	111	79	5	8.3	19
		0.38	200	122	87	5	8.3	15
		0.42	200	119	85	4	6.7	25
4	A8 (02-26st, 04st)	0.44	100	61	87	6	20	38
		0.46	110	64	83	4	12	14
		0.41	150	90	86	4	8.9	11
		0.42	120	70	83	1	2.8	17
		0.39	150	92	88	5	11	21
5	A9 (02-26st, 04st)	0.58	200	108	77	7	12	40
		0.71	200	103	74	6	10	32
		0.69	200	104	74	3	5	25
8	AA (10-26st, 04st)	1.25	160	81	72	3	6.3	78
		1.39	200	106	76	17	28	45
		1.48	210	121	82	15	24	108

55Y, Harmonic complex tone discrimination task								
Order	Targets	Threshold	#Trials	#Hit	HR	#FA	FAR	# Error
1	HC A2 (08-80st, 12st)	1.89	200	113	81	18	30	63
		2.67	200	104	74	10	17	40
		1.93	200	113	81	9	15	42
2	HC A3 (04-52st, 08st)	1.33	200	109	78	8	13	38
		0.87	200	122	87	12	20	59
		1.20	200	111	79	6	10	22
3	HC A5 (02-26st, 04st)	0.48	190	108	81	6	11	88
		0.81	200	96	69	11	18	42
		0.86	200	98	70	13	22	58
		0.54	200	109	78	4	6.7	31

Table 3. M55Y: Experiment records for each qualified session (subject: M55Y).

45

13W, Pure tone discrimination task								
Order	Targets	Threshold	#Trials	#Hit	HR	#FA	FAR	# Error
4	A3 F0 (16-112st, 16st interval)	2.667	200	121	86	12	20	41
		2.400	200	125	90	15	25	50
		2.756	200	119	85	8	14	36
1	A4 F0 (12-84st, 12st interval)	2.240	200	106	75	4	7	62
		2.100	122	64	76	3	9	46
		2.923	200	94	67	4	7	27
		2.244	200	95	67	5	9	37
2	A5 F0 (08-56st, 08st interval)	2.000	170	93	78	9	18	46
		2.067	190	104	76	11	19	120
		1.644	200	111	79	14	24	63
		1.733	200	105	75	13	21	33
3	A6 F0 (08-56st, 08st interval)	2.178	130	61	67	9	24	47
		1.933	180	101	77	9	17	109
		1.813	190	100	75	7	13	41
		1.939	200	104	75	4	7	28
6	A7 F0 (01-25st, 04st interval)	0.407	200	102	72	10	17	60
		0.331	200	112	80	9	15	37
		0.252	170	107	88	10	20	79
		0.478	120	63	75	4	12	8
5	A8 F0 (02-26st, 04st interval)	0.413	150	85	80	5	11	35
		0.440	200	118	84	12	20	60
		0.497	200	115	82	10	17	37
		0.436	200	119	85	10	17	34
7	A9 F0 (04-16st, 02st interval)	0.608	140	77	78	7	16	64
		0.667	200	106	75	12	20	79
		0.733	120	61	72	8	22	12
		0.638	150	82	78	6	13	34

13W, Harmonic complex tone discrimination task								
Order	Targets	Threshold	#Trials	#Hit	HR	#FA	FAR	# Error
4	A2 HC (04-76st, 12st interval)	1.275	200	102	72	5	9	29
		1.114	200	112	80	7	12	36
		1.246	200	98	70	7	12	42
		1.147	200	109	77	2	4	30
3	A3 HC (04-52st, 08st interval)	1.027	130	64	70	8	20	24
		0.800	170	93	77	6	11	28
		0.933	180	95	76	8	15	54
		0.947	200	112	80	11	19	90
1	A4 HC (02-14st, 02st interval)	0.389	140	77	77	11	25	58
		0.333	190	101	76	7	12	76
		0.393	200	95	68	11	19	78
2	A5 HC (03-39st, 06st interval)	0.600	110	59	76	3	9	32
		0.733	183	92	72	7	13	31
		0.800	130	67	74	5	13	22
		0.856	170	80	67	5	10	34

Table 4. M13W: Experiment records for each qualified sessions (subject: M13W).
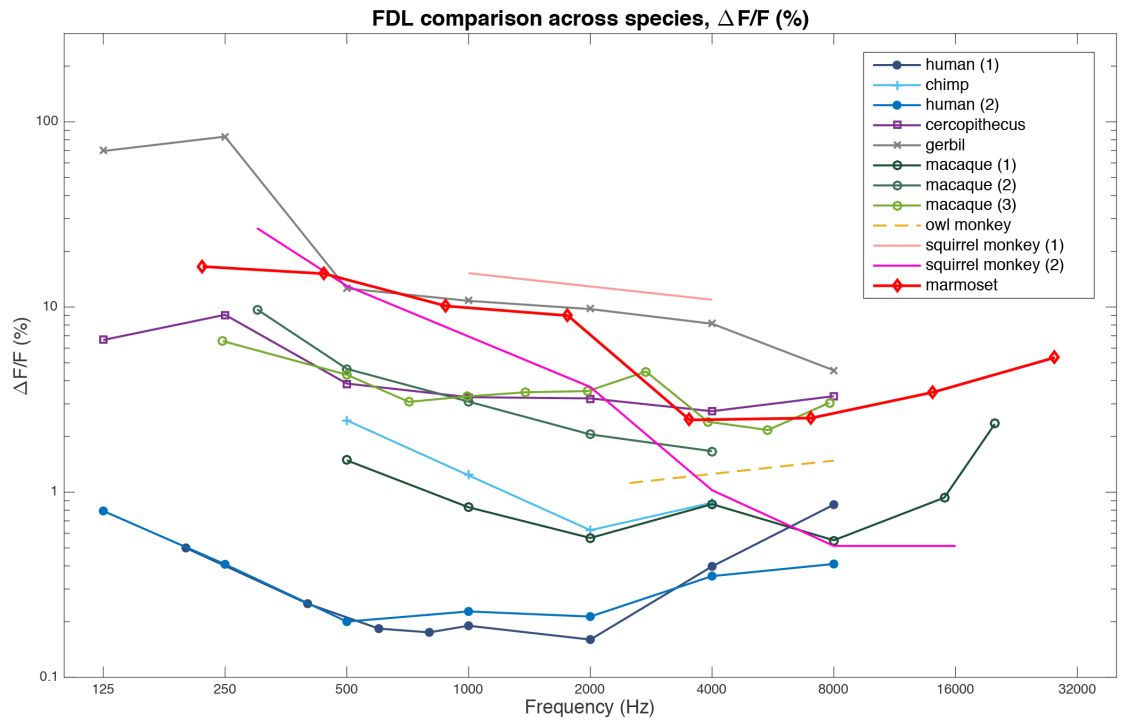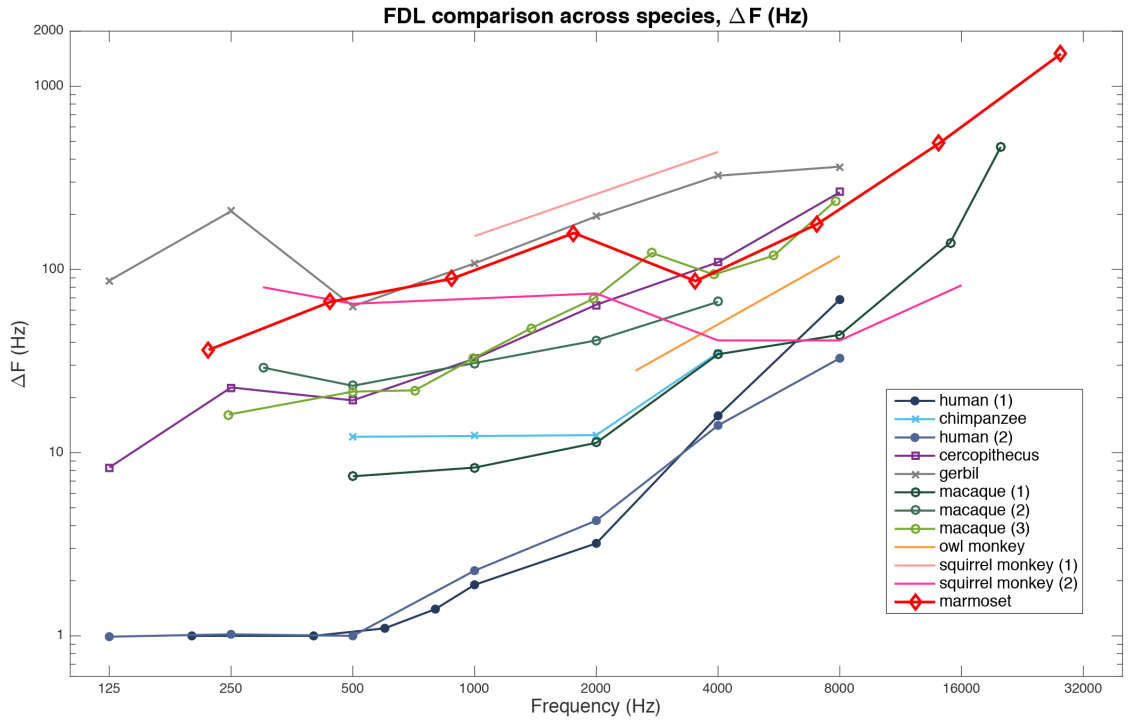
Figure 11. Comparison of FDL (absolute value and relative value) across different species.

Legend	Source Reference	Source Data	Sound Level
human (1)	Wier 1977	numbers in original paper	40dB SL
chimpanzee	Kojima 1990	digitized figure from Sinnott 1992	70dB SPL
human (2)	Sinnott 1992	digitized figure from original paper	60dB SPL
cercopithecus	Sinnott 1992	digitized figure from original paper	60dB SPL
gerbil	Sinnott 1992	digitized figure from original paper	60dB SPL
macaque (1)	Stebbins 1970	digitized figure from original paper	60dB SPL
macaque (2)	Moody 1986	digitized figure from Wienicke 2001	Not Available
macaque (3)	Prosen 1990	digitized figure from original paper	≥ 40dB SL
owl monkey	Recanzone 1991	numbers in original paper	~70dB SPL
squirrel monkey (1)	Capps 1968	digitized figure from Wienicke 2001	70DB SPL
squirrel monkey (2)	Wienicke 2001	numbers in original paper	50±2dB SPL
marmoset	*current study*	*numbers from current study*	*35~40dB SL*

Table 5. Source of the data points in figure 11.
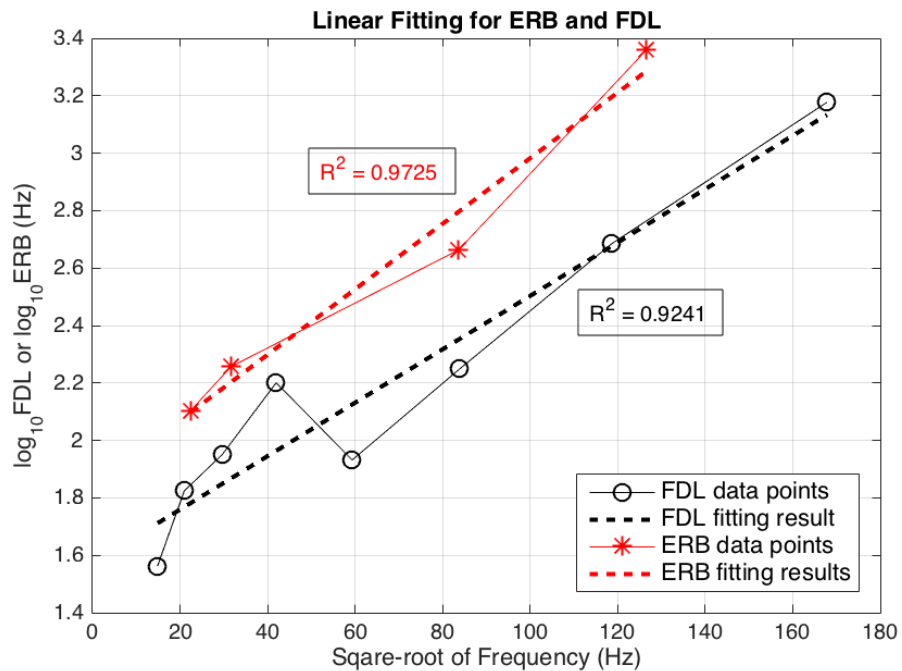
Figure 12. Fitting results of marmoset's ERB and FDL data according to Wier's (1977) model.
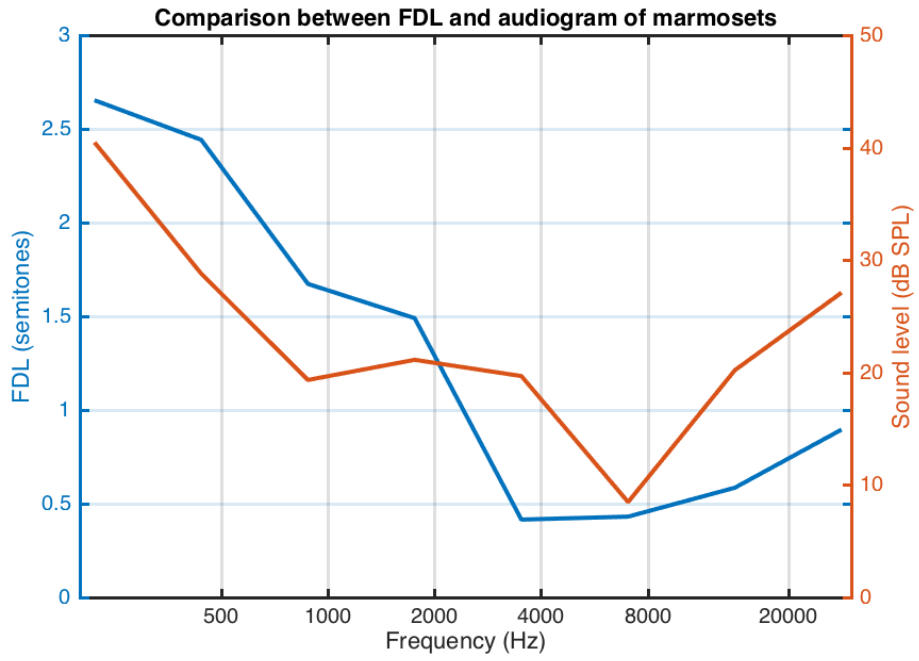
Figure 13. Comparison between relative FDL and audiogram of marmosets.
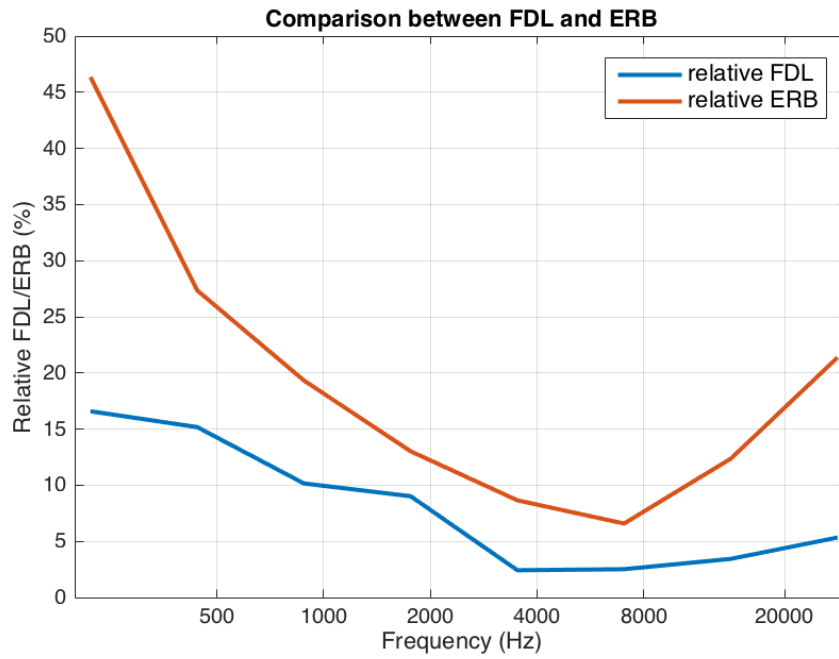
Figure 14. Comparison between relative FDL and relative ERB of marmosets.

Marmoset				Human			
Freq (Hz)	ERB (interpolated, Hz)	FDL (Hz)	ERB/FDL	Freq (Hz)	CB (Hz)	FDL (Hz)	CB/FDL
220	102.0	36.5	2.79	250	100	0.7	143
440	120.3	66.8	1.80	500	114	0.83	137
880	170.1	89.4	1.90	1000	160	1.8	89
1760	229.2	158.7	1.44	2000	300	3.8	79
3520	304.7	85.5	3.56	4000	660	12.9	51
7040	463.8	177.5	2.61	6000	1130	38	30
14080	1741.3	484.4	3.59	8000	1650	96	17
28160	6020.0	1503.9	4.00				

Table 6. Comparison of ERB/FDL ratio between human and marmoset.
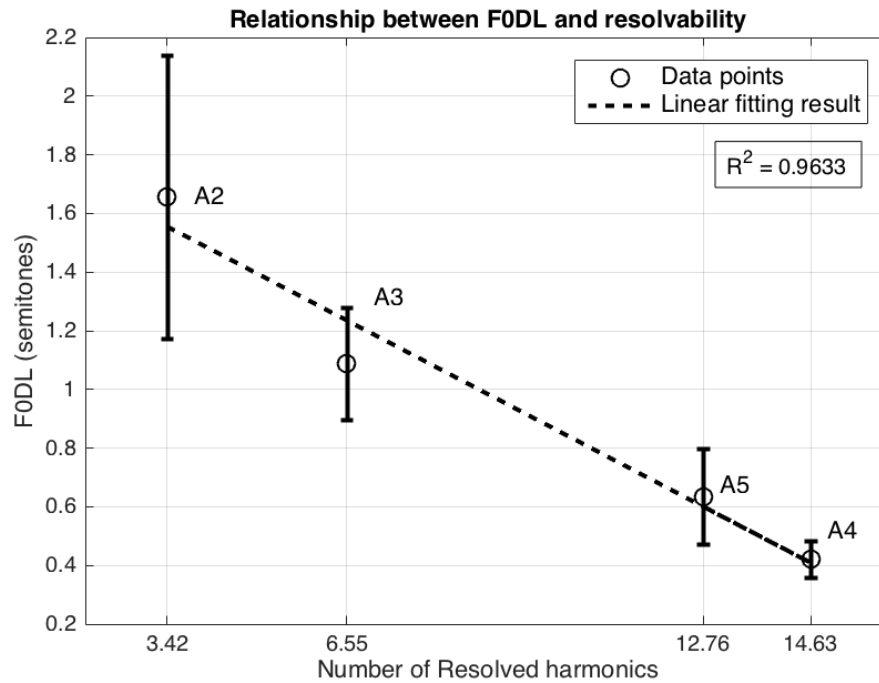
Figure 15. Relationship between marmoset's F0DL and resolvability at four frequencies.

Chapter 3: Consonance vs. dissonance preference by Marmoset Monkeys

1 Background

1.1 Conclusions from previous studies on auditory preference

It has been a controversial question whether music is unique for human. From evolutionary perspective, the conflicting theories on origin of music dates back to the age of Charles Darwin and Herbert Spencer (Kleinman 2015). As far as we know, humans find some sounds more pleasant than others, the fact of which may underlie our enjoyment of music. It is nevertheless unclear whether this preference is hard-wired in the brain, or acquired by adaptation to music. Behavioral studies on animals are necessary to separate these two possibilities, as the animals are not normally exposed to music, and the acoustic environment in the laboratory can be controlled and manipulated to an extent not possible in humans. In order to investigate whether animals are also able to "enjoy", or "appreciate" music, researchers have done experiments using various behavioral paradigms and types of sound stimulus, on different species ranging from rats to primates and humans (Figure 16). Here I summarized some relevant studies that tested animals' aesthetic responses to music, to provide some rationale behind the experimental design of this current study.

Generally, researchers were asking two types of questions: (1) Is music unique to humans? Or does human music possess some common features that are also preferred by

animals? (2) What kind of music appeals to animals? What can we learn from this regarding how their brains process sounds?

Although this chapter focuses on preference for consonance in music, not all the reviewed studies concern consonance. Researchers were not only exploring the consonant-dissonant dimension, but other auditory dimensions as well, in order to search for effective auditory enrichments for animals, or to test animals' preference on other dimensions. These studies will also be reviewed because the behavioral paradigm and stimulus design may provide useful insights into the question under the current study.

1.1.1 *Consistent conclusions about preference for consonance on humans*

There have been a lot of studies showing that adult human prefer consonant intervals to dissonant ones in music (McDermott, Lehr, and Oxenham 2010; Cousineau and McDermott 2012). When subjects were asked to rate the level of pleasantness for different music intervals, they always rate highest for perfect consonant chords like perfect fifth and octave, lower for imperfect consonant chords (e.g. major third, major sixth), lowest for dissonant intervals (e.g. minor second, tritone). These different emotional responses are also the perceptual basis for music theory and music composition—dissonance introduces tension or unpleasantness, and consonance expresses joy or a feeling of closure by resolving the dissonance. However, it is still not known whether this preference for consonance is nature or nurture.

Human infants are appropriate subjects to study because of their minimum exposure to music. Studies on human infants lead to the consistent conclusion that they prefer consonance to dissonance, no matter what specific stimuli were used to test (Zentner and

Kagan 1998; Trainor and Heinmiller 1998; Trainor, Tsang, and Cheung 2002; Masataka 2006; Kastner and Crowder 1990). However, the limitation is that it was not feasible to control the acoustic environment during pregnancy, considering that human fetus is able to hear at 24 weeks, providing 4 months of constant sound exposure prior to birth (Snowdon and Teie 2013a). Although natural sounds are filtered so that only low-frequency sound get through the womb, the influence of exposure to music could not be completely eliminated. Thus, the evidence was insufficient to support the hypothesis that musical preference is hard-wired in our auditory brain, rather than a form of adaptation to the sound that we were exposed to.

1.1.2 *Inconsistent conclusions on animals' preference for consonance*

In view of the limitation of controlling human acoustic environments, researchers also used different animal models to test their preference for consonance, trying to answer the question whether the preference is evolutionary innate or unique to human society. However, inconsistent conclusions were drawn from different animal models, ranging from rodents (rats) to non-human primates (apes, old-world and new-world monkeys). Despite the fact that these studies may have used different stimulus and behavior paradigms, the conclusions are summarized here: species that showed preference for consonance includes chimpanzee (Sugimoto et al. 2010), Java sparrows (Watanabe and Nemoto 1998), chicks (Chiandetti and Vallortigara 2011) and rats (Borchgrevink 1975); species tested that did not show preference are Moloch Gibbons (Wallace et al. 2013), Campbell monkeys (Koda, Basile, Olivier, Remeuf, Nagumo, Blois-Heulin, and Lemasson 2013a), tamarins and marmosets (McDermott and Hauser 2004; McDermott

and Hauser 2007). However, not all the conclusions are solid enough, depending on the specific behavior paradigm and acoustic stimuli they used.

1.1.3 Animals' preference on other acoustic dimensions

Researchers also explored animal's auditory preference in other acoustic dimensions, mainly for two reasons. 1) Given the lack of evidence for consonance preference, some species, however, did show preference for some certain types of music. It is thus necessary to test other acoustic dimensions that could possibly elicit this preference. Watanabe (1998) showed that Java Sparrows preferred Bach to Schoenberg (the music stimulus used are shown in Figure 17E.), and this preference also generalized to Vivaldi and Carter, which is also a contrast between classical music and modern music; McDermott (2007) also showed tamarins and marmosets preferred a Russian lullaby to a techno song ("Nobody gets out alive"). These pieces of music also differ in other acoustic parameters, suggesting a combination of musical components such as pitch, timbre and tempo can modulate affective and behavioral states in animals. In order to find which parameters contribute to the preference behavior, more experiments have been done by controlling all other parameters except for the one that of interest.

Firstly, McDermott (2004) showed sound intensity is one of the parameters that elicit preference – tamarins and marmosets preferred silence when presented with silence vs. noise, or silence vs. music. Also, when presented with 60dB and 90dB white noise, tamarins showed significant preference for softer noise. These results suggested that these species prefer softer sound intensity, which provides the basis for the necessity that the sound energy be equalized when testing on other acoustic dimensions.

54

The second parameter is the tempo of acoustic events. When tested with faster and slower click trains, the tamarins showed preference for the slower ones (McDermott and Hauser 2007). In addition, one alternative simple explanation for the preference is that the abrupt onset of clicks elicits aversive responses, thus the slower tempo is less aversive, considering the smaller number of aversive events per unit time. To eliminate this possibility, they also tested tamarins on silence vs. chirp (presumably non-aversive vocalization produced when presented with food) trains. The results showed that they did not have significant preference when presented with the choices of slow chirps (50 pulses/min) and silence, while they did again prefer silence when the other side was fast chirps (250 pulses/min). Noticeably, the rate of 250 pulses/min is approximately the maximum rate of chirps in their normal behavior when food is presented, which should not be more aversive than slower chirps. These results suggest that marmosets and tamarins prefer slower tempo.

Another factor that influence their preference behavior is the meaning of the sound. Different vocalizations in a specific species are associated to different events, such as food-related chirps and distress-related screams. Thus, no matter how the physical features of these sounds differ, they elicit different emotional states by evoking the related positive or negative events. McDermott (2004) showed that tamarins have a significant preference for "chirps" (produced when presented with food) compared to the "screams" (produced when they are distressed).

1.2 Typical behavioral paradigms used in preference studies

It is tricky to develop a reliable behavioral paradigm to measure animals' preference. Researchers have been using different methods depending on their goal of study and the resources they have. The most widely used behavior measurements are summarized below: (1) sound-associated location preference; (2) self-triggered sound playback time; (3) observations of natural behavior; (4) physiological parameters.

1.2.1 Sound-associated location preference

The basic assumption is that the animal tends to stay in places that they prefer. Thus, if two specific locations are associated with different sound while other factors including visual field, shape and scent are identical, their preference for location is an indicator of their preference for the associated sound. McDermott (2004, 2007) used a V-shaped maze for tamarins and marmosets. The monkeys were placed in a maze with two branches through a door and were allowed to move freely in the maze. The maze has two branches such that two different sounds were played from two speakers placed at the end of the maze branches. At any given time, only one of stimuli was played depending on the monkey's position (left side or right side). In the end, the times spent on each side were calculated and compared. If they stayed on one side significantly longer than the other, they could infer that the animal prefer the sound playing on that side. Koda (2013) also used a similar location-based paradigm for Campbell monkeys. Alternatively, the experiments were conducted in their habitual enclosure without any animal manipulation, reducing the stress caused by handling the animal and putting them in an unfamiliar environment, which could affect the animal's spontaneous behavior. The chick study (Chiandetti and Vallortigara 2011) also adopted a similar setup, in which the chicks were

placed in an arena with two speakers on opposite sides. Additionally, two imprinting objects were hung above the speakers to attract the chicks to move towards the speakers. However, the speakers were playing different sound at the same time, which could possibly cause unwanted interactions. For those studies on birds (Ikkatai and Watanabe 2010; Watanabe and Nemoto 1998), perches were used as location measurements. Several optical sensors were set up to detect the birds' landing on the perches. Once the bird landed on a perch, the associated sound started playing.

This location-based measurement is simple to set up, with no need to train the animals. Nevertheless, irrelevant factors must be strictly controlled to eliminate the possibility that they prefer something else rather than the associated sound. Also, it requires large contrast between the two sounds for the animals to show significant preference, and the individual difference could be large as well.

1.2.2   Self-triggered sound playback

Another behavior paradigm allows the subjects to trigger the sound by themselves, with their natural behaviors or some trained behaviors. By comparing the total time that each sound was played, the preference for a particular sound could be determined. This paradigm was used on human infants (DeCasper and Fifer 1980; Sullivan and Lewis 2003). DeCasper (1980), exploiting their natural behavior of sucking a nipple: they were given a rubber nipple, and the rate of sucking determined what sound they heard. In this particular study, infants were presented with either their mom's voice or a stranger woman's voice. A threshold of inter-pulse interval (IBI, each suction action was defined as a "pulse") was set to determine which sound they will hear—for some of the infants, higher IBIs triggered their mom's voice, for the others lower IBIs triggered their mom's

voice. Preference for the mom's voice was shown by an increase in the proportion of IBI's capable of producing the mom's voice. Another study on human infants (Sullivan and Lewis 2003) utilized their spontaneous limb movement, to examined how different frustration contexts affect the instrumental and emotional responses of infants. These infants wore an elastic bracelet attached to a ribbon. Pulling on the ribbon controlled a switch that triggers music and corresponding slides. It turned out that babies can learn the association between pulling and the expected outcome, and that once the outcome did not match their intention they would get angry. With the similarly set up, Sugimoto (2010) did experiments on a baby chimpanzee to test its preference for consonance. The baby chimpanzee's arm was attached to a string, and arm movement triggered a switch for music. The pulling activity is part of its spontaneous repertoire and it happened more frequently when there was auditory feedback. By programming the computer controlling the sound, preference for a particular sound could be inferred from the temporal pattern of pulling. A more difficult task for rats was described by Borchgrevink et al. (1975), in which the rats were trained to press pedals to trigger sounds. Two pedals produce either a consonant sound or dissonant one. They tested 34 rats and the results showed that the number of presses on the two pedals had small and unreliable difference during the initial period, while at the end the pedal producing the consonant sound was pressed almost twice often as the other one.

The self-triggered sound paradigm may require more time of training. However, once the subject learnt to associate an action with sound, this paradigm is more stable and reliable compared to measuring spontaneous location preferences.

1.2.3 *Observations of natural behaviors*

A simple but more subjective method is to observe and record animals' response to different sound (Wallace et al. 2013; Snowdon and Teie 2009; Snowdon, Teie, and Savage 2015). Wallace (2013) investigated the possibility of using music as auditory enrichment for Moloch Gibbons. They evaluated the musical effects by observing gibbons' natural behaviors during music-playing weeks and control weeks in the zoo. The behaviors taken into account were activity, brachiation, affiliative behavior (e.g. giving and receiving grooming) and anxiety behaviors (e.g. self-scratching and self-grooming). Among the 8 captive gibbons they observed, only 2 of them showed significant differences in some behaviors between music and control conditions, suggesting that music may not be an effective enrichment for them, and that individual differences must be considered.

Snowdon's group has studied musical aesthetic responses on tamarins (Snowdon and Teie 2009) and cats (Snowdon, Teie, and Savage 2015). They have composed affiliation-based and threat-based tamarin-specific music, which were designed to elicit different emotional states in these animals. Their natural behaviors were classified into 5 groups and were used as indicators for different internal emotional states: Head and body orientation to speaker indicates interest in the stimulus; foraging (eating, drinking) and social behaviors (grooming, huddling, sex) are considered as calm behaviors; behaviors like piloerection, urination, scent marking are indicators of anxiety; head shaking and stretching implies an arousal state. By counting the number of these behavioral events before and after the music was played, they concluded that threat-based music increased movement, anxious behavior and social behavior compared to affiliation-based music,

and these differences were significant only when species-specific music was played—
they were indifferent for human music. A similar paradigm was used in the cat study, in
which the cats showed more significant affective responses when presented with cat
music.

Observation of behaviors can be tricky because of its subjective nature, and strict
standards have to be set to classify each related behaviors. In addition, the correlation
between different behaviors and emotional states requires a thorough research in the
species' natural behavior. On the other side, the advantage might be that it is easier to
pick out the subtle difference when the animals are responding differently to the stimuli.

1.2.4 Physiological parameters

The problem for most behavior-based preference experiments was that the contrast
between stimuli may not be robust enough to generate differences at the behavioral level.
Instead, some physiological parameters may have already changed but not observable by
eye. Therefore, measuring physiological parameters that are related to emotional states
(e.g. heartbeat, blood pressure, hormone levels) can also serve as a measurement of the
animal's state. Akiyama et al. investigated the effects of music on rats' blood pressure
(Akiyama and Sutoo 2011). They restrained the rats and measured their blood pressure
every hour, using a tail-cuff method with a programmed sphygmomanometer. The results
suggested that music lowered the blood pressure of spontaneously hypertensive rats, and
that only the music filtered in their vocalization frequency range (high-frequency band
pass filters) had this effect. Some other parameters related to affect could also be
measured such as heart rate and hormone levels, although it's usually difficult to conduct
the measurement when the animal is freely moving.

1.3   Sound stimuli used in preference studies

Because of the various definitions for consonance, different types of stimuli were used in the previous studies. Some of them were carefully controlled so that the degree of consonance is the only variable. Some of them are more musical, and there are some other acoustic dimensions that could possibly b attributes to the preference.

1.3.1   *Random selection from sets of consonant or dissonant chords*

There are 12 possible different combinations for two-tone chords within one octave. When being asked to rate the degree of pleasantness, normal human adults usually rate higher for consonant chords such as perfect 5th and octave, lower for dissonant chords such as minor 2nd and minor 7th (McDermott, Lehr, and Oxenham 2010). The simplest way to create contrast between consonance and dissonance is to compare two different chords, or chords randomly chosen from a consonant set and a dissonant set. This type of stimulus contrast has been used for human infants (Kastner and Crowder 1990) and also for tamarins and marmosets (McDermott and Hauser 2004). The consonant set contains perfect fourth, perfect fifth and octave, and the dissonant set contains minor second, tritones and minor ninth (Figure 17A). It was shown that both adult humans and human infants have preference for the consonant set of chord, while tamarins and marmosets do not. The drawback of this type of stimulus design is that it was not in a "musical" context, and the contrast only happens in the harmonic dimension, not in the melodic dimension, which could be insufficient to generate significant contrast between consonance and dissonance.

1.3.2 Original music versus modified music

Another design put the stimuli in a musical context, and the contrast was manipulated by changing one of the notes in the chords of an original music piece. As shown in Figure 17B, compared to those chords randomly chosen from a "chords pool", it exhibits some basic musical elements like the dynamic tempo pattern. However, it is still not very "musical" because of the flat melody line and the invariant tempo patterns. The consonant version contains two octaves and two fifth, and the dissonant version contains two tritones and two minor ninth. With this type of contrast, Trainor (2002) showed that human infants prefer the consonant version, while Koda (2013) showed Campbell monkeys did not have preference.

In contrast to the stimuli in Figure 17B, the stimuli in Figure 17D have a more variable temporal pattern and a more musical melody line. In the dissonant version, all minor thirds and major thirds were changed to minor seconds, perfect fifths were changed to major sevenths. By changing one note in each chord, the melody line was also changed, establishing a contrast in the melodic dimension besides the harmonic dimension. This manipulation is supposed to create bigger contrast between consonance and dissonance than stimuli in Figure 17B. Zentner (1998) first used this design on infants, showing that infants prefer the consonant version. Chiandetti (2011) also claimed that chicks prefer the consonant version using the same stimulus design. I also adopted the same stimuli shown in Figure 17B and used the on marmoset monkeys (experiment 3) in this thesis.

Figure 17C is an example used by Sugimoto (2010) on the chimpanzee infant. Instead of composing their own sound stimuli, they used three pieces of existing classical music, and modified the original music according to some rules (for example, change all Gs to G

flat) to create a dissonant version. This type of stimuli are also musical and contrast in both harmonic and melodic dimensions.

1.3.3   Different styles of original music recordings

Another set of experiments brought the concept of consonance to a much broader sense. Different music styles were used and they asked whether the animals prefer one style to the other (McDermott and Hauser 2007; Watanabe and Nemoto 1998). McDermott (2007) used a Russian flute lullaby and a German techno, and demonstrated that tamarins and marmosets prefer the lullaby. Watanabe (1998) tested Java Sparrows with Bach French Suite BVW816 and Schoenberg suite for piano op.25 (Figure 17E), and showed two subjects preferred Bach to Schoenberg. They further showed that this preference generalized to another two pieces of music by Vivaldi and Carter, suggesting that these animals prefer classical period music over 21th century music. Although these original music recordings differ extensively in the dimension of consonance, they also have distinct temporal patterns, timbres, etc. Thus, it is not sufficient to conclude they prefer consonance, but these results suggest animals at least have preference for some particular types of music that are considered more consonant in general, while it's unclear which musical elements elicit the preference. I adopted the Bach and Schoenberg pieces in experiment 4 to test if marmosets have preference for the classical style.

1.3.4   Specifically composed music for animals

Playing human music to other species assumes that the species has the same auditory capacities as humans, which is not necessarily true. The tamarin music and cat music used by Snowdon (2009, 2015) was composed by a musician, and the assumption was

that the affective effects in music are largely attributed by their vocalization features. They asked musicians to cluster the tamarins' call into 5 types based on the acoustic structures and match these categories with the meanings of calls. Then according to the relationship between acoustic features and meanings, the musician composed tamarin affiliation music with cello that was designed to capture the quality of pure, high frequency sounds common to the range and timbre of tamarin affiliation vocalizations. Similarly, another type of music—tamarin threat music played by two guitars, was designed to reflect the repeated motives consisting of minor seconds and thirds using quick-onset picked notes in complex timbre. Basically these sounds were designed to mimic their vocalizations, but in a musical way. Indeed, in human music, there are some similarities to human speech in terms of emotional content. For example, angry is usually expressed by faster tempo, higher pitch and dissonant intervals (e.g. minor second), and calm is usually featured slower tempo, lower intensity, moderate pitch and consonant intervals (e.g. unison or third, fifth). It is important to consider the acoustic environment that a species is adapted to—human music may fall into a frequency range about which other species do not care or actually dislike. For example fundamental frequencies may be so low that the music sounds annoying to monkeys, no matter how soothing it is to us. This is also a possible factor causing the non-preference in McDermott's studies (Mcdermott 2004, 2007) on marmosets and tamarins.

2 Methods and experimental designs

2.1 Behavioral paradigm and subjects

The experiments were conducted in a sound isolation chamber, as shown in Figure 18, A transparent V-shaped maze was place on the floor, with a square middle compartment and two branches. The subject was able to move freely to any part of the entire space. The bottom, in white color, and the frames were made of Plexiglass. The sides and tops were covered with transparent plastic mesh. The visual field from inside of the maze was unblocked by any part of the maze. We transferred the subject from the carrier box into the maze through a door on the maze that could be lifted up. The maze was enclosed by four identical thick curtains (dark blue color), so that the visual field was symmetric. The square-shaped enclosed space was illuminated by four lights on the top of the curtain stands to maintain the whole field of view with uniform luminance. The speaker was located on one side of the sound chamber, approximately 4m far away from the maze. Sound stimuli were always played from the same speaker. To monitor the subject's behavior, a camera was hung on the roof of the chamber, right above the maze. Figure 18B shows the view from the camera. By adjusting the luminance and contrast of the camera settings, the inside of the maze can be nearly white and the outside is black.

The experiment was controlled by a customized MATLAB program (Mathworks). This program first captured a snapshot of the background with no monkey in the maze, which was used as a reference to do image analysis. Then the maze area was divided into two parts by drawing two regions of interest (ROIs) on the background snapshot (as shown in Figure 18B). Two sound files were then loaded into the program and associated with the

two ROIs. After the subject was placed into the maze, testing sessions started. The program started to record video and conduct online analysis. For each frame, the number of "dark pixels" within the two ROIs were calculated by counting the number of pixels with values lower than a threshold (the threshold was set to 128, with a possible value range 0~255). As shown in Figure 18C, the y-axis shows the percentage of dark pixels within the left ROI (red) and right ROI (blue). By comparing the two numbers, the subject's location could be determined, as the monkey appeared in the video as a dark blob. For example, if the blue curve is above the red curve (more dark pixels on the right side), the program will make a judgment that the monkey is on the right side, and vice versa.

Once the monkey's location was detected, the associated sound was played. As long as the animal stayed on one side, the stimulus for that side kept playing continuously. As the animal switched side, the sound also switched immediately. Four major experiments were conducted under the following testing conditions: (1) Silence vs. Silence; (2) Silence vs. White Noise; (3) Consonance vs. Dissonance (Figure 17 D); (4) Classical music vs. Modern music (Figure 17 E).

The stimuli were played from the "audioplayer" object in MATLAB (Mathworks), through a sound amplifier (Crown Audio, Model D-75A). Sound intensity was measured at the location of the maze by a hand-held sound level meter (Bruel & Kjaer, type 2270), and calibrated to roughly 60dB.

A total of 3 subjects (2~3 years old) were used in the preference tests: M64A, M65A, M29A. For each experiment, 2 subjects were tested. Each session lasted for 20 minutes to 1 hour (detailed information about each session is listed in Table 7 and Table 8). In

66

experiment 3, only the 1 hour sessions were counted into analysis. In experiment 4, all sessions are 30 minutes. Each session was separated by at least a full day, and the sound-side paring was reversed at least once.

2.2   Stimulus generation

The music scores used in experiment 3 is shown in Figure 17 D, and the stimuli used in experiment 4 are shown in Figure 17E.

In experiment 3, the sound was generated by writing a MIDI file, and then converting it to a WAVE file with MATLAB (Mathworks). The tempo was 92 beats per minute. For each note, a harmonic complex tone was used with 16 harmonics. The sound waves of the first beat in the consonant and dissonant melodies are shown in Figure 19, left column. The right column demonstrates the spectrograms for the entire consonant (above) and dissonant (below) melodies. The sampling rate was 44100Hz.

In experiment 4, the two piano pieces were extracted directly from the original recordings: Arnold Schoenberg, Piano Suite, Op. 25: Gavotte (played by Christopher Oldfather) and Johann Sebastian Bach, French Suite No. 5 in G Major, BWV 816: I. Allemande (played by Glenn Gould). The sound waves (left column) as well as the spectrograms (right column) of the first 5 seconds for both music pieces are shown in Figure 20.

In both experiment 3 and experiment 4, sound intensity was normalized according to peak intensity. The average sound level was adjusted to ~60dB SPL measured outside of

the middle compartment of the maze. In experiment 2, the white noise sound level was also adjusted to ~60dB SPL.

3 Results

3.1 Feasibility of behavioral paradigm

The first and second experiment is to confirm that this behavioral paradigm is feasible to measure their preference to sound. In experiment 1, no sound was playing, and the monkey was running freely in the maze. As shown in Figure 21A, they tend to spend equal time on each side (two-tailed t-test, $t(10) = 1.8837$, $p = 0.089$), suggesting that the marmosets do not have a preference for one side of the maze. However, the p value is relatively small, and there is a tendency that they prefer the right side. That may be caused by the location of the door of the maze—the monkey was put into the maze from a door on the left side and they may try to avoid the door in the first several sessions. This silence test also gave them some time to get adapted to the new environment so that they can have relatively stable behaviors in the following testing sessions, minimizing the effects of anxiety on the results.

In experiment 2, there was a significant preference to the silence side (two-tailed t-test, $t(10) = 11.8371$, $p < 0.0001$). Although the 60dB White Noise is a moderate level, it drove the subjects to the silence side. When observing their behaviors during experiments, I found that the monkeys tend to escape from the noise once the sound is on. This result is consistent with that in the study on tamarins (McDermott and Hauser 2004).

3.2 Response to consonance and dissonance music

In experiment 3, in order to test whether the marmosets have preference to consonance when the stimuli have contrast on both harmonic and melodic dimensions, I used the stimuli (Figure 17 D) that were used in the preference test on chicks (Chiandetti and Vallortigara 2011) and human infants (Zentner and Kagan 1998). In both of these two studies, the authors claimed that the subjects prefer the consonant stimulus. However, as shown in Figure 21 B, the marmosets do not significantly prefer the consonant music (two-tailed t-test, $t(20) = 0.4679$, $p = 0.6449$). From Figure 22 C, one can tell that the variance across different sessions is very large (standard deviation = 25.5%).

3.3 Response to classical music and contemporary music

It is possible that even the stimuli used in experiment 3 did not provide enough contrast for them, as the two music pieces have the same melody contour and timbre and tempo patterns. Thus, I tested with a bigger contrast between consonance and dissonance in experiment 4 with two pieces of original piano music recordings of Bach's work and Schoenberg's work. These two stimuli was used in a study on Java Sparrows, in which they showed that 2 of the 4 subjects tested prefer Bach over Schoenberg, and that this preference generalized to other Classical period composers and modern period composers (Watanabe and Nemoto 1998). However, in this current study, there is no significant preference for Bach's work (two-tailed t-test, $t(23) = 1.8959$, $p = 0.0706$). Again, the variance among individual sessions is big (standard deviation = 18.2), compared to that in experiment 1 and 2.

4 Conclusion and Discussion

4.1 Observations of behaviors in the current setting

This behavioral paradigm adopted from McDermott's studies (McDermott and Hauser 2004; McDermott and Hauser 2007) turned out to be sufficient to measure these animals' preference for sound. There were some small differences in terms of the shape of the maze, also the location and number of speakers. The visual environment was more controlled by enclosing the maze with identical curtains, so that the only difference between two sides was the associated sound stimulus. Some observations of their behaviors in the maze are listed below:

1) Adapting to new environments: The experiment 1 with no sound playing also served as an adapting procedure for the subjects. The behaviors may differ from monkey to monkey. When M64A was first placed into the maze, he started to run fast from left to right. The number of transitions from one side to another decreased as he became more adapted to the maze. The other two subjects M65A and M29A were very cautious in the beginning, and moved slowly to explore the new environment. After several days of adapting, they become more active.

2) Preference for location: Figure 23 and Figure 24 show the heat maps of the monkey's location in the maze during experiments. The heat maps were generated in the following way: For each frame of the video recording, the program detects the monkey's position and calculates the center of mass for the body. Then the monkey's location was represented by a round dot with smoothed edge and a similar size of the monkey's body,

centered at that center of mass. Then the heat map was generated by averaging the round

dots across every frame. Brighter color indicates longer time staying at that location. The

shaded gray area is the maze area. When there was no sound playing, different subjects

seem to have their own preferred locations in the maze. Figure 23 A shows a heat map of

M64A's location from one session of experiment 1 (only one session of video recording

was available under silence condition for this subject), and Figure 23 B shows the

averaged heat map of M65A's location from 4 sessions of experiment 1. These two maps

have different patterns. The first column in Figure 24 are 4 example heat maps for

individual 1-hour sessions from experiment 3, where consonance was associated with left

side. The second column are 4 example heat maps of the same subject, when the sound

association was swapped. There were several spots that they like to stay: the corners of

the middle compartment, and the relatively large space in the middle. Notice the

preferred locations were different across different sessions.

3) Learning effects: In experiment 2, when the acoustic condition was different

depending on their locations, the monkeys showed a learning process indicating that they

are capable of associating the sound with location. Figure 25 and Figure 26 show 4

example sessions of experiment 2 from M64A and M65A, respectively. The upper row

shows bar plots of inter-transition intervals (ITIs). Blue bars indicate the duration of time

segments staying on the left side, that is, the time between a transition from right to left

and a transition from left to right. Red bars are time segments staying on the right side. It

is obvious that the ITIs of the silence-paired side is much longer than those of the noise-

paired side. The bottom row of each figure shows how the percentage of time staying on

the noise side change during each testing sessions. Each point is the percentage calculated

from a 5-minute period. The first example session is the first day when experiment 2 started. These figures show that marmosets have a very quick learning process: in the first 5 minutes they tend to spend equal time on both sides, then the time spending on the noise side began to drop gradually. At the end of the first session, the percentage dropped nearly to zero. In the following days of experiment 2, they started to escape from the noise side from the beginning of the session, so that the percentage of time spending on noise side remains low (usually below 30%) throughout the entire session.

4.2   Marmosets showed no preference for consonance

Experiment 1 and 2 demonstrated that the behavioral paradigm was feasible to measure animal's spontaneous preference for sound. However, in experiment 3, marmosets showed no preference for the consonant version, contrary to the studies in chicks and infants in which the same stimuli were used (Chiandetti and Vallortigara 2011; Zentner and Kagan 1998). When the data are grouped for the data by maze branch, instead of consonance and dissonance, one of the subject (M65A) actually showed a preference for the right side (Figure 27, t-test, ***$p < .001$), providing another evidence of no-preference for sound.

In experiment 4, these animals did not show preference to the Bach's piano piece, which is also contrary to the study in Java Sparrows where the same pieces of music were used (Watanabe and Nemoto 1998). Again, when the data are grouped by left or right side, M65A showed significant preference for the left side instead (Figure 27, t-test, **$p < .01$). In this section I will discuss some of the possible explanations for these discrepancies.

72

4.2.1 Species differences

Assuming the conclusions from the previous studies and the current study are valid, the most parsimonious explanation is that different species have different capacities of appreciating music. Humans have a higher evolutionary level than marmosets and musical preference may have developed after New-World monkey. As for Java Sparrows and chicks, they belong to the "sauropsids" branch of evolution, which separated from the mammals more than 300 million years ago. It is possible that they develop the singing ability and musical preference via independent but convergent evolution. In addition, individual difference also plays a role. As described in Watanabe's study, 2 of the 4 subjects showed preference for the Bach piece, while the other two did not. Here I tested 2 subjects for each experiment, which is a small number. It cannot rule out the possibility that other marmosets have preference. This inconsistency could also due to different behavioral measurement settings. Both Watanabe and Chiandetti adopted the location-based behavioral paradigm, the birds were tested in a room with 3 perches, and the chicks were tested in a large arena with two speakers playing sounds at the same time from two sides. These details in the experimental settings may have influence on the animal's behavior, thus the results cannot be compared directly.

4.2.2 Problem of statistical tests

The other factor may account for this discrepancy is the statistical methods used in different studies. Here I used a one-sample two-tailed t-test, and chose 95% confidence interval ($*p < 0.05$) as a criterion for significant difference, basically testing whether the difference between the staying times of the two conditions deviated significantly from zero. In the previous preference tests, some of them also used ANOVA or Chi-square test.

Different tests may have different results, and lead to different conclusions. Indeed, when performing a significance test using ANOVA on the data from experiment 4 (Music of Bach/Schoenberg as within-subject factors), the statistical result was $F_{1,46} = 7.19$, $p = 0.0101$, which leads to the interpretation that they prefer Bach to Schoenberg. The same problem happens when using ANOVA to test the results from experiment 1 ("Silence" vs. "Silence"), the 2 marmosets prefer the right side significantly ($F_{1,20} = 7.0969$, $p = 0.0149$). Chiandetti et al. used one-sample t-test in the chicks study, same as what I used here. Watanabe et al used Chi-square test on 3 groups of data (3 perches associated with Bach, Schoenberg and Silence respectively) in the Java Sparrow study. To compare experiment 4 results with that in the Java Sparrow study, in which the same stimuli were used, I test the percentage numbers from experiment 4 with Chi-square test, and it still leads to a non-significant difference ($\chi^2 = 399$, $p = 0.0820$). Thus, the discrepancy between conclusions still exists. However, ANOVA and Chi-square tests are not appropriate in this study, because of their assumption that the two groups are independent (in this case, the percentage of left and right sums up to 100 and each session actually generates only one outcome, so a one-sample test is more proper). In addition, the small sample size makes it tricky to interpret the significance test results.

4.2.3 *Problem of using low-frequency sound stimulus*

The major problem of the previously tested stimulus is that those are music composed by human and for human listeners. However, marmosets have dramatically different auditory characteristics from humans, in terms of hearing range, frequency difference limen, vocalization frequency range and most sensitive frequency range (as shown in

chapter 2 of this thesis). So it is necessary to modify the human versions of music to the salient range of marmosets' auditory perceptual space.

As suggested by Snowdon (Snowdon and Teie 2013b), the animals tend to care only about those sounds with significant energy in the frequencies of their natural vocalization, or the frequencies that they are most adapted to. The fundamental frequency of marmosets' typical vocalizations ranges approximately from 4kHz to 7kHz, which is far beyond the fundamental frequency of the music I used in experiment 3 and experiment 4. Moreover, the aversive vocalization type ("Ek", "Cough" etc. produced when the animal is anxious or angry) is characterized by low frequencies (*www.marmosetcare.com*). Thus, it is possible that the consonant version and dissonant version both sound aversive to the marmosets, and they were trying to escape from both sides, not showing preference to either one. Generally speaking, from the evolutionary perspective, low frequency sounds may be alarming to a small animal, as larger animals emit lower frequency sounds. To test whether the low frequency sounds elicit aversive effects, I shifted the frequency up to their vocalization range, which is also their most sensitive hearing range according to their audiogram, and tested 2 marmosets (4 sessions for each subject) on the shifted version and un-shifted version of consonant melodies used in experiment 3. The range of fundamental frequency for the original consonant music is 207Hz to 554Hz, with a median of 392Hz. By shifting the fundamental frequencies of all the notes up by 51 semitones, the range of F0 becomes 3.95kHz to 10.55kHz, with a median of 7.458kHz that fall into their most sensitive hearing range. The results from this test are shown in Figure 29. There was no significant difference between the time spending on high-

frequency side and low-frequency side, suggesting that the frequency range is not a major factor that have influenced their musical perception.

4.2.4 Problem of animal's behavior in current apparatus

Deciding the duration of each session is tricky. Although the average time between transitions is approximately 5 seconds, which is a reasonable length of time for them to listen to the sound, there are two extreme conditions. Generally, my observation is that in the first several minutes they tend to be too anxious and run from one branch to the other continuously. In that case, the two stimuli were switching too rapidly and what they heard was not a segment of continuous music, but separated notes. After exploring in the maze for some time, they tend to sit still at one place for a very long time. It usually happens after 30mins. That period of time was classified as one of the two sides, even though they may not be caring about the sound anymore, leading to a huge bias to one side in the final result. Again, individual differences have to be considered. The behavioral patterns vary across subjects, also across different sessions.

In addition, preference for different music may not be strong enough to show significant difference. The behavioral paradigm may not be capable of measuring subtle preferences. It was obvious that the marmosets prefer the silence side when the other side was noise, and there was a 60dB SPL sound intensity contrast. This contrast between the two sides may be much bigger than the contrast only on the consonant dimension. In experiment 3, the only differences are the degree of consonance on harmonic and melodic dimensions, while sound intensity, tempo and timbre are all the same. In experiment 4, although the contrast on consonance dimension is presumably bigger than that in experiment 3, and another contrasting dimension tempo was added, the timbres were still the same and

sound intensity was equalized. In addition, this location-based behavioral paradigm requires a relatively large degree of preference to show significant difference in staying times.

4.3 Discussion and future work

4.3.1 Alternative behavioral measurements

The preference of the monkeys for consonance may not be obvious enough to be reflected by the current settings. Further studies could be done by developing new behavioral paradigms, like the self-triggered methods, which requires training but measures preference more accurately and stably.

Another alternative way of measuring preference is by observing their natural behaviors under different sound stimuli, like the observing methods used in tamarins (Snowdon and Teie 2009) and Campbell monkeys studies (Koda, Basile, Olivier, Remeuf, Nagumo, Blois-Heulin, and Lemasson 2013b). Although this method is more subjective and harder to quantify, it can capture some subtle differences. However, this method requires some experienced animal behaviorists to watch and take notes.

4.3.2 Future directions

Compared to McDermott's studies on marmosets, the same conclusion was drawn despite of the fact that different stimuli were used, that is, they do not prefer consonance to dissonance as we humans do. It was also pointed out that human infants' preference for consonance may be caused by the pre-natal exposure to musical environments. Thus, it is natural to think about the next stage of experiments to do some developmental studies. By controlling the acoustic environments for multiple groups of marmosets during their

critical period, and comparing their preference to music, we can get better idea about whether preference for consonance is nature or nurture.

Figures and tables for Chapter 3
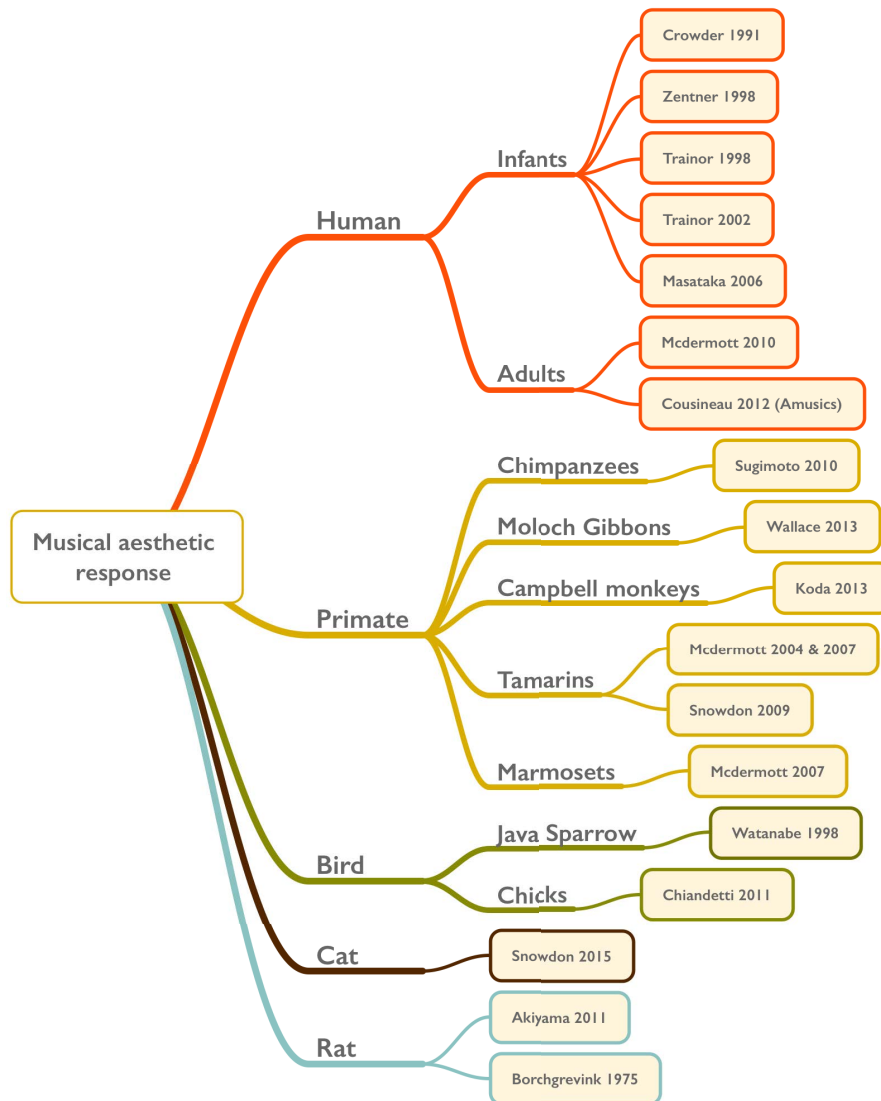

Figure 16. Summary of the major literature concerning musical aesthetic responses measured on difference species.

Figure 17. Music stimulus used in preference studies (examples).

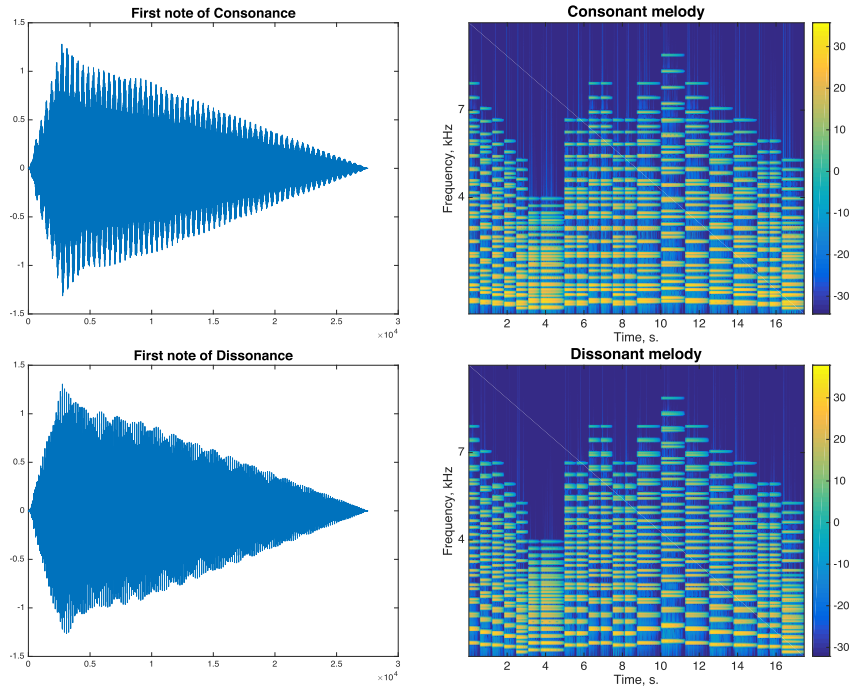Figure 18. Schematic behavior paradigm settings.

Figure 19. Sound waves (first note) and spectrograms of consonant and dissonant melodies used in experiment 3.
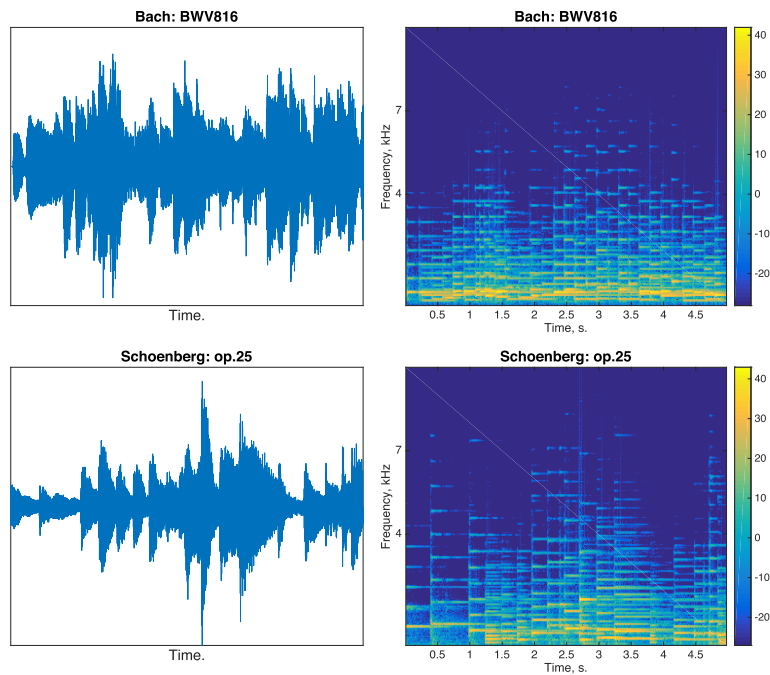

Figure 20. Sound waves (first 5 seconds) and spectrograms of Bach and Schoenberg melodies used in experiment 4.

81

Exp.1: Silence on both sides. (p = 0.089)						
Monkey ID	Date	Duration	Silence		Silence	
			Side	Percentage	Side	Percentage
64A	7/28/2014	1hr	Left	34.0	Right	66.0
	7/29/2014	1hr	Left	42.9	Right	57.1
	8/4/2014	1hr 34mins	Left	39.1	Right	60.9
	8/5/2014	1hr 15mins	Left	36.1	Right	63.9
	8/8/2014	42mins	Left	34.8	Right	65.2
	8/10/2014	53mins	Left	28.0	Right	72.0
	8/12/2014	1hr	Left	47.9	Right	52.1
	Sum./Ave.	7hr 24mins	37.5 ± 6.5		62.5 ± 6.5	
65A	9/3/14	30mins	Left	78.5	Right	21.5
	9/5/14	1hr10mins	Left	37.2	Right	62.8
	9/6/14	1hr15mins	Left	46.1	Right	53.9
	9/20/14	30mins	Left	42.5	Right	57.5
	Sum./Ave.	3hrs 25mins	51.1 ± 18.6		48.9 ± 18.6	
	Sum./Ave.	10hrs 49mins	42.5 ± 13.3		57.5 ± 13.3	

Exp.2: Silence v.s. White noise. (p < 0.0001)						
Monkey ID	Date	Duration	Silence		White Noise	
			Side	Percentage	Side	Percentage
64A	8/19/14	40mins	Left	66.8	Right	33.2
	8/21/14	20mins	Left	75.7	Right	24.3
	8/22/14	40mins	Right	76.5	Left	23.5
	8/24/14	24mins	Right	72.6	Left	27.4
	8/26/14	20mins	Left	75.2	Right	24.8
	Sum./Ave.	144mins	73.4 ± 3.9		26.6 ± 3.9	
65A	9/22/14	40mins	L	73.9	R	26.1
	9/23/14	40mins	L	86.6	R	13.4
	9/24/14	40mins	L	85.1	R	14.9
	9/25/14	40mins	R	89.0	L	11.0
	9/26/14	40mins	L	76.3	R	23.7
	9/27/14	40mins	R	93.9	L	6.1
	Sum./Ave.	240mins	86.2 ± 6.5		13.8 ± 6.5	
	Sum./Ave.	6hrs 24mins	79.2 ± 8.2		20.8 ± 8.2	

Exp.3: Consonance v.s. Dissonance. (p = 0.645)						
Monkey ID	Date	Duration	Consonance		Dissonance	
			Side	Percentage	Side	Percentage
64A (30mins)	9/5/14	30mins	Right	47.3	Left	52.7
	9/6/14	30mins	Right	48.6	Left	51.4
	9/15/14	30mins	Right	54.6	Left	45.4
	9/20/14	30mins	Right	51.3	Left	48.7
	9/22/14	30mins	Right	48.7	Left	51.3
	Sum./Ave.	150mins	50.1 ± 2.9		49.9 ± 2.9	
64A (1hr)	9/23/14	1hr	Left	58.1	Right	41.9
	9/24/14	1hr	Left	64.0	Right	36.0
	9/25/14	1hr	Left	66.5	Right	33.5
	9/26/14	1hr	Left	29.7	Right	70.3
	9/27/14	1hr	Left	89.0	Right	11.0
	9/28/14	1hr	Left	81.0	Right	19.0
	9/29/14	1hr	Left	60.0	Right	40.0
	9/30/14	1hr	Right	83.3	Left	16.7
	10/1/14	45mins	Right	13.2	Left	86.8
	10/14/14	1hr	Right	7.0	Left	93.0
	10/17/14	1hr	Right	9.7	Left	90.3
	10/18/14	1hr	Right	88.7	Left	11.3
	10/19/14	1hr	Right	13.5	Left	86.5
	Sum./Ave.	12hrs 45mins	51.1 ± 32		48.9 ± 32	
65A (1hr)	9/23/14	1hr	Left	20	Right	80
	9/24/14	1hr	Left	32.0	Right	68.0
	9/25/14	1hr	Left	28.7	Right	71.3
	9/26/14	1hr	Left	16.5	Right	83.5
	9/27/14	1hr	Left	30.2	Right	69.8
	9/30/14	1hr	Right	70.3	Left	29.7
	10/1/14	1hr	Right	70.3	Left	29.7
	10/14/14	1hr	Right	57.3	Left	42.7
	Sum./Ave.	8hrs	40.7		59.3	
	Sum./Ave.	20hrs 45mins	47.7 ± 25.5		52.3 ± 25.5	

Table 7. Records of preference tests from experiment 1~3.

Exp.4: Classical v.s. Contemporary piano music (p = 0.071)						
Monkey ID	Date	Duration	Classical (Bach)		Contemporary (Schoenberg)	
			Side	Percentage	Side	Percentage
64A	3/31/15	30mins	Left	83.2	Right	16.8
	4/1/15	30mins	Left	60.3	Right	39.7
	4/2/15	30mins	Left	71	Right	29.0
	4/3/15	30mins	Left	59.8	Right	40.2
	4/5/15	30mins	Right	29.4	Left	70.6
	4/6/15	30mins	Right	41.6	Left	58.4
	4/7/15	30mins	Right	94.2	Left	5.8
	4/10/15	30mins	Right	78.8	Left	21.2
	4/11/15	30mins	Left	82.2	Right	17.8
	4/13/15	30mins	Left	80.8	Right	19.2
	4/14/15	30mins	Right	32.4	Left	67.6
	4/15/15	30mins	Right	45.5	Left	54.5
	Sum./Ave.	6hrs	63.3 ± 21.8		36.7 ± 21.8	
29A	4/7/15	30mins	L	64.1	R	35.9
	4/10/15	30mins	L	47.9	R	52.1
	4/11/15	30mins	L	59.5	R	40.5
	4/13/15	30mins	L	50.7	R	49.3
	4/14/15	30mins	R	30.5	L	69.5
	4/15/15	30mins	R	51.7	L	48.3
	4/18/15	30mins	R	45.8	L	54.2
	5/22/15	30mins	R	40.7	L	59.3
	5/23/15	30mins	L	69.4	R	30.6
	5/29/15	30mins	L	62.7	R	37.3
	5/30/15	30mins	R	36.3	L	63.7
	5/31/15	30mins	R	50.9	L	49.1
	Sum./Ave.	6hrs	50.9 ± 11.7		49.2 ± 11.7	
	Sum./Ave.	12hrs	57.1 ± 18.2		42.9 ± 18.2	

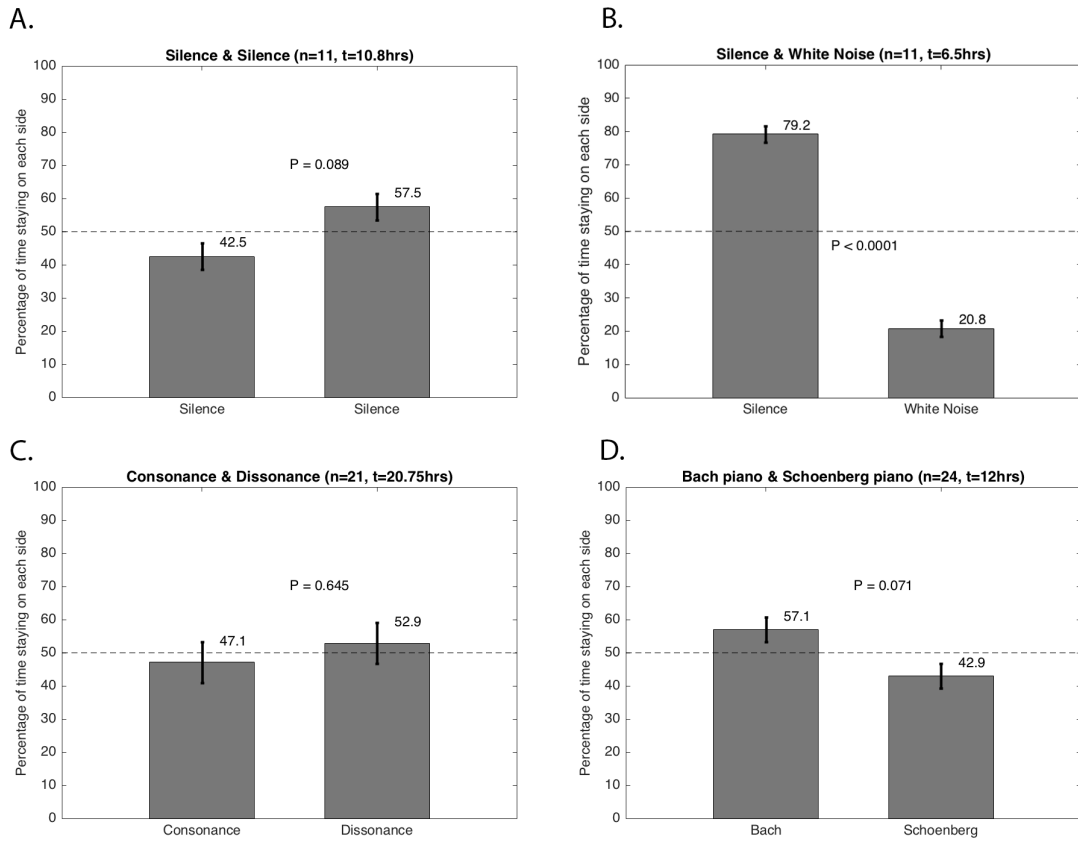Table 8. Recordings of preference test from Experiment 4.

Figure 21. Statistical results from all four experiments of preference tests (error-bars represent standard error).
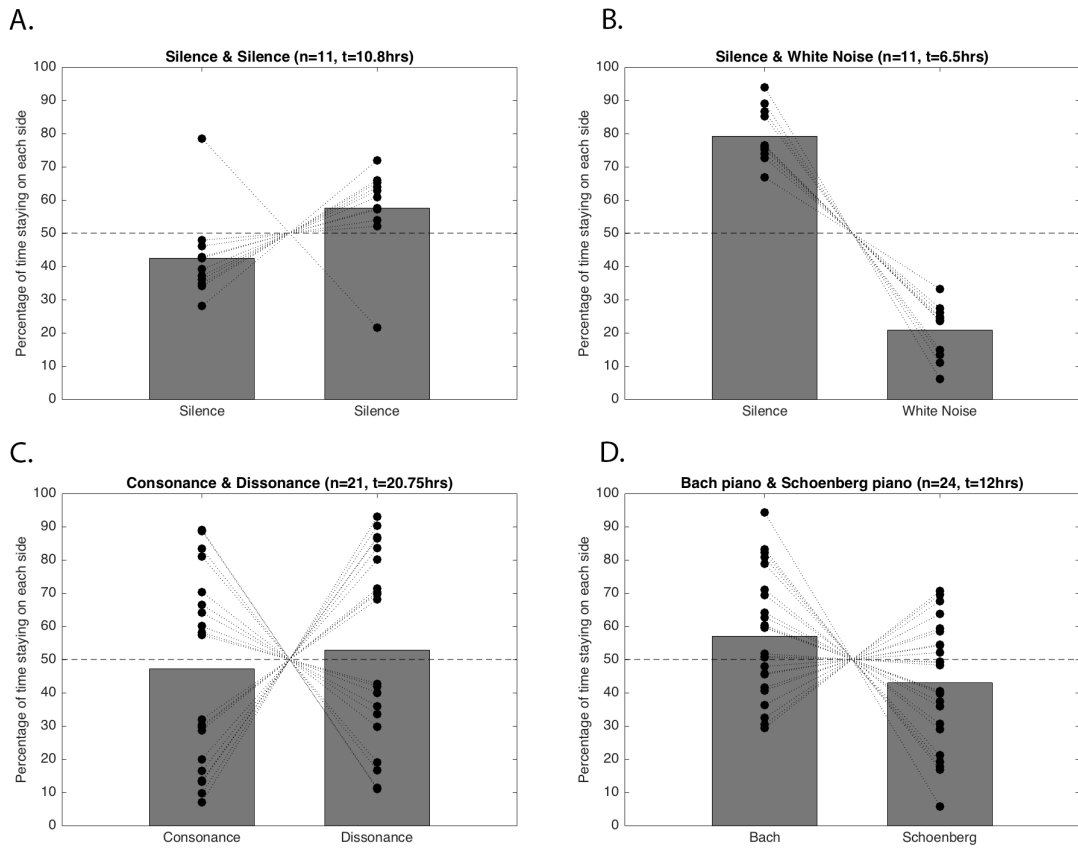
Figure 22. Results from all four experiments of preference test: individual sessions represented with dots; connected dots are from the same session.
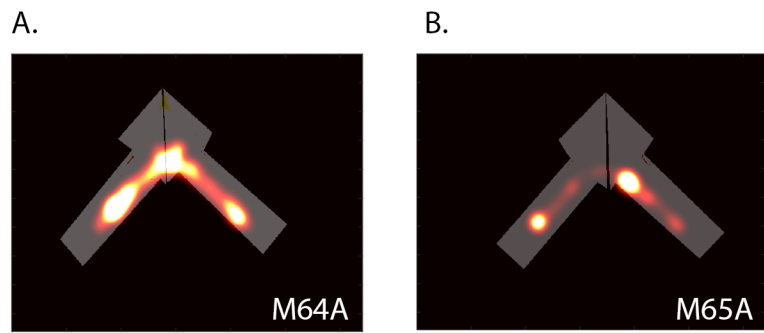

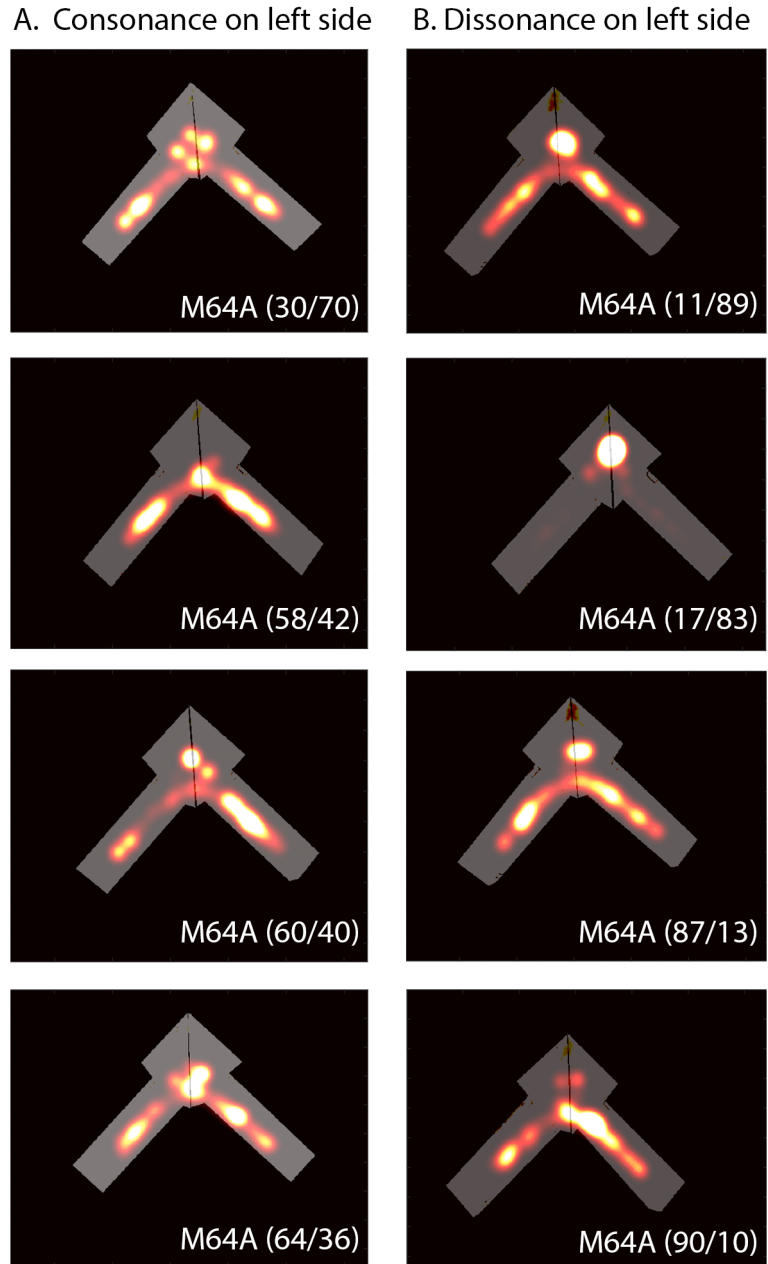Figure 23. Heat maps of the marmoset's location when no sound was playing

A. Consonance on left side   B. Dissonance on left side

M64A (30/70)   M64A (11/89)

M64A (58/42)   M64A (17/83)

M64A (60/40)   M64A (87/13)

M64A (64/36)   M64A (90/10)

Figure 24. Example heat maps of the marmoset (M64A) when one side pairs with consonant music, and the other side pairs with dissonant music.
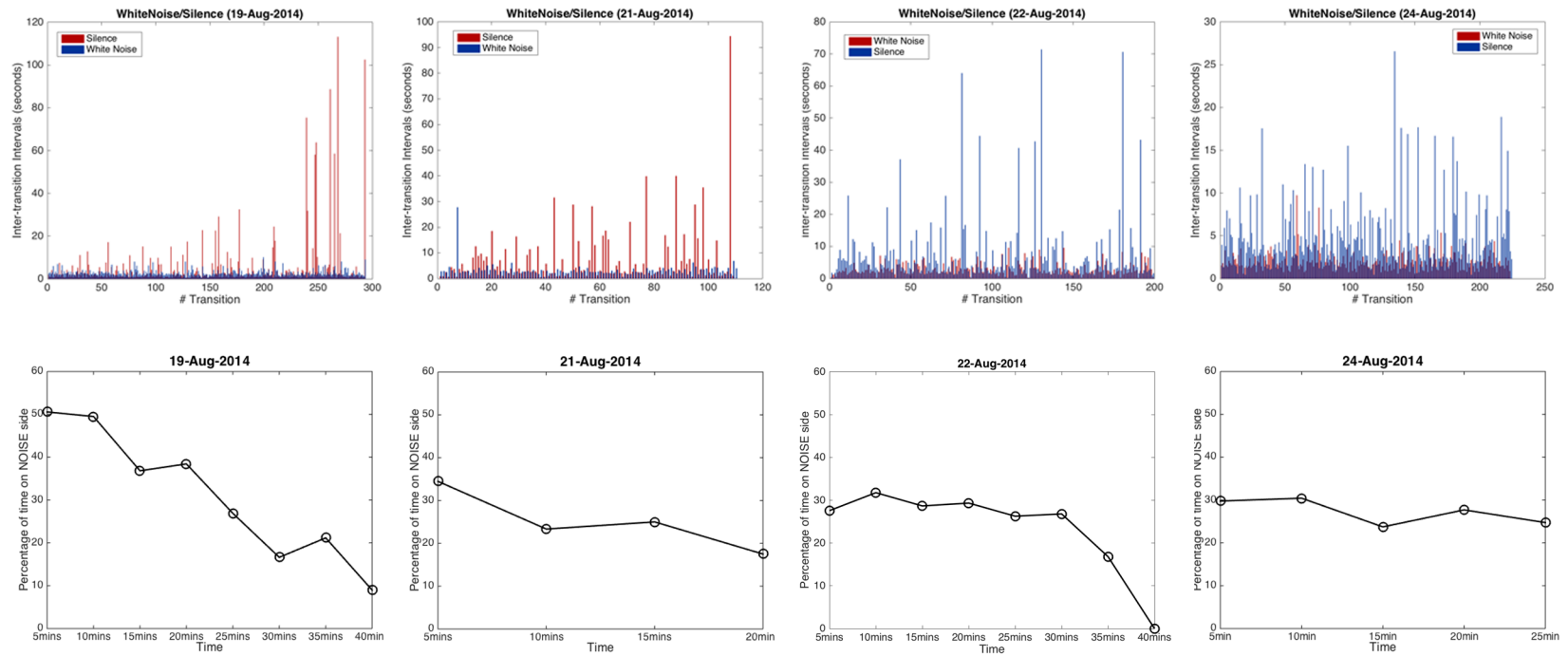
86

Figure 25. Experiment 2: four example sessions from M64A showing the learning process
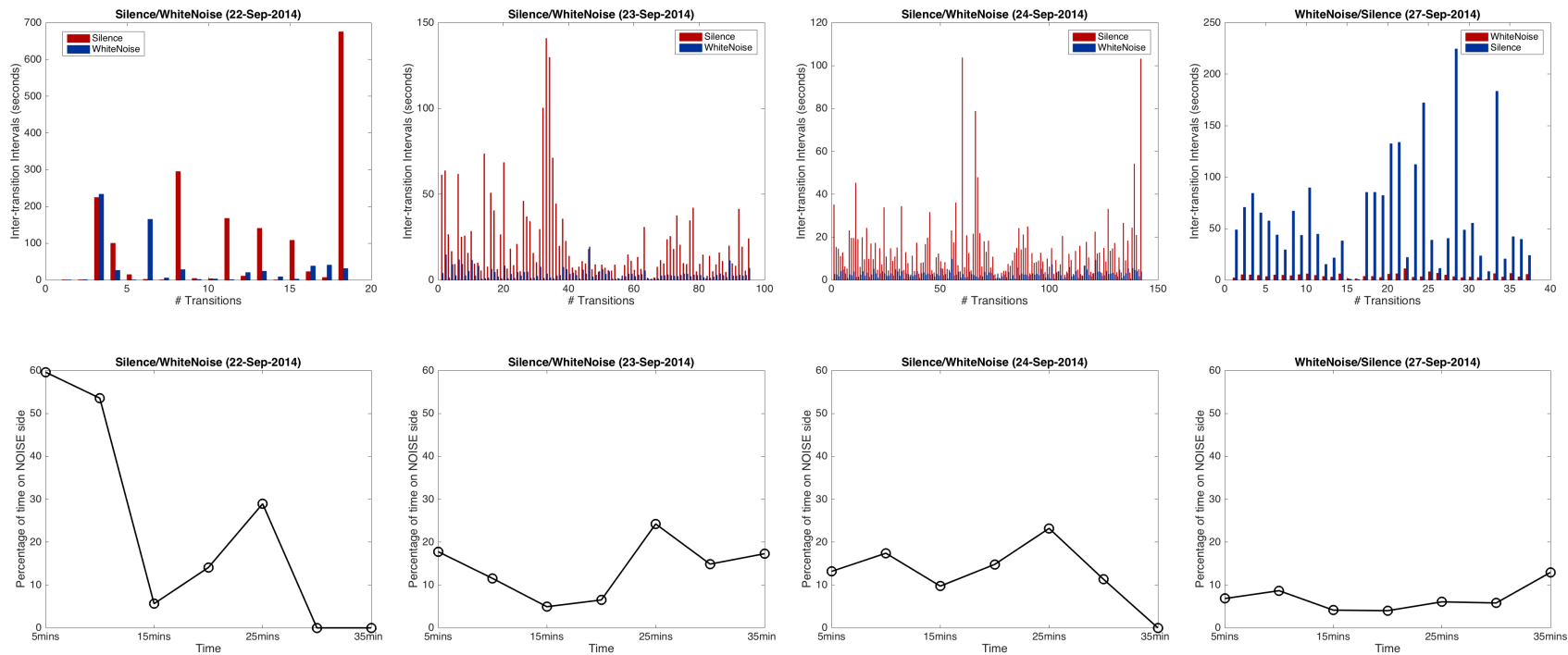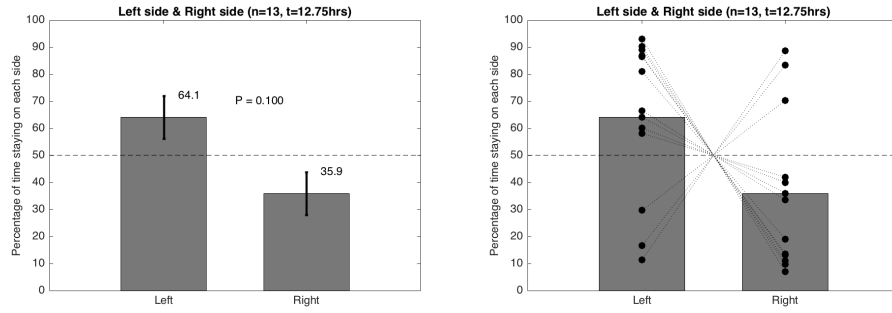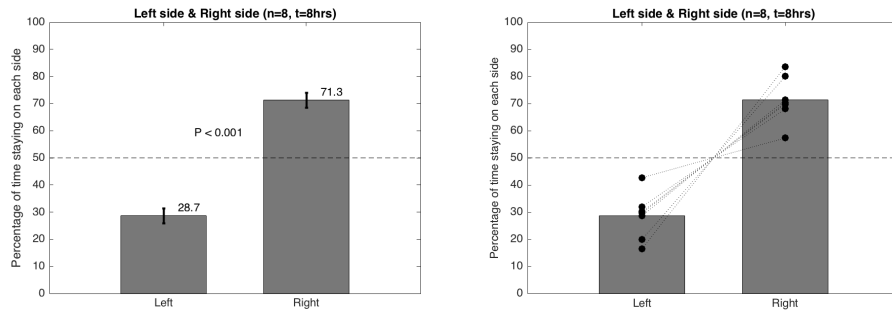
Figure 26. Experiment 2: four example sessions from M65A showing the learning process

Experiment 3: Data grouped by side (64A)

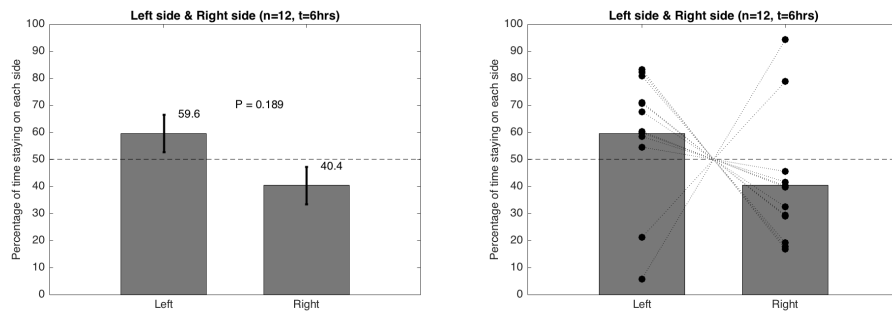# Experiment 3: Data grouped by side (65A)

Experiment 4: Data grouped by side (64A)
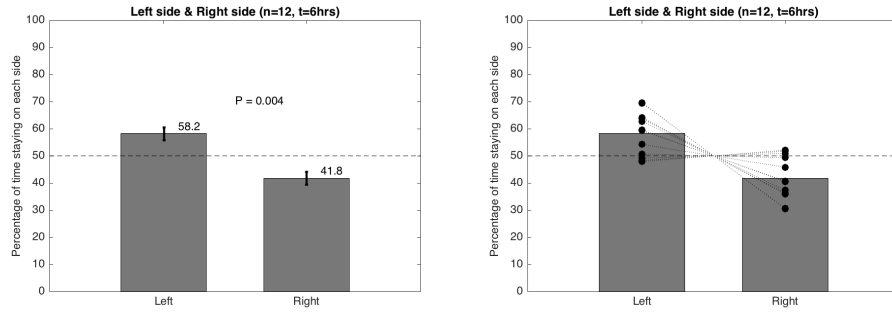

Experiment 4: Data grouped by side (65A)

Figure 27. Results from experiment 3 & 4 grouped by left or right side.
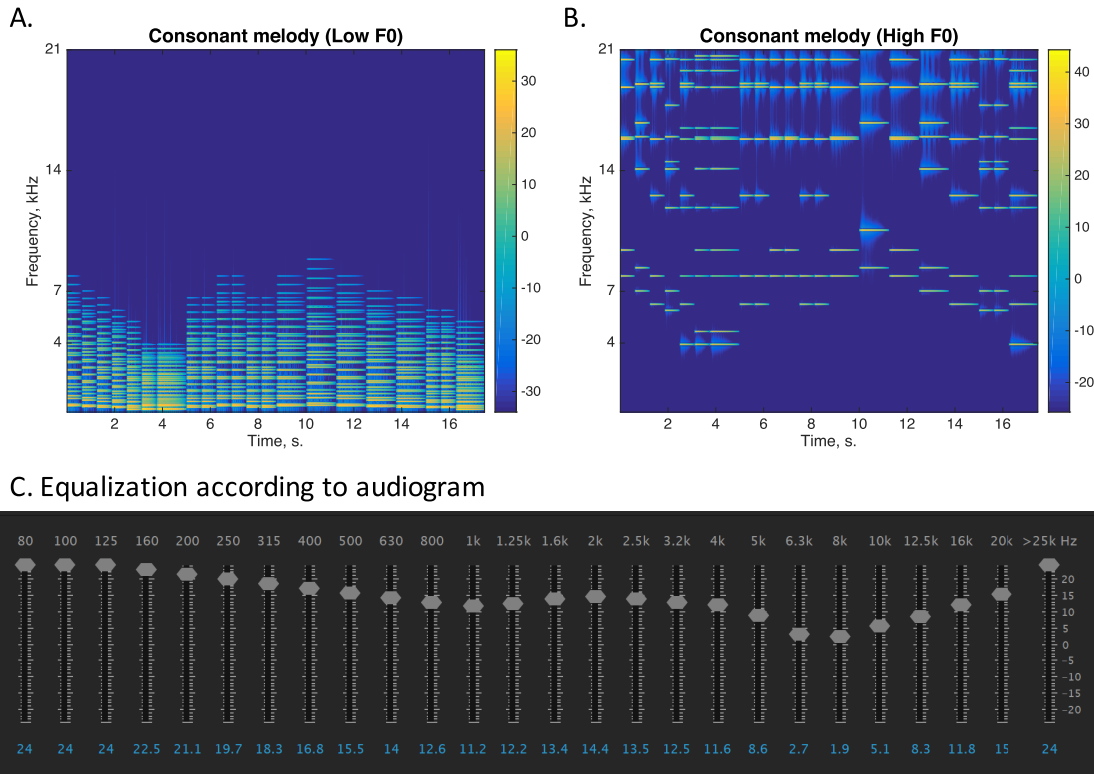
Figure 28. Spectrograms of original (low-F0) consonant melody and shifted (high-F0) consonant melody.
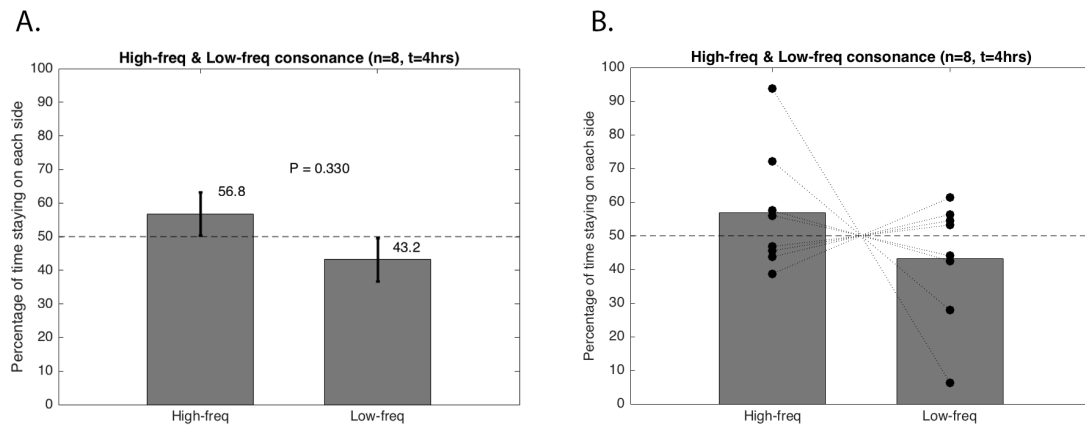
Figure 29. Results from experiment 5: shifted vs. un-shifted consonant melodies.

Exp.5: High frequency v.s. Low frequency Consonance (p = 0.330)						
Monkey ID	Date	Duration	High-frequency		Low-frequency	
			Side	Percentage	Side	Percentage
64A	6/23/15	30mins	R	72.1	L	27.9
	6/24/15	30mins	R	38.7	L	61.3
	6/26/15	30mins	L	56	R	44
	6/27/15	30mins	L	93.7	R	6.3
	Sum./Ave.	2hrs	65.1 ± 23.4		34.9 ± 23.4	
29A	6/23/15	30mins	R	45.5	L	54.5
	6/24/15	30mins	R	43.7	L	56.3
	6/26/15	30mins	L	46.8	R	53.2
	6/27/15	30mins	L	57.5	R	42.5
	Sum./Ave.	2hrs	48.4 ± 6.2		51.6 ± 6.2	
	Sum./Ave.	4hrs	56.8 ± 18.2		43.3 ± 18.2	

Table 9. Experiment records for experiment 5.

References:

Akiyama, Kayo, and Den'etsu Sutoo. 2011. "Effect of Different Frequencies of Music on Blood Pressure Regulation in Spontaneously Hypertensive Rats." *Neuroscience Letters* 487 (1). Elsevier Ireland Ltd: 58–60. doi:10.1016/j.neulet.2010.09.073.

Bendor, D, M S Osmanski, and X Wang. 2012. "Dual-Pitch Processing Mechanisms in Primate Auditory Cortex." *Journal of Neuroscience* 32 (46): 16149–61. doi:10.1523/JNEUROSCI.2563-12.2012.

Bendor, Daniel, and Xiaoqin Wang. 2005. "The Neuronal Representation of Pitch in Primate Auditory Cortex." *Nature* 436 (7054): 1161–65. doi:10.1038/nature03867.

Bernstein, Joshua G, and Andrew J Oxenham. 2003. "Pitch Discrimination of Diotic and Dichotic Tone Complexes: Harmonic Resolvability or Harmonic Number?." *The Journal of the Acoustical Society of America* 113 (6): 3323–12. doi:10.1121/1.1572146.

Burkart, Judith M, and Christa Finkenwirth. 2015. "Marmosets as Model Species in Neuroscience and Evolutionary Anthropology." *Neuroscience Research* 93 (April). Elsevier Ireland Ltd and Japan Neuroscience Society: 8–19. doi:10.1016/j.neures.2014.09.003.

Carlyon, Robert P. 1994. "Comparing the Fundamental Frequencies of Resolved and Unresolved Harmonics: Evidence for Two Pitch Mechanisms?." *The Journal of the Acoustical Society of America* 95 (6): 3541. doi:10.1121/1.409971.

Chiandetti, Cinzia, and Giorgio Vallortigara. 2011. "Chicks Like Consonant Music.." *Psychological Science* 22 (10): 1270–73. doi:10.1177/0956797611418244.

Cousineau, M, and J H McDermott. 2012. "The Basis of Musical Consonance as Revealed by Congenital Amusia." In. doi:10.1073/pnas.1207989109/-/DCSupplemental.

Cousineau, Marion, Andrew J Oxenham, and Isabelle Peretz. 2015. "Congenital Amusia_ a Cognitive Disorder Limited to Resolved Harmonics and with No Peripheral Basis." *Neuropsychologia* 66 (C). Elsevier: 293–301. doi:10.1016/j.neuropsychologia.2014.11.031.

DeCasper, A, and W Fifer. 1980. "Of Human Bonding: Newborns Prefer Their Mothers' Voices." *Science* 208 (4448): 1174–76. doi:10.1126/science.7375928.

DiMattina, Christopher, and Xiaoqin Wang. 2006. "Virtual Vocalization Stimuli for Investigating Neural Representations of Species-Specific Vocalizations.." *Journal of Neurophysiology* 95 (2): 1244–62. doi:10.1152/jn.00818.2005.

Elemans, Coen P H. 2014. "The Singer and the Song: the Neuromechanics of Avian Sound Production.." *Current Opinion in Neurobiology* 28 (October): 172–78. doi:10.1016/j.conb.2014.07.022.

Glasberg, B R, and BCJ Moore. 1990. "Derivation of Auditory Filter Shapes From Notched-Noise Data." *Hearing Research* 47 (1-2): 103–38. doi:10.1016/0378-5955(90)90170-T.

Goldstein, J L. 1973. "An Optimum Processor Theory for the Central Formation of the Pitch of Complex Tones." *The Journal of the Acoustical Society of America* 54 (6): 1496. doi:10.1121/1.1914448.

Henning, G B, and S L Grosberg. 1968. "Effect of Harmonic Components on Frequency Discrimination.." *The Journal of the Acoustical Society of America* 44 (5): 1386–89.

Ikkatai, Yuko, and Shigeru Watanabe. 2010. "Discriminative and Reinforcing Properties of Paintings in Java Sparrows (Padda Oryzivora)." *Animal Cognition* 14 (2): 227–34.

doi:10.1007/s10071-010-0356-3.

Johnson, Luke A, Charles C Della Santina, and Xiaoqin Wang. 2012. "Temporal Bone Characterization and Cochlear Implant Feasibility in the Common Marmoset (Callithrix Jacchus).." *Hearing Research* 290 (1-2): 37–44. doi:10.1016/j.heares.2012.05.002.

K Ohgushi. 1993. "Audibility of Partials in Inharmonic Complex Tones.." *The Journal of the Acoustical Society of America* 93 (1): 452–61.

Kameoka, A, and M Kuriyagawa. 1969a. "Consonance Theory Part I: Consonance of Dyads.." *The Journal of the Acoustical Society of America* 45 (6): 1451–59.

Kameoka, A, and M Kuriyagawa. 1969b. "Consonance Theory Part II: Consonance of Complex Tones and Its Calculation Method.." *The Journal of the Acoustical Society of America* 45 (6): 1460–69.

Kastner, M P, and R G Crowder. 1990. "Perception of the Major/Minor Distinction: IV. Emotional Connotations in Young Children." *Music Perception: an Interdisciplinary Journal* 8 (2): 189–201. doi:10.2307/40285496.

Kleinman, Kim. 2015. *Darwin and Spencer on the Origin of Music: Is Music the Food of Love? Music, Neurology, and Neuroscience: Evolution, the Musical Brain, Medical Conditions, and Therapies*. 1st ed. Vol. 217. Elsevier B.V. doi:10.1016/bs.pbr.2014.11.018.

Koda, Hiroki, Muriel Basile, Marion Olivier, Kevin Remeuf, Sumiharu Nagumo, Catherine Blois-Heulin, and Alban Lemasson. 2013a. "Validation of an Auditory Sensory Reinforcement Paradigm: Campbell's Monkeys (Cercopithecus Campbelli) Do Not Prefer Consonant Over Dissonant Sounds.." *Journal of Comparative Psychology* 127 (3): 265–71. doi:10.1037/a0031237.

Koda, Hiroki, Muriel Basile, Marion Olivier, Kevin Remeuf, Sumiharu Nagumo, Catherine Blois-Heulin, and Alban Lemasson. 2013b. "Validation of an Auditory Sensory Reinforcement Paradigm: Campbell's Monkeys (Cercopithecus Campbelli) Do Not Prefer Consonant Over Dissonant Sounds.." *Journal of Comparative Psychology* 127 (3): 265. doi:10.1037/a0031237.

Köppl, C. 1997. "Phase Locking to High Frequencies in the Auditory Nerve and Cochlear Nucleus Magnocellularis of the Barn Owl, Tyto Alba.." *The Journal of Neuroscience* 17 (9): 3312–21.

Laudanski, J, Y Zheng, and R Brette. 2014. "A Structural Theory of Pitch." *eNeuro* 1 (1): 1–18. doi:10.1523/ENEURO.0033-14.2014.

Licklider, J C R. 1951. "A Duplex Theory of Pitch Perception," March, 1–8.

Loeb, G E, M W White, and M M Merzenich. 1983. "Spatial Cross-Correlation. a Proposed Mechanism for Acoustic Pitch Perception.." *Biological Cybernetics* 47 (3): 149–63.

Masataka, Nobuo. 2006. "Preference for Consonance Over Dissonance by Hearing Newborns of Deaf Parents and of Hearing Parents." *Developmental Science* 9 (1). Wiley Online Library: 46–50.

McDermott, Josh H, Andriana J Lehr, and Andrew J Oxenham. 2010. "Individual DifferencesReveal the Basis of Consonance." *Current Biology* 20 (11). Elsevier Ltd: 1035–41. doi:10.1016/j.cub.2010.04.019.

McDermott, Josh, and Marc D Hauser. 2007. "Nonhuman Primates Prefer Slow Tempos but Dislike Music Overall.." *Cognition* 104 (3): 654–68. doi:10.1016/j.cognition.2006.07.011.

McDermott, Josh, and Marc Hauser. 2004. "Are Consonant Intervals Music to Their Ears?

Spontaneous Acoustic Preferences in a Nonhuman Primate..” *Cognition* 94 (2): B11–B21. doi:10.1016/j.cognition.2004.04.004.

Micheyl, Christophe, and Andrew J Oxenham. 2007. “Across-Frequency Pitch Discrimination Interference Between Complex Tones Containing Resolved Harmonics.” *The Journal of the Acoustical Society of America* 121 (3): 1621–11. doi:10.1121/1.2431334.

Micheyl, Christophe, Li Xiao, and Andrew J Oxenham. 2012. “Characterizing the Dependence of Pure-Tone Frequency Difference Limens on Frequency, Duration, and Level..” *Hearing Research* 292 (1-2): 1–13. doi:10.1016/j.heares.2012.07.004.

Moore, Brian C J. 1973. “Frequency Difference Limens for Short-Duration Tones.” *The Journal of the Acoustical Society of America* 54 (3): 610. doi:10.1121/1.1913640.

Moore, Brian C J. 1974. “Relation Between the Critical Bandwidth and the Frequency-Difference Limen..” *The Journal of the Acoustical Society of America* 55 (2): 359.

Moore, Brian C J, and Stephan M A Ernst. 2012. “Frequency Difference Limens at High Frequencies: Evidence for a Transition From a Temporal to a Place Code.” *The Journal of the Acoustical Society of America* 132 (3): 1542–46. doi:10.1121/1.4739444.

Moore, Brian C J, Moore, BCJ, and G A Moore. 2003. “Discrimination of the Fundamental Frequency of Complex Tones with Fixed and Shifting Spectral Envelopes by Normally Hearing and Hearing-Impaired Subjects.” *Hearing Research* 182 (1-2): 153–63. doi:10.1016/S0378-5955(03)00191-6.

Nelson, David A. 1983. “A General Equation Describing Frequency Discrimination as a Function of Frequency and Sensation Level.” *The Journal of the Acoustical Society of America* 73 (6): 2117. doi:10.1121/1.389579.

Osmanski, M S, X Song, and X Wang. 2013. “The Role of Harmonic Resolvability in Pitch Perception in a Vocal Nonhuman Primate, the Common Marmoset (Callithrix Jacchus).” *The Journal of Neuroscience* 33 (21). Society for Neuroscience: 9161–68. doi:10.1523/JNEUROSCI.0066-13.2013.

Osmanski, Michael S, and Xiaoqin Wang. 2011. “Measurement of Absolute Auditory Thresholds in the Common Marmoset (Callithrix Jacchus).” *Hearing Research* 277 (1-2): 127–33. doi:10.1016/j.heares.2011.02.001.

Plack, Christopher J, Andrew J Oxenham, and Richard R Fay. 2006. *Pitch*. Springer Science & Business Media. doi:10.3758/bf03207484.

Plomp, R, and W J Levelt. 1965. “Tonal Consonance and Critical Bandwidth..” *The Journal of the Acoustical Society of America* 38 (4): 548–60.

Remington, Evan D, Michael S Osmanski, and Xiaoqin Wang. 2012. “An Operant Conditioning Method for Studying Auditory Behaviors in Marmoset Monkeys.” Edited by Alain Claude. *PLoS ONE* 7 (10): e47895–98. doi:10.1371/journal.pone.0047895.

Shamma, S A. 1985a. “Speech Processing in the Auditory System. I: the Representation of Speech Sounds in the Responses of the Auditory Nerve..” *The Journal of the Acoustical Society of America* 78 (5): 1612–21.

Shamma, Shihab A. 1985b. “Speech Processing in the Auditory System II: Lateral Inhibition and the Central Processing of Speech Evoked Activity in the Auditory Nerve.” *The Journal of the Acoustical Society of America* 78 (5): 1622. doi:10.1121/1.392800.

Slee, Sean J, and Eric D Young. 2010. “Sound Localization Cues in the Marmoset Monkey.” *Hearing Research* 260 (1-2): 96–108. doi:10.1016/j.heares.2009.12.001.

Snowdon, C T, and D Teie. 2009. "Affective Responses in Tamarins Elicited by Species-Specific Music." *Biology Letters* 6 (1): 30. doi:10.1098/rsbl.2009.0593.

Snowdon, C T, and D Teie. 2013a. "Emotional Communication in Monkeys: Music to Their Ears?." *... Evolution of Emotional Communication: ....*

Snowdon, Charles T, and David Teie. 2013b. "Emotional Communication in Monkeys: Music to Their Ears?." In *Evolution of Emotional Communication*, 133–52. From Sounds in Nonhuman Mammals to Speech and Music in Man. Oxford University Press. doi:10.1093/acprof:oso/9780199583560.003.0009.

Snowdon, Charles T, David Teie, and Megan Savage. 2015. "Cats Prefer Species-Appropriate Music." *Applied Animal Behaviour Science* 166 (May). Elsevier B.V.: 106–11. doi:10.1016/j.applanim.2015.02.012.

Sugimoto, Tasuku, Hiromi Kobayashi, Noritomo Nobuyoshi, Yasushi Kiriyama, Hideko Takeshita, Tomoyasu Nakamura, and Kazuhide Hashiya. 2010. "Preference for Consonant Music Over Dissonant Music by an Infant Chimpanzee.." *Primates; Journal of Primatology* 51 (1): 7–12. doi:10.1007/s10329-009-0160-3.

Sullivan, Margaret W, and Michael Lewis. 2003. "Contextual Determinants of Anger and Other Negative Expressions in Young Infants.." *Developmental Psychology* 39 (4): 693–705. doi:10.1037/0012-1649.39.4.693.

Talwar, S K, and G L Gerstein. 1999. "A Signal Detection Analysis of Auditory-Frequency Discrimination in the Rat.." *The Journal of the Acoustical Society of America* 105 (3): 1784–1800.

Terhardt, Ernst. 1974. "Pitch, Consonance, and Harmony," October, 1–9.

Trainor, L J, and B M Heinmiller. 1998. "The Development of Evaluative Responses to Music:: Infants Prefer to Listen to Consonance Over Dissonance." *Infant Behavior and Development* 21 (1): 77–88. doi:10.1016/S0163-6383(98)90055-8.

Trainor, Laurel J, Christine D Tsang, and Vivian H W Cheung. 2002. "Preference for Sensory Consonance in 2- and 4-Month-Old Infants." *Music Perception* 20 (2): 187–94. doi:10.1525/mp.2002.20.2.187.

Wallace, Emma K, Mark Kingston-Jones, Matthew Ford, and Stuart Semple. 2013. "An Investigation Into the Use of Music as Potential Auditory Enrichment for Moloch Gibbons (Hylobates Moloch)." *Zoo Biology* 32 (4): 423–26. doi:10.1002/zoo.21074.

Watanabe, S, and M Nemoto. 1998. "Reinforcing Property of Music in Java Sparrows (< I> Padda Oryzivora</I>)." *Behavioural Processes* 43 (2): 211–18. doi:10.1016/S0376-6357(98)00014-X.

Wienicke, A, U Hausler, and U Jurgens. 2001. "Auditory Frequency Discrimination in the Squirrel Monkey." *Journal of Comparative Physiology a: Sensory, Neural, and Behavioral Physiology* 187 (3): 189–95. doi:10.1007/s003590100189.

Wier, C C, W Jesteadt, and D M Green. 1977. "Frequency Discrimination as a Function of Frequency and Sensation Level." *The Journal of the Acoustical Society of America* 61 (1): 178–84.

Zeitlin, L R, and R L Cramer. 1957. *Further Studies on Frequency Discrimination of Pure and Complex Tones*.

Zentner, M R, and J Kagan. 1998. "Infants' Perception of Consonance and Dissonance in Music." *Infant Behavior and Development*. doi:10.1016/s0163-6383(98)90021-2.

Curriculum Vitae
Yueqi Guo

yguo29@jhu.edu   720 Rutland Ave, Traylor 412. Baltimore, MD 21205

EDUCATION

8/2020 (expect)	**Johns Hopkins University, Baltimore, MD, U.S.A** · PhD, Biomedical Engineering.
8/2015 (expect)	**Johns Hopkins University, Baltimore, MD, U.S.A** · Master of Science and Engineering, Biomedical Engineering Department. ***GPA: 4.0/4.0***
7/2013	**Tsinghua University, Beijing, China** · Bachelor of Biomedical Engineering, School of Medicine. ***GPA: 86.4/100***

RESEARCH EXPERIENCE

Behavior

10/2013-present **Pitch perception mechanisms on marmoset monkeys**
Johns Hopkins University –PI: Dr. Xiaoqin Wang; Co-worker:Xindong Song, Michael Osmanski
- Goal: to investigate whether marmosets have the same pitch perception mechanisms as humans.
- Trained marmoset monkeys on pitch discrimination tasks with "licking food tube" behavioral paradigm.
- Conducting behavioral experiments daily to measured *fundamental frequency difference limen (F0DL)* for 3 animals, under pure tones and different manipulations of harmonic complex sounds.
- Demonstrated pitch perception mechanisms on marmosets are similar to those on humans.

10/2013-present **Preference for consonant music in marmoset monkeys**
Johns Hopkins University –PI: Dr. Xiaoqin Wang; Co-worker:Xindong Song
- Goal: to determine if marmosets prefer consonant over dissonant music or not.
- Designed and constructed experimental setup (two-branch maze enclosed by symmetric visual environment, with webcam monitor) for behavioral test.
- Programmed custom software for tracking animal's movement using image analysis, and for real-time closed-loop sound presentation based on animal's location.
- Conducting daily behavioral testing, musical stimulus generating, and behavioral data analysis.
- Demonstrated marmosets prefer silence to white noise; proved feasibility of the experimental system.

Physiology and computation

10/2012-6/2013 **Modeling of the spectra-temporal receptive fields (STRF) of neurons in the inferior colliculus (IC) of rats**
Tsinghua University –PI: Dr. Bo Hong; Co-worker: Wenbo Tang
- Goal: to classify and parameterize STRFs of IC neurons, provide a tool for online adaptive stimulus generation based on STRF features, to investigate selectivity and invariance of IC neurons
- Conducted single unit extracellular recording on rat's IC with Michigan multichannel probe electrode, estimated STRFs with dynamic moving ripple stimulus from 190 neurons.
- Programmed a software for modeling STRFs based on singular value decomposition method with Garbor filters; Extracted parameters that quantify different types (single-peak, multi-peak, oblique) of STRFs.
- Obtained consistent statistical data of STRF parameters with that in previous literatures.

8/2014-present **Two-photon imaging preparations in the auditory cortex of awake behaving marmoset monkeys**
Johns Hopkins University –PI: Dr. Xiaoqin Wang; Co-worker: Xindong Song
- Goal: to develop a chronic two-photon *in vivo* imaging technique in marmoset monkey's auditory cortex that enables investigation of functional organization in the auditory cortex.
- Developed reliable manufacture procedure of silicone Artificial Dura (50-60 μm thick membrane, to prevent tissue growth after removing the dura) and craniotomy maintainance procedures.
- Conducted acute surgery to practice craniotomy and Artificial Dura application.
- Conducted injection of AAV5 virus expressing GFP and Gcamp6s in the auditory cortex; Evaluated the

speed of liquid injection with micropipettes of different inner diameters.
- Obtained GPF expressing two-photon images with cellular structures.

Psychophysics

10/2013-
present

Consonance and Dissonance perception on humans
Johns Hopkins University –PI: Dr. Xiaoqin Wang; Co-worker: Dr. Juan Huang
- Goal: to investigate melodic consonance perception mechanisms in humans.
- Designed and conducted psychophysical experiments to measure subjects' preference for consonant and dissonant triads and melodies by rating for melodies with systematically manipulated structures.

Programming

7/2012-
9/2012

Cortical representation of sound localization in awake marmosets
Johns Hopkins University –PI: Dr. Xiaoqin Wang; Co-worker: Evan Remington
- Goal: to investigate responses in auditory cortex to single and multiple sound sources from a full, free field speaker array in awake marmosets using extracellular single-unit recording technique;
- Programmed with MATLAB to classify units with monotonic and non-monotonic sound level responses, analyzed neural responses to RLS (Random Location Stimulus) to estimate first order spatial receptive fields for monotonic units.

7/2012-
9/2012

Neuronal representation of harmonic spectral structure in auditory cortex of awake marmosets based on intracellular recording data
Johns Hopkins University –PI: Dr. Xiaoqin Wang; Co-worker: Lixia Gao
- Goal: to measure input currents in harmonic neurons through intracellular recording on marmosets' auditory cortex; investigate neuronal representation of harmonic complex sound.
- Programmed a user-friendly software in MATLAB to process and analyze sub-threshold membrane potentials**.**

2/2012-
7/2012

National College Students Original Experiment Plan: "Quantification of micro-bleeds in brain based on Magnetic Resonance Imaging"
Tsinghua University –PI: Dr. Hua Guo and Dr. Xihai Zhao
- Goal: to develop an accurately quantified method to detect micro-bleeds in the brain, thus help evaluation of the risk of stroke.
- Investigated literature and reported feasibility of using quantitative susceptibility map (QSM) to image cereberal micro-bleeds.
- Reconstruct MRI image with QSM to analyze the amount of cereberal micro-bleeds.

PRESENTATIONS
- X. Song, M. S. Osmanski, **Y. Guo**, X. Wang. Possible origins of human-like pitch perception mechanisms. *2014 Society for Neuroscience*

TECHNICAL SKILLS
- *Programming:* proficient in MATLAB; experience in C, Verilog, image processing with ITK and VTK
- *Signal processing:* experimental capabilities of electric circuit and MCU programming
- *Electrophysiological experiments:* experience of single unit recording on rats inferior colliculus
- *Animal behavior*: experience in training marmoset monkeys on behavioral tasks

HONORS AND AWARDS
- Academic Scholarship from Tsinghua University (roughly 10%)
- Scholarship for ethnic minorities from Tsinghua University (as Manchu)
- Silver award, Summer social practice team (5/450 teams)
- "Excellent student leader" of Student Union, Tsinghua University (2/80)
- "First-level Excellence" award in piano, Tsinghua Art Specialty Student test (8/300+)

MEMBERSHIP
- Student Member: *Association for Research in Otolaryngology (ARO)*
- Student Member: *Society for Neuroscience (SfN)*
- Committee & Student Member: *Advances and Perspective in auditory neurophysiology (APAN)*

Edited abstract book for APAN 2014

STANDARD TESTS

2/2011	GRE General (paper-based): Verbal 490, Quantitative 800, Analytical Writing 3.0
9/2012	TOEFL iBT: 103 (Reading 27, Listening 24, Speaking 24, Writing 28)

LEADERSHIP

7/2010-	**Leader, social practice team investigating gaze-limiting policy, in Inner Mongolia**
8/2010	· Lead the 10-member team work on social research, field work, and forum with the government
9/2010-	**Captain, Clavier Team, THU Student Art Group**
6/2011	· Organized the Clavier Team concerts (twice) and entertainment activities for team members.
	· Responsible for the professional work, provided guidance about performance for team members.
9/2011-	**Social Chair of the Student Union, School of Medicine**
7/2012	· Directed the Annual Art Festival, the New Year Party, and Singing Competition of the School of Medicine