Mechanisms for Coding Pitch

A Dissertation
SUBMITTED TO THE FACULTY OF THE
UNIVERSITY OF MINNESOTA
BY

Kelly L. Whiteford

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Advisor: Andrew J. Oxenham, Ph.D.

May 2018

Acknowledgments

My undergraduate mentors at UC Berkeley, particularly Bill Prinzmetal, Karen Schloss, Erv Hafter, and Steve Palmer, provided early support for pursing my graduate degree. They gave me a sense of purpose and well-being at a period in my undergraduate when I needed it most.

I give full credit to Erv Hafter for encouraging me to apply to Minnesota. I liked it here so much that I have decided to stay a bit longer.

As a first-year graduate student, Neal Viemeister allowed me to have my desk in his N625 lab. Because of our proximity, he had to entertain many of my naïve questions. He always made time to give me advice on signal detection theory or provide a critical eye to the interpretation of my data. Chapter 6 began from one of these conversations. I am thankful for his input.

The large-sample studies described in Chapters 3 and 4 were completed in collaboration with Heather Kreft, who recruited, screened, and ran most of the participants. Heather is a data-collecting machine and provided a lot of helpful advice for working with older and hearing-impaired listeners. I am very fortunate to have been able to work with her.

There were numerous undergraduates who worked with me the past 5 years, making a substantial portion of the data collection possible: Jiao Xu (Chapters 2, 5, and 6), Lindsey Ma (Chapters 2 and 5), Jeffrey Boucher (Chapters 5 and 6), Nathan Torunsky (Chapter 6), Cella Westray (Chapter 6), Ana Benson (Chapter 6), and Kristen Chalmers (Chapter 6).

The experiments conducted were supported by Andrew Oxenham's NIH Grant No. R01 DC005216 and by an Eva O. Miller Fellowship.

My friends and colleagues in the Auditory Perception and Cognition Lab made the journey fun and go by quickly: Emily Allen, Jordan Beim, Andy Byrne, Marion David, Coral Dirks, Lei Feng, Jackson Graves, Daniel Guest, Heather Kreft, Bonnie Lau, Hao Lu, Sara Madsen, Anahita Mehta, Juraj Mesik, Erin O'Neill, Chhayakanta Patro, Dorea Ruggles, and Magda Wojtczak. I've been incredibly fortunate to be surrounded by so many intelligent, hard-working, and kind people who all have similar interests.

Jackson, Emily, and Anahita were my go-to experts when I needed advice or just wanted to bounce around ideas. They are fantastic scientists and friends.

It is easy to be fueled by inspiration on days that go well. On more difficult days, Anahita helped keep my head on straight. Thank you for your positivity, your levelheadedness, and your friendship.

I am especially thankful for my committee members, Dan Kersten, Magda Wojtczak, Peggy Nelson, and my advisor, Andrew Oxenham. They took time out of their hectic schedules to review this dissertation, attend my defense, and provide their input.

I am grateful to have had the opportunity to work with Andrew. It has been extremely convenient to have an advisor with a seemingly endless memory span for psychoacoustics literature. But he is as kind and patient as he is brilliant. Thank you, Andrew, for being calm and coolheaded and always having confidence in me.

Matthew Scauzillo assisted with creating the figures for Chapter 2, and then helped hone my Adobe Illustrator skills so I could do the rest of the chapters myself. More importantly, he provided moral support, home-cooked meals, and transportation to Elliott Hall during (and in between) Minnesota's frigid winters. Thanks, Matt, for being my biggest supporter, my most reliable subject, and my best friend.

Lastly, I would like to thank my family, especially my parents, Rita and Joseph Whiteford, for their unconditional love and support.

Abstract

Pitch, the perceptual correlate of a sound’s frequency, is a fundamental attribute in speech and melody perception. We utilized individual differences across listeners with normal and disordered hearing to better understand how pitch is represented in the auditory system. Results from young, normal hearing listeners and listeners varying in age suggested the bulk of variability in sensitivity to modulations in frequency (FM) and amplitude (AM) likely reflects central, rather than peripheral, limitations. For listeners varying in degree of sensorineural hearing loss, however, sensitivity to FM was directly related to the fidelity of tonotopic (place) coding within the cochlea. This was contrary to the widely accepted understanding that FM is represented by precise, phase-locked spike times in the auditory nerve. To test the role of central processes on pitch perception, several experiments were conducted on listeners with congenital amusia, a neurogenetic disorder characterized by poor fine-grained pitch perception, unrelated to peripheral coding. We found that amusic deficits extend beyond poor pitch discrimination, including poor discrimination for high frequencies as well as poor detection for FM and AM tones. Despite the long-held understanding that amusia is a life-long deficit for pitch and music, impervious to training, we found rapid learning for pitch and melody discrimination in amusia. The learning effects were large and maintained for at least one-year. Overall, the findings suggest peripheral place coding is important for the fidelity of pitch, but many processes beyond the periphery can also contribute to variability in pitch perception.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AM: Amplitude modulation

AMDL: AM difference limen

F0: Fundamental frequency

FDL: Frequency difference limen

FFR: Frequency following response

FM: Frequency modulation

FMDL: FM difference limen

HI: Hearing impaired

HL: Hearing level

ILD: Interaural level difference

IPD: Interaural phase difference

ITD: Interaural time difference

NH: Normal hearing

PCA: Principal components analysis

RAU: Rationalized arcsine unit

rIFG: Right inferior frontal gyrus

SNHL: Sensorineural hearing loss

STG: Superior temporal gyrus

TFS: Temporal fine structure

CHAPTER 1: INTRODUCTION

Pitch is a psychological attribute of sound that allows listeners to order sounds from low to high and is closely related to the frequency or rate of periodic patterns of pressure fluctuations over time. This characteristic of ordered sound is fundamental to our ability to perceive and recognize contours in speech and music, as well as perceive harmonies between multiple pitches. One criterion for whether or not a sound has a pitch is whether one can form a recognizable melody out of it (Burns and Viemeister, 1976; Plack et al., 2005). By this common definition, pitch is a necessary attribute for music perception. In addition to communicating melodies and harmonies, pitch is essential for identifying voices and communicating speech prosody. In the English language, prosody can differentiate questions from statements, while in tonal languages, such as Mandarin, changes in pitch convey semantic meaning of words. Changes in pitch are also used to help convey emotion, such as happiness versus sadness (e.g., Schellenberg et al., 2000; Coutinho and Dibben, 2013). When there are multiple talkers in an environment (or multiple melodic lines), pitch is a strong cue for segregating objects, allowing us to attend to one speaker in a crowded cocktail party or pick out one melodic line from a trio of accompanying instruments (Bregman, 1990). Because pitch is vitally important in many everyday activities, changes in pitch perception due to natural aging or hearing disorders may profoundly affect one's quality of life.

One source of variability in the perception and discrimination of pitch may be related to the efficacy of frequency coding in the peripheral auditory system (i.e., the cochlea and auditory nerve). Before sound can be processed by the brain, the signal must first be filtered and transformed to neural impulses, so that any loss of information due to poor peripheral coding will result in a noisier neural signal. There is a great deal of literature indicating that frequency coding degrades with normal aging, sensorineural hearing loss, and auditory neuropathy/synaptopathy (e.g., Moore, 2008; Rance and Starr, 2015). It is less clear whether differences in pitch perception in the normal hearing (NH) population can be accounted for by differences in peripheral coding fidelity.

Identifying the specific mechanism(s) involved in peripheral coding for frequency is a challenging question that has vexed psychoacousticians for some time. One reason for this complication is there are two possible physiological codes available to the brain: a rate-place code (tonotopy) and a temporal code (phase-locking) (discussed further in section 1.1.1). One goal of this dissertation is to utilize large-sample individual differences in frequency coding in normal and disordered hearing to identify the peripheral mechanisms responsible for coding pitch.

However, pitch is by no means a one-to-one mapping between the peripheral frequency representation and the pitch salience (described further in section 1.2). For example, people with congenital amusia, a neurogenetic disorder in melody perception, have normal peripheral coding yet poor fine-grained pitch perception (Cousineau et al., 2015). Furthermore, pitch discrimination is susceptible to rapid auditory learning (e.g., Micheyl et al., 2006), perhaps suggesting changes in central rather than peripheral coding. A second, related goal is to better understand the central processes involved in pitch representation in normal and disordered perception. Understanding the central mechanisms involved in pitch is important because these factors are less likely to be amenable to intervention by sensory devices, such as hearing aids, but are more likely to be affected by auditory training paradigms.

This chapter provides a brief review of background information on peripheral and central mechanisms for coding pitch, beginning with a review of the basic peripheral mechanisms for coding pitch of varying types of stimuli (section 1.1). A substantial portion of this section is devoted to the background of frequency and amplitude modulated pure tones, as these well-controlled stimuli provide an efficient means of dissociating pitch coding mechanisms while controlling for task demands. Section 1.2 describes central factors that also contribute to variability in pitch perception and how amusics can be compared to non-amusic controls to better understand central influences on pitch perception. Section 1.3 discusses how large-sample studies on individual differences in audition can be used to compare the two possible peripheral mechanisms for coding pitch. The last section of this chapter outlines a series of experiments that were conducted using both large-scale samples, utilizing individual differences in normal and disordered hearing, and smaller-scale samples of special populations with amusia and matched controls, in order to better understand how pitch is represented in the auditory system.

1.1 Review of Peripheral Frequency Coding

1.1.1 Basic mechanisms for coding pure-tone pitch

How frequency is coded in the periphery is an open question, even for pure tones, the simplest of sounds (Plack et al., 2005; Oxenham, 2013). There are two primary physiological cues available to the brain: 1) Frequency may be mapped to the *place* of maximal excitation along the basilar membrane, leading to an increase in the *rate* of firing in the corresponding auditory nerve fibers (rate-place code), or 2) frequency may be mapped to the temporal phase-locked firing of action potentials in the auditory nerve, providing precise *timing* information about the periodicity (temporal code). There is evidence that the brain may use one or both of these mechanisms, either alone or in combination.

The general consensus is that for pure tones, lower frequencies (f < ~4-5 kHz, but perhaps as high as 8 kHz) may be coded via phase-locking to the temporal fine structure (TFS) of the waveform, while higher frequencies may be coded via a rate-place code (e.g., Moore, 1973; Moore and Ernst, 2012). Evidence for temporal coding at lower frequencies is supported by a combination of physiological studies in animals and behavioral data in humans. In animals, more direct measures of phase-locking can be attained using physiological recordings of the auditory nerve, where phase-locking is quantified as the synchrony index. In many mammals, phase-locking begins to degrade quite rapidly above about 1-2 kHz, and the exact limit of phase-locking depends on the species (Rose et al., 1967; Johnson, 1980; Palmer and Russell, 1986; Taberner and Liberman, 2005). For example, the upper limit of neural synchrony in guinea pigs and mice, ~3.5-4 kHz, (Palmer and Russell, 1986; Taberner and Liberman, 2005) is about 1-1.5 kHz lower than the upper limit of phase-locking in cats, ~5 kHz (Johnson, 1980). Physiological recordings of auditory nerve synchrony are too invasive for use in humans, so there is currently no direct, physiological measure in humans to quantify the limits of phase-locking. However, it is generally assumed that the phase-locking limits in the human auditory nerve are similar to those found in other mammals.

In contrast to the temporal code, the rate-place code should be less dependent on frequency: many studies of cochlear tuning have shown that it is the filter sharpness that underlies the rate-place code (e.g., Flanagan and Guttman, 1960; Bernstein and Oxenham, 2006a, 2006b; Oxenham, 2012). If filter sharpness is the limiting factor in coding pure-tones, then pitch discrimination in humans should not be related to the limits of phase-locking. In a classic behavioral study, Moore (1973) measured frequency difference limens (FDLs) in humans for pure-tone frequencies between 250 and 8000 Hz. The FDL is the smallest change in frequency that a participant can discriminate at a predefined threshold (e.g., the 70.7% correct point), where smaller FDLs indicate better pitch perception. A change in frequency of 6% roughly corresponds to a semitone, or the smallest pitch change in western music. Moore found that FDLs were as low as .14% at frequencies between 1-2 kHz, but FDLs increased (worsened) quite steeply between 4-8 kHz. Moore suggested that the good performance at low frequencies may be mediated by a temporal code, and that the poorer performance above about 4 kHz may be due to the breakdown in phase-locking. More recent measures of pure-tone FDLs at higher frequencies found that FDLs saturate at 8 kHz, with equally high thresholds up to frequencies of 14 kHz (Moore and Ernst, 2012). This was interpreted as suggesting that the contribution of timing information gradually decreases from 4-8 kHz, with only place information remaining available at and above 8 kHz. The saliency of pitch may be directly linked to the limits of phase-locking, at least for pure tones. For example, Ward (1954) found that people were unable to recognize pure-tone musical intervals when frequencies were above about 4-5 kHz. Similarly, recognition of familiar melodies also degrades above this limit (Attneave and Olson, 1971). Hence, for pure tones, the existence region of pitch appears to be closely linked to the limits of phase-locking to TFS cues in the auditory nerve.

1.1.2 Complex pitch

Place versus time theories of pure tone pitch perception can be extended to understand the pitch of complex tones, which make up the majority of sounds we encounter in our everyday environment. Complex tones are a sum of two or more pure tones (partials). Complex tones can be harmonic or inharmonic; the former has harmonics spaced at integer multiples of F0, while the latter does not. In the case of harmonic complex tones, the pitch usually corresponds to the repetition rate, or fundamental frequency (F0), of the waveform. The components of a complex tones can vary in resolvability, which is generally linked to harmonic number (e.g., Plomp, 1964; Bernstein and Oxenham, 2003). “Resolved” versus “unresolved” harmonics are distinct from the psychological definition of resolvability, which refers to the stability of particular chords in a musical phrase. In the present context, resolved harmonics are those components that fall into separate filters along the basilar membrane (i.e., frequencies with separate place representations); these are generally the lower-order harmonics (i.e., 1-5, but perhaps as high as 10) and produce peaks in the excitation pattern. Unresolved harmonics occur when two frequencies stimulate the same auditory filter along the basilar membrane (i.e., their “place” representations overlap); higher-order harmonics are usually unresolved. When this happens, the frequencies interact within the cochlea to create an envelope repetition rate corresponding to the frequency difference between the harmonics. Conveniently, these envelope fluctuations are equivalent to the F0 when the complex tone is harmonic. Unresolved harmonic complex tones are interesting for models of pitch because the F0 can only be inferred by using a timing code, while resolved harmonic complex tones contain both place and timing information. Pitch perception of unresolved harmonics suggest we are capable of perceiving pitch even when there is no useful place information of the F0, again supporting a timing code. However, unresolved complex tones tend to have a much weaker pitch salience, suggesting they use a different mechanism from resolved complex tones (Houtsma and Smurzynski, 1990; Carlyon and Shackleton, 1994), although some evidence suggests differences may be attributed to harmonic number rather than resolvability, per se (e.g., Bernstein and Oxenham, 2003).

Another interesting aspect of complex pitch perception is the phenomenon of the pitch of the missing fundamental (also known as periodicity pitch or virtual pitch). For harmonic complex tones, the F0 does not need to be present to perceive a pitch corresponding to F0. In fact, F0 can be perceived even when masking noise is presented in the frequency region of F0, ensuring the perception of the F0 is not due to cochlear distortion products (Licklider, 1956). At face value, the perception of the missing fundamental seems problematic for place models. How can one perceive a pitch at F0 when there is no place representation at F0? But, as long as there are at least two, resolved, successive harmonics, information about the F0 is present in both the *pattern* of the spectral content (i.e., the pattern of vibration along the basilar membrane) and in the pooled firings of action potentials in the auditory nerve. Either place or timing codes could represent the missing F0 of a resolved complex tone.

Pattern-matching models were developed in part to address the mechanisms for coding periodicity pitch. Pattern models need both 1) a representation of the spectral pattern of the input stimulus and 2) a set of harmonic templates stored in the brain (de Cheveigné, 2005). Some pattern models suggest that pitch is represented based on the pattern of excitation along the basilar membrane, consistent with place coding, which is preserved in the rate of firings of the auditory nerve fibers (Goldstein, 1973; Wightman, 1973; Terhardt, 1974). Other pattern matching models could be applicable to a purely time-based code (Goldstein, 1973; Srulovicz and Goldstein, 1983) or a combination of place and timing cues (e.g., Shamma and Klein, 2000). Pattern models are contingent on the assumption that the frequency spectrum of the input can be compared to a stored harmonic template. The pitch will correspond to the best matching template, even when the F0 is missing. One problem with pattern matching models is that it is not clear where/how harmonic templates are implemented in the brain. In addition, pattern models work well for resolved complex tones but work less well for explaining the pitch of unresolved harmonics (e.g., Bernstein and Oxenham, 2003) or amplitude-modulated noise (Burns and Viemeister, 1976, 1981).

Temporal models tend to be applicable to both resolved and unresolved harmonics (e.g., de Cheveigné, 2005). One method for extracting time-based periodicity is the autocorrelation function, which can calculate F0 based on phase-locking to TFS or envelope cues. Licklider (1951) proposed that the F0 can be calculated by comparing the input signal to a time-delayed representation of the same signal. This means that correlations of the time-delayed representations will be highest for delays at integer multiples of the period of the waveform. Most autocorrelation models are capable of calculating periodicity pitch and the F0 of unresolved harmonics. The strength of the autocorrelation model, its application to both resolved and unresolved harmonics, is also its weakness. Autocorrelation works *too well* for predicting pitch perception of unresolved complex tones (Shackleton and Carlyon, 1994). F0DLs tend to be much worse for unresolved compared to resolved harmonics (Houtsma and Smurzynski, 1990; Bianchi et al., 2015), suggesting that time-based models that work equally well regardless of resolvability cannot be the entire explanation for how the brain computes F0.

Some models have been developed that rely on a combination of place and timing cues (e.g., Loeb et al., 1983; Shamma and Klein, 2000). In these models, the extraction of phase-locked timing information may rely on cochlear place cues by assuming an array of coincidence detectors calculates the instantaneous correlation between the outputs of each of the cochlear filters. In Shamma and Klein's (2000) model, even if given just broadband noise, correlations in the array of coincidence detectors will be highest at detectors that correspond to harmonically spaced places along the basilar membrane. Shamma and Klein's model is a nice extension to pattern matching models in that it provides a realistic model for how harmonic templates might be learned without needing a lot of exposure to specific kinds of stimuli. Their model is physiologically plausible; for example, it is possible that an array of coincidence detectors may be implemented in the cochlear nucleus, although no direct evidence exists yet.

Perhaps one of the stronger arguments for temporal models, or combined models such as Shamma and Klein's, is the purported necessity of temporal coding for forming pure-tone melodies (Attneave and Olson, 1971) and for the perception of periodicity pitch of complex tones (Ritsma, 1962). More recent findings, however, suggest that temporal coding may not be necessary or sufficient for pitch perception (Oxenham et al., 2004, 2011). Oxenham et al. (2004) measured pitch discrimination and pitch-matching for transposed stimuli. Transposed stimuli have the TFS of lower-frequency tones presented to a higher-frequency place along the basilar membrane, thereby dissociating the natural covariation between place and temporal cues. Participants were considerably worse at pitch discrimination and pitch-matching for the transposed stimuli relative to their non-transposed counterparts, demonstrating that accurate place cues are necessary for complex pitch perception. A later study suggested that place cues alone may be sufficient for perceiving the periodicity pitch of complex tones (Oxenham et al., 2011). Oxenham et al. (2011) found accurate pitch matching and melody discrimination was possible for harmonic complex tones, even when all the frequency components were presented above the purported limit of phase-locking. In their study, the F0 was always below the limit of phase-locking but never present, so that participants had to hear out a periodicity pitch. Broadband threshold equalizing noise was presented, preventing the presence of distortion products. Oxenham et al. manipulated the F0 across trials, so that in some cases (e.g., F0 = 2000 Hz), some of the lower-harmonics present (e.g., H3 and H4) were resolved on the cochlea but all harmonics were above the limit of phase-locking. Melody discrimination in this case was better than melody discrimination of the same complex tones frequency-shifted (i.e., unresolved), where participants presumably used temporal coding to the envelope rather than place cues. This suggests that either the limit of phase-locking in humans is higher than that observed in many animals or that phase-locking is not sufficient for periodicity pitch.

1.1.3 Peripheral coding for frequency modulated pure tones

Another method for studying frequency coding is by measuring detection thresholds for frequency-modulated (FM) and amplitude-modulated (AM) pure tones. FM tones change in frequency over time and are described by the following equation:

$$x(t) = \sin\{2\pi f_c t + \beta \sin(2\pi f_m t)\}$$

where, $x(t)$ refers to pressure variation over time t, f_m is the modulation rate, or the number of cycles of changes per second, f_c is the carrier frequency, and β is the modulation index, defined as $\Delta f / f_m$, where Δf is the frequency excursion from the carrier. For AM tones, f_c is constant in frequency but changes in amplitude over time, as described by the following equation:

$$x(t) = A\{1 + m * \sin(2\pi f_m t + \varphi)\} \sin(2\pi f_c t)$$

A is the peak amplitude, m is the modulation depth, and φ is the starting phase. FM difference limens (FMDLs) correspond to the smallest peak-to-peak frequency change that a participant can detect at a pre-defined threshold. Analogously, AM difference limens (AMDLs) correspond to the smallest detectable modulation depth (m), where m is a proportion of peak amplitude varying from 0 (no modulation) to 1 (100% modulation). Both the waveforms of AM and FM tones contain temporal fine structure (TFS) cues, the fine-grained changes in pressure over time. AM tones also have envelope cues, corresponding to the overall shape of the waveform. Before cochlear filtering, the envelope of FM is flat. Near-threshold detection of FM and AM tones are interesting for models of pitch because in some instances, but not others, FM and AM may be coded via the same, peripheral rate-place cues, utilizing detection of envelope cues. This is because cochlear filtering may transform FM to AM. Whether and under what circumstances FM uses a place-rate or a timing code has been a point of contention for several decades and is still an open question.

Zwicker (1956, 1970) and Maiwald (1967a, 1967b) proposed that FM may be detected entirely via a place code, similar to AM. In Zwicker's model, FM can be detected whenever there is a change in the excitation pattern greater than approximately 1 dB. As the frequency sweeps back and forth across the characteristic frequency filter, the corresponding auditory nerve fibers will systematically change their firing rate. Envelope cues may be extracted from FM by monitoring changes in the overall firing rate of the auditory nerve, similar to AM detection. Based on the asymmetry of the auditory filters at medium to high levels (>~ 40 dB SPL), changes in the excitation pattern should be greatest on the low-frequency side of the response to the tone, while changes on the high-frequency side will be less pronounced. Whether the auditory system monitors the output of just the low frequency side of the characteristic frequency filter (Zwicker, 1956, 1970, Maiwald, 1967a, 1967b) or multiple, neighboring filters (e.g., Moore and Sek, 1992), the outcome may be that FM is transformed to AM in the cochlea (e.g., Moore and Sek, 1992, 1994; Saberi and Hafter, 1995).

An alternative theory proposes that FM and AM may be detected via independent mechanisms, at least in some circumstances, using a temporal code for FM and a rate-based code for AM (e.g., Feth, 1972; Coninx, 1978a; Hartmann and Hnath, 1982; Demany and Semal, 1986; Moore and Glasberg, 1989). Based primarily on work by Moore, Sek and colleagues (e.g., Moore and Sek, 1992, 1996; Sek and Moore, 1995a; Moore and Skrodzka, 2002), the current consensus is that the nature of FM coding may depend on the frequency of the carrier (f_c) and the rate of the modulation (f_m). At low carrier frequencies (f_c < ~4-5 kHz) and slow modulation rates (f_m < ~10 Hz), FM is believed to be coded via neural phase-locking to the TFS cues (e.g., Moore and Sek, 1995, 1996; Moore and Skrodzka, 2002). Lower carriers at faster rates and higher carriers at all modulation rates (up to but not exceeding the rate that produces detectable, resolved sidebands, e.g., Hartmann and Hnath, 1982) are believed to be coded via a rate-place code, similar to AM detection (e.g., Moore and Peters, 1992; Moore and Sek, 1994; Saberi and Hafter, 1995). There are several pieces of indirect, behavioral evidence that support the dual-mechanism model for sinusoidal FM.

Perhaps the most convincing evidence for a dual-mechanism model of FM comes from comparing average FMDLs to average AMDLs at different carriers and modulation rates. First, sensitivity for FM is better for low carrier frequencies when the modulation rate is less than about 10 Hz compared to when the modulation rates are equal or faster than around 10 Hz (Moore and Glasberg, 1989; Moore and Sek, 1995, 1996; Sek and Moore, 1995; Moore and Skrodzka, 2002). Better sensitivity for slow FM with low-carrier frequencies suggests that slow FM is coded by precise phase-locked timing to the TFS of the waveform. Faster rates, even for low-carrier FM, may be too fast for phase-locking to TFS cues to be useful. While cat models indicate that f_m = 10 Hz should be slow enough for the auditory nerve to extract phase-locking information (Khanna and Teich, 1989), Moore and Sek (1995) suggest that the central extraction of phase-locking cues may be sluggish, similar to the binaural system (Blauert, 1972; Kollmeier and Gilkey, 1990). It may be that this central mechanism integrates across multiple time windows to calculate the period of the waveform. FM cycles become shorter as the modulation rate increases, but the time window remains constant. Once the time window integrates over one or more cycles of FM (perhaps around f_m >= 10 Hz), the central mechanism will no longer be able to detect changes in frequency from TFS cues. In the absence of a viable temporal code, the auditory system may track changes in fast FM based on changes of the excitation pattern along the basilar membrane (e.g., Zwicker, 1970; Moore and Sek, 1992, 1994; Edwards and Viemeister, 1994a, 1994b; Saberi and Hafter, 1995; Sek and Moore, 1995). The opposite trend in threshold and modulation rate is observed for AM detection with gated carriers (Viemeister, 1979; Sheft and Yost, 1990; Moore and Sek, 1995; Whiteford and Oxenham, 2015). In this case, AMDLs either remain constant or decrease (improve) with increases in modulation rate. This effect is typically attributed to the increase in the number of cycles of AM per second, providing the listener with more chances to sample the modulation cycles (e.g., a gated carrier with f_m = 20 Hz will have a ten-fold number of cycles per second relative to the same carrier with an f_m = 2 Hz). Viemeister (1979) notes these trends are consistent with an increase in sensitivity for AM with increased signal duration, which has the greatest effect at slow modulation rates. This trend may also be attributed to the gating at the carriers in addition to the fewer number of cycles. Ramping on a tone, for example, adds additional modulation, potentially creating a forward masking effect that masks a greater proportion of the modulation cycles for AM tones with a slow rate relative to AM tones with a fast rate. The opposite trend of thresholds as a function of modulation rate for FM and AM remains one of the strongest indicators that slow, low-carrier FM is coded via a different (presumably temporal) mechanism.

A second line of evidence supporting a dual-mechanism model for FM comes from detection thresholds for low versus high carrier frequencies at slow modulation rates. Similar to pure tone FDLs (Moore, 1973; Moore and Ernst, 2012), slow FMDLs increase (worsen) for carrier frequencies until around 4-5 kHz (Sek and Moore, 1995; Moore and Sek, 1996; He et al., 2007), while fast-rate FMDLs do not. Again, the outcome is that only envelope cues are available for high-carrier FM. These same trends are not observed with slow AM at different carrier frequencies (Moore and Sek, 1995). Moore and Sek found that AMDLs were moderately better for a 1000 Hz carrier compared to the 250 Hz or 6000 Hz carrier.

A third line of evidence comes from studies of FM with an AM masker (FM+AM), whereby both FM and AM are applied to the same carrier. Participants' task is to detect the FM+AM tone, where the reference tone is AM. In instances where FM and AM use the same rate-place code, applying a fixed amount of AM to FM should wipe out excitation pattern cues, leading to elevated FM+AM thresholds. Moore and Sek (1996) had subjects listen to two tones, each with a fixed amount of AM (m = .33), and the task was to detect the FM+AM tone. Subjects were poor at detecting FM with added AM for the highest carrier (f_c = 6 kHz) relative to FM detection alone (without AM), and the amount of interference from AM did not interact with modulation rate ($f_{m,}$ = 2, 5, 10, and 20 Hz). Moore and Skrodzka (2002) and Ernst and Moore (2010) found the same results for FM+AM detection at f_c = 6 kHz. The assumption is that the added AM interferes with FM detection whenever FM undergoes a cochlear transformation to AM, such as when the carrier frequency is above the limit of phase-locking. For low carriers (f_c <= 4 kHz), detection of FM with added AM worsened as the modulation rate increased (Moore and Sek, 1996; Moore and Skrodzka, 2002; Ernst and Moore, 2010). Again, this supports evidence that FM is converted to AM for all carriers when the modulation rate is fast (f_m >=~ 10 Hz).

Findings from FM+AM studies, however, could be considered as presenting contradictory evidence to a two-mechanism model of FM. Multiple studies (Moore and Sek, 1996; Moore and Skrodzka, 2002; Ernst and Moore, 2010) found that added AM *still* interfered with low-carrier, slow-rate FM detection. While this interference was not as severe as for FM with faster modulation rates, the presence of any interference when a fixed amount of AM is added to FM could indicate that excitation pattern cues are also important for slower-rate, lower-carrier FM. This would suggest that low carrier, slow FM uses both phase-locking and place cues. However, a more parsimonious explanation is that the presence of detecting FM amongst AM is a more cognitively demanding task than detecting FM without the presence of AM, irrespective of the type of peripheral code used for FM. Poorer FM+AM detection relative to detecting FM alone (without added AM) is weak evidence at best for the use of excitation pattern cues in low-carrier, slow-rate FM.

More puzzling evidence suggesting some amount of rate-place coding for low-carrier, slow FM comes from studies of FM detection at varying sensation levels (SLs) in NH and HI listeners. Ernst and Moore (2010) had NH listeners detect FM in the presence of a fixed amount of AM (m = .33) both at a normal SL (60 dB SL) and at a low SL (20 dB SL). At the normal SL with low carrier frequencies (f_c = 1 kHz and f_c = 4 kHz), Ernst and Moore replicated previous findings of a greater deleterious effect of added AM as modulation rate increased, with the smallest interference at the slowest modulation rate (f_m = 2 Hz). With the highest carrier (f_c = 6 kHz), added AM equally impaired FMDLs, regardless of modulation rate. Interestingly, the effects of added AM on FM detection were different at the 20 dB SL compared to the 60 dB SL. Instead, there was an equal amount of interference of added AM at slow (f_m = 2 Hz) and fast (f_m = 10 Hz) modulation rates for the low carrier frequencies, similar to that found with f_c = 6 kHz at 60 dB SL. For f_c = 6 kHz at 20 dB SL, the deleterious effect of added AM was greater for the faster modulation rate. Results at the low SL can be considered as inconsistent with a completely separate, independent mechanism for low-carrier, slow FM. Ernst and Moore proposed that such a low SL will provide a very small excitation pattern, which may, perhaps, limit the amount of temporal information in the auditory nerve and/or enhance the role for place cues (i.e., because the excitation pattern is very sharp). It could be that at low SLs, the precision of phase-locking decreases (Johnson, 1980; Palmer and Russell, 1986). Hence, low carriers with slow modulation rates at low SLs may rely primarily on a place code or combined place and temporal information. However, if this were the case, then one might expect the FM detection trends (in quiet, with no AM) to *flip* and mirror trends observed in fast versus slow AM at low SLs. Ernst and Moore found that for the low SL, similar to the high SL, FM detection in quiet was best at the lowest carrier frequency with the slowest modulation rate- an argument that typically supports phase-locking for slow FM at low carriers with a low SL. Zwicker (1952) also measured FM detection at different phon units and found similar FMDL trends across modulation rate regardless of level. Data from FM detection in quiet at low SLs suggests that slow, low carrier FM relies on phase-locking to TFS cues. FM+AM results at low SLs are not consistent with phase-locking (Ernst and Moore, 2010). Therefore, exactly what mechanism codes FM at low SLs is unclear. One possible alternative explanation for the conflicting trends in Ernst and Moore's results is that their small sample (n=6) does not have enough power to detect an effect of f_m on FM+AM detection for low carriers at 20 dB SL. Low power can increase the chance of incorrectly accepting the null hypothesis (i.e., Type II error).

At face value, effects of FM+AM detection at low SLs in normal-hearing listeners are similar to Moore and Skrodzka's (2002) FM+AM findings in older, HI listeners. Moore and Skrodzka measured FM detection with a fixed amount of AM (m = .33) at 70 dB SPL in young, NH listeners and 85 dB SPL in older, HI listeners. FM detection in quiet and in the presence of AM was measured at multiple f_cs (.25, .5, 1, 2, 4, and 6 kHz) and f_ms (2, 5, 10, and 20 Hz). For the three lowest carriers (.25, .5, and 1 kHz), HI listeners were equally worse at detecting FM with added AM regardless of the modulation rate. Young listeners showed similar trends to previous studies, with more interference for low carriers at faster modulation rates, indicating that added AM disrupted the excitation pattern information of fast FM more than slow FM for low carriers. These results can be interpreted two ways. Either slow FM with low carriers uses both place and temporal coding, or cochlear hearing loss effects both place and temporal coding. Moore and Skrodozka's results are difficult to compare directly with Ernst and Moore's (2010) FM+AM results at the low SL because the HI listeners had varying degrees of hearing loss for different frequencies. This means for a given carrier frequency, SL was not held constant within or across HI listeners. Conflicting results with HI listeners detecting FM with added AM could be a consequence of measuring FM+AM detection at low SLs for some listeners. Adding a fixed amount of AM becomes especially problematic with HI listeners who often have better AMDLs than NH listeners, and even more so at low SLs (Moore, 2007; Ernst and Moore, 2012; Wallaert et al., 2017). Adding more detectable amounts of AM to FM for the HI listeners may have increased the task difficulty (i.e., AM cues were more distracting for the HI listeners), thus making their results incomparable to NH listeners.

Another study by Ernst and Moore (2012) measured FM, AM, and FM+AM detection in five HI listeners at 20 dB SL and 90 dB SPL. The benefit of this study over others is that Ernst and Moore took into account SL and SPL. Again, difference limens were assessed at varying carrier frequencies (f_c = 1, 4, and 6 kHz) and modulation rates (f_m = 2 and 10 Hz). The effect of added AM on FM detection was smallest for the lowest carrier with the slowest modulation rate, consistent with the use of temporal coding at low carriers with slow rates. Contrary to NH listeners (Ernst and Moore, 2010), this effect was consistent at 20 dB SL and the higher, 90 dB SPL. Ernst and Moore (2012) suggested their results are consistent with Ernst and Moore (2010) because auditory filter shapes do not change much with level in HI listeners compared to NH listeners. NH listeners may have relied more on place cues for low-carrier, slow-rate FM at the low SL because filter shapes in NH listeners become sharper at lower SLs (e.g., Moore and Glasberg, 1987; Baker and Rosen, 2006). Filter shapes in HI listeners are relatively broad and do not change much with level (e.g., Stelmachowicz et al., 1987; Baker and Rosen, 2002); hence, the conclusion is that TFS cues are utilized for low carrier, slow-rate FM at low SLs in HI but perhaps not in NH listeners.

In summary, the weight of the evidence suggests that for lower carrier frequencies, the type of peripheral code for FM may depend on the rate of frequency changes over time. For slower rates ($f_m <= \sim 10$ Hz) and lower carriers ($f_c < 4$-5 kHz), the brain may be able to utilize phase-locking to TFS cues to calculate the frequency.

At faster rates (f_m >= ~10 Hz) and higher carriers (f_c >4-5 kHz) at all rates, FM may be transformed to AM via cochlear filtering, leading the brain to rely on place cues. However, evidence discussed so far has been indirect, relying on assumptions based on mean differences in FMDLs compared to AMDLs, or differences in FMDLs at low versus high carriers at different modulation rates. Results from FM+AM studies further complicate a two-mechanism model for FM, suggesting that low carrier, slow FM at low SLs may also rely on place coding in NH listeners (but not HI listeners). Most of the previous work on FM, including the FM+AM results, relies on small sample sizes in well-trained listeners. Using small samples of experts has the benefit in that the subjects are presumably highly motivated, well trained (preventing confusions related to task demands), and can complete many different conditions for a given experiment. However, the trade-off is low-power statistical analyses with a biased sample. Low power increases the chance of incorrectly accepting the null hypothesis (Type II error). Restricting the sample to expert and/or well-trained listeners removes the natural across-listener variability in FMDLs and AMDLs. If the stark trends in low carrier FMDLs at different modulation rates compared to AMDLs are truly due to different peripheral coding mechanisms, then between-subject variability in tasks thought to measure TFS coding versus place coding should be an informative means of further investigating the peripheral code for FM (discussed further in Section 1.3).

1.2 Central Pitch Processing

The focus so far has been on a place vs. time peripheral representation of pitch, but there are many stops along the auditory pathway at which frequency information may be transformed before it reaches the auditory cortex. The upper limits of phase locking decrease substantially at higher-levels of the auditory pathway, suggesting that most timing information must eventually be transformed to a place code. For example, the upper limit of phase locking in the inferior colliculus (IC; located in the midbrain) is estimated to be ~1000 Hz (Liu et al., 2006). This is substantially lower than the existence region for pitch. Primary auditory cortex has an even lower upper limit of phase locking, estimated at just ~100 Hz (Lu and Wang, 2000). High-field imaging studies with humans have shown that peripheral place (tonotopic) coding, on the other hand, is preserved up through the auditory cortex (Moerel et al., 2014), suggesting a rate-place code is implicated at the cortex.

1.2.1 Where is the “pitch center” of the brain?

As discussed in the section on complex pitch (1.1.2), two stimuli with different spectra can sometimes elicit the same pitch, as in the case of complex tones with and without a missing F0. The phenomenon of periodicity pitch is highly suggestive of a pitch center somewhere in the brain that extracts the pitch of the F0. A true “pitch neuron” should therefore selectively respond to the pitch of the F0, even if it is not spectrally present. Recent studies have suggested such pitch-selective neurons exist in primates (Bendor and Wang, 2005, 2010). Bendor and Wang (2005) used single-electrode recordings in the marmoset monkey to measure cortical responses to complex tones with a missing F0 and to pure tones. They found pitch-selective neurons near primary auditory cortex that responded similarly to pure tones and to complex tones with a missing F0. For example, if a neuron had a best frequency of 500 Hz for pure tones, it would also have a best F0 frequency of 500 Hz for the complex tones with a missing F0, even though no spectral energy was present at 500 Hz. These neurons also preferred low more than high harmonic numbers.

Interestingly, marmosets also have harmonic template neurons (Feng and Wang, 2017). These are neurons that prefer harmonic over inharmonic complex tones, providing a possible physiological implementation for pattern-matching models. Harmonicity-preferring neurons, however, are distinct from pitch-selective neurons in that they do not always respond similarly to pure tones of the same pitch or even two-tone complexes. While pitch-preferring neurons were found over a small region for just lower F0s (< ~ 1000 Hz), harmonic template neurons were more prevalent and spanned a wide range of F0s (.4-12 kHz). Marmosets are useful animal models for pitch because they are highly vocal and have a similar hearing range to humans, but there are obvious limitations when generalization anatomical findings between species.

Whether there is an analogous human pitch center, and where it is located, is somewhat controversial (Bendor, 2012; Plack et al., 2014). Most evidence suggests such a pitch center may reside in lateral Heschl’s gyrus, located in a small region of non-primary auditory cortex (e.g., Patterson et al., 2002; Penagos et al., 2004; Hall et al., 2006). This region is considered analogous to the region for pitch-selective neurons Bendor and Wang (2005) found in marmosets (Bendor and Wang, 2006). However, when fMRI responses were measured to varying types of pitch-evoking stimuli, distributed parts of planum temporal showed activation across the stimuli (Hall and Plack, 2009). These results were likely affected by including unresolved complex tones as part of the stimuli set (Norman-Haignere et al., 2013). Unresolved complex tones have a weaker pitch salience and produce overall weaker activation in auditory cortex compared to resolved complex tones. Furthermore, conflicting results could arise because some studies have averaged responses across pre-defined anatomical regions while others have used functionally defined regions.

1.2.2 Pitch processing in congenital amusia

One classic technique for understanding how the brain processes information is to compare behavioral performance in participants with a specific, neurological impairment to those with no known neurological impairments. People with congenital amusia, more commonly known as “tone deafness”, have a disorder in melody perception but no history of traumatic brain injury and no known problems with hearing or cognitive ability (Peretz, 2001; Ayotte et al., 2002). The disorder is referred to as congenital because it is believed to be present at birth or developed very early in childhood. This contrasts with acquired amusia, where musical deficits are caused from brain damage to areas important for music perception, typically after a right-hemispheric stroke (Sihvonen et al., 2016a). Brain imaging studies attempting to reveal the neural correlates of congenital amusia have shown increased grey matter in the auditory cortex and right inferior frontal gyrus (rIFG) (Hyde et al., 2006, 2007). Connectivity within and between these two areas, on the other hand, appears to be decreased, suggesting a problem with the fronto-temporal pathway (Hyde et al., 2011; Albouy et al., 2013a, 2015b). This has led to the description that amusia is a “disconnection syndrome” (Loui et al., 2009; although see Chen et al., 2015).

Amusia is a unique population for understanding central mechanisms for pitch processing because fine-grained pitch discrimination is impaired (Peretz et al., 2002; Foxton et al., 2004; Cousineau et al., 2015; Vuvan et al., 2015) but there are no known deficits in peripheral coding (Cousineau et al., 2015). On average, amusics tend to be able to detect changes in pitch larger than a semitone, whereas they are not able to differentiate smaller changes in pitch, although there are substantial individual differences (Foxton et al., 2004; Vuvan et al., 2015). This means amusic's frequency difference limens (FDLs) tend to be much larger than non-amusic, matched controls with comparable amounts of musical training. In addition to pitch discrimination deficits, amusics have also demonstrated deficits in short-term pitch memory (Tillmann et al., 2016). It is unclear how much of the pitch deficits are driven by problems with short-term or working memory, which must be used to some extent on pitch discrimination tasks. And vice versa, impairments in short-term memory for pitch in amusics may be partly conflated by poor underlying sensitivity to pitch (Jiang et al., 2013).

A recent study examined whether or not amusic deficits can be attributed to abnormal pitch coding within auditory cortex (Norman-Haignere et al., 2016). Norman-Haignere et al. (2016) used fMRI to find areas of the brain that had a greater response to harmonic complex tones relative to frequency-match noise. Such areas were deemed "pitch-selective" because the noise and the harmonic complex tones were spectrally similar, so presumably the difference in activity reflected neural activity for regions with a preference for pitch. They found that pitch-selective voxels in amusics were no different in selectivity and location than non-amusic matched controls. This could suggest that pitch coding is normal in amusics up to at least auditory cortex. Instead, pitch deficits in amusia may be related to abnormal communication between auditory cortex and other areas of the brain, such as the rIFG, in line with studies implicating problems in the fronto-temporal pathway (Hyde et al., 2011; Albouy et al., 2013a, 2015b). The implication is that amusia could be a deficit in awareness or memory for pitch, while pitch encoding remains normal. An alternative explanation is that the resolution was too course to detect group differences in pitch-selective regions between amusics and controls, or that group differences are from differences in the tuning or temporal properties of the neurons. It could, therefore, still be possible that amusics have an underlying pitch coding deficit, but this effect may be small.

Interestingly, several studies have suggested that the pitch and melody-related impairments in amusia are not malleable to training (Hyde and Peretz, 2004; Mignault Goulet et al., 2012; Peretz et al., 2012; Wilbiks et al., 2016) or only limited to pitch-contour identification (Liu et al., 2017) or vocal production (Anderson et al., 2012). This has led to the common report that amusia is a “life-long” disorder. But an inability to improve pitch perception with practice is especially surprising given that the non-amusic literature has shown pitch discrimination in naïve listeners is incredibly plastic (Wright and Zhang, 2009). Micheyl et al. (2006) demonstrated that pitch discrimination in professional musicians is a factor of 6 greater than pitch discrimination for non-musicians. Despite this advantage, the non-musicians required only 4-8 hours of laboratory training to perform on par with the professional musicians. Amitay et al. (2006) found that active practice was not even necessary to improve pitch discrimination, as participants improved at their 1-kHz pure-tone discrimination thresholds through only passive exposure to 1-kHz tones while playing a game of Tetris. Given that pitch discrimination is highly susceptible to improvements through laboratory training, it is surprising that amusics have so far not been able to learn on pitch or melody-related tasks. However, no studies have specifically trained amusics on pitch discrimination using paradigms similar to the psychoacoustical literature, so a direct comparison is not possible.

1.3 Individual Differences

Many auditory tasks exhibit a wide range of across-listener variability (e.g., Johnson et al., 1987). Such differences may arise from many different sources, including differences in cortical coding, differences in peripheral coding, measurement error, or differences in task-relevant cognitive functioning, such as auditory working memory. Although the causes can be difficult to distinguish, across-listener differences can be used to help answer some basic questions involving auditory coding.

Evidence suggests that performance on auditory tasks can be broken down into a subset of tasks that each reflect different auditory abilities, usually by employing factor analytic techniques on performance across a wide variety of tasks in a large sample of listeners (e.g., Karlin, 1941, 1942; Mcleish, 1950; Elliott et al., 1966; Stankov and Horn, 1980; Festen and Plomp, 1983; Spiegel and Watson, 1984; Johnson et al., 1987; Kidd et al., 2007; Conzelmann and Süß, 2015). The nature and number of auditory factors (or components) that can explain the variability across auditory tasks varies widely across studies (see Johnson et al., 1987 for review of the earlier literature), likely reflecting the different tasks employed across studies as well as the different data-reduction techniques (e.g., factor analysis vs. principal components analysis; PCA). Across 10 individual differences factor analytic studies, Johnson et al. (1987) concluded that there were 4 auditory factors that tended to occur most often: (1) auditory memory, (2) sensitivity to changes in pitch/frequency, (3) sensitivity to changes in loudness/intensity, and (4) sensitivity to changes in duration. A more recent study testing a large sample of listeners (n=340) on a set of 19 auditory discrimination and identification tasks found a different set of four factors explaining auditory performance: loudness and duration discrimination, sensitivity to temporal envelope variation (AM noise), identification of familiar sounds (including speech), and pitch and time discrimination (Kidd et al., 2007). Structural equation modeling indicated that these four components were subsumed by an underlying ability in general auditory intelligence, although support from other studies suggesting a general factor in auditory intelligence has been mixed (e.g., Karlin, 1941, 1942; Mcleish, 1950; Martin and Martin, 1973; Conzelmann and Süß, 2015).

While individual differences can be used to examine both auditory intelligence and separate, auditory abilities, individual differences can also be used to examine specific, basic mechanisms of auditory perception (e.g., place versus temporal coding). Festen and Plomp (1981), for example, aimed to investigate whether across-listener variability could provide insight into which psychophysical tasks utilize similar peripheral coding mechanisms. They had a group of 50 NH listeners complete various psychophysical tasks related to temporal resolution, frequency resolution, and nonlinearity. Even though their test-retest scores were quite reliable, they found low correlations between many of the tasks. There were some exceptions; for example, auditory filter widths (measured via simultaneous masking) were negatively correlated with the low frequency slope of psychophysical tuning curves, suggesting both measures reflect frequency resolution. Factor analysis indicated no interpretable underlying structure to the dataset, perhaps suggesting that most of their measures used reflect different auditory functions. Another possibility is that the variance across young, normal-hearing listeners does not directly reflect variability in peripheral coding and instead reflects variance in higher-level processing. A follow-up study using 22 HI listeners was conducted with the aim to increase the variability related to peripheral coding (Festen and Plomp, 1983). Festen and Plomp (1983) used a variety of psychophysical tasks believed to reflect frequency resolution, temporal coding, audibility, and speech perception. Many of the tasks thought to reflect frequency resolution were well correlated. However, some tasks believed to measure separate mechanisms were also correlated (e.g., auditory filter bandwidth measured with non-simultaneous masking and temporal resolution). PCA suggested two components could account for 65% of the variance in HI listeners: (1) absolute thresholds and (2) frequency resolution. Both measures of temporal coding and frequency resolution loaded highly on the same component, providing no evidence for dissociations of these mechanisms in the hearing impaired.

While most of individual differences studies discussed so far were largely exploratory, correlational measures across large samples of participants can also be used to test specific, hypothesis-driven questions. McDermott et al. (2010), for example, utilized the variability in consonance preferences (i.e., “pleasantness” ratings of musical intervals) across large samples of young participants varying in musicianship to test whether harmonicity or lack of beating was more important for consonance perception. They had over 250 participants rate the pleasantness of musical intervals as well as dichotic and diotic stimuli designed to isolate either harmonicity or beating cues. They found that preference for harmonicity predicted preference for consonance but preference for lack of beating did not. Dissociations such as these are a powerful means of understanding which acoustic properties contribute most to perceptual experiences.

Correlational measures between tasks may also help understand how the efficacy of peripheral coding affects performance on higher-level tasks, such as selective attention (e.g., Ruggles et al., 2011; Bharadwaj et al., 2015). Some listeners with audiometrically “normal” hearing and can perceive speech in quiet, but have difficulty understanding speech in more complex environments, such as a crowded restaurant. Is poor speech in noise performance related to variability in peripheral coding, or variability in cognitive abilities, such as selective attention? Ruggles et al. (2011) tested a group of 42 normal-hearing adults ranging from 18-55 years on a spatial selective attention task. The selective attention task involved identifying speech from a single talker amongst two other competing talkers. All three talkers had the same voice, thus requiring listeners to make use of interaural time difference (ITD) cues (and, therefore, TFS coding). They found wide variability across listeners in both anechoic and reverberant listening environments. Furthermore, performance across the top and bottom quartile listeners was correlated with slow-rate, low-carrier FM detection (a task believed to require TFS coding, see section 1.2) and with the strength of the frequency following response (FFR), a physiological measure of subcortical phase-locking. This suggests that peripheral coding acts as a bottleneck for auditory information, so that declines in sensory processing will inevitably affect abilities to perform higher-level, cognitive tasks (such as selective auditory attention) in the same modality (Ruggles et al., 2011; Humes et al., 2013). Consistent with these results, Bharadwaj et al. (2015) found that variability in performance on a spatial selective attention task in 26 young, NH listeners was also related to a physiological measure of subcortical coding. In addition, ITD sensitivity and detection of AM embedded in noise were related to their physiological measure of subcortical phase-locking. Together, findings from Bharadwaj et al. (2015) and Ruggles et al. (2011) both suggest that variability in NH listeners is at least in part related to variability in peripheral phase-locking fidelity.

If there is extensive variability in peripheral phase-locking even in young, normal-hearing listeners, then tasks that utilize TFS coding should be well correlated with one another, while tasks that utilize a different code, such as place coding or level coding, should not be as highly correlated with TFS tasks. Ochi et al. (2014) used individual differences across a combination of monaural and binaural tasks to test the importance of peripheral coding for level and timing cues in ITDs and interaural level differences (ILDs) in low and high frequency ranges. One expected result would be that low-frequency, monaural coding for TFS cues should correlate with low-frequency binaural coding of ITD cues, while level coding for the same monaural and binaural coding conditions should be correlated. The low-frequency stimulus was a complex tone (F0 = 100 Hz) bandpass filtered around 1000 Hz. To assess monaural efficiency for TFS coding, listeners completed frequency discrimination with the low-frequency stimulus. Similarly, intensity discrimination for the same low-frequency stimulus was a marker for monaural level coding efficiency. Monaural efficiency of low-frequency level coding was correlated with low-frequency

ILD sensitivity, but monaural efficiency for low-frequency time coding was not correlated with low-frequency ITD sensitivity. One possibility for this discrepancy is that the instructions for the binaural task (indicate the direction of a change) were different form the monaural task (determine which stimulus changed), and across-listener variability for ITDs and TFS coding could have been swamped by the differences in task demands. Another possibility is that the low frequency “TFS” stimulus may have also had useable spectral cues, meaning that their monaural frequency discrimination task may not solely reflect TFS coding. In addition, the sample size (n=22) was quite small for assessing individual differences in NH listeners.

1.4 Overview of Chapters

The overarching goal of this dissertation was to better understand the mechanisms involved in pitch perception by examining the different factors involved in coding frequency at both the peripheral and central levels. Given the variability observed in previous studies thought to reflect peripheral coding in NH listeners (e.g., Ruggles et al., 2011; Ochi et al., 2014; Bharadwaj et al., 2015), we used similar correlational measures to understand the role for temporal versus place coding in FM detection across large samples of listeners. Our samples included listeners that were young and NH (Chapter 2), listeners that varied in age (Chapter 3), and listeners that varied in degree of SNHL (Chapter 4). The purpose of including the group varying in age was to increase the variability due to temporal coding to TFS, as older adults are thought to have poorer TFS coding (Moore, 2014). SNHL listeners, on the other hand, have shallower filter slopes (Glasberg and Moore, 1986), so that increasing the variability in SNHL should increase the variability in place coding.

To better understand the role of central processing on pitch perception, we examined FM and AM detection in a sample of listeners with congenital amusia and a group of matched controls (Chapter 5). Amusics are known to have problems with short-term memory for pitch (e.g., Tillmann et al., 2016), but it is unclear to what extent this confounds their performance in pitch discrimination or how specific their deficit is to pitch. Therefore, we also examined their pitch perception in tasks that had a low (one-interval FM detection) and high memory load (three-interval frequency discrimination). Lastly, we examined the malleability of pitch and melody discrimination deficits in amusia (Chapter 6). Chapter 7 discusses the primary findings across the studies and the open areas for future research.

CHAPTER 2: FM DETECTION IN YOUNG LISTENERS

Chapter 2 is reprinted from:

Whiteford, K. L., & Oxenham, A. J. (2015). Using individual differences to test the role of temporal and place cues in coding frequency modulation. *Journal of the Acoustical Society of America, 138*, 3093-3104.

Abstract

The question of how frequency is coded in the peripheral auditory system remains unresolved. Previous research has suggested that slow rates of frequency modulation (FM) of a low carrier frequency may be coded via phase-locked temporal information in the auditory nerve, whereas FM at higher rates and/or high carrier frequencies may be coded via a rate-place (tonotopic) code. We tested this hypothesis in a cohort of 100 young normal-hearing listeners by comparing individual sensitivity to slow-rate (1-Hz) and fast-rate (20-Hz) FM at a carrier frequency of 500 Hz with independent measures of phase-locking (using dynamic interaural time difference discrimination), level coding (using amplitude modulation, AM, detection), and frequency selectivity (using forward masking patterns). All FM and AM thresholds were highly correlated with each other. However, no evidence was obtained for stronger correlations between measures thought to reflect phase-locking (e.g., slow-rate FM and ITD sensitivity), or between measures thought to reflect tonotopic coding (fast-rate FM and forward masking patterns). The results suggest that either psychoacoustic performance in young normal-hearing listeners is not limited by peripheral coding, or that similar peripheral mechanisms limit both high- and low-rate FM coding.

2.1 Introduction

Periodic sounds represent an important category of natural sounds, including voiced speech, song, and many animal vocalizations. Despite their importance, there is very little consensus regarding how periodic sounds are coded in the auditory system (Plack et al., 2005; Oxenham, 2013). At the most peripheral level (in the cochlea) and for the simplest periodic sounds (sinusoids), two classical theories exist. Pitch may be coded based on the *place* of maximal excitation on the cochlea, leading to changes in the rate of firing in auditory nerve fibers (rate-place code), or by the stimulus-driven *timing* of phase-locked action potentials in the auditory nerve (temporal code).

It is generally believed that low-frequency pure tones are coded by the more precise temporal code, whereas higher frequencies are coded primarily via a rate-place code. The evidence for this conjecture is indirect and comes from different sources. First, auditory-nerve phase-locking (as quantified by measures such as the synchrony index) in mammals, such as cat and guinea-pig, is strong at low frequencies but degrades rapidly at frequencies higher than about 1-2 kHz, with the exact cut-off frequency depending on the species (Rose et al., 1967; Johnson, 1980; Palmer and Russell, 1986), suggesting that temporal coding is not viable at higher frequencies. Second, human behavioral pure-tone frequency discrimination (and detection of slow frequency modulation) is relatively good at low frequencies, but becomes dramatically worse above about 3-4 kHz, leading to poorer difference limens (e.g., Moore, 1973; Moore and Sek, 1995) and a reduced ability to recognize even familiar melodies (Attneave and Olson, 1971; Oxenham et al., 2011). Despite the general consensus about the role of the temporal code at low frequencies, studies do not agree on the exact frequency above which the rate-place code becomes dominant, with estimates ranging from around 4 kHz (e.g., Moore, 1973) to above 8 kHz (Moore and Ernst, 2012). Third, studies have found little to no relationship between pure-tone frequency discrimination at low or high frequencies and frequency selectivity, suggesting that a rate-place code based on tonotopic representation is unlikely to limit performance (Tyler et al., 1983; Moore and Peters, 1992).

Another approach to elucidating the coding of frequency has involved the detection of changes in frequency over time, known as FM. Here again, indirect evidence has been used to suggest a distinction between temporal coding and rate-place coding, depending on the conditions. Sensitivity to FM tends to be greatest at low carrier frequencies (f_c < 4000 Hz) and at slow modulation rates (f_m < ~10 Hz) (Moore and Sek, 1995, 1996; Moore and Skrodzka, 2002). This pattern of results can be explained if it is assumed that low carrier frequencies are coded via a temporal code that is “sluggish,” in that it can only follow relatively slow rates of frequency change (Sek and Moore, 1995; Moore and Sek, 1996; Plack et al., 2005). At higher carrier frequencies and higher modulation rates, poorer performance is explained through a reliance on rate-place coding of the temporal-envelope fluctuations induced by the FM. Although FM tones do not have any inherent envelope fluctuations (i.e., the envelope is flat), envelope cues can potentially be extracted from FM via cochlear filtering, such that FM is converted to AM, which is then detected by the fluctuations in firing rate (rather than the timing of individual spikes) in the auditory nerve (e.g., Zwicker, 1970; Coninx, 1978a, 1978b; Moore and Sek, 1992, 1994; Edwards and Viemeister, 1994a, 1994b; Saberi and Hafter, 1995).

Additional support for a two-mechanism model for FM comes from a variety of behavioral studies on FM and AM detection, alone and in combination. First, there is an added benefit for quasi-trapazoidal FM detection at low carriers compared to quasi-trapzoidal AM detection, indicating that more time spent at the modulation extremes is more beneficial for detecting slow FM (i.e., where phase-locking may occur) than for detecting slow AM (Moore and Sek, 1995). Second, when a fixed amount of AM is added to FM, the added AM interferes more with the detection of fast-rate than slow-rate FM at low carrier frequencies, suggesting that slow-rate FM is coded differently from slow-rate AM. In contrast, the amount of interference of AM on FM detection at high carrier frequencies (e.g., 6000 Hz, where phase-locking in unlikely to be a strong cue) is similar at all modulation rates, suggesting similar cues for both AM and FM detection (Moore and Sek, 1996). Third, the discriminability of AM from FM decreases with increasing modulation rate, suggesting that AM and FM may use similar (and hence confusable) mechanisms at fast modulation rates, but more separate mechanisms at slower modulation rates (Edwards and Viemeister, 1994b).

More direct evidence for the role of temporal and rate-place codes may come from correlations in performance between different tasks thought to rely on the same peripheral code. Ochi et al. (2014) tested the role of a phase-locking in frequency coding by correlating performance in a frequency-discrimination task with the discrimination of interaural time differences (ITDs), which are known to be represented via a temporal code. Both tasks used a bandpass-filtered tone complex centered around 1000 Hz, with a fundamental frequency (F0) of 100 Hz. Contrary to predictions, no positive correlation (and a slight non-significant negative correlation) was found between the monaural frequency-discrimination task and the binaural ITD task. One reason for the lack of the expected correlation may be because of the difference in the procedures used: the frequency discrimination task involved identifying which of two intervals included changes in the stimulus frequencies, whereas the ITD task involved not only detecting an ITD, but determining the direction of ITD change from one interval to the next. In addition, the number of participants (22) was rather small for identifying correlations based on individual differences between young normal-hearing listeners, especially when compared to recent studies using individual differences paradigms (Kidd et al., 2007; McDermott et al., 2010). Large samples are likely to be necessary to accurately measure performance variance within the normal-hearing population.

Previous work has assessed individual differences on a variety of psychoacoustical tasks within both normal-hearing and hearing-impaired populations to reveal potential underlying coding mechanisms (e.g., Festen and Plomp, 1981, 1983; Johnson et al., 1987; Kidd et al., 2007; McDermott et al., 2010; Watson et al., 1996). Our experiment used a similar paradigm, involving 100 young normal-hearing listeners. We aimed to minimize differences in task procedures and stimuli, with different tasks designed to tap into different underlying codes. Both diotic and dichotic AM and FM detection were tested. Phase-locked sensitivity to TFS cues was measured using a dichotic FM disparity task, where differences in the instantaneous phase between each ear result in ITDs. Dichotic and diotic detection performance for slow (f_m = 1 Hz) and fast (f_m = 20 Hz) modulation rates was measured for both FM and AM of a 500-Hz carrier. In addition, frequency selectivity was estimated using forward-masking patterns centered around 500 Hz, along with absolute thresholds at and around 500 Hz. If slow-rate FM detection is based on phase-locking, then performance in the slow-rate (diotic) FM detection task should be strongly correlated with performance in the slow-rate dichotic FM (ITD) detection task. Similarly, if fast-rate FM detection is based on a rate-place code, then performance in the fast-rate diotic FM detection task should be correlated with both fast-rate diotic AM (representing intensity coding) and the slopes of the forward-masking patterns (representing frequency selectivity).

2.2 Methods

2.2.1 Participants

One-hundred young adults (25 male, *M* = 21.06 years, range: 18-32) were recruited through the Research Experience Program at the University of Minnesota. All participants provided written informed consent and had NH, defined as audiometric thresholds of 20 dB hearing level (HL) or better for pure tones at octave frequencies between 0.25 and 8 kHz. Participants were compensated with course credit or hourly payment for their time. The protocols were approved by the University of Minnesota Institutional Review Board.

2.2.2 Stimuli

Stimuli were presented over open-ear headphones (Sennheiser HD650) in a sound-attenuating chamber. All FM and AM stimuli, diotic and dichotic, were presented at 60 dB sound pressure level (SPL). The FM and AM tasks involved either detection of frequency or amplitude modulation (FM and AM detection, respectively), or the detection of an interaural disparity in phase or level (dichotic FM detection and dichotic AM detection, respectively). In all cases, the carrier was a 500-Hz pure tone, 2 s in duration, including 50-ms raised-cosine onset and offset ramps. The FM difference limens (FMDLs) and AM difference limens (AMDLs) were measured for slow (f_m = 1 Hz) and fast (f_m = 20 Hz) sinusoidal modulation rates. For diotic FM, the starting phase of the modulator began with either an increase or a decrease in frequency excursion from the carrier (Δf), with 50% a priori probability. For the diotic AM detection task, the target tone randomly began at an amplitude peak or trough. The listeners’ task was to identify which of two intervals contained the modulated, as opposed to the unmodulated, tone.

For the dichotic FM detection tasks, the target tone was an FM tone, with an opposite modulator starting phase in each ear. One ear was presented with an FM tone beginning with an increase in Δf, while the opposite ear was presented with an FM tone beginning with a decrease in Δf. Because the modulator starting phases are different, the two tones shift in and out of phase with each other over time, creating a moving, intracranial image when f_m = 1 Hz. Fig. 2.1 plots an example of dynamic ITDs as a function of time when Δf = .06% and f_m = 1 Hz, calculated based on the running phase difference between the signal in each ear. The reference tone was a 2-s diotic FM tone, randomly beginning with either an increase or a decrease in Δf. The starting instantaneous frequency for all tones was the carrier frequency of 500 Hz. The carrier, modulation rates, level, and duration were identical to those in the diotic FM tasks. An analogous design was used for the dichotic AM disparity tasks, with the target tone containing opposite modulator starting phases in each ear. One ear was presented with an AM tone beginning at an envelope peak, while the other ear was presented with an AM tone beginning at an envelope trough. The reference tone was a diotic 2-s AM tone, randomly beginning with either an envelope peak or an envelope trough.

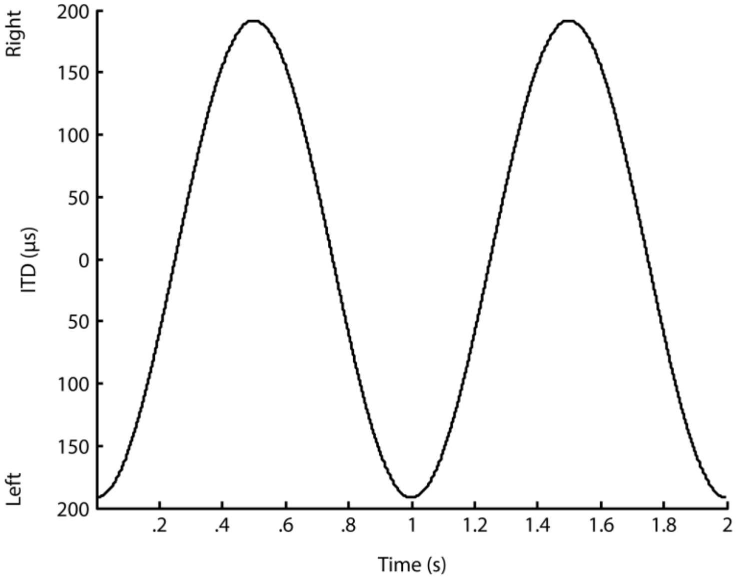

Figure 2.1

Example of dynamic ITDs as a function of time when Δf = .06% and f_m = 1 Hz. The black curve corresponds to the ITD at each point in time for a dichotic FM tone with Δf = .06% (the average dichotic FMDL at the 1-Hz rate across all subjects). Note that whether the tone began as a left-lateralized percept or a right-lateralized percept depends on the (randomized) starting phase of the modulator.

For the forward-masking task, the forward masker was a 500-Hz pure tone, presented at 70 dB SPL for a total duration of 500 ms. The signal was 20 ms in total duration, and both the masker and the signal had 10-ms raised-cosine onset and offset ramps. The onset of the signal was contiguous with the offset of the masker, resulting in a 10-ms gap between the offset of the masker and the onset of the signal at the half-amplitude points of their respective envelopes. Thresholds were measured for signal frequencies of 400, 430, 460, 490, 510, 540, 570, and 600 Hz. The slope of masking function (signal threshold as a function of masker-signal frequency difference in octaves, calculated separately for signals below and above the masker frequency) provided an estimate of frequency selectivity.

2.2.3 Procedures

Participants completed ten tasks across 2-3 sessions, with a maximum duration of 2 hours per session. In order to avoid fatigue, participants were instructed to take breaks as needed. All participants ran the tasks in the same order, as is typical of individual-difference paradigms (e.g., Kidd et al., 2007). All tasks used a two-alternative forced-choice paradigm with a three-down, one-up adaptive procedure, converging to the 79.4% correct point (Levitt, 1971). The target was randomly presented in either the first or second interval, and participants clicked a virtual button on the computer screen corresponding to the interval that they thought contained the target (i.e., “1” or “2”). Feedback was presented after each response, indicating whether the response was “correct” or “incorrect.”

All FM and AM tasks, dichotic and diotic, had a 500-ms inter-stimulus-interval (ISI). The slow (f_m = 1 Hz) condition was always run before the fast (f_m = 20 Hz) condition. For all FM and AM tasks, participants completed three adaptive runs. For each run, threshold was defined as the geometric mean of the tracking values at the last six reversal points. If the standard deviation across the runs was greater than or equal to 4, participants completed three additional runs, and the first three runs were regarded as practice. In order to discount learning effects, only the last three runs were included in analyses. About 8% of conditions resulted in the completion of additional runs. All subsequent FM and AM tasks used this same criterion to help control for learning effects. The procedures for each task are described below in the order in which they were presented to subjects.

2.2.3.1 Tasks 1 and 2: Dichotic FM disparity

First, participants completed the slow-rate (1-Hz) dichotic FM disparity detection task. Participants were instructed that they would hear two tones, one at a time, and their task was to pick the tone that sounded as though it was "moving in their head." They were reminded to look at the screen throughout the task, as they would receive visual feedback based on their response. In order to perceive lateralization and avoid confusion, the peak-to-peak frequency change must be sub-threshold but high enough for running phase to be accurately coded. Thus, each run began with a frequency excursion from the carrier (Δf) of .2%, slightly below most frequency modulation difference limens (FMDLs). The maximum value of the tracking variable was Δf = 1%, as pilot data indicated that lateralized percepts were no longer salient with larger Δf s. If the maximum value was reached for more than 10 consecutive trials, no threshold was recorded and listeners had to repeat three additional runs. One listener was not able to perform this task, and needed a higher starting value. This listener was able to perform the task with a starting value set to Δf = .6%[1]. Initially, Δf varied by a factor of 2. After the first two reversals, the step size was reduced to a factor 1.4 for the following two reversals, and was then set to the final step size of 1.19 for last six (measured) reversals. All subsequent FM tasks used the same series of step sizes.

Second, subjects completed the fast-rate (20-Hz) dichotic FM disparity detection task. Participants were instructed to pick the tone that had the "broader auditory image." Again, participants were reminded to look at the feedback after each trial to help them decide how to identify the target tone. The starting value was set to Δf = 1%, based on pilot data, with Δf never exceeding 100% throughout each run.

2.2.3.2 Tasks 3 and 4: FM detection

For both slow and fast FM detection, participants were instructed to pick the tone that was modulated, and that the modulated tone will sound like it is "changing." The initial value of the tracking variable was set to Δf = 2.51% and never exceeded Δf = 100%.

2.2.3.3 Task 5: Absolute threshold

Absolute thresholds were measured for all signal frequencies tested in the forward-masking task: 400, 430, 460, 490, 510, 540, 570, and 600 Hz. Participants completed one adaptive run at each signal frequency, with the signal frequency randomized between runs. The duration of the signal was the same as in the forward-masking task: 20 ms, including 10-ms onset and offset ramps (no steady state). Initially, the signal was presented at 40 dB SPL, and the initial step size was 8 dB. After two reversals, the step size was reduced to 4 dB and then to the final step size of 1 dB after two more reversals. Absolute threshold for each signal frequency was defined as the mean of the last 6 reversal points at the final step size. Participants were instructed to determine whether the first or second time interval, marked by lights on the virtual response box on a computer screen, “had a click in it.” The duration of each time interval was designed to be analogous to the forward-masking task. Each trial began with 500 ms silence, followed by either a 20-ms signal (target interval) or 20 ms of silence (reference interval). The two intervals were separated by 400 ms silence. If the standard deviation of the six reversal points within any of the runs was greater than or equal to 4, then one more run was completed at the corresponding signal frequency. At least one additional run was obtained in 23 of the 100 participants. Of the original runs, 3.4% were repeated. In the event that additional runs were needed from more than one signal frequency, the order of the additional runs was randomized.

2.2.3.4 Tasks 6 and 7: Dichotic AM disparity detection

Instructions for the slow (1-Hz) dichotic AM disparity task were identical to the slow dichotic FM disparity task. The initial modulation depth, in units of 20log(m), was -8 dB. The step size was 6 dB for the first two reversals, and was 2 dB for the next two reversals, until the final step size was of 1 dB was reached for the final six reversals. Threshold was defined as the mean depth at the last six reversal points.

Task instructions for fast (20-Hz) dichotic AM disparity were the same as the fast dichotic FM disparity task. Other than the instructions, the procedures were identical to those used for the slow dichotic AM task.

2.2.3.5 Tasks 8 and 9: AM detection

Participants were instructed to pick the modulated (i.e., “changing”) tone. Otherwise, all procedures were the same as in the dichotic AM tasks.

2.2.3.6 Task 10: Forward masking patterns

The 500-Hz masker was presented in both intervals of a trial, and the 20-ms signal was presented in one. Participants were instructed to pick the interval that had the “click” following the tone. The ISI was the same as in the absolute threshold task. At the beginning of each run, the signal level was 60 dB SPL. The initial step size of the adaptive procedure was 8 dB. After two reversals, the step size was decreased to 4 dB for the following two reversals, before reaching its final value of 2 dB for the final six reversals. Threshold was defined as the mean signal level at the last six reversal points.

Participants completed two runs for each of the eight target frequencies, totaling 16 runs, and the order of the runs was randomized. If the standard deviation across the runs for any of the signal frequencies was greater or equal to 4, then participants completed 2 more runs for the corresponding signal frequency. At least one additional run was obtained in 50 of the 100 participants. Of the original runs, 11.6% were repeated. In the event that participants had to repeat runs for two or more signal frequency conditions, the order of subsequent runs was also randomized.

2.3 Results

2.3.1 Comparisons of performance in FM and AM tasks

Results in the FM and AM tasks are presented as boxplots in Fig. 2.2. Repeated-measures ANOVAs were conducted on the log-transformed thresholds [$10\log(\%\Delta f)$ and $20\log(m)$] for all (diotic and dichotic) FM and AM tasks. For FM, a 2x2 within-subjects ANOVA revealed a main effect of modulation rate (1 Hz vs. 20 Hz) [$F(1,99) = 825$, $p < .0001$, $\eta_p^2 = .893$], a main effect of task-type (diotic vs. dichotic) [$F(1,99) = 216$, $p < .0001$, $\eta_p^2 = .686$], and a significant interaction [$F(1,99) = 457$, $p < .0001$, $\eta_p^2 = .822$]. *Post hoc* Bonferroni-corrected *t*-tests ($\alpha = .0083$) indicated significant differences between all comparisons except for diotic and dichotic fast FM tasks ($p = .312$). As expected, thresholds for slow dichotic FM were substantially and significantly smaller (better) than thresholds for slow diotic FM ($p <$

.0001), indicating that slow dichotic FM disparity detection was based on the dynamic ITDs that were not available in the diotic conditions. The average threshold for slow dichotic FM is Δf = .06%, which corresponds to a maximum instantaneous ITD of 192 μs (see Fig. 2.1).

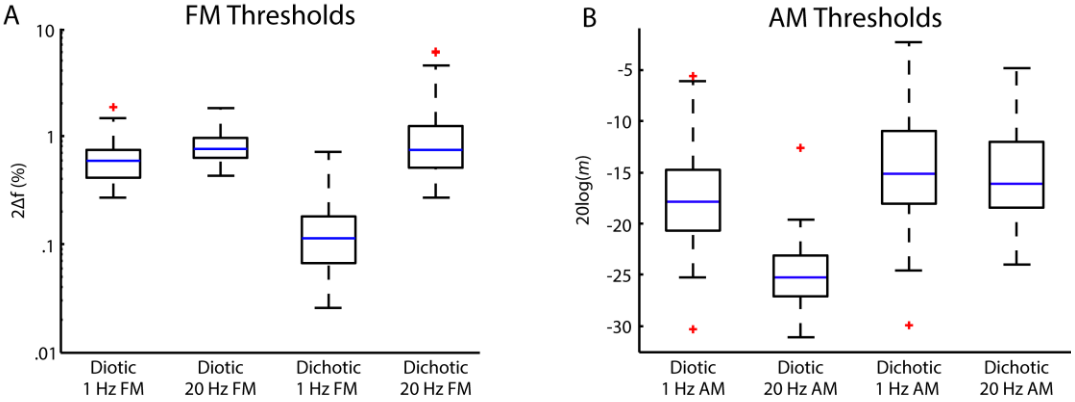

Figure 2.2

Boxplots for diotic and dichotic (A) FM detection and (B) AM detection thresholds across all participants. The two boxes closest to the y-axis represent performance on diotic FM (A) and diotic AM (B) tasks. Center lines within each box represent the median of each group (color online). Whiskers correspond to the lowest and highest data points within 1.5 times the lower and higher inter-quartile ranges, respectively. Small crosses represent individual data points outside the range of the whiskers, considered outliers.

Analyses of the AM results were conducted using a 2x2 (modulation rate vs. task-type) within-subjects ANOVA. There was a main effect of modulation rate [$F(1,99) = 127$, $p < .0001$, $\eta_p^2 = .562$], a main effect of task-type [$F(1,99) = 354$, $p < .0001$, $\eta_p^2 = .782$], and a significant interaction [$F(1,99) = 166$, $p < .0001$, $\eta_p^2 = .626$]. Differences between AM tasks were examined using *post hoc* Bonferroni-corrected *t*-tests. All pair-wise comparisons were significant except for slow versus fast dichotic AM ($p = .138$). Consistent with previous findings with gated carriers (Viemeister, 1979; Sheft and Yost, 1990; Moore and Sek, 1995), AMDLs were significantly better for fast AM detection compared to slow AM detection ($p < .0001$). This effect has been ascribed to the effects of gating stimuli with low modulation rates, where the duration of the stimulus is only a small number of modulation cycles (2 in the case of our 1-Hz modulation rate). In addition, slow diotic AM detection was significantly better than slow dichotic AM detection ($p < .0001$), and fast diotic AM detection was significantly better than fast dichotic AM detection ($p < .0001$). Thus, for AM (but not for FM), listeners were more sensitive to the detection of modulation than to the discrimination of interaural differences in modulation.

2.3.2 Within-subjects versus between-subjects variance

As the analyses described below are correlational, it is important to examine the within-subjects versus the between-subjects variance across each of the modulation tasks. This is because correlations will be limited if the within-subjects variance is high relative to the between-subjects variance (Altman and Bland, 1983). The within-subjects variance was calculated by taking the pooled estimated variance across all three runs for all of the subjects; this is equivalent to the mean-squared error from a one-way, repeated-measures ANOVA where run is the independent variable and threshold is the dependent variable. The square-root of the within-subjects variance (i.e., the within-subjects standard deviation, SD) was compared to the between-subjects SD for each of the modulation tasks, listed in Table 2.1. The ratio of between- vs. within-subjects SD ranged from a factor of 2.63 to 1.4, indicating that the variance across subjects was greater than the variance within subjects.

Table 2.1

Between- and within-subjects standard deviation for each modulation task. Ratio represents the ratio of the between- and within-subjects SD.

Task	**Between-Subjects SD**	**Within-Subjects SD**	**Ratio**
Dichotic AM (1 Hz)	5.61	2.13	2.63
Diotic AM (1 Hz)	4.64	2.26	2.05
Dichotic AM (20 Hz)	4.44	2.26	1.97
Dichotic FM (1 Hz)	3.38	1.71	1.98
Dichotic FM (20 Hz)	3	1.69	1.78
Diotic AM (20 Hz)	2.95	1.66	1.77
Diotic FM (1 Hz)	1.84	1.24	1.48
Diotic FM (20 Hz)	1.46	1.04	1.4

To estimate the highest possible correlation our methods are capable of detecting, we calculated the average correlation based on 100,000 simulated test-retest correlations. First, six runs (three “test” and three “retest” runs) were sampled from each individual subject’s estimated distribution, based on their actual mean and standard deviation for a given modulation task. Next, a simulated test-retest correlation was calculated using the average simulated test and retest mean for each subject. This iteration was completed 100,000 times, producing 100,000 simulated test-retest correlations. The test-retest correlations were transformed using Fisher’s r to z transformation, averaged, and then the average was transformed back to *r*. This process was completed for all modulation tasks, with the average simulated correlations ranging from $r = .96$ for slow dichotic AM to $r = .86$ for fast diotic FM (average across conditions was $r = .92$). The high average simulated test-retest correlations indicate the ratio of between-subjects to within-subjects variance should be large enough for our correlations between tasks to be sensitive to individual differences between subjects.

2.3.3 Correlational analyses of FM and AM thresholds

We expected all diotic tasks to correlate with their dichotic counterpart, as monaural processing of TFS or envelope cues should be related to performance on binaural tasks that utilize these same cues. As predicted, slow diotic FM thresholds correlated positively with slow dichotic FM thresholds ($r = .42$, $p < .0001$) and fast diotic FM thresholds correlated positively with fast dichotic FM thresholds ($r = .54$, $p < .0001$); see Fig. 2.3A. Correlations for the AM data were similar to those found for the FM data: slow diotic AM correlated with slow dichotic AM ($r = .57$, $p < .0001$) and fast dichotic AM correlated with fast diotic AM ($r = .56$, $p < .0001$); see Fig. 2.3B.

Taken at face value, the strong correlation between slow diotic FM thresholds and slow dichotic FM thresholds could be interpreted as support for the hypothesis that phase-locked temporal information underlies performance in both tasks. A similarly strong correlation was also observed between the fast AM thresholds and fast FM thresholds for both diotic ($r = .5$, $p < .0001$) and dichotic ($r = .5$, $p < .001$) conditions, as would be expected if fast FM were detected via FM-to-AM conversion by cochlear filtering (e.g., Zwicker, 1970; Saberi and Hafter, 1995). Unfortunately, our results do not provide support for the dichotomy between temporal coding for slow

FM and rate-place coding for fast FM, because most of the other modulation thresholds were also correlated with each other. In particular, if the correlation between diotic and dichotic slow FM thresholds reflects a common underlying (temporal) mechanism, then we would expect the correlations between thresholds for stimuli that do not share the same underlying mechanism to be lower. In fact, essentially all the modulation detection tasks were highly correlated with each other. For instance, slow FM and fast FM thresholds were correlated ($r = .56$, $p < .0001$), as were slow FM and slow AM ($r = .5$, $p < .0001$) and slow FM and fast AM ($r = .43$, $p < .0001$) (see Fig. 2.3C and 2.3D), despite the fact that these pairs are often regarded as being coded by different peripheral mechanisms. Thus, our results provide no clear support for the idea that performance in slow FM detection tasks is limited by different mechanisms than performance in fast FM or AM detection tasks.

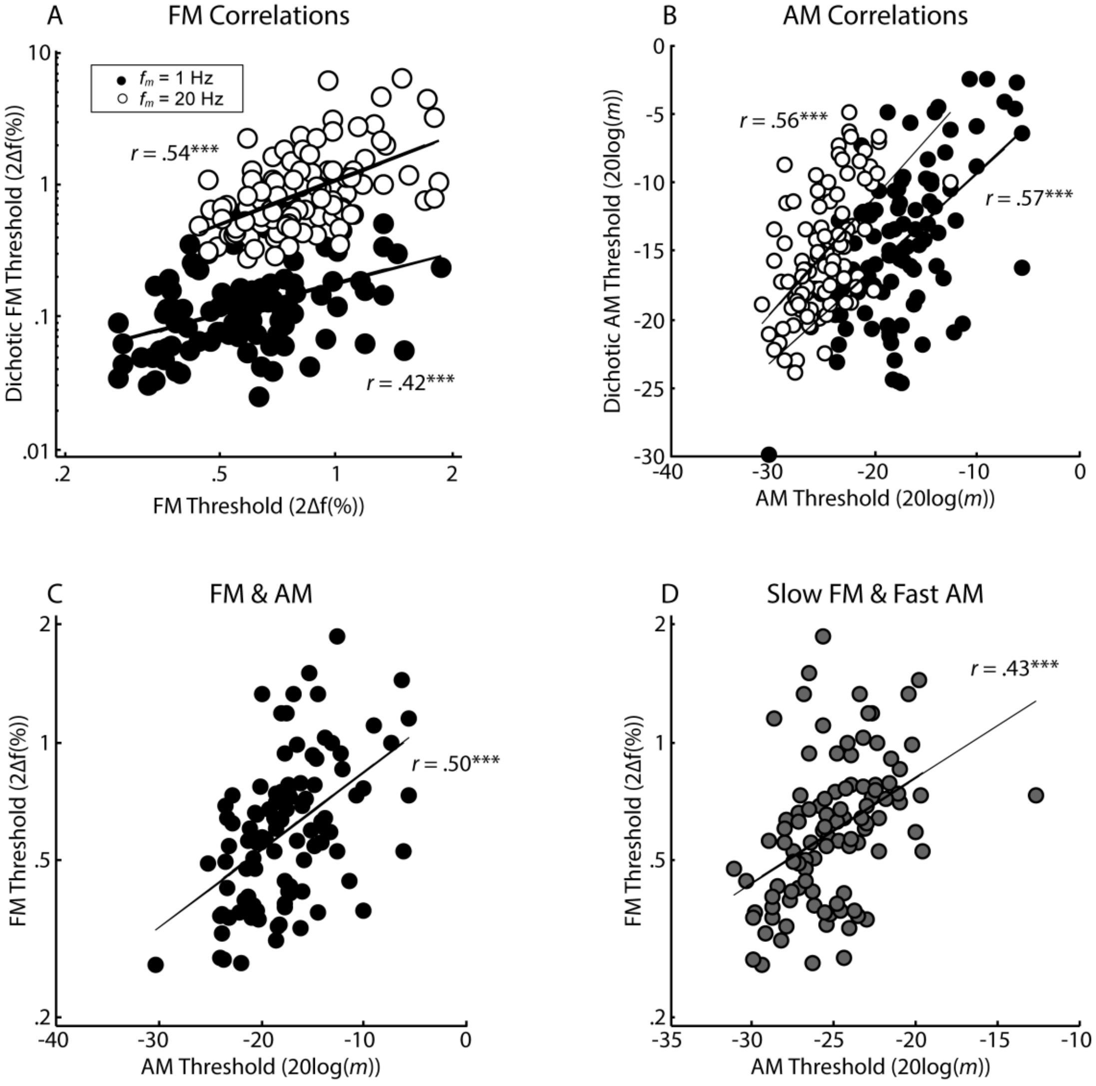

Figure 2.3

Correlations between diotic and dichotic (A) FM and (B) AM. Black circles correspond to f_m = 1 Hz, and white circles correspond to f_m = 20 Hz. Grey circles (D) correspond to different modulation rates on the x and y-axis. The black lines are the lines of best fit. For (A), both the x and y-axis are plotted in peak-to-peak frequency change (2Δ*f*(%)), where Δ*f* is the frequency excursion from the carrier (in percent). (B) The x and y-axis are plotted in 20log(*m*), where *m* is the modulation index. Panel (C) plots the correlation between slow diotic FM detection and slow diotic AM detection, while panel (D) plots the correlation between slow diotic FM and fast diotic AM. The *** indicates correlations that were highly significant ($p < .0001$).

2.3.4 Frequency selectivity and FM detection

Mean thresholds (and standard deviations across the 100 subjects) for detection of the 20-ms signal in quiet and in the presence of the 500-Hz forward masker are shown in Fig. 2.4. Mean absolute thresholds for each subject were obtained by averaging thresholds across the eight signal frequencies, and mean masker effectiveness was estimated for each subject by averaging all eight forward-masked thresholds. Frequency selectivity was estimated for each subject by calculating the slope of the masking functions below and above the masker frequency separately using masker threshold as a function of the frequency separation of the masker and target in octaves. The linear regression resulted in slope estimates in units of dB/octave below and above 500 Hz. Boxplots of the lower and upper slopes of the forward masking pattern are presented in Fig. 2.5. As expected based on numerous studies of frequency selectivity (e.g., Patterson, 1976; Glasberg and Moore, 2000; Shera et al., 2002), the median slopes were relatively steep, and the slope of the lower side of the masking pattern was significantly steeper than the slope of the higher side (paired t-test; $t = 39.3$, $p < .0001$).

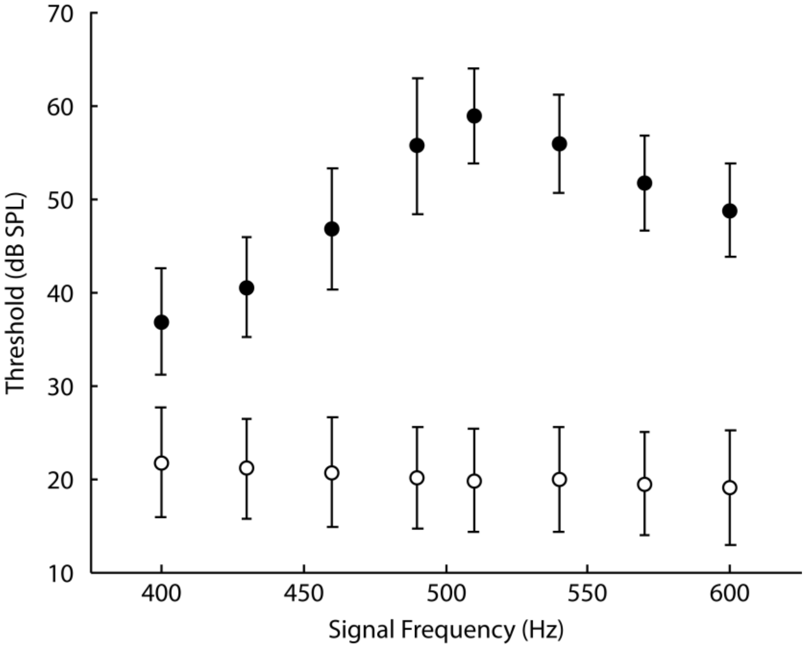

Figure 2.4

Average signal frequency detection thresholds. Open circles represent the average absolute threshold for each of the signal frequencies when no masker is present. Filled circles represent the average detection threshold for each of the signal frequencies when preceded by a 500-ms, 500-Hz pure-tone masker. Error bars correspond to standard deviations across the 100 subjects.

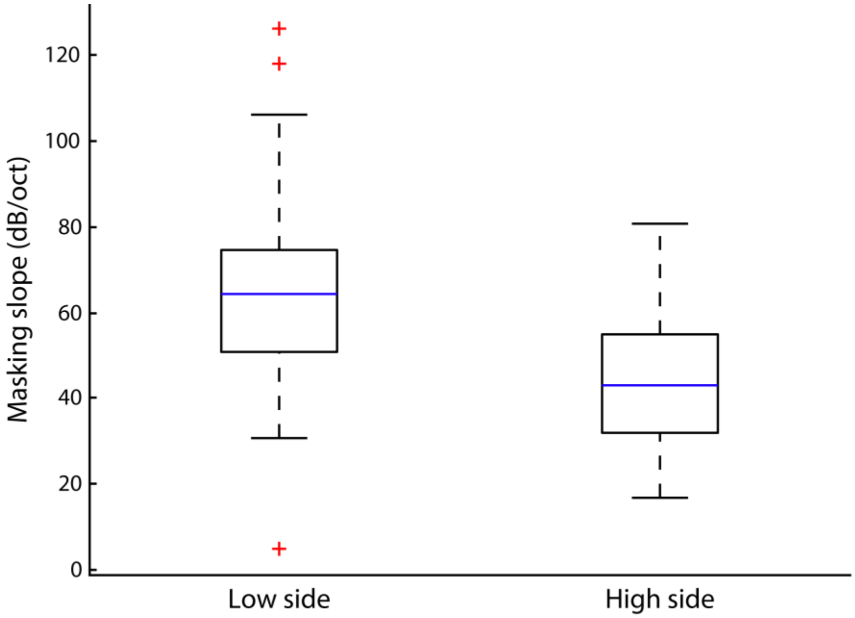

Figure 2.5

Boxplots of slopes from the forward-masking patterns. The low-side values represent the estimated slope of the masking pattern below the masker frequency. The high-side values represent the absolute value of the estimated slope of the masking pattern above the masker frequency. Center lines represent the median of each group (color online). Whiskers correspond to the lowest and highest data points within 1.5 times the lower and higher inter-quartile ranges, respectively. Small crosses represent individual data points outside the range of the whiskers, considered outliers.

If fast-rate FM detection relies on detecting the AM induced by passing the FM stimulus through the auditory filters, then FM thresholds should be predicted by the combination of sensitivity to AM and the auditory filter slopes. Specifically, the FMDLs should approximate the smallest detectable change at the output of the characteristic frequency filter, divided by the slope of that filter (Zwicker, 1956; Moore and Glasberg, 1986; Lacher-Fougère and Demany, 1998). Predicted fast- and slow-rate FM thresholds were based on the individual subjects' fast- and slow-rate AM thresholds and their steeper masking-pattern slopes, which was the lower slope for 88 of the 100 subjects. Correlations between the measured and predicted FMDLs were significant for both slow ($r = .41$, $p < .0001$) and fast ($r = .38$, $p < .0001$) modulation rates, although the magnitude of this effect was moderate. Again, at face value, the result appears to indicate that frequency selectivity is related to both slow and fast FM detection, but that frequency selectivity does not explain the majority of the inter-individual variance in FM detection. Again, there was no clear difference between the correlations for slow- and fast-rate FM thresholds, inconsistent with idea that fast- and slow-rate thresholds are governed by different mechanisms. Most importantly, these moderate correlations between predicted FM and measured FM are confounded by the high correlations between slow FM and slow AM ($r = .5$) and fast FM and fast AM ($r = .5$). Because predicted FM thresholds are calculated based on AM sensitivity divided by the steeper filter slope, and FM and AM are well correlated, a correlation between measured and predicted FM would likely be present regardless of the steepness of the filter slopes. In fact, correlations between predicted and actual FM thresholds are actually lower than just the raw correlations between FM and AM thresholds, suggesting that adding the filter slopes provides no additional information to the predictions. Thus, the correlations between measured and predicted FM are driven by the high correlations between AM and FM, rather than the individual differences in frequency selectivity.

Although the correlations between measured and predicted FM are clearly driven by the correlations between AM and FM, rather than masking-pattern slopes, the group averages between measured and predicted FMDLs may still provide useful information. A 2x2 (threshold type vs. modulation rate) within-subjects ANOVA was conducted on the log-transformed thresholds for predicted and measured FMDLs [i.e., $10\log(\%\Delta f)$]. Results indicated a main effect of threshold type [$F(1,99) = 492$, $p < .0001$, $\eta_p^2 = .833$], a main effect of modulation rate [$F(1,99) = 76.9$, $p < .0001$, $\eta_p^2 = .437$], and a significant interaction [$F(1,99) = 423$, $p < .0001$, $\eta_p^2 = .81$]. A *post hoc t*-test indicated that predicted slow FMDLs ($M = 1.1$, $SD = 2.43$) were significantly larger (i.e., poorer) than measured slow FMDLs ($M = -5.27$, $SD = 1.84$) ($p < .0001$). Consistent with previous literature (e.g., Moore and Glasberg, 1986; Lacher-Fougère and Demany, 1998), rate-place information alone, based on the single largest change in the excitation pattern, far underestimates listeners' actual ability to detect slow FM.

More surprisingly, a *t*-test showed a similar trend between predicted fast FMDLs (*M* = -2.7, *SD* = 1.77) and measured fast FMDLs (*M* = -3.94, *SD* = 1.46) ($p < .0001$), although the difference between the predicted and measured means for fast FMDLs was smaller.

If FM relies on excitation pattern information, then FM should be related to the steepness of the auditory filter slopes. However, neither the low slope, high slope, or the overall steepness of both filter slopes (calculated as the low slope summed with the absolute value of the high slope) were correlated with slow diotic FM (low slope: $r = .17$; high slope: $r = -.14$; overall steepness: $r = .2$) using a one-tailed test. In fact, each of the correlations were opposite of the predicted direction, as steeper low slopes are positive (bigger numbers), which should be negatively related to better (smaller) FMDLs, and steeper high slopes are negative (smaller numbers), which should be positively related to better FMDLs. There was no correlation between fast diotic FM and filter slopes (low slope: $r = .04$; high slope: $r = -.04$; overall steepness: $r = .05$).

It is possible that the correlations between diotic FM and filter slopes are unobservable because FM thresholds are overshadowed by variability in sensitivity to AM, assuming FM is converted to AM in the cochlea. In order to account for sensitivity to AM, both slow-rate and fast-rate FMDLs and AMDLs were z-transformed so that they were on the same scale. The z-transformed AMDLs were subtracted from FMDLs at the corresponding modulation rate. These difference scores were then correlated with the z-transformed filter slopes. Correlations were conducted between the difference scores and the low slope, high slope, and overall steepness for both slow and fast FM. Although the correlations between slow difference scores and filter slopes were in the predicted direction, they were not significant (low slope: $r = -.02$, $p = .42$; high slope: $r = .12$, $p = .12$; overall steepness: $r = -.08$, $p = .21$). The correlations between fast difference scores and frequency selectivity were slightly better, but still very weak (low slope: $r = -.09$, $p = .19$; high slope: $r = .2$, $p = .02$; overall steepness: $r = -.17$, $p = .045$), with the high slope and overall steepness reaching significance without correcting for multiple comparisons. Assuming the correlations between slow-rate difference scores and the overall steepness is reflective of the true population, one would need a sample size of n=427 to reach significance using a one-tailed test. Overall, there was very little evidence for a relationship between either slow- or fast-rate FM and frequency selectivity, even when controlling for sensitivity to AM.

2.3.5 Principal components analysis

Given the large number of measures involved in our study, we conducted a principal components analysis (PCA), using the average thresholds for all 100 participants on each of the tasks. The average absolute threshold across the signal frequencies was included in the PCA analysis, as listeners with good sensitivity would have, on average, lower absolute thresholds across the signal frequencies. Similarly, the average forward masking threshold across the signal frequencies was included as a measure of the overall effect of masking. Listeners with a higher overall effect of masking would have, on average, higher thresholds across all of the frequencies, regardless of slope. The overall steepness of the filter slopes from the forward-masking patterns task was included as a measure of frequency selectivity.

PCA is an important exploratory analysis to conduct because it could reveal a different structure in the dataset that is not intuitively obvious from the full correlation matrix. This is because PCA takes into account the relationship of each task with every other task when performing the dimension reduction. With 55 possible correlations, the dataset is too large to safely intuit the multivariate structure by simply inspecting the correlation matrix. In addition, PCA should isolate any common variance across conditions based, for instance, on “attentiveness,” or other non-sensory factors that may be shared by many or all of the measures. Based on our initial hypotheses, the PCA should produce separate components that reflect peripheral rate-place coding (i.e., tonotopy) and time coding (i.e, phase-locking to TFS cues). For example, if slow FM is coded via phase-locking to TFS cues, but fast FM is not, then slow diotic and dichotic FM would load on to one component. Slow and fast AM, diotic and dichotic, would load onto a second component, reflecting sensitivity to envelope cues. If fast FM is converted to AM via cochlear filtering, than fast FM would load onto a component with frequency selectivity as well as fast AM. Because some of the tasks were measured in different units (e.g., dB/oct for filter slopes vs. 20log(m) for amplitude modulation), PCA was conducted using an eigen-decomposition on the correlation matrix. Because PCA was performed on the correlation matrix, the amount of variance accounted for by each component reflects the variance accounted for when each task has a standardized variance ($s^2 = 1$). This ensures that tasks measured in larger units (with, consequently, arbitrarily larger variances) do not dominate the component loadings.

The results from the PCA did not reflect our predicted results. The PCA (varimax rotated) produced three interpretable components, accounting for approximately 63.3% of the standardized variance (Fig. 2.6). Component 4 only accounted for an additional 11% of the standardized variance, and was not clearly interpretable, so was not included in the analysis.

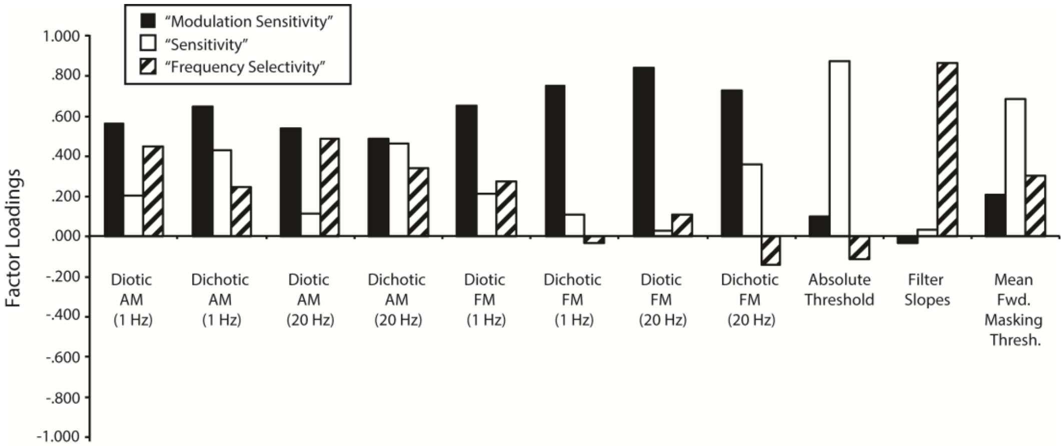

Figure 2.6

Principal components analysis suggesting that three factors account for the majority of the variance. The x-axis groups tasks based on the component for which they had the greatest factor loading. Solid bars correspond to the "Modulation Sensitivity" component, open bars correspond to the "Sensitivity" component, and striped bars correspond to the "Frequency Selectivity" component.

All of the FM and AM, diotic and dichotic, tasks loaded onto the first component, which accounted for 32% of the standardized variance. Thus, component 1 was named the "Modulation sensitivity" component, as it appears to reflect a general ability to perform AM- and FM-related tasks in both diotic and dichotic situations at both slow and fast rates. The names of components 2 and 3 were determined based on the task that loaded most onto the components. Component 2 (accounting for 17% of the standardized variance) had the highest loadings by average absolute and average forward-masked threshold, and so was termed “Sensitivity”, while component 3 (accounting for 14.3% of the standardized variance) most strongly reflected filter slopes and so was termed “Frequency Selectivity.”

Consistent with the earlier analysis based on paired correlations, the PCA provided no evidence for separable coding mechanisms reflecting phase-locking for slow FM and rate-place coding for fast FM. Frequency selectivity appeared to be related to neither, while binaural sensitivity to temporal fine structure (as reflected in the slow dichotic FM thresholds) was equally related to diotic FM, as well as AM, at both slow and fast rates. Although the PCA reiterates the observed patterns in the correlational analyses, it provides a parsimonious description of the dataset and confirms that no other interpretable structure appears to exist within the dataset.

2.4 Discussion

2.4.1 Summary of results

The aim of this study was to use individual differences in a large cohort of young normal-hearing listeners to test the hypothesis that the coding of slow-rate FM is based on temporal (phase-locked) information, whereas the coding of fast-rate FM is based on rate-place information, through the transformation of FM to AM via peripheral auditory filtering. Strong correlations were observed between most of the modulation detection and discrimination tasks. The two main findings were not consistent with the predictions of the hypothesis. First, the correlation between a measure known to reflect timing information (dichotic FM disparity detection) was not more strongly correlated with slow FM detection than with any of the other monaural (or binaural) modulation detection tasks. Second, the measure of frequency selectivity combined with the measure of AM sensitivity did not predict fast-rate (or slow-rate) FM thresholds any better than just the measure of AM sensitivity, suggesting no clear relationship between frequency selectivity and either slow- or fast-rate FM detection thresholds. An exploratory PCA approach resulted in the same conclusions: the diotic and dichotic, slow- and fast-rate, FM and AM detection thresholds were all related to one another, but were generally unrelated to measures of absolute threshold, masked threshold, or frequency selectivity.

2.4.2 Comparisons with previous studies

One concern is that the large number of subjects precluded extended practice in any of the tasks before measurement. It may be, therefore, that the thresholds obtained by our listeners do not reflect the sensory limits of FM or AM detection, but rather reflect more cognitive or procedural limitations that might have been overcome by further training. To test for this possibility, we compared our thresholds with those reported in the literature from smaller, more practiced, groups of subjects.

In general, although listeners completed only 3-6 runs for each condition, average diotic FMDLs and AMDLs are comparable to those from well-trained listeners in the psychoacoustic literature. For example, the average peak-to-peak frequency change ($2\Delta f$) across all 100 participants was 0.6% for slow FM and 0.81% for fast FM. Using three well-trained listeners with a 500-Hz carrier and similar slow and fast modulation rates, the average FMDLs for Moore and Sek (1996) were 0.9% for $f_m = 2$ Hz and 1.3% for $f_m = 20$ Hz. It is possible that our FMDLs may be slightly better than the trained listeners in the earlier study because the durations of our stimuli were twice as long (2 s, as opposed to 1 s). Demany and Semal (1989) also used a 2-s duration, and trained their participants until thresholds were stable, and obtained similar average FMDLs [$f_m = 1$ Hz: $M = 0.732\%$; $f_m = 16$ Hz: $M = 0.902\%$]. Notably, our listeners were first exposed to FM and AM tones via the dichotic tasks, so they were not completely untrained with respect to exposure to FM or AM.

Our AMDLs are not as straightforward to compare with values in the literature, as most previous studies have used a different carrier frequency, modulation rate, and/or methods of measurement (e.g., constant stimuli procedures to calculate d'). To make comparisons across studies, measures of d' from previous literature were transformed to approximate the 79% correct point. AMDLs are reported in $20\log(m)$. All studies reported used a 2-interval, 2-alternative forced choice procedure. The average AMDLs across the 100 listeners from the current study was -17.6 dB for slow AM and -25.0 dB for fast AM. This is roughly comparable to that found by Moore and Sek (1996) using a higher carrier frequency (e.g., $M = -23.0$ dB for $f_c = 1$ kHz and $f_m = 2$ Hz). With the same 1-kHz carrier and three listeners with “extensive practice,” Moore and Sek (1995) reported average AMDLs of approximately -17.7 dB for f_m = 2 Hz and -26.0 dB for f_m = 10 Hz, very similar to our mean results.

To our knowledge, only one other study has measured dichotic FMDLs (Grose and Mamo, 2012). In this study, young listeners trained on dichotic FM until thresholds "appeared stable," and stable performance was achieved with just an average of 1.6 practice runs. Grose and Mamo used a similar low carrier (frequency roved: $460 <= f_c <= 540$) and slow modulation rate (f_m = 2 Hz), and obtained comparable thresholds [$2\Delta f$ = .17% in Grose and Mamo vs. $2\Delta f$ = .12% in the present study]. There are several methodological differences that may account for better thresholds in our study: Grose and Mamo 1) roved the carrier frequency, 2) had a pure tone reference instead of an FM tone as a reference, and 3) used a three-interval task as opposed to two-interval.

In summary, despite the relative small amount of practice provided to our subjects, the average FM and AM thresholds obtained in our study are very comparable to those reported in earlier studies. It therefore seems unlikely that the lack of clear differences based on underlying coding mechanisms reflects generally poor performance on the part of our subjects.

2.4.3 Analyses of subsets of data

Another way to address the potential effects of generally poor performance is to examine the results from a subset of better performers. The rationale is that the subjects with the lowest thresholds are most likely to have reached their sensory limits and so are more likely to reflect variance based on sensory limitations. To test this hypothesis, we reanalyzed the data from the 'best' 30 and 'worst' 30 listeners, based on their values for the first component in the PCA, and retested the idea that fast FMDLs should be predicted by a combination of fast AMDLs and masking slope. If the best 30 listeners' results more closely reflect sensory limitations, then the correlations should be higher in that group than in the whole group. In fact, the correlations between measured and predicted FMDLs did not reach significance for either the best 30 listeners (slow FM: $r = .15$, $p = .21$; fast FM: $r = .22$, $p = .12$) or the worst 30 listeners (slow FM: $r = .18$, $p = .17$; fast FM: $r = .09$, $p = .32$). The lack of correlation can probably be explained by the reduced range of thresholds in the relevant modulation-detection tasks, but also suggests that the main findings of this study are not due solely to poor performers being limited by non-sensory factors.

2.4.4 Variability in peripheral coding

Although the correlations between all FM and AM tasks are high, correlations between tasks thought to utilize the same peripheral coding mechanisms are not as high as one might expect given the supposed importance of phase-locking in slow FM detection and frequency selectivity in fast FM detection. One possible explanation is that the variability in peripheral coding in young, normal-hearing listeners is too small to exert a large influence on thresholds. The variability in our measure of phase-locking to TFS cues [expressed as $2\Delta f$(%)] is, in fact, more than a factor of three smaller than the variability observed in 12 older listeners from Grose and Mamo [SD = .516% in Grose and Mamo vs. SD = .163% in the present study]. It has been suggested that coding of TFS declines with both age (e.g., Hopkins and Moore, 2011; Grose and Mamo, 2012; Moore et al., 2012) and degree of hearing loss (e.g., Hopkins and Moore, 2007, 2011; Lorenzi et al., 2009), which would contribute to the increased variability of TFS coding.

In addition, it is well known that auditory filter slopes become shallower with sensorineural hearing loss (e.g., Glasberg and Moore, 1986), and the variability in bandwidth across hearing-impaired listeners is quite wide (e.g., Moore et al., 1999). Although the variability in the steepness of the auditory filter slopes was quite large in our young, normal-hearing listeners (see Fig. 2.5), this between-subjects variability would certainly increase if the subject pool were expanded to include older and hearing-impaired listeners. Future studies including a large sample of different ages and degrees of hearing impairment may elucidate whether variability in peripheral coding can be made large enough to outweigh other factors determining individual differences in performance.

2.4.5 Limitations of correlational studies

Taken at face value and out of the context of the other results, the strong correlation between dichotic FM disparity detection and diotic FM detection at slow rates could have been interpreted as evidence that phase-locking to temporal fine structure dominates for slow FM detection. It was only the equally strong correlations between measures not thought to be related to phase-locking (such as fast-rate FM and AM detection) that cast doubt on this interpretation. Similar correlational analyses have become a popular method for examining questions of underlying neural coding in normal and impaired hearing, and as a function of age (e.g., Strelcyk and Dau, 2009; Ruggles et al., 2011; Ochi et al., 2014; Bharadwaj et al., 2015). Caution is required in interpreting the results from such studies, as they rarely include measures of performance that are similar in task nature but are thought to reflect different underlying neural mechanisms. In other words, it can be important to provide measures that are *not* correlated with the others in order to demonstrate specificity of the putative mechanisms, and to ensure that the correlation does not reflect higher-level central processing that is not specific to particular underlying mechanisms. In a related domain, studies predicting FMDLs or frequency difference limens based on sensitivity to level changes and frequency selectivity need to account for possible high correlations between frequency discrimination/detection thresholds and intensity discrimination/detection thresholds (Moore and Glasberg, 1986; Dai et al., 1995).

2.4.6 Explaining the high sensitivity to slow FM

Although our correlational approach provided no strong evidence for two distinct coding mechanisms for slow- and fast-rate FM, the fact remains that thresholds for slow-rate FM are generally lower (better) than fast-rate FM thresholds. In contrast, our slow-rate AM thresholds were considerably higher (worse) than the fast-rate AM thresholds. How can this difference be explained if both slow and fast FM detection are governed by the same underlying mechanisms?

One potential explanation lies in a recent solution to the long-standing problem for why sensitivity to intensity changes and frequency selectivity seem unable to account for sensitivity to frequency changes. Micheyl et al. (2013) proposed that pure-tone frequency-discrimination performance could be explained by a cortical population rate code, relying entirely on a neural population rate-place code, which could also explain human intensity discrimination performance. Their model relied on some correlation between spike counts of neurons with similar characteristic frequencies. This correlation resulted in a deterioration in intensity coding and a relative improvement in frequency coding, leading to reasonable predictions of thresholds in both dimensions. As spike correlations rely on a certain time window over which to count spikes, the effects of correlations between neural units will decrease with decreasing analysis duration. Thus, the relative benefit of neural correlations for frequency coding will be observed more for long durations (or slow FM rates) than for short durations (or fast FM rates). This explanation may provide the basis for an account of the different sensitivity between slow and fast FM without the need for a separate neural code. Similarly, the decrease in frequency discrimination abilities at high frequencies may reflect cortical coding limitations (perhaps based on the tonotopic distribution of responses), rather than peripheral limitations based on phase-locking (e.g., Oxenham et al., 2011). However, further modeling work is required to test this conjecture.

[1] Due to a programming error, 5 additional participants also began each run with Δf = .6%.

CHAPTER 3: AGING AND FM DETECTION

Chapter 3 is reprinted from:

Whiteford, K. L., Kreft, H. A., & Oxenham, A. J. (2017). Assessing the role of place and timing cues in coding frequency and amplitude modulation as a function of age. *Journal of the Association for Research in Otolaryngology, 18*, 619-633.

Abstract

Natural sounds can be characterized by their fluctuations in amplitude and frequency. Ageing may affect sensitivity to some forms of fluctuations more than others. The present study used individual differences across a wide age range (20-79 yr) to test the hypothesis that slow-rate, low-carrier frequency modulation (FM) is coded by phase-locked auditory-nerve responses to temporal fine structure (TFS), whereas fast-rate FM is coded via rate-place (tonotopic) cues, based on amplitude modulation (AM) of the temporal envelope after cochlear filtering. Using a low (500-Hz) carrier frequency, diotic FM and AM detection thresholds were measured at slow (1 Hz) and fast (20 Hz) rates in 85 listeners. Frequency selectivity and TFS coding were assessed using forward masking patterns and interaural phase disparity tasks (slow dichotic FM). Comparable interaural level disparity tasks (slow and fast dichotic AM and fast dichotic FM) were measured to control for effects of binaural processing not specifically related to TFS coding. Thresholds in FM and AM tasks were correlated, even across tasks thought to use separate peripheral codes. Age was correlated with slow and fast FM thresholds in both diotic and dichotic conditions. The relationship between age and AM thresholds was generally not significant. Once accounting for AM sensitivity, only diotic slow-rate FM thresholds remained significantly correlated with age. Overall, results indicate stronger effects of age on FM than AM. However, because of similar effects for both slow and fast FM when not accounting for AM sensitivity, the effects cannot be unambiguously ascribed to TFS coding.

3.1 Introduction

Dynamic changes in pitch are fundamental for communicating contour in speech and music. The ability to detect changes in pitch is in part related to the efficacy with which the cochlea and auditory nerve transduce sound into neural impulses. For sinusoidal FM, two potential peripheral cues involve rate-place (tonotopic) and temporal (phase-locking) information. According to the rate-place coding theory, FM is detected by fluctuations in the firing rate of auditory neurons as the instantaneous frequency of the tone changes, resulting in shifts of the excitation pattern. In this way, FM is converted to AM by cochlear filtering (Zwicker 1956; Maiwald 1967a,b; Zwicker 1970; Moore and Sek 1992; Moore and Sek 1994; Saberi and Hafter 1995). In contrast, the temporal code relies on neural spikes that are phase-locked to the vibrations of the basilar membrane, providing the auditory system with time-interval based information relating to the TFS to convey the presence of FM (e.g., Moore and Sek 1995; Moore and Sek 1996).

It has been proposed that FM with a low-frequency carrier (f_c < 4-5 kHz) at slow modulation rates (f_m < ~10 Hz) utilizes a temporal code (Demany and Semal, 1989; Moore and Sek, 1995, 1996; Lacher-Fougère and Demany, 1998; Moore and Skrodzka, 2002), whereas FM at faster rates (f_m ≥ ~10 Hz) for the same low-frequency carriers and higher carriers at all rates (up to rates at which the sidebands become spectrally resolved; e.g., Hartmann and Hnath, 1982) utilizes a rate-place code (e.g., Moore and Sek 1992; Moore and Sek 1994; Saberi and Hafter 1995). There is some evidence suggesting that performance in tasks relying on neural phase locking may degrade with age. For example, older listeners generally perform more poorly than younger listeners on interaural phase and time difference (IPD and ITDs, respectively) detection/discrimination (Grose and Mamo, 2010; Moore and Ernst, 2012; Füllgrabe, 2013; Gallun et al., 2014; King et al., 2014; Füllgrabe et al., 2015). Age effects on IPDs are present even when controlling for audiometric thresholds between young and older participants (Füllgrabe et al., 2015), and may be present as early as middle age (Ross et al., 2007; Grose and Mamo, 2010; Füllgrabe, 2013). Discrimination of harmonic from inharmonic stimuli, believed to require TFS coding, is also poorer in older participants (Hopkins and Moore, 2011; Füllgrabe, 2013). Although there is consensus that phase locking is required for IPD/ITD-based tasks for pure tones, it remains possible that tasks involving frequency coding, including frequency discrimination and FM tasks, are coded via a rate-place mechanism (e.g., Oxenham et al. 2009; Oxenham et al. 2011; Micheyl et al. 2013), or some combination of rate-place and time coding (e.g., Loeb et al. 1983; Shamma 1985; Shamma and Klein 2000; Loeb 2005).

Several studies have found that older participants perform more poorly at low-carrier, slow-rate FM detection than do younger (He et al., 2007; Strelcyk and Dau, 2009; Grose and Mamo, 2012; Paraouty et al., 2016; Wallaert et al., 2016; Paraouty and Lorenzi, 2017) or middle-aged participants (Grose and Mamo, 2012), even when all participants have NH at the carrier frequency (He et al., 2007; Strelcyk and Dau, 2009; Grose and Mamo, 2012; Paraouty et al., 2016; Wallaert et al., 2016; Paraouty and Lorenzi, 2017). These results are also consistent with the theory that temporal coding of TFS degrades with age in the absence of audiometric loss at the test frequency. However, some of these studies did not include comparable measures (such as AM detection or FM detection at fast rates) that are not thought to involve temporal coding of TFS. The lack of such "control" measures makes it difficult to rule out more general effects of ageing, such as changes in cortical sensory coding or cognitive function (e.g., attention or processing speed). For example, Grose and Mamo (2012) found that older, NH adults were worse at slow-rate, low-carrier FM detection relative to younger adults, but it is unclear whether they would have found the same effect for fast-rate FM detection at the same carrier. Even among studies that have used exclusively slow-rate FM, the outcomes have not been completely consistent. For instance, Schoof and Rosen (2014) measured slow FM difference limens (FMDLs) ($f_m = 2$ Hz; $f_c = 1$ kHz) in young (range: 19-29 years) and older (range: 60-72 years) listeners, but found no difference between the age groups. In this respect, as well as in several other measures examined by Schoof and Rosen (2014), their results are unusual in finding no perceptual deficits associated with ageing, perhaps in part because of their strict definition of NH for the older group.

A few recent studies have used correlational measures in NH listeners to examine what peripheral code may be responsible for low-frequency carrier FM and have found conflicting evidence for the presence of TFS coding (e.g., Whiteford and Oxenham 2015; Otsuka et al. 2016; Paraouty and Lorenzi 2017). These studies have revealed high multi-collinearity across modulation-detection tasks, including those thought to use separate mechanisms (e.g., low carrier, slow-rate FM and slow-rate AM). For instance, Whiteford and Oxenham (2015) used binaural modulation tasks to assess the fidelity of TFS coding based on IPDs (slow-rate dichotic FM) and level cues (slow- and fast-rate dichotic AM and fast dichotic FM), and found that many pairs of tasks (e.g., slow-rate dichotic FM and slow-rate dichotic AM) were correlated as strongly as pairs thought to share the same peripheral code (e.g., slow-rate dichotic FM and slow-rate diotic FM). The non-specific correlations could indicate that the variability between young NH listeners is driven primarily by non-peripheral factors, or that FM and AM use the same peripheral code. Paraouty and Lorenzi (2017) used a large sample of listeners varying in age to potentially increase the variability in peripheral TFS coding, and found that thresholds for low carrier FM with a 5 Hz modulation rate and low carrier AM of the same rate were no longer correlated once AM was added to FM in order to disrupt potential excitation-pattern cues. This could suggest that slow FM uses a combination of temporal and place coding; however, it is also possible that even at 5 Hz there may be less viable TFS cues, given that the upper limit of extracting TFS cues in low carrier FM is estimated to be around 10 Hz (e.g., Moore and Sek 1995).

The present study measured slow-rate (f_m = 1 Hz) and fast-rate (f_m = 20 Hz) FM and AM detection in both diotic and dichotic conditions, along with a measure of frequency selectivity based on forward masking, in a large cohort of participants whose ages ranged from 20 to 79 years. The paradigm was similar to that used in our earlier study of only young NH listeners (Whiteford and Oxenham, 2015). The purpose of the dichotic tasks was to assess performance for disparity detection when TFS cues are necessary to complete the task (slow dichotic FM) relative to performance on the same task when TFS cues are not thought to be available (slow dichotic AM, but also fast dichotic FM and fast dichotic AM). The prediction of the study was that a selective deficit in the temporal coding of TFS should lead to poorer detection of slow-rate FM in both diotic and dichotic conditions, in ways that are unrelated (or at least less related) to AM or fast-rate FM detection. It was expected that age effects would be less likely to occur for diotic AM detection, given that previous studies have found either no effect (for f_m = 5 Hz: He et al. 2008; Paraouty and Lorenzi 2017) or small effects (Füllgrabe et al., 2015; Wallaert et al., 2016) of age on sinusoidal AM detection at modulation rates comparable to that used in the present study.

3.2 Methods

3.2.1 Participants

Eighty-five adults (25 male, 60 female, mean age of 48.5 years, range: 20.1-79.5) from the University of Minnesota and surrounding community participated in this study. There were 15 participants from each decade of age between 20 and 69 years, and 10 participants between 70 and 79 years. Audiometric thresholds were assessed at octave-spaced frequencies between 250-8000 Hz. All participants had NH for low frequencies, defined as a low-frequency pure tone average (PTA) (250 Hz, 500 Hz, and 1000 Hz) ≤ 20 dB hearing level (HL) in both ears, with no low-frequency PTA asymmetries greater than 10 dB. Across-ear average low-frequency PTA tended to increase (worsen) with age (*r* = .56, *P* < .0001, two-tailed). Average audiometric thresholds for each decade are plotted in Fig. 3.1. Participants provided written informed consent and were compensated with course credit or hourly payment for their time. The experimental protocols were approved by the Institutional Review Board of the University of Minnesota.

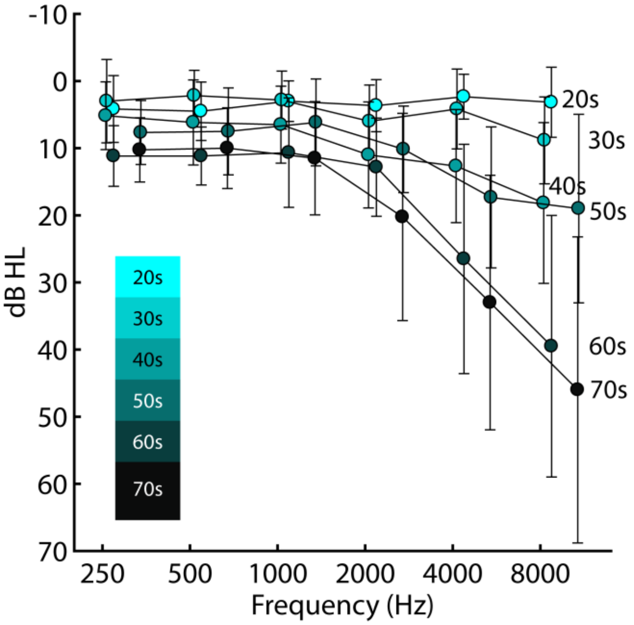

Figure 3.1

Average audiometric thresholds for each decade of age. Error bars represent ±1 standard deviation.

3.2.2 Stimuli

With the exception of the first task (absolute thresholds for the carrier frequency), all stimuli and procedures were identical to those used by Whiteford and Oxenham (2015), and are described below.

Absolute threshold for the 500-Hz carrier frequency was measured separately in each ear using a 500-ms pure tone with 10-ms raised-cosine onset and offset ramps. FMDLs and AM difference limens (AMDLs) were measured diotically at this same frequency (f_c = 500 Hz) for a slow (f_m = 1 Hz) and fast (f_m = 20 Hz) modulation rate. Both the target and the reference tones were 2 s in duration with 50-ms raised-cosine onset and offset ramps. The reference tone was always a 500-Hz pure tone, and the target tone was modulated. On a given trial, the modulator starting phase was set so that the FM target began with either an increase or a decrease in frequency excursion from the carrier frequency, with 50% *a priori* probability. For the diotic AM tasks, the AM target began at either an amplitude peak or an amplitude trough. Stimuli for the dichotic FM and dichotic AM tasks were similar to their diotic counterparts, except that both the reference and target intervals were modulated. The difference was that the reference interval consisted of diotic stimuli, whereas the target interval had a starting modulator phase that was inverted in one ear, leading to dichotic stimulation that created the percept of a moving inter-cranial image for the slow modulation rate, based on ITDs in the case of FM and ILDs in the case of AM. At the fast modulation rate of 20 Hz, the dynamic ITDs and ILDs were too fast to induce the perception of movement. Instead, the target spatial image was perceived as more diffuse and less punctate than that of the reference (diotic) stimulus. For all FM and AM tasks (diotic and dichotic), the target and the reference tones were separated by a 500-ms inter-stimulus interval (ISI). All FM and AM stimuli were presented at 60 dB sound pressure level (SPL) in each ear.

Detection thresholds were also measured for a brief tone (20-ms total duration) with frequencies of 400, 430, 460, 490, 510, 540, 570, and 600 Hz, both in quiet and in the presence of a 500-Hz, 500-ms, pure-tone forward masker. The target frequency was only presented to the right ear, but the forward masker was presented diotically to reduce potential “confusion effects” (e.g., Neff 1986). Both the target and the masker had 10-ms raised-cosine onset and offset ramps. The forward masker level was fixed at 70 dB SPL, while the level of the target varied adaptively. The onset of the target occurred directly after the offset of the masker, resulting in a 10-ms gap between the half-amplitude points of masker and target envelopes. The slope of the masking function (target threshold as a function of masker-target frequency difference in octaves, calculated separately for targets below and above the masker frequency) provided an estimate of frequency selectivity.

All stimuli were generated digitally, converted to analog at a sampling rate of 48 kHz via a LynxStudio L22 soundcard, and presented over open-ear headphones (Sennheiser HD650) in a sound-attenuating chamber.

3.2.3 Procedure

Participants completed eleven tasks across 3-4 sessions. Each session lasted no longer than 2 hours, and most participants completed the entire study within 3 sessions. The only difference in the procedures from Whiteford and Oxenham (2015) was the inclusion of absolute thresholds for a 500-Hz, 500-ms pure tone in each ear. The purpose of including this task was to obtain a more accurate estimate of audibility of the carrier frequency. The first task was the measurement of absolute thresholds for the 500-ms tones at the carrier frequency. All tasks used a two-interval, two-alternative forced-choice procedure, with an adaptive tracking rule that tracks the 79.4% correct point of the psychometric function (Levitt, 1971). The target was randomly presented in either the first or second interval, and the task was to select the interval containing the target by clicking the corresponding virtual button on the screen (labeled “1” or “2”). Feedback (“Correct” or “Incorrect”) was provided after each response. Each task is described in the order it was presented to the listeners. The order is consistent with that used by Whiteford and Oxenham (2015).

3.2.3.1 Task 1: Absolute Threshold for 500-ms Tone

The target was a 500-ms, 500-Hz pure tone, while the reference was 500 ms silence. The target and reference were separated by a 400-ms ISI. Participants were instructed to indicate whether they heard a tone in the first or second time interval, marked by red lights on the virtual response box on the screen. The target was presented at 40 dB SPL in the first trial, and the target level varied by a step size of 8 dB for the first two reversals. The step size was reduced to 4 dB for the third and fourth reversals, and then to 2 dB for the following six reversals. Absolute threshold was defined as the mean level of the last six reversal points. At least three adaptive runs were measured for each ear. If the standard deviation (SD) across the first three runs was ≥ 4 dB, thresholds from an additional three runs were collected, and only the last three runs were used in analyses. The order of the presentation ear (left vs. right) was randomized between runs.

3.2.3.2 Tasks 2 and 3: Dichotic FM Detection

Slow (f_m = 1 Hz) dichotic FM detection was completed first, followed by fast (f_m = 20 Hz) dichotic FM detection. For slow dichotic FM, participants heard two tones, one at a time, and were instructed to pick the tone that sounded as though it were “moving in your head.” Participants were encouraged to view the feedback on the screen to ensure they were listening for the correct feature. The frequency excursion from the carrier (Δf) was varied adaptively in the same manner as in Whiteford and Oxenham (2015). Each run began with the peak-to-peak frequency excursion ($2\Delta f$) set to 0.4%, just below most average diotic FMDLs. The maximum value of the tracking variable was $2\Delta f$ = 2% so that pitch cues would not interfere with the task (i.e., so that the target was perceived as a moving pure tone rather than one that was modulated in pitch). If the adaptive procedure called for a value of $2\Delta f$ that exceeded the maximum allowable value in more than 10 trials within a run, no threshold was recorded and listeners had to complete three additional runs. The value of Δf varied by a factor of 2 for the first two reversals and a factor of 1.4 for the third and fourth reversals. The step size for the final six reversals was reduced to a factor of 1.19. Threshold was defined as the geometric mean value of $2\Delta f$ at the last six reversal points. All subsequent FM tasks use the same series of step sizes. If the SD across the first three runs was greater than or equal to 0.4 log units, those runs were regarded as practice, and thresholds from three additional runs were collected. The same SD criterion was used for all following FM tasks. Two participants could not differentiate the target from the reference tone with the standard starting value. For both participants, the starting value of the tracking variable was adjusted to $2\Delta f$ = 1.2%. One participant was able to complete the task with the higher starting value. The other participant was unable to complete the slow dichotic FM task, even with an adjusted start value (age = 68). This participant was able to complete every other task, and so the ceiling value ($2\Delta f$ = 2%) was used as their threshold for slow dichotic FM.

Next, participants completed fast (f_m = 20 Hz) dichotic FM detection. The instructions were to select which interval contained the tone with the "broader auditory image." Again, participants were instructed to look at the feedback to help them identify the correct feature. Each run began with $2\Delta f$ = 2%, with $2\Delta f$ never exceeding 200% throughout each run.

3.2.3.3 Tasks 4 and 5: Diotic FM Detection

For both slow and fast FM detection, participants were instructed to pick the tone that was "modulated," and that the modulated tone would sound like it is "changing." Slow FM was always measured before fast FM. The starting value of the tracking variable was $2\Delta f$ = 5.02% and never exceeded $2\Delta f$ = 200%. The adaptive step sizes and the number of reversals used to define threshold were the same as in tasks 2 and 3.

3.2.3.5 Task 6: Absolute Thresholds for 20-ms Tones

Participants completed one adaptive run at each target frequency (400, 430, 460, 490, 510, 540, 570, and 600 Hz), and the order of the target frequency was randomized between runs. These were the same target frequencies as used in the forward masking patterns task (Task 11) but without the presence of the pure-tone forward masker. The instructions were to indicate whether the first or second time interval, marked by lights on the virtual response box on the screen, "had a click in it." The design was analogous to the forward masking patterns task, so that the target interval was 500 ms silence, directly followed by the 20-ms target. The reference interval was 520 ms silence. The reference and target intervals were separated by a 400 ms ISI. The level of the target frequency was varied adaptively. The target was presented at 40 dB SPL during the first trial, and the level was varied by a step size of 8 dB for the first two reversals. The step size was reduced to 4 dB for the third and fourth reversals, and then to 2 dB for the following six reversals. Absolute threshold was defined as the mean target level at the last six reversal points. If the SD within a given run was ≥ 4 dB, one additional run was completed for the corresponding frequency, and only the second run was used in analyses.

3.2.3.6 Tasks 7 and 8: Dichotic AM Detection

The instructions for dichotic AM detection tasks were the same as for the dichotic FM tasks. Slow (f_m = 1 Hz) dichotic AM was always measured before fast (f_m = 20 Hz) dichotic AM. The starting value of the tracking variable, in units of 20log(m), was -8 dB. The step size was 6 dB for the first two reversals, 2 dB for the next two reversals, and 1 dB for the final six reversals. Participants with SDs ≥ 4 dB for their first three runs completed three additional runs, and only the subsequent runs were used in analyses. The same SD criterion was used for diotic AM detection (tasks 9 and 10).

3.2.3.7 Tasks 9 and 10: Diotic AM Detection

The task instructions for AM detection were the same as FM detection. Slow (f_m = 1 Hz) AM was always measured before fast (f_m = 20 Hz) AM. The modulation depth (m) was varied adaptively in the same manner as the dichotic AM tasks.

3.2.3.8 Task 11: Forward Masking Patterns

The 500-Hz pure-tone forward masker was presented in both intervals. In one of the intervals, a 20-ms target directly followed the masker. The instructions were to pick the tone that had a “click” after it. The ISI was 400 ms. The level of the masker was fixed at 70 dB SPL, while the target level varied adaptively. The starting value of the target was 60 dB SPL. Initially, the step size was 8 dB for the first two reversal points. The step size was decreased to 4 dB for the following two reversals, and then decreased to 2 dB for the last six reversals. Threshold was defined as the average target level at the final six reversal points.

Participants completed 2 runs for each of the 8 target frequencies (400, 430, 460, 490, 510, 540, 570, and 600 Hz; 16 runs total), and the order of the target frequencies was randomized between runs. If the SD across any of the 2 runs was ≥ 4 dB, participants completed 2 additional runs for the given target frequencies. If a participant had to repeat runs for two or more target frequencies, the order of the subsequent target frequencies was randomized. The recorded threshold was the mean (in dB) of thresholds across the final two runs in each condition.

3.3 Results

3.3.1 Diotic and dichotic frequency and amplitude modulation detection thresholds

Boxplots of the FMDLs across listeners are presented in Fig. 3.2A. A 2-way repeated-measures analysis of variance (ANOVA) was conducted on the average log-transformed thresholds for each subject, with modulation rate (slow vs. fast) and task type (diotic vs. dichotic) as within-subjects factors. Results indicated a main effect of modulation rate [$F_{(1,84)} = 434$, $P < 0.0001$, $\eta_p^2 = 0.838$], a main effect of task-type [$F_{(1,84)} = 63$, $P < 0.0001$, $\eta_p^2 = 0.428$], and a significant interaction [$F_{(1,84)} = 268$, $P < 0.0001$, $\eta_p^2 = 0.761$]. *Post hoc* Bonferroni-corrected *t* tests ($\alpha = 0.0083$) showed significant differences between all combinations of FM tasks, diotic and dichotic at both modulation rates, except for fast diotic and fast dichotic FM. There was a slight trend for better performance on fast diotic FM relative to fast dichotic FM, but this was not significantly different once correcting for multiple comparisons [$t_{(84)} = -2.11$, $P = 0.03$]. Thresholds for slow dichotic FM detection were significantly lower (better) than those for slow diotic FM detection ($P < 0.0001$). This result is consistent with the idea that subjects were using phase-locking to detect dynamic IPDs in the dichotic task, as IPD thresholds for static 500-Hz tones (e.g., Yost 1974) are far smaller than those produced by detectable values of $2\Delta f$ in the diotic FM detection tasks. Slow diotic FM detection was also significantly better than fast diotic FM ($P < 0.0001$). These trends are consistent with a number of previous studies that have implicated the use of phase-locking for slow, but not fast, FM detection (e.g., Moore and Sek 1995; Moore and Sek 1996; Lacher-Fougère and Demany 1998; Strelcyk and Dau 2009).

Boxplots of the AMDLs are presented in Fig. 3.2B. A 2-way repeated-measures ANOVA with modulation rate (slow vs. fast) and task-type (diotic vs. dichotic) as within-subjects factors revealed a main effect of modulation rate [$F_{(1,84)} = 52.2$, $P < 0.0001$, $\eta_p^2 = 0.383$], a main effect of task-type [$F_{(1,84)} = 518$, $P < 0.0001$, $\eta_p^2 = 0.861$], and a significant interaction [$F_{(1,84)} = 63.6$, $P < 0.0001$, $\eta_p^2 = 0.431$]. *Post hoc* Bonferroni-corrected *t* tests ($\alpha = 0.0083$) for all possible pairwise comparisons demonstrated significant differences between all AM tasks, diotic and dichotic, except for slow vs. fast dichotic AM [$t_{(84)} = 1.28$, $P = 0.203$]. As has been found several times for sinusoidal AM with gated carriers (Viemeister, 1979; Sheft and Yost, 1990; Moore and Sek, 1995; Whiteford and Oxenham, 2015), thresholds for fast diotic AM were significantly better than those for slow diotic AM ($P < 0.0001$), possibly due to the increased number of cycles in fast-rate compared to slow-rate AM (e.g., 2 cycles in slow AM compared to 40 cycles in fast AM) (e.g., Wallaert et al. 2016). Both slow and fast diotic AM thresholds were significantly better than their dichotic counterparts ($P < 0.0001$ in both cases).

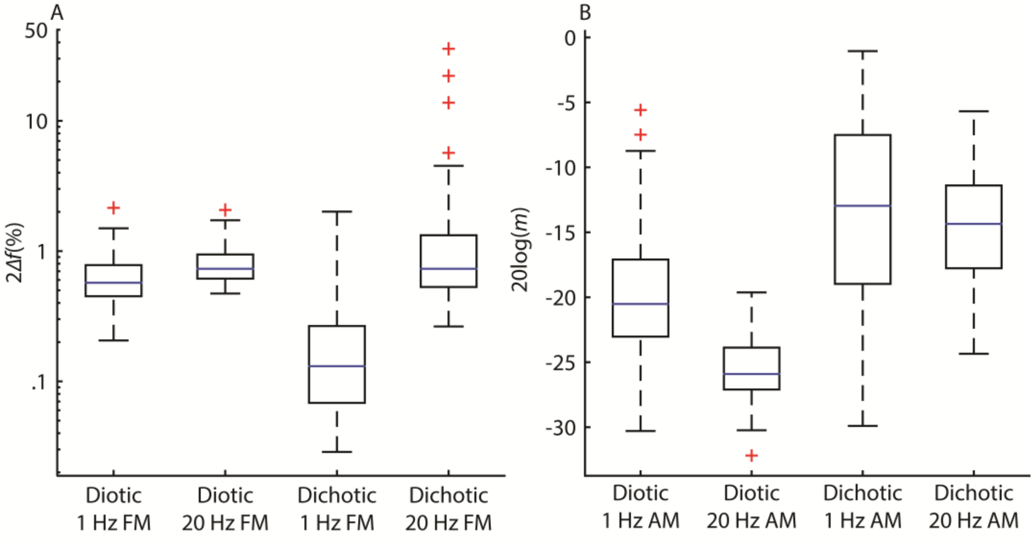

Figure 3.2

Boxplots of diotic and dichotic (A) FMDLs and (B) AMDLs. Solid lines within each box represent the median, and the whiskers are the lowest and highest data points within 1.5 times the lower and higher inter-quartile ranges. Crosses represent individual data points outside the whiskers, considered outliers. All data points, including outliers, were included in the analyses.

3.3.2 Correlations between age, FMDLs, and AMDLs

One way to examine the use of phase-locking in slow and fast FM is to correlate age with FM detection, given the large body of evidence suggesting TFS coding degrades with age (e.g., Hopkins and Moore 2011; Moore et al. 2012b; Füllgrabe 2013). One possibility is that performance on all modulation-detection tasks may degrade with age, regardless of the peripheral code involved. In fact, all of the FM detection tasks, diotic (slow FM: $r = 0.404$, $P < 0.0001$, one-tailed; fast FM: $r =$

0.372, $P = 0.00025$, one-tailed) and dichotic (slow FM: $r = 0.42$, $P < 0.0001$, one-tailed; fast FM: $r = 0.42$, $P < 0.0001$, one-tailed), were significantly correlated with age, even after using Bonferroni correction for running 16 multiple comparisons (i.e., all correlations run with age; $\alpha = 0.003$); see Fig. 3.3. This was not the case for all of the AM tasks; see Fig 4. In fact, only fast dichotic AM was significantly correlated with age ($r = 0.367$, $P = 0.0003$, one-tailed), while slow dichotic AM ($r = 0.174$, $P = 0.056$, one-tailed) and slow ($r = -0.069$, $P = 0.265$, one-tailed) and fast diotic AM ($r = 0.154$, $P = 0.08$, one-tailed) were not.

The lack of correlation between diotic AM and age could not be accounted for by subclinical hearing loss and age co-varying (where reduced cochlear compression in older listeners might provide a benefit in AM detection), as neither diotic slow AM nor diotic fast AM correlated with age once partialling out average absolute thresholds at 500 Hz (slow AM: $r_p = -.148$, $p = .088$, one-tailed; fast AM: $r_p = .173$, $p = .057$, one-tailed). However, both slow and fast diotic FM still correlated with age after controlling for audibility at 500 Hz (slow FM: $r_p = .335$, $P = .0009$, one-tailed; fast FM: $r_p = .33$, $P = .001$, one-tailed). Even though fast diotic FM was significantly correlated with age and fast diotic AM was not, it is important to note that the *difference* between these two correlations, assessed using Steiger's Z-test (Steiger, 1980), was not significant ($Z = 1.75$, $P = 0.08$, two-tailed), while the difference between the correlation of slow-rate diotic FM with age and the correlation of slow-rate diotic AM with age was highly significant ($Z = 4.23$, $P < 0.0001$, two-tailed).

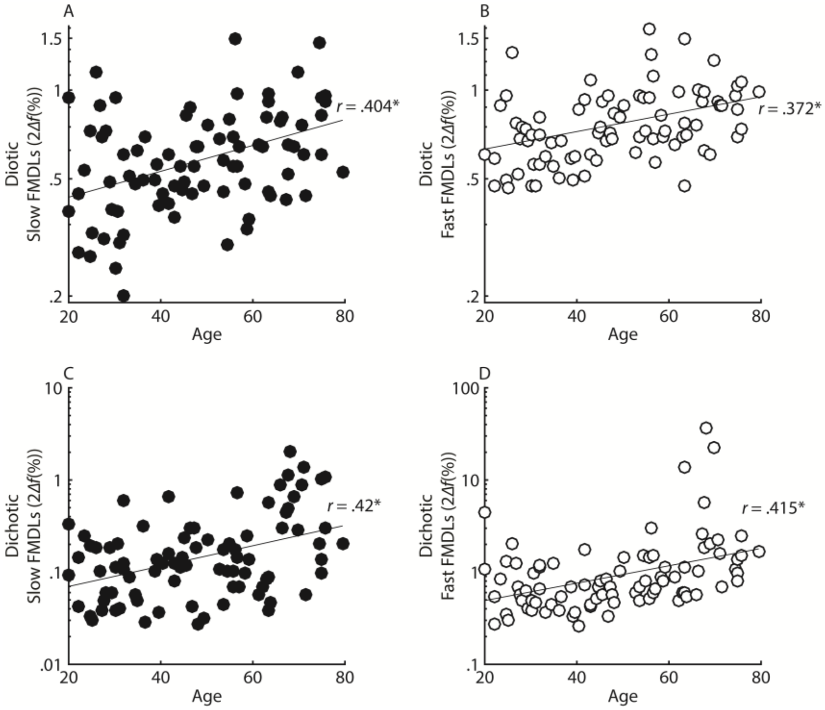

Figure 3.3

Correlations between (A & B) diotic FM and age and (C & D) dichotic FM and age at slow (f_m = 1 Hz; black dots) and fast (f_m = 20 Hz; white dots) modulation rates.

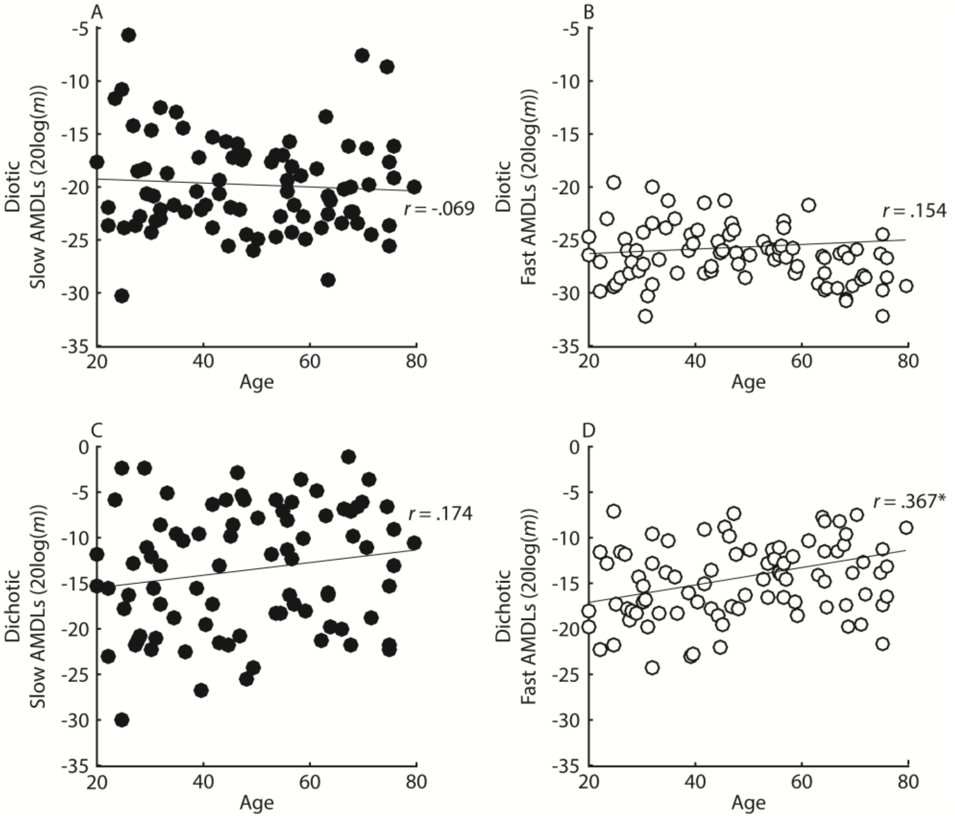

Figure 3.4

Correlations between (A & B) diotic AM and age and (C & D) dichotic AM and age at slow (f_m = 1 Hz; black dots) and fast (f_m = 20 Hz; white dots) modulation rates.

To control for any shared factors involved in modulation detection, including non-peripheral factors, we computed difference scores for each subject. First, the average FMDLs and AMDLs across subjects were z-transformed so that they were in the same units. Then, a difference score for each subject was calculated, subtracting their z-scored threshold in an AM detection task from their z-scored threshold in the corresponding FM detection task (e.g., slow diotic FM – slow diotic AM). Participants who perform better on FM detection relative to AM detection will have lower difference scores, and vice versa. Once controlling for sensitivity to diotic AM detection, slow diotic FM detection thresholds were still significantly correlated with age (r = 0.458, P < 0.0001, one-tailed), whereas fast diotic FM detection thresholds were not (r = 0.184, P = 0.046, one-tailed); see Fig. 3.5A. The difference between these two correlations was significant (*Z* = 2.34, *P* = 0.019, two-tailed). Assuming TFS coding degrades with age, this outcome provides some evidence for a role for TFS coding in slow diotic FM detection and perhaps no role, or a smaller role, for TFS coding in fast diotic FM.

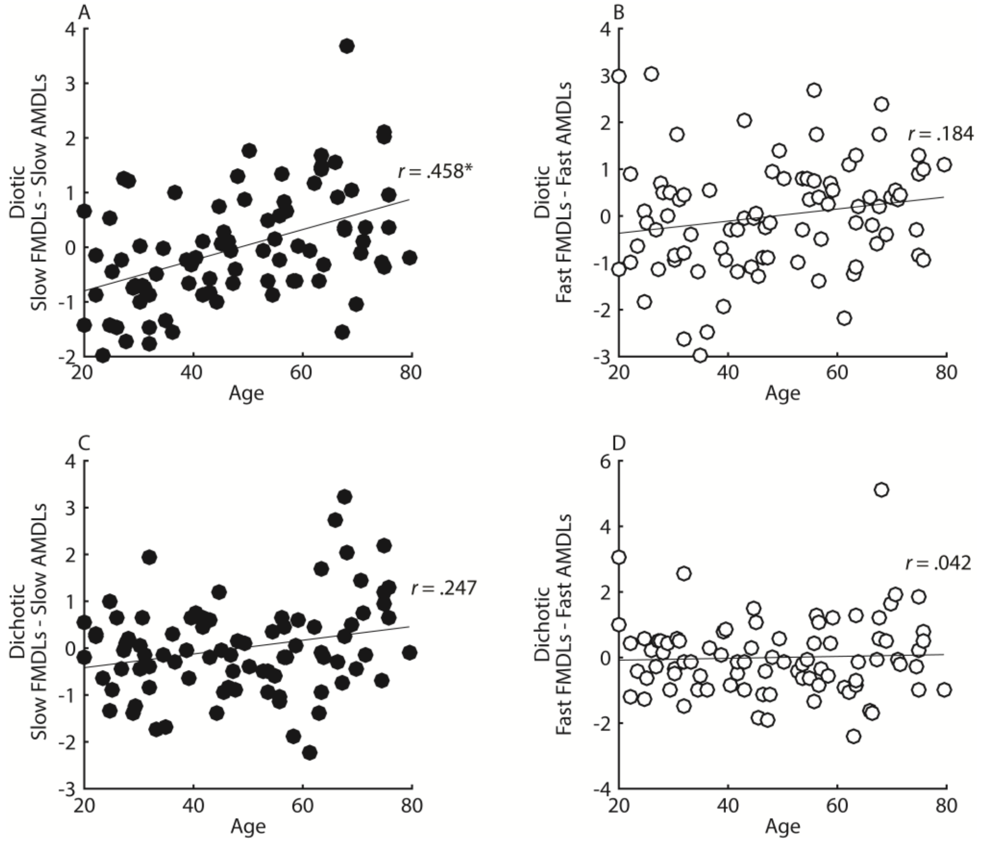

Figure 3.5

Correlations between diotic (A & B) and dichotic (C & D) FM detection and age, once controlling for sensitivity to AM.

If the variability in slow dichotic FM is driven by variability in TFS coding, then slow dichotic FM should be correlated with age, once controlling for sensitivity to slow dichotic AM. However, there was no significant correlation between the difference scores for slow dichotic FM once correcting for multiple comparisons (*r* = 0.247, *P* = 0.011, one-tailed); see Fig. 3.5C. As expected, there was also no correlation between fast dichotic FM difference scores and age (*r* = 0.042, *P* = 0.352, one-tailed).

3.3.3 Correlations between FM and AM tasks

If the variability between participants in FM and AM diotic and dichotic tasks is largely driven by variability in peripheral coding, then thresholds in tasks that measure similar peripheral codes should be highly correlated. Similarly, thresholds in tasks thought to rely on separate peripheral codes should not be highly correlated. In line with the peripheral coding hypothesis, slow diotic FM detection thresholds were correlated with slow dichotic FM detection thresholds ($r = 0.426$, $P < 0.0001$, one-tailed), a task that is believed to rely on TFS coding; see Fig. 3.6A. This correlation was significant even after using Bonferroni correction for seven multiple comparisons ($\alpha = 0.007$). Tasks believed to reflect excitation pattern information, such as fast diotic FM and fast dichotic FM ($r = 0.634$, $P < 0.0001$, one-tailed) and fast diotic AM and fast dichotic AM ($r = 0.529$, $P < 0.0001$, one-tailed), were also significantly correlated. However, tasks believed to reflect separate peripheral codes were also well correlated. For instance, thresholds in the slow diotic FM condition and the slow diotic AM condition were correlated ($r = 0.468$, $P < 0.0001$, one-tailed); see Fig 6C. In the same vein, dichotic tests believed to reflect similar peripheral mechanisms, such as fast dichotic FM and fast dichotic AM ($r = 0.34$, $P = 0.001$, one-tailed), as well those thought to reflect separate mechanisms, such as slow dichotic FM and slow dichotic AM ($r = 0.502$, $P < 0.0001$, one-tailed), were also moderately correlated; see Fig. 3.7. Because thresholds in most of the modulated tasks were well correlated with each other, our correlational measures provide no strong evidence for either the use of place or temporal coding in FM detection at different rates.

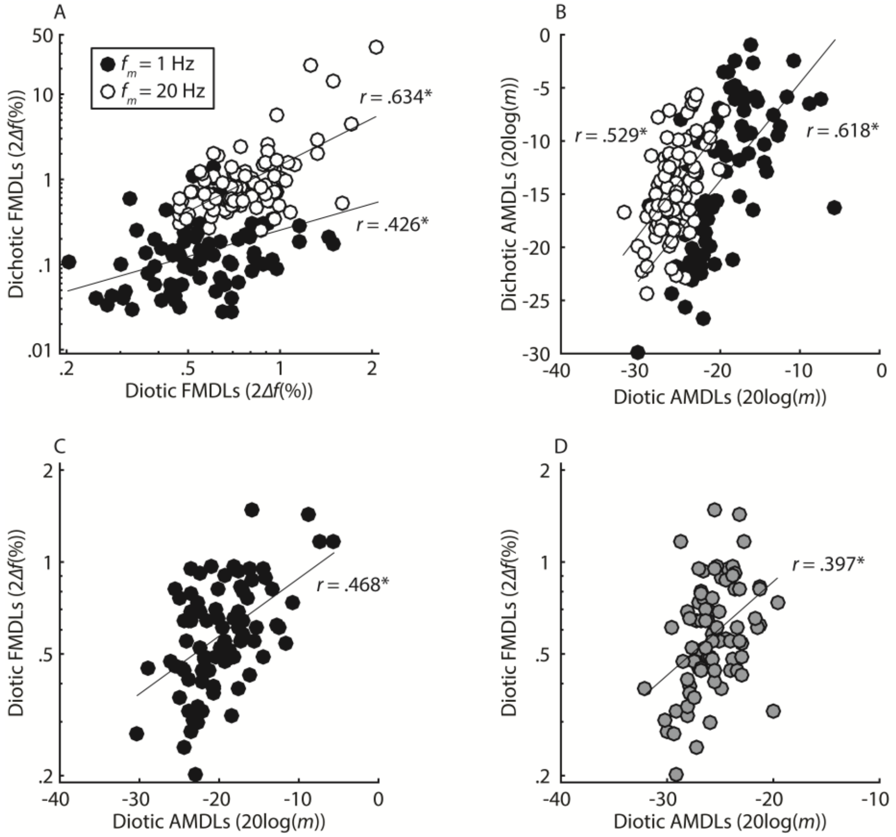

Figure 3.6

Correlations between FM and AM, diotic and dichotic, tasks. Filled circles correspond to individual thresholds at slow rates (f_m = 1 Hz) and open circles correspond to individual thresholds at fast rates (f_m = 20 Hz). Grey circles represent conditions where the x-axis represents thresholds with the fast rate and the y-axis represents thresholds with the slow rate. (A and B) Modulated tasks believed to measure similar peripheral codes are well correlated; (C and D) but tasks believed to measure different peripheral codes are also well correlated.

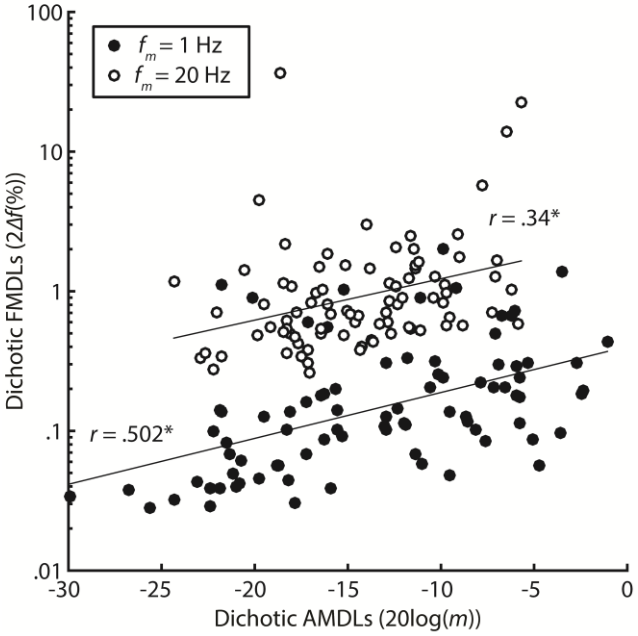

Figure 3.7

Correlations between dichotic FM and AM at two modulation rates. Filled circles represent individual thresholds with f_m = 1 Hz, while open circles represent thresholds with f_m = 20 Hz.

3.3.4 Frequency selectivity and FM detection

In order to estimate the steepness of the low- and high-frequency sides of the excitation pattern for each subject, we first calculated the threshold for each of the 20-ms target frequencies when preceded by the 500-Hz pure-tone masker. Then, two linear regression analyses were conducted to estimate the low and high slopes of the forward masking patterns. The low slope was based on a regression using the thresholds from the four lowest target frequencies (400, 430, 460, and 490 Hz) and the high slope was based on a regression using the thresholds from the four highest target frequencies (510, 540, 570, and 600 Hz). Under this arrangement, steep low slopes are increasingly positive, whereas steep high slopes are increasingly negative.

If fast, but not slow, diotic FM detection relies on place coding, then fast diotic FM thresholds should be correlated with the steepness of the filter slopes from the forward masking patterns but slow FM thresholds should not. However, neither slow (low slope: r =0.003, P = 0.489; high slope: r =0.039, P = 0.362; low slope + |high slope|: r = -0.02, P = 0.428) nor fast (low slope: r =0.064, P = .28; high slope: r

=0.069, *P* = 0.265; low slope + |high slope|: *r* =0.006, *P* = .478) diotic FMDLs were correlated with frequency selectivity. Subtractive measures controlling for the sensitivity to AM at the same modulation rate are plotted in Fig. 3.8 (top row: slow FM; bottom row: fast FM). As can be seen from the top row in Fig. 3.8, there was also no correlation between slow FMDLs and filter slopes once controlling for sensitivity to slow AMDLs (low slope: *r* = -0.015, *P* = 0.446; high slope: *r* =0.139, *P* = 0.103; low slope + |high slope|: *r* = -0.089, *P* = 0.209). There was only a very weak correlation between fast FM and filter slopes (low slope: *r* = -0.209, *P* = 0.028; high slope: *r* =0.17, *P* = 0.06; low slope + |high slope|: *r* = -0.242, *P* = 0.013); see bottom row of Fig. 3.8. Note that these correlations only reach significance without corrections for multiple comparisons.

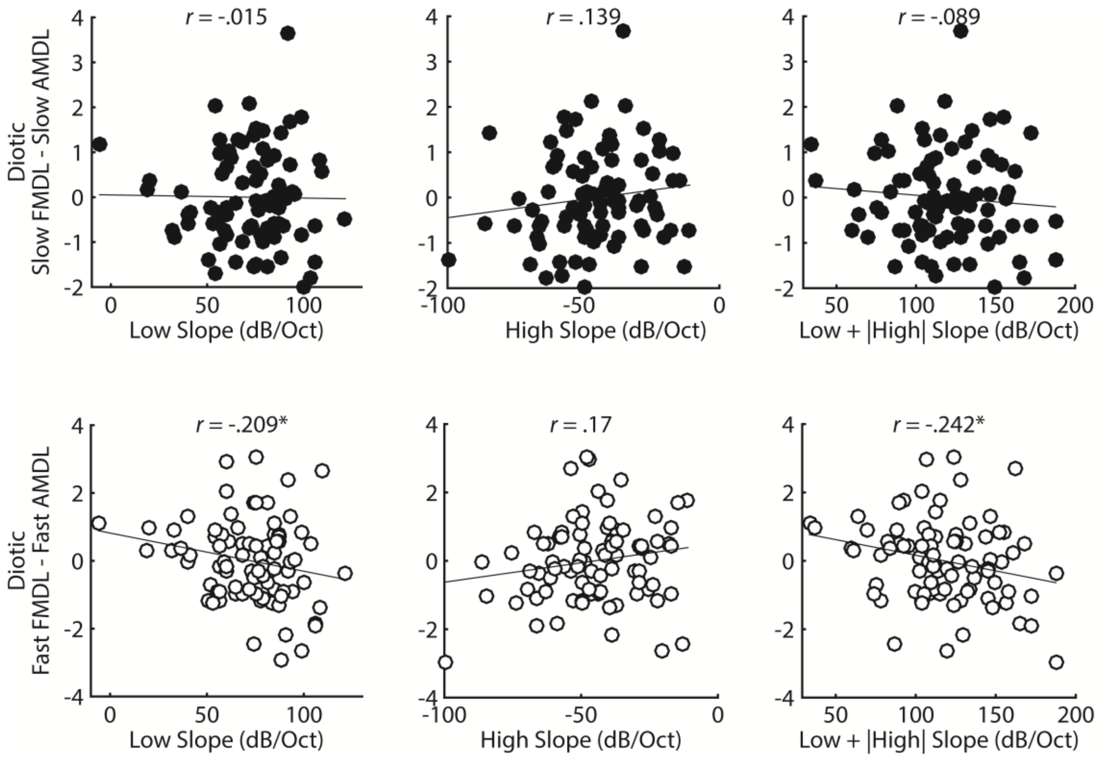

Figure 3.8

Correlations between frequency selectivity and diotic FM detection, controlling for sensitivity to AM detection. The top row (black dots) corresponds to f_m = 1 Hz, while the bottom row (white dots) corresponds to f_m = 20 Hz. Diotic fast FM was weakly, significantly correlated with the steepness of the filter slopes, while diotic slow FM was not.

3.4 Discussion

3.4.1 The role of TFS coding in FM detection

Results from this study provided evidence both for and against the hypothesis that slow FMDLs require temporal coding of TFS, and that TFS coding degrades with age. The primary piece of supporting evidence was that slow, but not fast, diotic FMDLs were correlated with age, after controlling for sensitivity to AMDLs at the same rate. These two correlations were also significantly different from one another. AMDLs at the same rate make an ideal control for diotic FM because the task procedures and demands were very similar, with only the stimulus differing. Hence, any non-TFS factors should be well controlled. Second, as found in many previous studies (e.g., Moore and Sek 1996; Lacher-Fougère and Demany 1998; Whiteford and Oxenham 2015), detection of slow diotic FM was better than detection of fast diotic FM, while the opposite trend was found for AM. Together, the correlations with age and the patterns of average detection thresholds could suggest that slow, low-carrier FM uses a separate peripheral code, based on phase-locking to TFS cues. However, even a dissociation between AM and FM does not necessarily imply TFS processing, as these same results could be a product of differences in the correlational properties of cortical neurons (Micheyl et al., 2013). The present findings might also be explained by a combined place and temporal model for slow, low carrier FM, as suggested by some studies measuring slow-rate FMDLs in the presence of an AM masker (e.g., Paraouty et al. 2016; Paraouty and Lorenzi 2017). For example, consistent with our findings, Paraouty and Lorenzi (2017) found that slow diotic FM and slow diotic AM were correlated across a large group of listeners. However, slow diotic FM in the presence of added AM, used to minimize place information, did not correlate with slow diotic AM but did correlate with slow diotic FM in quiet (with no added AM).

Other aspects of the results from the present study are less easily interpreted in terms of different coding mechanisms for FM and AM. First, slow dichotic FMDLs were not correlated with age, once slow dichotic AMDLs were factored out. Thus, for the one task where temporal coding of TFS must play a role (dichotic FM, or dynamic

IPD, detection), no strong correlation with age was found, raising questions as to why it was found for the diotic FM, which may or may not be represented via temporal coding. One possibility is that performance in the binaural tasks was limited by more central constraints, during or following the binaural integration of information. Second, similar correlations were observed between age and both slow and fast diotic FM without controlling for sensitivity to AM. It could be that the variability was driven by non-TFS coding factors in fast-rate diotic FM (as suggested by the lack of correlation with age once controlling for sensitivity to fast diotic AM) but not slow-rate diotic FM. Alternatively, both slow- and fast-rate diotic FM detection may use a similar peripheral code. The moderate correlations between many of the modulation-detection tasks, even those thought to use separate peripheral codes, found both in the present data and several other studies (Whiteford and Oxenham, 2015; Otsuka et al., 2016; Paraouty and Lorenzi, 2017), could also suggest similar peripheral codes for slow/fast FM and AM. For example, Otsuka et al. (2016) interpreted significant correlations between slow FM, slow AM, and IPDs to mean that neural phase-locking to TFS cues might be used for both types of slow, low carrier sinusoidal modulation detection. However, this interpretation would not explain why detection for slow rate FM is better than fast rate FM, while the opposite trend exists for AMDLs.

A number of non-peripheral factors might also be responsible for the high multicollinearity between many FM and AM tasks. First, the primary limiting factor in diotic FMDLs and AMDLs may arise from central coding. For example, a shared cortical rate-place code for frequency and intensity (Micheyl et al., 2013) might account for correlations in behavioral FM/AM data, although the model has yet to be applied to FMDLs and AMDLs in the same manner as frequency and intensity difference limens. Second, all of the diotic and dichotic FM and AM tasks in the present study have very similar task demands, and some multicollinearity could be from cognitive aspects required to detect modulation, such as sustained attention (e.g., Füllgrabe et al. 2015). Given that more general cognitive measures were not assessed in the present study, their influence on the present measures remains speculative.

3.4.2 Effects of age or high-frequency hearing loss?

It is important to note that any effects of age found in the present study could, in fact, be driven by high-frequency hearing loss. This is because age and average high-frequency (2 kHz, 4 kHz, and 8 kHz) audiometric thresholds were highly correlated (r =0.741, $P < 0.0001$). In addition, Schoof and Rosen (2014) did not find a relation between poor slow, low-carrier FMDLs and age. One difference in their study is the more stringent NH criterion relative to the current study or previous studies (He et al., 2007; Strelcyk and Dau, 2009; Grose and Mamo, 2012; Paraouty et al., 2016). Schoof and Rosen (2014) required older and younger subjects to have audiometric thresholds in both ears ≤ 25 dB HL up to and including 4 kHz (two octaves above the carrier frequency, f_c = 1 kHz), and then thresholds ≤ 25 dB HL in at least one ear at 6 kHz (two and a half octaves above the carrier). Fig. 3.9 plots the present diotic FMDLs and AMDLs from a subset of our subjects, selected using the same age and an analogous audiometric criteria. Average FMDLs are plotted for younger (age < 30; n=15) and older (age ≥ 60; n=11) participants. Average FMDLs from Schoof and Rosen for a 1-kHz carrier with a 2-Hz modulation rate are plotted for reference (squares). Even when using a similar, more stringent NH criterion, our results demonstrate significantly poorer FMDLs in the older adults [$F_{(1,24)} = 4.24$, $P = 0.046$, $\eta_p^2 = 0.155$], a main effect of modulation rate [$F_{(1,24)} = 15.754$, $P = 0.001$, $\eta_p^2 = 0.396$], but no interaction between modulation rate and group [$F_{(1,24)} = 0.693$, $P = 0.413$, $\eta_p^2 = 0.028$]. Importantly, the same group trends were not observed for AMDLs: there was no main effect of group [$F_{(1,24)} = 0.343$, $P = 0.563$, $\eta_p^2 = 0.014$], but there was a main effect of modulation rate [$F_{(1,24)} = 44.875$, $P < 0.0001$, $\eta_p^2 = 0.652$], and no group by rate interaction [$F_{(1,24)} = 1.9$, $P = 0.181$, $\eta_p^2 = 0.073$]. The fact that group differences remain in the FM but not the AM tasks, despite the more stringent criteria, is consistent with the poorer FMDLs being due to age rather than age-related high-frequency hearing loss.

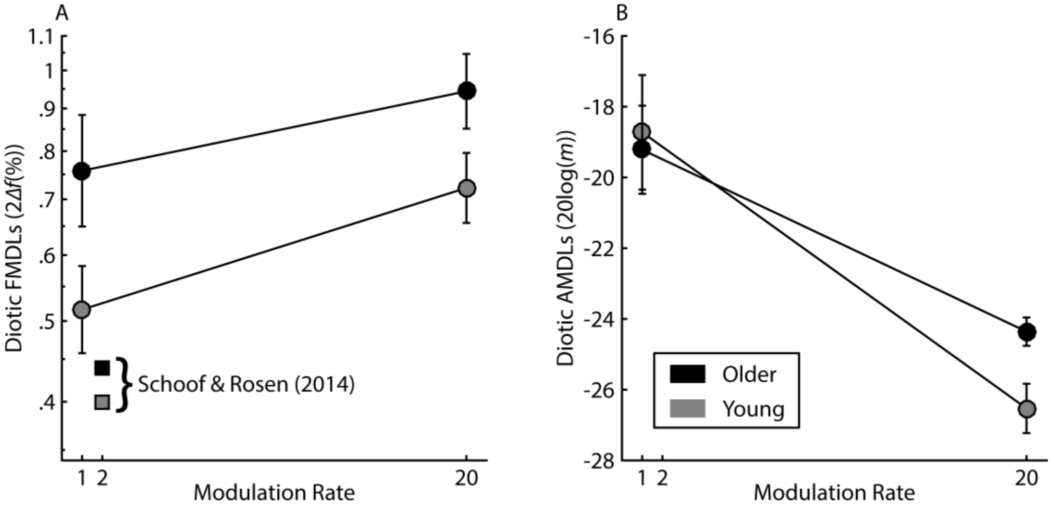

Figure 3.9

Average (A) diotic FMDLs and (B) diotic AMDLs for young (age < 30; grey) vs. old (age ≥ 60; black) participants. Circles correspond to the present data, measured for a 500 Hz carrier at slow (f_m = 1 Hz) and fast (f_m = 20 Hz) modulation rates. Squares correspond to average FMDLs from Schoof and Rosen (2014), measured for a 1 kHz carrier at a 2 Hz modulation rate.

It remains to be explained why Schoof and Rosen (2014) did not find an age effect for slow, low carrier FM. The average age of older and young adults was comparable between studies when using a similar NH criterion (Schoof and Rosen (2014): mean age young = 23.7, mean age older = 64.1; present study: mean age young = 25.1, mean age older = 66.3). Apart from a few methodological differences (e.g., 3AFC instead of 2AFC, f_c = 1 kHz instead of f_c = 500 Hz, and minor differences in the use of practice trials), their procedures were quite similar to those of the present study. Schoof and Rosen's (2014) FMDLs are substantially better than were found in the current study, but this was expected, as FMDLs tend to be better at 1 kHz relative to 500 Hz when quantified as percent peak-to-peak frequency change (Demany and Semal, 1989). It is possible that there may have been differences in audiometric thresholds at the test frequency between the current study and Schoof and Rosen (2014) accounting for these effects. However, this seems unlikely given that all but one participant in the present study had audiometric thresholds ≤ 20 dB

HL at 500 Hz, and this participant had 25 dB HL in just one ear. One last explanation is that our group may have had greater absolute hearing loss at 6 kHz (not measured in the present study) or 8 kHz, and perhaps hearing loss specifically at 6 or 8 kHz may influence the results. This could be a meaningful difference if high frequency hearing loss is a marker for age-related synaptopathy throughout the cochlea, including lower-frequency areas where absolute thresholds are normal, as was found in CBA/CaJ mice (Sergeyenko et al., 2013). However, there is currently no direct, physiological measure to confirm age-related synaptopathy in humans, so it remains unclear whether high-frequency hearing loss co-varies with a loss of synaptic terminals at lower-frequency places along the basilar membrane. Furthermore, simulations using signal detection theory suggest that very severe synaptopathy may be necessary to noticeably affect frequency coding, whether or not a temporal code is used (Oxenham, 2016).

3.4.3 The role of frequency selectivity in FM detection

Even after controlling for sensitivity to AM, fast-rate diotic FM was only very weakly correlated with the steepness of the auditory filter slopes but slow diotic FM was not, suggesting a possible small role for frequency selectivity specific to fast diotic FM. This trend was only significant when no corrections were made for multiple comparisons. Yet, if cochlear filtering were responsible for the average trends observed in Fig. 3.3, then one might expect moderate to strong correlations between fast-rate diotic FM and frequency selectivity. Hence, the evidence for a role of place coding in fast FM detection is weak. These results are consistent with the findings of Whiteford and Oxenham’s (2015) study of young, NH listeners, where they found similarly weak correlations between FMDLs and filter slopes. It may be that stronger correlations will be observed in a population that spans a wide range of sensorineural hearing loss at the carrier frequency, and hence a wide range of frequency selectivity. According to the initial hypothesis of this study, increasing the variability in place coding should have a greater influence on fast, but not slow, FM detection.

3.4.4 Changes in between-subject variability with age

The current study examined low-carrier FM and AM detection at different rates in diotic and dichotic conditions across a large cohort of listeners varying in age, potentially providing insight into which tasks utilize temporal coding of TFS, given the number of earlier studies indicating an effect of age on tasks thought to employ temporal coding of TFS (e.g., Grose and Mamo 2010; Grose and Mamo 2012; Ruggles et al. 2012; Füllgrabe 2013; Gallun et al. 2014). The variability in performance on slow dichotic FM was considerably larger in the current sample than in Whiteford and Oxenham's (2015) cohort of young NH listeners (see Table 3.1). This could reflect an increase in variability in TFS coding due to our inclusion of older listeners. However, the variability on slow dichotic AM and fast dichotic FM was also much higher in the current sample relative to the younger listeners. Because the increase in variability in the dichotic tasks was not specific to slow dichotic FM, the specific source of this variability is unclear and may not, in fact, be driven by variability in peripheral temporal coding. The implications of increased variability in several binaural tasks with age, even non-TFS binaural tasks, are important and suggest that assuming performance variability on binaural tasks largely reflects differences in peripheral coding may not be appropriate. These findings, in addition to the lack of significant correlation between slow dichotic FM and age once controlling for performance on slow dichotic AM (Fig. 3.5C), could mean that previous attempts to measure TFS processing using IPDs in older NH or HI listeners may have reflected variability in more central binaural coding (e.g., Grose and Mamo 2010; Hopkins and Moore 2011; Moore et al. 2012a).

Table 3.1

SD for each modulated task in the current study, with a wide age range (20.1-79.5) of participants, and Whiteford and Oxenham's (2015) study, with a smaller age range (18-32) of strictly NH participants.

	Current Study (n = 85)	**Whiteford & Oxenham (2015) (n=100)**
Slow Dichotic FM	4.43	3.38
Fast Dichotic FM	3.83	3
Slow Diotic FM	1.84	1.84
Fast Diotic FM	1.34	1.46
Slow Dichotic AM	6.76	5.61
Fast Dichotic AM	4.4	4.44
Slow Diotic AM	4.54	4.64

Fast Diotic AM	2.38	2.95

3.5 Conclusions

Findings from the present study across a large cohort of listeners varying in age provide mixed evidence for the role of temporally based TFS coding in slow, but not fast, FM detection. Slow but not fast diotic FMDLs correlated with age, but this specific trend was only observed once controlling for sensitivity to diotic AM. FMDLs and AMDLs were correlated, even for tasks thought to use different peripheral codes, potentially implicating a role of central processing, including more central sensory coding, as well as non-sensory cognitive factors, such as sustained attention, on FM and AM detection. Overall, the effects of age on peripheral coding may be outweighed by variability in non-peripheral factors.

CHAPTER 4: SENSORINEURAL HEARING LOSS AND FM DETECTION

Abstract

Many natural sounds, including speech and music, convey information via modulations in frequency and amplitude. Frequency modulation (FM) at the rates most relevant for both speech and music has been thought to be coded via the precise timing of action potentials in the auditory nerve, whereas faster rates of FM are thought to be coded via a transformation of FM into amplitude modulation (AM) via cochlear filtering. The present study tested this long-held belief by studying individual differences in a group of listeners (N = 49) with hearing losses ranging from no loss to moderate-to-severe loss. Hearing loss typically results in poorer cochlear tuning, which should affect FM-AM transformation but not the timing code. Listeners were assessed on their sensitivity to FM and AM tones at slow and fast rates, as well as on a behavioral measure of cochlear tuning using forward masking. Moderate correlations were observed between all FM and AM tasks, including those thought to use separate codes. Sensitivity to FM at both slow and fast rates was strongly correlated with the sharpness of cochlear tuning on the low-frequency slope of the masking pattern, even after controlling for factors known to influence FM detection, such as hearing thresholds, age, and sensitivity to AM. Contrary to long-held beliefs concerning the role of timing information for slow-rate FM, the results suggest a unitary code for FM, based on cochlear FM-to-AM conversion.

4.1 Introduction

Modulations in frequency (FM) and amplitude (AM) are prominent in natural sounds, including speech and music, and are also biologically relevant for animal communication (Attias and Schreiner, 1997; Nelken et al., 1999). In humans, AM is crucial for understanding speech in quiet (Shannon et al., 1995; Smith et al., 2002), while FM is particularly important for perceiving melodies, recognizing talkers, determining speech prosody and emotion, and segregating speech from other competing background sounds (Zeng et al., 2005; Strelcyk and Dau, 2009; Sheft et al., 2012). The perception of FM is often degraded in older listeners and people with hearing loss (Grant, 1987; Lacher-Fougère and Demany, 1998; Moore and Skrodzka, 2002; He et al., 2007; Strelcyk and Dau, 2009; Grose and Mamo, 2012; Paraouty et al., 2016; Wallaert et al., 2016; Paraouty and Lorenzi, 2017; Whiteford et al., 2017). This deficit, along with the effects of loss of audibility, likely contributes to the communication difficulties experienced by older people and people with hearing loss in noisy real-world environments, which in turn may help explain why age-related hearing loss has been associated with decreased social engagement, a greater rate of cognitive decline, and an increased risk of dementia (Lin et al., 2011, 2013; Lin and Albert, 2014; Deal et al., 2017; Thomson et al., 2017). Current assistive listening devices, such as hearing aids and cochlear implants, have been generally unsuccessful at reintroducing viable FM cues to the auditory system (Chen and Zeng, 2004; Ives et al., 2013). This lack of success is partly related to a gap in our scientific understanding regarding how FM is transduced from sound to neural impulses in the earliest stages of the auditory system.

The coding of AM is well established, beginning in the auditory nerve with periodic increases and decreases in the firing rate of auditory nerve fibers that are time-locked to the temporal envelope of the stimulus (Schreiner and Langner, 1988; Joris et al., 2004). The coding of FM is less straightforward. For a pure tone with FM, the temporal envelope is flat; however, the changes in frequency lead to dynamic shifts in the tone's representation along the basilar membrane, leading to a transformation of FM into AM at the level of the auditory nerve (Zwicker, 1956; Moore and Sek, 1995; Saberi and Hafter, 1995; Sek and Moore, 1995).

Although this FM-to-AM conversion provides a unified and neurally efficient code for both AM and FM (Saberi and Hafter, 1995), it falls short of explaining human behavioral trends in FM sensitivity, specifically at low carriers (f_c < ~4-5 kHz) and slow rates (f_m <~ 10 Hz) within the range of human speech, where sensitivity tends to be better than at higher carrier frequencies or high modulation rates (Demany and Semal, 1989; Moore and Sek, 1995, 1996; Sek and Moore, 1995; He et al., 2007; Whiteford and Oxenham, 2015; Whiteford et al., 2017). This enhanced sensitivity for slow FM at low carrier frequencies has been explained in terms of an additional time-based neural code that accurately codes the temporal fine structure of the waveform (Moore and Sek, 1995). One drawback of the time code is that there is no clear consensus of exactly how or where the extraction of auditory-nerve timing information occurs in the brain. The upper limit of phase locking decreases progressively along the auditory pathways, and is estimated to be ~1 kHz in the inferior colliculus (Liu et al., 2006) and ~100 Hz in the auditory cortex (Lu and Wang, 2000), so most timing information must be transformed to a place code well before the auditory cortex.

To date, the strongest evidence for time-based FM coding comes from comparing average thresholds across relatively small samples of listeners (Moore and Sek, 1994, 1995, 1996; Ernst and Moore, 2010). Recent studies with larger sample sizes have revealed high multicollinearity between FM and AM (Otsuka et al., 2014; Whiteford and Oxenham, 2015; Paraouty and Lorenzi, 2017; Whiteford et al., 2017), even between tasks thought to use separate peripheral codes, such as slow and fast FM, or between AM and slow FM (Whiteford and Oxenham, 2015; Whiteford et al., 2017). This outcome raises the possibility that perhaps AM and FM do share a common neural code, even at the low rates that are most critical for human and animal communication. One way to test this hypothesis would be to determine whether FM thresholds at low rates are dependent on cochlear filter tuning. If FM thresholds depend on an FM-to-AM conversion, the depth of the resultant AM (and hence its detectability) will depend on the slopes of the cochlear filters, with steeper filters resulting in greater AM depth for a given depth of FM. Studies using normal-hearing listeners have not demonstrated such a correlation (Whiteford and Oxenham, 2015; Whiteford et al., 2017). However, this failure to find a correlation may be due to lack of variability in cochlear filtering within a normal-hearing population. People with cochlear hearing loss often have poorer frequency selectivity, due to a broadening of cochlear tuning (Glasberg and Moore, 1986; Moore et al., 1999). In contrast, damage to the cochlea is not thought to lead to a degradation of auditory-nerve phase locking to temporal fine structure for sounds presented in quiet (Henry and Heinz, 2012). If coding of FM at fast rates relies on FM-to-AM conversion, whereas FM coding at slow rates relies on a timing code, then detection thresholds for fast-rate FM should be correlated with measures of cochlear tuning, whereas thresholds for slow-rate FM should not.

The present study measured FM and AM detection at slow (f_m = 1 Hz) and fast (f_m = 20 Hz) rates in a large sample of listeners with hearing thresholds at the carrier frequency (f_c = 1 kHz) ranging from normal (~0 dB sound pressure level, SPL) to severely impaired (70 dB SPL). The fidelity of cochlear frequency tuning was examined using psychophysical methods to estimate the steepness of the forward masking function around 1 kHz. Contrary to predictions of time models, there was a direct relation between the estimated sharpness of cochlear tuning (place coding) and FM detection at both the fast and the slow rate. This relationship remained significant even after controlling for audibility, sensitivity to AM, and age. Our results indicate that slow FM is represented by a place or a combined place-time code. Findings suggest medical and prosthetic interventions may be more successful if they focus on fine-tuning cochlear place, rather than time, coding.

4.2 Results

4.2.1 Effects of hearing loss on masking functions

The fidelity of place coding at the test frequency (1 kHz) was measured using pure-tone forward-masking patterns. Participants heard two tones, one at a time, and were instructed to select the tone that had a short, 20-ms tone pip directly following it. The masker tones were fixed in frequency (1 kHz) and level, while the target level adapted to measure the lowest sound level that the participant could detect. Without the presence of a masker, the level of the target reflects the absolute threshold (Appendix Fig. A1, unfilled circles). In the presence of a pure-tone forward masker, the level of the target depends on the target's frequency proximity to the masker and the shape of the individual's cochlear filters, where detection for targets close in frequency to the masker are much poorer (i.e., the level must be higher) than for targets farther away in frequency. For each participant, the steepness of the low and high slopes of the masking function were estimated by calculating two linear regressions between the thresholds for the four lowest target frequencies (800, 860, 920, and 980 Hz) and the four highest target frequencies (1020, 1080, 1140, and 1200 Hz), with target frequency transformed to $\log_2$ units for the regression. The range of masking function slopes in the present study, spanning 154.6 dB/Oct for the low slope and 114.8 dB/Oct for the high slope (Fig. 4.1, y-axis), was greater than that observed in normal hearing listeners at a similar characteristic frequency (Whiteford and Oxenham, 2015; Whiteford et al., 2017), demonstrating wider variability in place coding when listeners vary in degree of hearing loss. In line with evidence that cochlear tuning deteriorates with SNHL (Glasberg and Moore, 1986), participants with more hearing loss at 1 kHz tended to have shallower masking functions (Fig. 4.1; low slope: *r* = -.595, *p* < .0001; high slope: *r* = .707, *p* < .0001). Absolute thresholds are not a one-to-one mapping of masking function slopes because absolute thresholds may also reflect inner hair cell loss, whereas behavioral estimates of frequency tuning primarily reflect outer hair cell loss.

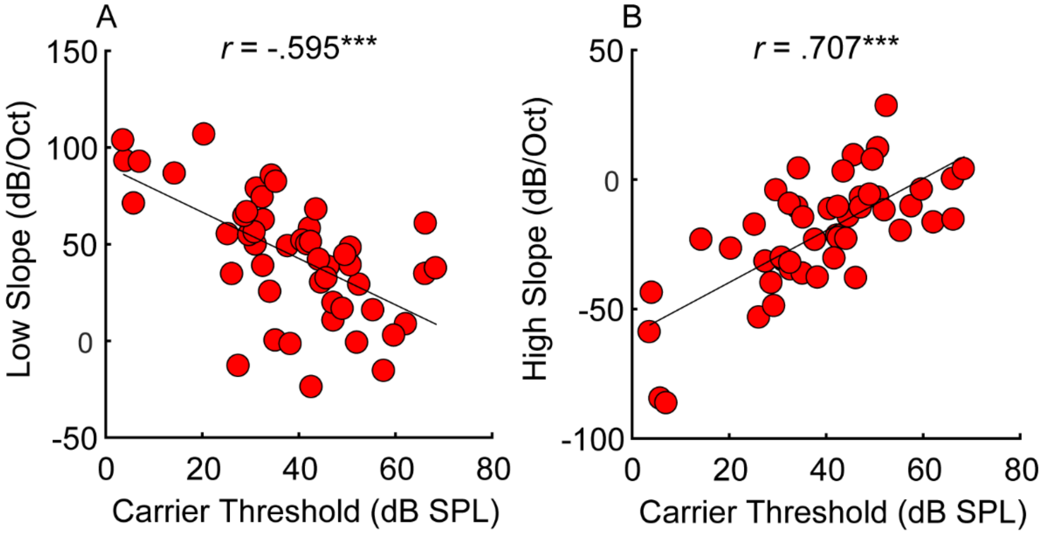

Fig. 4.1

Correlations between average absolute thresholds at 1 kHz (x-axis) and the steepness of the (A) low and (B) high side of the cochlear filter slopes. Participants with greater hearing loss at 1 kHz tended to have shallower filter slopes.

4.2.2 Average FM and AM detection thresholds

Participants heard two pure tones presented sequentially on each trial, and were instructed to select the tone that was varying. The target varied either in frequency (FM; tasks 2 and 3) or amplitude (AM; tasks 5 and 6), either slowly (tasks 2 and 5) or quickly (tasks 3 and 6) over time. The amount or depth of modulation was varied adaptively to measure the smallest modulation depth that the participant could perceive. The mean log-transformed thresholds (peak-to-peak frequency excursion, $2\Delta f$ (%), for FM and modulation index, *m*, for AM) were used in all analyses. When compared to earlier results from normal-hearing listeners varying in age (Whiteford et al., 2017), the range of FM detection thresholds, indicated by the upper and lower whiskers in Fig. 2A, was much wider in the present study, whereas the spread of AM detection thresholds (Fig. 2B) was comparable. This finding suggests that cochlear hearing loss may affect FM more than AM thresholds. The observation that AM thresholds were lower (better) for the high rate than for the low rate (two-tailed paired t-test, t = 16.7, p < .0001), whereas FM thresholds were more similar across the two rates (two-tailed paired t-test, t = -1.58, p = .122), is consistent with earlier studies (Viemeister, 1979; Sheft and Yost, 1990; Moore and Sek, 1995, 1996; Lacher-Fougère and Demany, 1998; Whiteford and Oxenham, 2015, 2017; Whiteford et al., 2017).

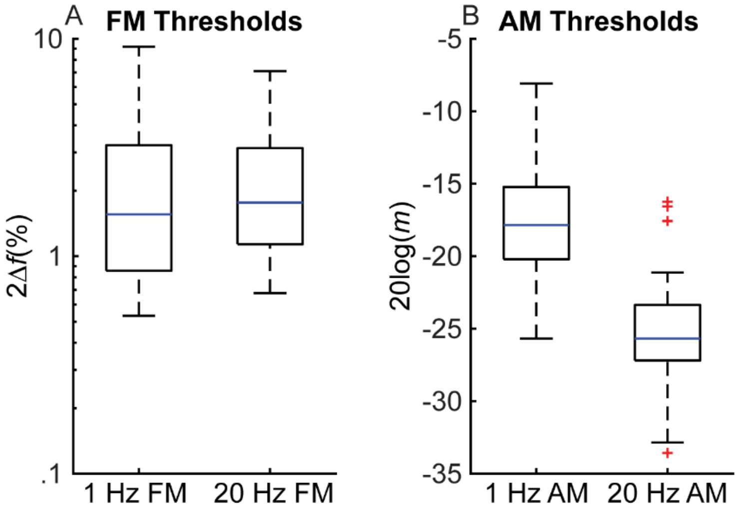

Fig. 4.2

Boxplots of sensitivity to modulations in frequency (A) and amplitude (B) across n=49 listeners. The y-axes correspond to the smallest amount of modulation detectable at threshold, where lower represents better performance. Blue bars are the median of each group. Whiskers are the lowest and highest data points within 1.5 times the lower and higher inter-quartile ranges. Red crosses are individuals whose performance fell outside this range.

4.2.3 Correlations between FM and AM detection

If slow FM utilizes a time code, then across-listener variability in slow FM detection should partly reflect variability in time coding. This means that across-listener correlations in tasks known to use a shared code (fast FM, slow AM, and fast AM) should be greater than tasks thought to use different codes (slow FM with any other task). Inconsistent with the time code hypothesis, slow and fast FM (*r* = .828, *p* < .0001) were strongly correlated, as were slow and fast AM detection (*r* = .609, *p* < .0001) and fast FM and fast AM detection (*r* = .4, *p* = .002) (Fig. 3; Bonferroni-corrected α = .0125). Slow FM and slow AM appeared to have a weaker, non-significant correlation (*r* = .28, *p* = .026), perhaps implicating an additional time code for slow FM, but this correlation was not significantly different from the fast FM and fast AM correlation (*p* = .514). Even though participants in the present study varied widely in peripheral place coding fidelity (Fig. 4.1), correlational trends between FM and AM thresholds mirrored those observed in normal-hearing listeners (Whiteford and Oxenham, 2015; Whiteford et al., 2017).

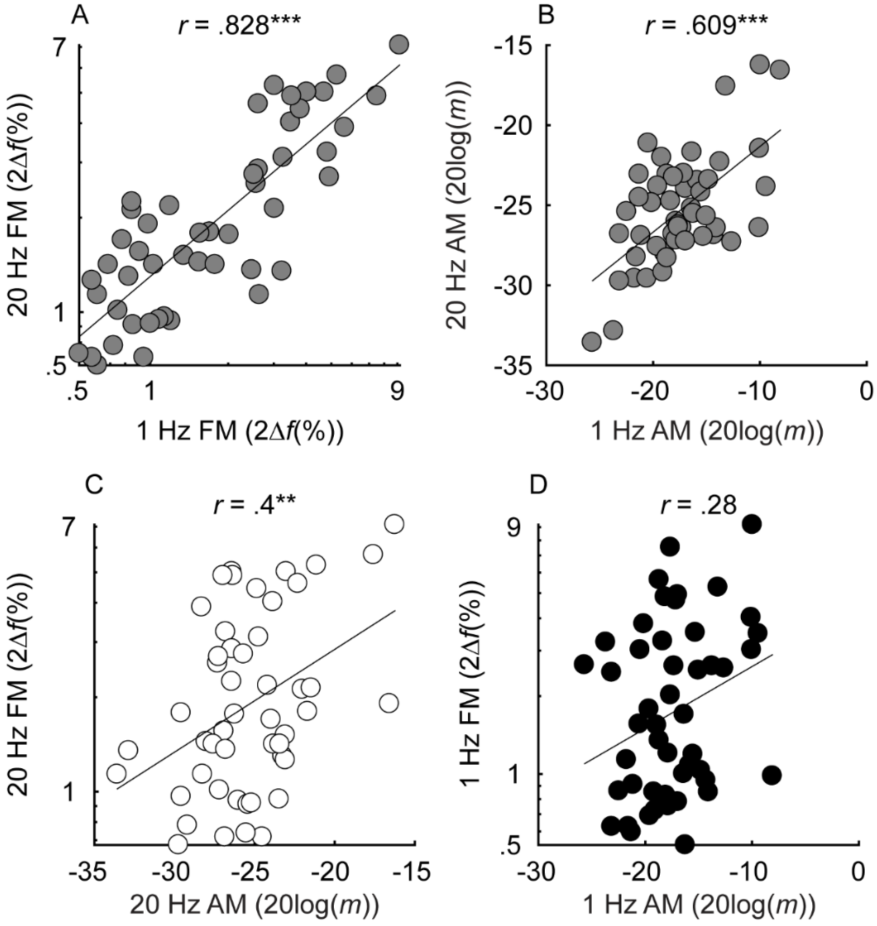

Fig. 4.3

Individual thresholds for slow (f_m = 1 Hz; black) and fast (f_m = 20 Hz; white) FM and AM detection. Grey circles represent different rates on the x- and y-axes. FM and AM thresholds are plotted in percent peak-to-peak frequency change (2Δ*f*(%)) and 20log(*m*), where Δ*f* is the frequency excursion from the carrier and *m* is the modulation depth (ranging from 0-1). For all tasks, lower on the x- or y-axis represents better thresholds. Correlations marked with an * are significant after Bonferroni correction.

4.2.4 The role of frequency selectivity in FM detection

A shared neural code for FM and AM should result in better FM detection in people with steeper masking functions, implying sharper cochlear tuning, as this will result in a greater change in the output between neighboring filters as the frequency shifts across the filters (i.e., better FM-to-AM conversion) (Zwicker, 1956). The current consensus is that cochlear filtering should matter for fast but not slow FM (e.g., Moore and Sek, 1995, 1996; Lacher-Fougère and Demany, 1998; Strelcyk and Dau, 2009). Surprisingly, both slow and fast FM detection were similarly strongly related to the masking function slopes (see Fig. 4), even after using Bonferroni correction for 12 multiple comparisons ($\alpha = .004$). Age and sensitivity to AM could confound effects of cochlear filtering because they are known to influence FM detection in normal hearing listeners (Whiteford and Oxenham, 2015; Whiteford et al., 2017). Audibility is not thought to effect FM for levels above about 25 dB SL (Zurek and Formby, 1981) but was included as a precaution, since a few listeners with greater SNHL had stimuli presented at 20 dB SL. Partial correlations between FM detection and masking function slopes were conducted, controlling for age, absolute thresholds at 1 kHz (task 1), and AM detection at the corresponding rate, thereby isolating the role of place coding in FM detection. The correlations between the residuals are shown in Fig. 5, and demonstrate a significant relation between the low slope and FM detection at both rates (slow FM: $r_p = -.462$, $p = .0005$; fast FM: $r_p = -.43$, $p = .001$), no relation between the high slope and FM (slow FM: $r_p = -.064$, $p = .337$; fast FM: $r_p = -.072$, $p = .317$), and a significant correlation between the sum of the low and absolute value of the high slope and FM (slow FM: $r_p = -.385$, $p = .004$; fast FM: $r_p = -.41$, $p = .002$). Because the low slope of the masking function (reflecting the upper slopes of the cochlear filters) is generally steeper than the high slope, it provides more stimulus information relative to the high side (Zwicker, 1956). The results therefore provide strong support for the hypothesis that place coding is utilized for FM detection at both slow and fast rates. These results were further confirmed using multiple linear regression analyses (see Appendix Text A1).

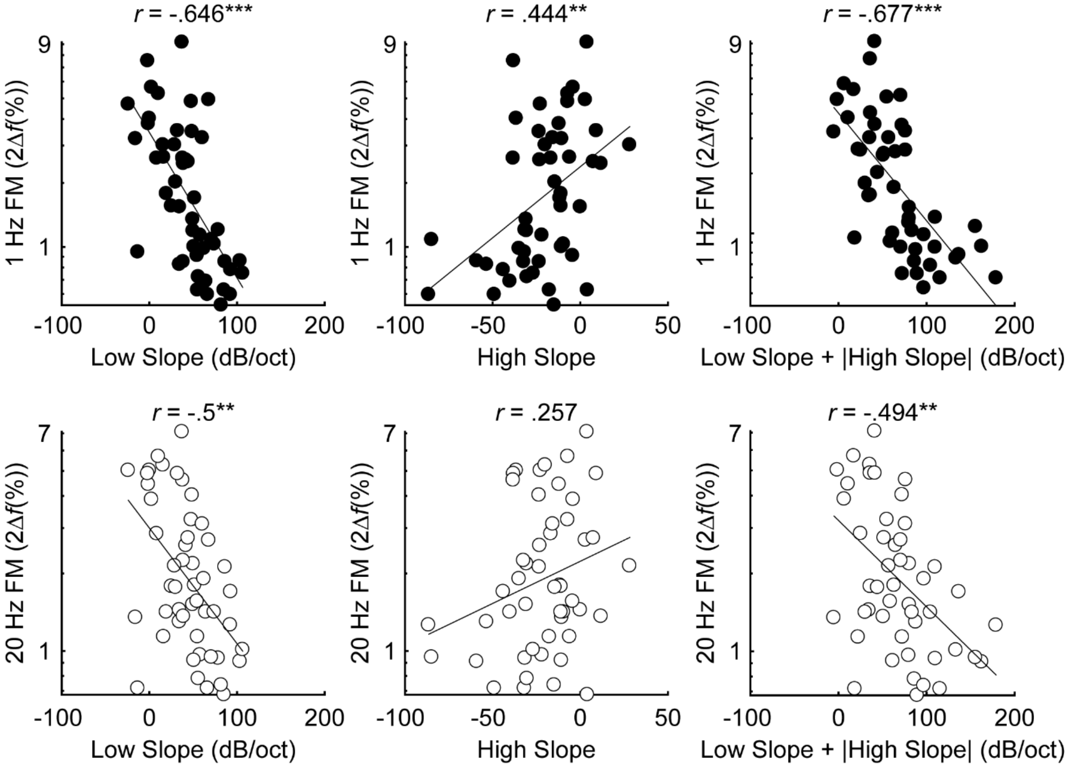

Fig. 4.4

Correlations between the low slope (leftmost column), high slope (middle column) and the low slope + |high slope| (rightmost column) and slow (f_m = 1 Hz; black) and fast (f_m = 20 Hz; white) FM detection.

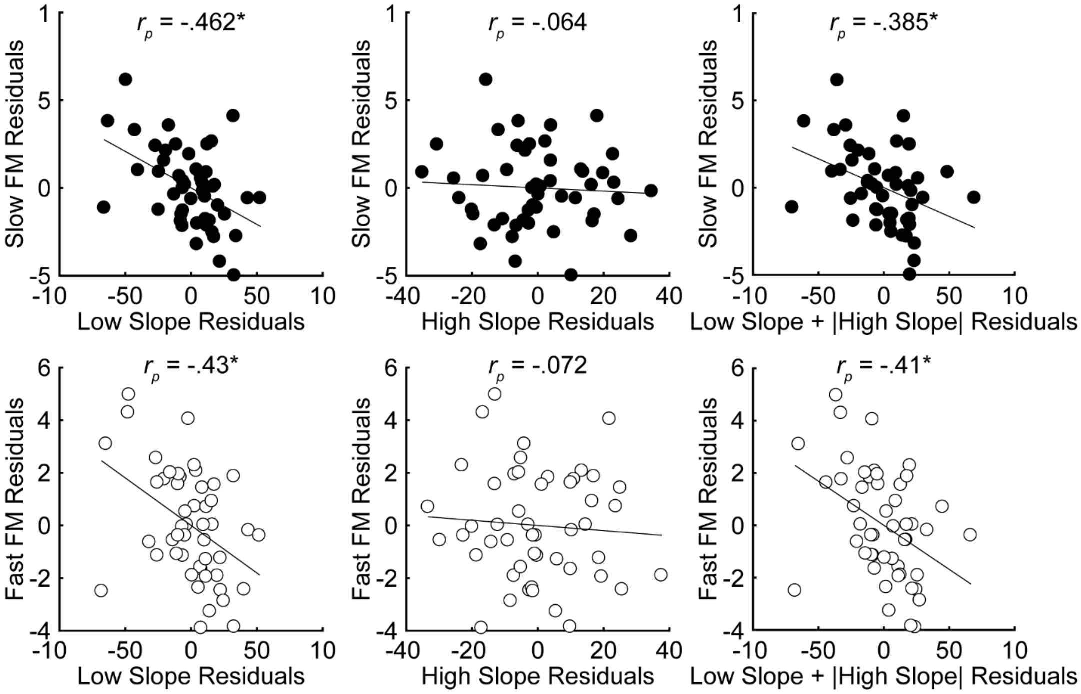

Fig. 4.5

Partial correlations between the steepness of the masking function slopes (x-axes) and FM detection (y-axes) after variance due to audibility, sensitivity to AM, and age has been partialled out. Units of the x- and y-axes are arbitrary because they correspond to the residual variance for slow (f_m = 1 Hz; black) and fast FM detection (f_m = 20 Hz; white).

4.3 Discussion

Our finding that cochlear place coding is equally important for both slow- and fast-rate FM detection was unexpected. Humans' acute sensitivity to slow changes in frequency at carriers important for speech and music has been thought to be a result of precise neural synchronization to the temporal fine structure of the waveform (Demany and Semal, 1989; Moore and Sek, 1995, 1996; Sek and Moore, 1995; Lacher-Fougère and Demany, 1998; Buss et al., 2004; Strelcyk and Dau, 2009; Ruggles et al., 2011). Previous large-sample studies with normal-hearing listeners using very similar methods saw essentially no role for frequency selectivity in slow or fast FM, presumably because all of the listeners had normal hearing, which would limit the variability in frequency selectivity (Whiteford and Oxenham, 2015; Whiteford et al., 2017). By using a sample with a wide range of hearing losses, we have revealed a clear role for place coding in FM, suggesting that either a place or a combined place-time code must be utilized. Indeed, the present study likely underestimates the importance of place coding, given the approximate nature of our estimate of cochlear filtering, based on only four data points on either side of the masker, and the approximation of the slope as a linear function.

4.3.1 Alternative explanations

One alternative explanation is that the present results are an epiphenomenon, resulting from co-occurring degradation of time coding to temporal fine structure with SNHL. There are several reasons why this is an unlikely explanation. First, the literature on whether time coding degrades with SNHL, particularly for tones in quiet, is mixed. Physiological studies with non-human animals have generally found no effects (Harrison and Evans, 1979; Miller et al., 1997) or very small effects (Henry and Heinz, 2012) of SNHL on time coding, with the exception of one study (Woolf et al., 1981). Support from human studies are based on poorer behavioral performance in hearing impaired listeners in tasks thought to use time coding (e.g., Lorenzi et al., 2006; Moore et al., 2006, 2012, Hopkins and Moore, 2007, 2011; Moore, 2014; Füllgrabe and Moore, 2017). For some of these tasks, whether one can truly isolate the temporal fine structure of a monaural stimulus from spectral cues is controversial (e.g., Oxenham et al., 2009). Even binaural tasks that are known to use precise time coding, such as localization based on microsecond differences in arrival times between the two ears (interaural time differences), are questionable indices of time coding because variability on these tasks may be dominated by variability in non-peripheral factors, such as binaural coding factors at stages after the initial integration of information from the two ears, task demands, aging, or processing efficiency (Whiteford et al., 2017).

A second reason why it is unlikely for the role of place coding in FM to be a byproduct of time coding degrading with hearing loss is that not all the listeners in the present study had SNHL, yet the trends between FM and masking function slopes persisted. Additionally, the relationships between FM and the slopes of the masking function were specific to the low-frequency side of the excitation pattern. Zwicker (1956) predicted over half a century ago that the steeper, low-frequency slope should play a larger role in FM-to-AM conversion. If the current findings were a spurious effect of time coding degrading with hearing loss, then the correlation should not be specific to the low-frequency slope, as the high slope is also strongly affected by hearing loss ($r = .707$, $p < .0001$). Lastly, it is unlikely that a spurious correlation between masking slopes and slow FM would have a similar magnitude to the correlation between fast FM and masking slopes. Equally strong relationships between slow vs. fast FM detection and the fidelity of cochlear tuning demonstrates that place coding has a vital role in FM detection.

4.3.2 Place vs. place-time models

A pure place model for FM proposes that FM is transduced to AM though cochlear filtering (Zwicker, 1952). As the frequency sweeps across the tonotopic axis, the auditory system monitors changes in the output of the cochlear filters. For a place-only model to explain FM, it would need to account for the rate-dependent trends in FM and AM sensitivity observed here (Fig. 2) and in many previous studies (Viemeister, 1979; Sheft and Yost, 1990; Moore and Sek, 1995, 1996; Lacher-Fougère and Demany, 1998; Whiteford and Oxenham, 2015, 2017; Whiteford et al., 2017). One possible explanation is that central auditory system's ability to compare changes in the output between neighboring cochlear filters is more efficient at slower rates, although this would not explain differential rate-dependent trends at low and high carrier frequencies in FM detection.

Alternatively, a combined place-time code would predict better sensitivity for slow, low-carrier FM relative to the same carrier at faster rates (Fig. 2). Place-time models purport that the extraction of timing information is place dependent. (Loeb et al., 1983; Shamma and Klein, 2000). There are various implementations, but place-time models generally rely on an array of coincidence detectors calculating the instantaneous cross-correlation between the phase-locked responses of auditory nerve fibers innervating different cochlear locations. Poor frequency tuning that occurs with hearing loss affects the traveling wave response, thereby disrupting this place-time relationship (Ruggero, 2013). A combined place-time code could account for the correlation between slow-rate FM and the low slope of the masking function.

Furthermoe, a place-time code could be considered as consistent with studies that have examined normal hearing sensitivity for FM in the presence of an AM masker (Moore and Sek, 1996; Moore and Skrodzka, 2002; Ernst and Moore, 2010, 2012; Paraouty and Lorenzi, 2017). For these studies, participants were presented with two AM tones, one at a time, and were instructed to pick the one that also had an FM tone present. The AM masker is thought to disrupt the excitation pattern cues (place information) while leaving the temporal fine structure timing cues intact. Several studies have shown that FM detection at low carriers is impaired at both slow and fast rates, but the degree of impairment is worse at the faster rate, which could suggest that place and timing cues are used for slow FM (Moore and Sek, 1996; Moore and Skrodzka, 2002; Ernst and Moore, 2010, 2012). These trends do not always hold for older listeners with SNHL or for normal hearing listeners at low sensation levels (Moore and Skrodzka, 2002; Ernst and Moore, 2010), whereby added AM equally impairs slow and fast FM detection. The combined results suggest either age, SNHL, and low sensation level all co-occur with a degradation of timing cues, or only place information is utilized for slow, low-carrier FM.

4.3.3 Unexplained variance

The MLR results showed that audibility, age, sensitivity to AM, and masking function slopes account for about 64.5% and 55.2% of the total variance in slow and fast FM, respectively (Appendix Text A1). What factors, then, account for the unexplained variance? Some of the leftover variance will be measurement noise, although this is unlikely to account for all the leftover variance. Additional variability, particularly in the asymmetric listeners, may be due to a “better ear” effect. When there are particularly large asymmetries, SNHL listeners may become accustomed to relying on their better-hearing ear, ignoring the worse ear. Hence, performance in the worse-hearing ear may be poor across tasks regardless of the peripheral coding mechanism involved, adding to the unexplained variance.

4.3.4 Implications

Using highly controlled, pure-tone stimuli and a large sample of listeners with low-frequency SNHL, we demonstrated that listeners’ sensitivity to changes in frequency is directly related to the fidelity of peripheral place coding. The clear role for place coding in slow FM is contrary to the widely accepted understanding that time coding is used. This has important implications for studies that quantify the fidelity of time coding to temporal fine structure using slow FM, as variability in slow FM at least partly reflects variability in place coding.

Another important implication is that the precision of cochlear filtering, rather than auditory-nerve phase-locked spike times, could explain problems with speech perception in listeners with SNHL. All sounds consist of a sum of one or more pure tones, and many natural sounds are modulated in frequency and amplitude. Our results suggest that prosthetic and medical interventions geared toward restoring the fidelity of place coding could be a successful means of improving communication in listeners with SNHL.

4.4 Methods

4.4.1 Participants

Experimental tasks were assessed on 60 ears (49 participants; 18 male, average age of 66.3, range: 19.4-78.5 years), with a range of SNHL at 1 kHz (3.66 – 68.5 dB SPL based on Task 1). Pure-tone audiometry was assessed at octave frequencies from 250-8000 Hz. Five participants had normal hearing, defined as audiometric thresholds <= 20 dB HL between 250-4000 Hz. SNHL participants had 1-kHz audiometric thresholds worse than 20 dB HL in at least one ear and air-bone gaps < 10 dB to preclude a conductive hearing loss. Ears with SNHL ≥ 70 dB SPL from Task 1 were not included in the study. Participants with symmetric hearing (n = 33; asymmetries ≤ 10 dB at 1 kHz from Task 1) completed all monaural experimental tasks in their worse ear. Six additional participants had SNHL at 1 kHz in both ears, but loss in the poorer ear exceeded the study criterion; for these subjects, tasks were completed in the better ear only. An additional three participants had one normal-hearing ear and one ear with SNHL at 1 kHz, and only measurements from the impaired ear were used in analyses. The final eight participants had asymmetric SNHL in both ears, defined as an asymmetry > 10 dB on Task 1. For these subjects, experimental tasks were completed for both ears separately. Unless otherwise stated, performance in the worse ear was used in analyses for asymmetric listeners. Participants provided informed consent and were given monetary compensation for their time. The Institutional Review Board of the University of Minnesota approved all experimental protocols.

4.4.2 Stimuli

Stimuli were generated in Matlab with a sampling rate of 48 kHz using a 24-bit Lynx Studio L22 sound card and presented over Sennheiser HD650 headphones in a sound-attenuating chamber. Tasks were measured monaurally with threshold equalizing noise (TEN) presented in the contralateral ear in order to prevent cross-talk between the two ears. TEN was presented continuously in each trial, with the bandwidth spanning 1 octave around the test frequency. Except for tasks that involved detection of a short (20 ms) target, the TEN level was always 25 dB below the target level, beginning 300 ms before the onset of the first interval and ending 200 ms after the offset of the second interval. Because less noise is needed to mask very short targets, the TEN was presented 35 dB below the target level for tasks that involved detection of a short, 20-ms target (Tasks 4 and 7), and began 200 ms before the onset of the first interval and ended 100 ms after the offset of the second interval.

To obtain a more precise estimate of sensitivity for the test frequency, pure-tone absolute thresholds were measured for each ear at 1 kHz. The target was 500 ms in duration with 10-ms raised-cosine onset and offset ramps. The reference was 500 ms of silence, and the target and the reference were separated by a 400-ms interstimulus interval (ISI). Modulated tasks were assessed for the same frequency (f_c = 1 kHz) at slow (f_m = 1 Hz) and fast (f_m = 20 Hz) rates. The target was an FM (Tasks 2 and 3) or AM (Task 4 and 5) pure tone while the reference was an unmodulated pure tone at 1 kHz. Both the target and the reference tones were 2 s in duration with 50-ms raised-cosine onset and offset ramps. In the FM tasks, the starting phase of the modulator frequency was set so that the target always began with either an increase or decrease in frequency excursion from the carrier frequency, with 50% probability determined a priori. The analogous manipulation was done for the AM tasks, so that the target always began at either an amplitude peak or an amplitude trough. Stimuli for the modulation tasks were presented at 65 dB SPL or 20 dB SL, whichever was greater, based on individualized absolute thresholds at 1 kHz from Task 1.

Detection for a short (20 ms), pure-tone target tone was measured with and without the presence of a 1-kHz, 500-ms pure-tone forward masker. Target frequencies were 800, 860, 920, 980, 1020, 1080, 1140, and 1200 Hz, and both the target and the masker had 10-ms raised cosine onset and offset ramps. The target was presented to one ear, directly following the offset of the masker, and the masker was presented to both ears to avoid potential confusion effects between the offset of the masker and the onset of the target (Neff, 1986). The masker was fixed in level at either 65 dB SPL or 20 dB SL, which ever was greater, based on absolute thresholds for the 500-ms test frequency in the target ear (Task 1). The starting level of the target was always 10 dB below the masker level in the masked condition. For unmasked thresholds, the starting level of the target was either 40 dB SPL or 20 dB SL, whichever was greater, and the target was preceded by 500 ms of silence.

4.4.3 Procedure

Procedures were adapted from Whiteford et al. (2017) and are described in full below. The experiment took place across 3-6 separate sessions, with each session lasting no longer than 2 hours. All tasks were carried out using a two-interval, two-alternative forced-choice procedure with a 3-down 1-up adaptive method that tracks the 79.4% correct point of the psychometric function (Levitt, 1971). The target was presented in either the first or second interval with 50% a priori probability, and the participant's task was to click the virtual button on the computer screen (labeled "1" or "2") corresponding to the interval that contained the target. Each corresponding response button illuminated red during the presentation of the stimulus (either reference or target). Visual feedback ("Correct" or "Incorrect') was presented on the screen after each trial. All participants completed the tasks in the same order, and the tasks are described below in the order in which they were completed by the participants.

4.4.3.1 Task 1: Absolute thresholds at 1 kHz

Participants were instructed to select the button on the computer screen that was illuminated while they heard a tone. The target was a 500-ms, 1-kHz pure tone presented to one ear, and the reference was 500 ms of silence. Three runs were measured for each ear, and the order of the presentation ear (left vs. right) was randomized across runs. Three participants were only assessed in their better ear, due to an extensive amount of hearing loss in the poorer ear according to their 1 kHz audiometric thresholds (all ≥ 80 dB HL). The remaining 40 participants completed monaural absolute thresholds for both ears.

On the first trial, the target was presented at 40 dB SPL. The target changed by 8 dB for the first reversal, 4 dB for the next 2 reversals, and 2 dB for all following reversals. Absolute thresholds were determined by calculating the mean level at the final 6 reversal points. If the standard deviation (SD) across the three runs was ≥ 4, then 3 additional runs were conducted for the corresponding ear, and the first three runs were regarded as practice.

Based on these thresholds, an additional 3 subjects had monaural absolute thresholds for the worse ear that was ≥ 70 dB SPL. For these subjects, the better ear was measured for all following tasks.

4.4.3.2 Tasks 2 and 3: FM detection

Participants were instructed to pick the tone that was “modulated” or “changing”. At the beginning of each run, the target had a peak-to-peak frequency change ($2\Delta f$) of 5.02%. The $2\Delta f$ varied by a factor of 2 for the first two reversal points, a factor of 1.4 for the third and fourth reversal points, and a factor of 1.19 for all following reversal points. The FM difference limen (FMDL) was defined as the geometric mean Δf at the final 6 reversal points.

Three runs were conducted for each modulation rate, and all three runs for slow-rate FM were completed before fast-rate FM. Asymmetric participants with two qualifying ears completed six runs (three runs per ear) for each modulation rate, and the order of the presentation ear was randomized across runs. If the SD across the three runs for a given ear was ≥ 4, the participant completed an additional three runs, and only the last three runs were used in analyses.

4.4.3.3 Task 4: Detection for 20-ms tones

Participants were instructed to select the button (labeled “1” or “2”) on the computer screen that was illuminated while they heard a short, 20-ms target tone. The target was presented at 40 dB SPL or 20 dB SL, whichever was greater, for the first trial of each run. The level of the target changed by 8 dB for the first two reversals, 4 dB for the following two reversals, and 2 dB for all following reversals.

The absolute threshold was defined as the mean target level at the final six reversal points.

Participants completed one run for each of the eight target frequencies: 800, 860, 920, 980, 1020, 1080, 1140, and 1200 Hz. The order of the target frequency condition was randomized across runs. Asymmetric participants with two qualifying ears had the order of the runs further blocked by presentation ear, so that 8 runs for the same ear had to be completed before the next presentation ear condition was measured. Whether the right or left ear was assessed first was randomized. One additional run was conducted for any conditions with an SD ≥ 4 dB, and only the final run for each condition was used in analyses.

4.4.3.4 Tasks 5 and 6: AM detection

The instructions for AM detection were the same as the instructions for FM detection. The first trial of each run had a target with an AM depth of -7.96, in 20log(*m*) units. The target modulation depth changed by 6 dB for the first two reversals, 2 dB for the next two reversals, and 1 dB for all following reversals. The AM difference limen (AMDL) was defined as the mean modulation depth (in 20log(*m*)) at the last 6 reversal points.

In the same manner as the FM tasks, all three runs for slow-rate AM ($f_m = 1$ Hz) were completed before fast-rate AM ($f_m = 20$ Hz). Asymmetric participants with two qualifying ears completed six runs (three runs per ear) for each modulation rate, and the order of the presentation ear was randomized across runs. If the SD across the first three runs for a given condition were ≥ 4 dB, then three additional runs were conducted, and only the final three runs were analyzed.

4.4.3.5 Task 7: Forward masking patterns

The task was to determine which of two tones was followed by a short, 20-ms target tone. Two runs were measured for each of the eight target frequencies (800, 860, 920, 980, 1020, 1080, 1140, and 1200 Hz), for a total of 16 runs, and the order of the target condition was randomized across runs. Asymmetric participants with two qualifying ears had the order of the runs further blocked by presentation ear, so that 8 runs for the same ear had to be completed before the next presentation ear condition was presented. The 1-kHz, 500-ms masker tones were fixed in frequency and level, presented binaurally at 65 dB SPL or 20 dB SL based on absolute thresholds from Task 1, whichever was greater. Within a trial, each masker was either directly followed by a 20-ms target tone, presented monaurally to the target ear, or 20-ms of silence. The starting level of the target tone was 10 dB below the masker level in the corresponding ear. The level of the target tone changed by 8 dB for the first two reversals, 4 dB for the third and fourth reversals, and 2 dB for the following reversals. The masked threshold for each frequency condition was calculated as the mean target level at the final 6 reversal points. For a given subject, if the SD of the masked threshold across the two runs was ≥ 4 dB, then the subject completed two additional runs for the corresponding target frequency. For these conditions, only the final two runs were used in analyses, and the first two runs were regarded as practice. The average across the final two runs for each target frequency was used in analyses.

CHAPTER 5: PITCH PERCEPTION IN AMUSIA

Chapter 5 is reprinted from:

Whiteford, K. L., & Oxenham, A. J. (2017). Auditory deficits in amusia extend beyond poor pitch perception. *Neuropsychologia, 99*, 213-224.

Abstract

Congenital amusia is a music perception disorder believed to reflect a deficit in fine-grained pitch perception and/or short-term or working memory for pitch. Because most measures of pitch perception include memory and segmentation components, it has been difficult to determine the true extent of pitch processing deficits in amusia. It is also unclear whether pitch deficits persist at frequencies beyond the range of musical pitch. To address these questions, experiments were conducted with amusics and matched controls, manipulating both the stimuli and the task demands. First, we assessed pitch discrimination at low (500 Hz and 2000 Hz) and high (8000 Hz) frequencies using a three-interval forced-choice task. Amusics exhibited deficits even at the highest frequency, which lies beyond the existence region of musical pitch. Next, we assessed the extent to which frequency coding deficits persist in one- and two-interval frequency-modulation (FM) and amplitude-modulation (AM) detection tasks at 500 Hz at slow (f_m = 4 Hz) and fast (f_m = 20 Hz) modulation rates. Amusics still exhibited deficits in one-interval FM detection tasks that should not involve memory or segmentation. Surprisingly, amusics were also impaired on AM detection, which should not involve pitch processing. Finally, direct comparisons between the detection of continuous and discrete FM demonstrated that amusics suffer deficits both in coding and segmenting pitch information. Our results reveal auditory deficits in amusia extending beyond pitch perception that are subtle when controlling for memory and segmentation, and are likely exacerbated in more complex contexts such as musical listening.

5.1 Introduction

Pitch, the psychological attribute that allows us to order sounds on a musical scale from low to high (ANSI, 2013), plays a fundamental role in our auditory worlds, helping us to segregate sounds, recognize voices, and enjoy our favorite music. Pitch is related to the repetition rate, or periodicity, of an acoustic waveform. For pure tones, or sinusoids that just have one frequency, the strength and accuracy of the tone's pitch depends on attributes such as its duration and frequency (Moore, 1973; Micheyl et al., 2012; Moore and Ernst, 2012).

Even with the same stimulus parameters, large differences in pitch discrimination abilities have been observed between individuals, due to factors such as experience (e.g., Carcagno and Plack, 2011; Micheyl et al., 2006), age (e.g., Moore and Peters, 1992), hearing loss (Moore and Peters, 1992; Arehart, 1994; Moore and Moore, 2003), and neurogenetic disorders (Peretz et al., 2007). One such disorder known to be related to deficits in pitch perception is congenital amusia (e.g., Vuvan et al., 2015). Amusia, commonly referred to as tone deafness, is a deficit in music perception that is believed to be independent of hearing loss, musical training, or intelligence (Peretz, 2001; Ayotte et al., 2002). Amusics typically report that they are "musically impaired," report difficulty recognizing familiar tunes when presented without lyrics, and cannot perceive when they or others are singing out of tune (Peretz et al., 2003, 2008). Amusia may be congenital (i.e., developed at birth or in early childhood) or can be acquired through brain injury to areas important for music perception (e.g., bilateral damage to superior temporal lobes) (Peretz et al., 1994; Peretz, 2001). Music perception deficits in amusia appear to be linked to an underlying impairment in fine-grained pitch perception (e.g., Foxton et al., 2004; Peretz et al., 2002; Vuvan et al., 2015), either due to deficits in pitch processing (e.g., Cousineau et al., 2015) or deficits in pitch memory (Gosselin et al., 2009; Tillmann et al., 2009, 2016; Williamson and Stewart, 2010; Williamson et al., 2010; Albouy et al., 2013a, 2015b).

The exact nature of the pitch-processing deficits remains unclear. Cousineau et al. (2015) found that amusics exhibited a deficit in pitch discrimination for complex tones that contained low-numbered harmonics (e.g., 1-6), which are thought to be resolved in the auditory periphery and coded via their spectro-temporal fine structure. However, no deficit was observed in pitch discrimination for complex tones that contained only high-numbered harmonics (> 10), which are thought to be unresolved and coded via their temporal envelope (Houtsma and Smurzynski, 1990; Bernstein and Oxenham, 2003). Because amusics exhibited normal abilities to detect interaural time differences (ITD) via temporal fine structure, as well as normal auditory filter shapes, Cousineau et al. (2015) concluded that the amusics' deficit in coding the spectro-temporal fine structure of resolved harmonics was not peripheral in nature.

It is not known why pitch deficits related to amusia were observed with low-numbered resolved harmonics but not high-numbered unresolved harmonics. One possibility is that only pitch coding via temporal fine structure is affected, so that the (generally weaker) pitch elicited by temporal envelope cues is not affected. In other words, if different mechanisms are involved in coding resolved and unresolved harmonics (Carlyon and Shackleton, 1994; but see Micheyl and Oxenham, 2004), then amusia may selectively affect the coding of resolved harmonics, which provides the dominant pitch percept in everyday life (Plack and Oxenham, 2005). Another possibility is that amusic deficits are limited to fine-grained pitch differences (e.g., Foxton et al., 2004; Peretz et al., 2002; Vuvan et al., 2015), and because pitch discrimination of complexes with only unresolved harmonics is relatively coarse, amusic deficits are no longer measurable in such conditions. Experiment 1 of our study addresses this question by measuring pitch discrimination for single pure tones over a range of frequencies, including a very high frequency (8 kHz), where discrimination is generally much poorer. If amusic deficits are linked to the poorer coding of individual harmonics and tones, then performance should remain poorer than normal even at high frequencies. On the other hand, if amusic deficits are limited solely to conditions where normal pitch discrimination is very fine, then the deficit should be reduced at 8 kHz, where normal pitch discrimination is degraded relative to discrimination at lower frequencies.

Another open question relating to amusia is the extent to which it affects perceptual (e.g., pitch processing) versus cognitive (e.g., short-term or working memory) processing. Attempts to find an anatomical or physiological marker for amusia have largely focused on cortical differences. One noted difference involves grey and white matter abnormalities in the right superior temporal gyrus (rSTG) and the right inferior frontal gyrus (rIFG) (Hyde et al., 2006, 2007), with amusics exhibiting increased grey matter in rSTG and rIFG relative to controls. This finding was somewhat counterintuitive, given that professional musicians have also been reported to have more grey matter relative to non-musicians in a subsection of the auditory cortex (e.g., anteromedial Heschl's gyrus; Schneider et al., 2002). Hyde et al. (2007) suggested that increases in grey matter in amusia may arise due to abnormal cortical migration during development. In line with the cortical migration theory, amusics have been found to exhibit impaired connectivity between the rIFG and the rSTG, confirmed via functional connectivity measures in fMRI (Hyde et al., 2011), resting-state fMRI (Lévêque et al., 2016), and Diffusion Tensor Imaging (Loui et al., 2009, although see Chen et al., 2015). More recently, abnormalities in connectivity patterns have been identified both within and between the auditory cortices using magneto-encephalography (MEG) (Albouy et al., 2013a, 2015b). These abnormalities within the frontotemporal network coincide with abnormal backward connectivity during pitch encoding (Albouy et al., 2013a) and abnormal forward connectivity during pitch retrieval (Albouy et al., 2015b). Taken together, these results support the hypothesis that amusia is related to dysfunctions in the frontotemporal network, rather than abnormalities in early cortical or pre-cortical processing. Furthermore, recent fMRI results show no difference in the proportion, location, and selectivity of pitch-responsive voxels of the auditory cortex in amusics compared to controls (Norman-Haignere et al., 2016). Thus, current imaging findings might implicate deficits in memory and/or segmentation, rather than the initial processing of pitch. Although numerous studies provide evidence for impaired retention for short-term pitch information in amusia, whether or not amusics have a specific deficit in working memory for pitch (i.e., a difficulty in "online" comparison of pitch information over time) is an important question that has received little attention (Tillmann et al., 2016). Experiments 2 and 3 in this study provide behavioral tests to distinguish between these possibilities by comparing frequency-modulation (FM) detection in a single-interval task with more traditional frequency-discrimination paradigms involving one or more comparison intervals.

5.2 Experiment 1: Pure-Tone Frequency Discrimination

The purpose of experiment 1 was to determine whether poor pitch perception in amusia extends to high frequencies. On one hand, amusia typically results in a poor ability to discriminate small frequency differences (for a review, see Vuvan et al., 2015), so one might expect this deficit to be present at all frequencies. On the other hand, there are at least two considerations to suggest that the deficit may be limited to lower frequencies. The first consideration is that musical pitch perception (i.e., pitch sufficiently salient to convey musical intervals and melodies) only typically extends up to about 4-5 kHz (Attneave and Olson, 1971; Oxenham et al., 2011). If amusia is a deficit specifically related to musical pitch, then the deficit may not extend to high frequencies. The second consideration is that pitch deficits in amusia are limited to discrimination of small frequency differences of a semitone or less (~6%; e.g., Foxton et al., 2004; Hyde & Peretz, 2004). It may be that the advantage of control participants over amusics only holds in conditions where very fine discrimination is normally possible. Because the ability of normal participants to discriminate the pitch of pure tones worsens dramatically at high frequencies (e.g., Micheyl et al., 2012; Moore, 1973; Moore & Ernst, 2012), it may be that amusics perform more poorly at low frequencies but more similarly at high frequencies, as was found with spectrally resolved and unresolved components within a complex tone (Cousineau et al., 2015). We addressed this question by comparing the ability of amusic and normal control participants to discriminate changes in frequency of pure tones with frequencies of 500, 2000, and 8000 Hz.

5.2.1 Methods

5.2.1.1 Participants

Twelve amusics and 12 matched controls participated in the study. Participants were matched in age, years of musical experience, and years of education (see Table 5.1). In addition, IQ was measured via the Vocabulary and Matrix Reasoning subtests of the Wechsler Abbreviated Scale of Intelligence – Second Edition (WASI-II; Wechsler, 2011). The scores from the two subtests were combined to calculate the full scale IQ scores, which did not differ significantly between the groups (see Table 5.1). Because the WASI-II includes a measure of verbal intelligence, all participants were native speakers of American English. Amusia was determined based on a global score on the MBEA (Peretz et al., 2003) that was at least two standard deviations below the mean of the general population. All except one amusic participant also performed below the same cutoff for the pitch-based subtasks on the MBEA. None of the participants reported a history of neurological conditions. All participants provided written informed consent and were compensated with course credit or hourly payment for their time. The experimental protocols were approved by the Institutional Review Board of the University of Minnesota.

Table 5.1

	Group		
	Amusics (n = 12)	**Controls (n = 12)**	***p*-value**
Age (years)	23.42 (9.83)	23.33 (8.84)	.983
Musical Experience (years)	1.46 (1.53)	1.17 (2.08)	.7
Education (years)	15 (1.26)	15.13 (1.72)	.841
IQ	106 (9.94)	112.17 (10.03)	.144
Gender	9 females, 3 males	9 females, 3 males	-
Pitch MBEA (%)	65.83 (4.65)	85.46 (4.46)	**< .0001***
Global MBEA (%)	67.5 (4.49)	86.3 (3.73)	**< .0001***
Audiogram- 4 Pure-Tone Average (dB HL)	3.65 (3.07)	6.09 (3.72)	.093

Demographic averages for 12 amusics and 12 controls. Standard deviation is in parentheses. Pitch MBEA represents the average percent correct on the pitch-based subtasks of the MBEA. Global MBEA represents the overall average percent correct across all tasks on the MBEA. All MBEA percentages were transformed to rationalized arcsine units for statistical analyses (Studebaker, 1985). The 4 pure-tone average is the audiometric thresholds between 500-4000 Hz, averaged across frequencies and ears. All *p* values correspond to independent-samples t tests. The *p* values less than 0.05 are marked with an asterisk and shown in bold.

Audiometric thresholds were measured at octave frequencies from 250 to 8000 Hz. All except one participant had NH up through 8 kHz, defined as audiometric thresholds no greater than 20 dB HL. The one participant who did not meet this criterion was in the control group and had NH up through 4 kHz, and a mild loss at 8 kHz (right ear: 25 dB HL; left ear: 35 dB HL). The average 8-kHz thresholds (across both ears) were not significantly different between the two groups ($t(22) = -.881$, $p = .388$, two-tailed).

5.2.1.2 Stimuli and procedures

The auditory tasks were administered in a double-walled sound-attenuating booth via Sennheiser HD650 headphones. All stimuli were generated in MATLAB via a 24-bit L22 soundcard (LynxStudio, Costa Mesa, CA) with a sampling rate of 48,000 Hz and were presented at 60 dB sound pressure level (SPL).

Pure-tone frequency difference limens (FDLs) were measured at 500, 2000, and 8000 Hz. The tones were 200 ms in duration, including 50-ms raised-cosine onset and offset ramps. FDLs were obtained using a three-interval, three-alternative forced-choice (AFC) task with a two-down, one-up adaptive procedure that tracks the 70.7% correct point on the psychometric function (Levitt, 1971). On each trial, three tones were presented sequentially, separated by inter-stimulus intervals of 500 ms. Two of the tones were reference tones, identical in frequency. The other tone was the target tone, which was always higher in frequency than the reference tones. The frequencies of the reference and target tones were geometrically centered around the nominal test frequency of 500, 2000, or 8000 Hz. The presentation order of the target tone relative to the reference tones was selected randomly on each trial with uniform distribution. Subjects were instructed to pick the tone that was different by selecting one of three virtual buttons on the computer screen and to look at the computer screen to monitor the feedback, which was presented after each trial with the word “Correct” or “Incorrect”. Throughout the task, the three buttons were labeled “1”, “2”, and “3”. Each button was illuminated red during the presentation of the corresponding tone (i.e., button 1 was illuminated red during the presentation of the first tone, etc.). The starting value of the frequency difference (Δf) between the target and the reference tones was 20% (i.e., slightly greater than three semitones). The value of Δf varied by a factor of 2 for the first two reversal points, by a factor of 1.41 for the following two reversal points, and by a factor of 1.12 for the final six reversal points. The threshold for each run was calculated as the geometric mean of the Δf values at the last 6 reversal points. Each participant completed 2 runs at each frequency condition, for a total of 6 adaptive runs. The testing order of the frequency conditions was randomized across subjects and across repetitions, such that all three frequencies were tested before any was repeated. For the full experimental protocol involving all three experiments, see Appendix Text B1.

5.2.2 Results

Pure-tone FDLs are shown for amusics and controls in Fig. 5.1. Individual differences were large, and there was considerable overlap in FDLs between the two groups (Fig. 5.1A). However, on average, amusics had thresholds around 1 semitone (i.e., 6%) or greater, which was considerably higher than the mean thresholds observed in the control group (Fig. 5.1B). Interestingly, this trend was observed at all three frequencies tested, including 8000 Hz – a frequency that is typically too high to form recognizable melodies. A mixed-design ANOVA was performed, with log-transformed values of Δ*f* (%) as the dependent variable, subject group as the between-subjects factor, and frequency as the within-subjects factor. Log-transforming FDLs is common practice to avoid compression close to zero and better approximate normality (e.g., Micheyl et al., 2012). Greenhouse-Geisser corrections were applied to correct for lack of sphericity, where appropriate (Frequency: $\hat{\varepsilon}$ = .791), and the corrected degrees of freedom are reported. The ANOVA revealed a main effect of group [$F(1,22) = 13.2$, $p = .001$, $\eta_p^2 = .375$], a main effect of frequency [$F(1.58,34.8) = 26.5$, $p < .0001$, $\eta_p^2 = .546$], and no interaction [$F(1.58,34.8) = .282$, $p = .705$, $\eta_p^2 = .013$] when using a criterion of α = .05. Bonferroni corrected *t*-tests (α= .017) indicated that thresholds at 8000 Hz were significantly higher (worse) than thresholds at 2000 Hz [$t(22) = 6.09$, $p < .0001$] and 500 Hz [$t(22) = 5.15$, $p < .0001$]. There was a non-significant trend for larger (worse) thresholds at 500 Hz compared to 2000 Hz [$t(22) = 1.80$, $p = .085$].

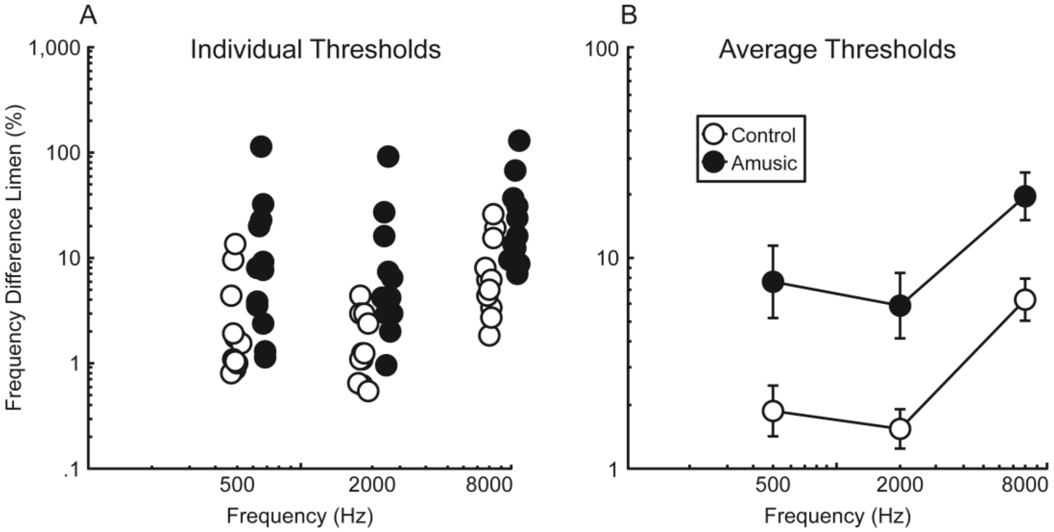

Figure 5.1

Frequency discrimination at low and high frequencies in amusics and controls. Filled and open circles represent data from amusic and control participants, respectively. A. Individual FDLs as a function of frequency. B. Group mean FDLs as a function of frequency. Error bars represent ±1 standard error of the mean. Amusic data points in (A) are offset horizontally for data visualization purposes.

One amusic participant had unusually poor pitch perception at all frequencies, with FDLs as large as an octave (one octave = 100% FDL) or more. When the value of Δf was one octave, the two frequencies present in a trial were half an octave below and above the nominal test frequency, i.e., about 5.65 and 11.3 kHz, respectively, for a test frequency of 8 kHz. This participant reported having trouble remembering the first two tones by the time the third tone in the sequence was presented. Removal of this participant with unusually high FDLs, along with his/her matched control, did not change the main effects or lack of interaction reported above.

5.2.3 Discussion

The results suggest that poor frequency discrimination in amusics extends to 8 kHz, and thus beyond the traditional existence region of musical pitch. Pure-tone frequencies of 8 kHz are generally too high to form recognizable melodies or musical intervals (Attneave and Olson, 1971), and are much higher than even the highest musical note produced by a grand piano or the highest orchestral instrument (i.e., the piccolo). Therefore, our results suggest that amusia is not a deficit selective to the musical attributes of pitch.

Poor frequency discrimination at 8 kHz in amusics relative to controls corroborates previous findings of normal ITD discrimination in amusics, which suggest that amusia does not exclusively affect any putative temporal coding of pitch information in the auditory nerve (Cousineau et al., 2015). Even if low-frequency tones are coded via phase-locked information, it is generally believed that this timing information is no longer available at high frequencies. Because no direct measurements of auditory-nerve phase locking in humans exist, the cut-off frequency is unknown, but different estimates have ranged from around 1.5 kHz, based on the limits of sensitivity to ITDs (Joris and Verschooten, 2013), up to 4-5 kHz (Moore, 1973; Sek and Moore, 1995) or even higher (Moore and Ernst, 2012). However, there is general consensus that phase-locked information is highly unlikely to be available at 8 kHz. Thus, the fact that amusics show the same pattern of results as the normal control subjects suggests that their deficit is not limited to tones coded via peripheral temporal fine structure information, but rather may reflect a problem with central coding of the spectro-temporal fine structure associated with pure tones and spectrally resolved harmonics.

5.3 Experiment 2: Assessing the Role of Memory Load

Performing frequency discrimination requires not only accurate coding of the frequency of each tone, but also memory storage and retrieval to allow successive tones to be compared (Durlach and Braida, 1969; Jesteadt and Sims, 1975; Zhang et al., 2016). It may be that poor FDLs in amusia are due not only to impaired frequency coding but also to deficits in short-term pitch memory (e.g., Albouy et al., 2013a; Gosselin et al., 2009; Williamson and Stewart, 2010) and/or manipulations of pitch information in working memory (Tillmann et al., 2016). There is conflicting evidence on whether amusia is related to problems with short-term memory for pitch. Several studies note that short-term memory deficits in amusia persist even for stimuli that are above threshold (Gosselin et al., 2009; Tillmann et al., 2009; Williamson et al., 2010; Albouy et al., 2013a, 2013b). However, presenting the same stimuli for both amusics and controls means that pitch changes will be closer to amusics' thresholds relative to controls (Jiang et al., 2013). Jiang et al. note it is difficult to interpret findings from memory studies that have not controlled for pitch sensitivity because pitch sensitivity and memory are confounded. Jiang et al. found that once controlling for each individual's perceptual sensitivity for pitch, there was no impairment in pitch memory for amusics relative to controls. In contrast, when perceptual sensitivity to pitch changes was not equated, Jiang et al. replicated findings from the previous studies (Gosselin et al., 2009; Williamson and Stewart, 2010; Williamson et al., 2010; Albouy et al., 2013a). They therefore concluded that amusia could not be attributed to deficits in short-term memory. However, even the paradigm of Jiang et al. (2013) relied to some extent on memory processes, as it still required a comparison of successive sounds across time. Auditory discrimination tasks may themselves utilize working memory if online comparisons are required to complete the task (Zhang et al., 2016).

Therefore, Jiang et al. (2013) could have over-corrected for pitch sensitivity if their pitch change detection task required even minimal memory demands, as suggested by Tillmann et al. (2016).

To further alleviate memory load, we measured participants' sensitivity to FM, using a one-interval yes/no task. After each tone, participants indicated "yes" if the tone was "changing" in pitch and "no" if the tone was not changing. Methods of signal detection theory were used to estimate sensitivity to FM in each group. An analogous task was used to measure AM detection – a task that does not rely on pitch perception but requires coding of temporal envelope cues. As a comparison, FM and AM detection were also measured using a standard two-alternative forced choice (2AFC) procedure. In this task, participants were presented with two tones and were instructed to pick the tone that was changing. The 2AFC task presents the listener with more information than a single-interval task, but accessing the additional information again requires that the information from both intervals be stored and retrieved. If amusia reflects only memory-related deficits, then no deficits should be observed for tasks involving the one-interval detection of either FM or AM. On the other hand, if amusia involves a specific deficit in frequency coding, then a deficit should be observed for FM detection, but not for AM detection.

5.3.1 Methods

5.3.1.1 Participants

The same participants from experiment 1 also completed experiment 2.

5.3.1.2 Stimuli and procedure

Difference limens for AM (AMDLs) and FM (FMDLs) were measured for a 500-Hz pure-tone carrier at slow (f_m = 4 Hz) and fast (f_m = 20 Hz) modulation rates. Both the target and the unmodulated reference tone were 1.5 s in duration, including 50-ms raised-cosine onset and offset ramps. The level was roved between intervals in the range from 57 and 63 dB SPL, with uniform distribution, to avoid cues related to overall level or loudness. The presentation order of the conditions, modulation type (AM vs. FM) and modulation rate (f_m = 4 Hz vs. f_m = 20 Hz), was counterbalanced across participants using a Latin Square design. Participants were yoked so that matched control-amusic pairs underwent the same counterbalanced order.

Before each adaptive run for the two-interval task, participants listened to an example of the corresponding target and reference tones. In order to ensure that the participants understood the type of modulation they would need to detect in the next experimental block of trials, they were required to correctly identify the example target by clicking a button labeled “Yes” after the modulated tone was presented. They were also required to correctly reject the reference tone by clicking a button labeled “No” after the pure tone was presented. Participants completed the example trials in the presence of the experimenter and were allowed to repeat the example trials as many times as necessary. The target for the AM example trials was presented at a modulation depth of 50% (m = 0.5), while the target for the FM example trials had a peak-to-peak frequency change of 12% (i.e., about 2 semitones). Most of the participants, including the amusics, did not need to repeat the example trials.

Initially, the two-interval, 2AFC task was used to estimate each individual’s sensitivity to AM and FM. Thresholds were measured using a three-down, one-up adaptive procedure that tracks the 79.4% correct point on the psychometric function (Levitt, 1971). On each trial, two tones were presented sequentially, separated by an interstimulus interval of 500 ms. One of the tones was the modulated target while the other was the reference tone with no modulation. Whether the target occurred in the first or second interval was selected randomly on each trial with equal *a priori* probability. The procedures for stimulus presentation and feedback were the same as for experiment 1. For the AM conditions, the starting value of the modulation depth was 50%. Modulation depth varied by a factor of 2 for the first two reversals, by a factor of 1.26 for the next two reversals, and by a factor of 1.12 for the final six reversals. For the FM conditions, the starting value of the frequency excursion from the carrier was 31.6 Hz. With a 500-Hz carrier, this corresponds to a peak-to-peak frequency change of approximately 12.6%, or 2 semitones. The frequency excursion varied by a factor of 2 for the first two reversals, by a factor of 1.41 for the next two reversals, and by a factor of 1.12 for the final six reversals. In all cases, the threshold for a given run was defined as the geometric mean of the tracking variable at the last 6 reversal points.

The thresholds from the 2AFC procedure were used to set the modulation depths for the one-interval AM task and the peak-to-peak frequency changes for the one-interval FM task. During the one-interval task, only one tone was presented in each trial. The task was to determine whether or not the tone was modulated by pressing the corresponding button on the screen. Subjects were instructed to press the “yes” button if they believed that the tone was modulated, and the “no” button if they did not believe that the tone was modulated. As before, example trials were presented before each modulation condition to ensure that the participants were familiar with the specific target they were to detect for the following block of trials. Because the target modulation was always presented near threshold in the experimental trials, the modulation depth and peak-to-peak frequency excursion of the target was half of that from the previous practice trials. After completing the corresponding example trials, subjects were instructed to “listen carefully, as the modulation will be very subtle.” In the same manner as the two-interval task, modulation condition (AM vs. FM and modulation rate) was counterbalanced across participants using a Latin Square design.

Both the AM and FM one-interval tasks each involved a total of 175 trials per subject. In 20% of these trials the modulation depth was set at the individual’s threshold (i.e., the 79.4% correct point, based on the 2-interval adaptive procedure). In another 20% of the trials the modulation depth was above threshold by a factor of 1.59; in another 20% of trials the modulation depth was below the measured threshold by a factor of 1.58; and in another 20% of trials the modulation depth was set to be below threshold by a factor of 2.51. The remaining 20% of the trials contained unmodulated pure tones to provide an estimate of the false-alarm rate (catch trials). The presentation order of the tones with their modulation depths was random within each block of 175 trials.

5.3.2 Results

5.3.2.1 Comparing sensitivity to one- and two-interval FM and AM detection

Analyses were conducted to compare one- and two-interval FM detection as well as one- and two-interval AM detection. Signal detection theory was used to calculate the sensitivity index, *d′*, for each observer on the one-interval FM and AM tasks. Assuming equal variance of the distributions underlying the response to the modulated and unmodulated stimuli, *d′* provides a criterion-free measure of sensitivity (Macmillan and Creelman, 2005). Proportions of 0 or 1 (i.e., all or no correctly identified modulation trials or no-modulation trials) were converted to 1/(2*N*) and 1-1/(2*N*), where *N* is the total number of trials on which the proportion is based (35 in this case), as suggested by Macmillan and Creelman (2005, pg. 8).

Four *d'* values were calculated for each participant, one for each of the four modulation indices (see Fig. 5.2 for an example), and were plotted as a function of the modulation index (20log(*m*) and 2*Δf*(%) for AM and FM, respectively). The modulation index that corresponded to a *d'* of 1.14 (i.e., the sensitivity equivalent to the 79% correct point for a two-interval AFC task as used in the first-half of experiment 2) (Macmillan & Creelman, 2005, pg. 429) was estimated for each subject using linear regression. The fits for the linear regression were adequate; the mean and median R^2 for each modulation rate was always greater than 0.9.

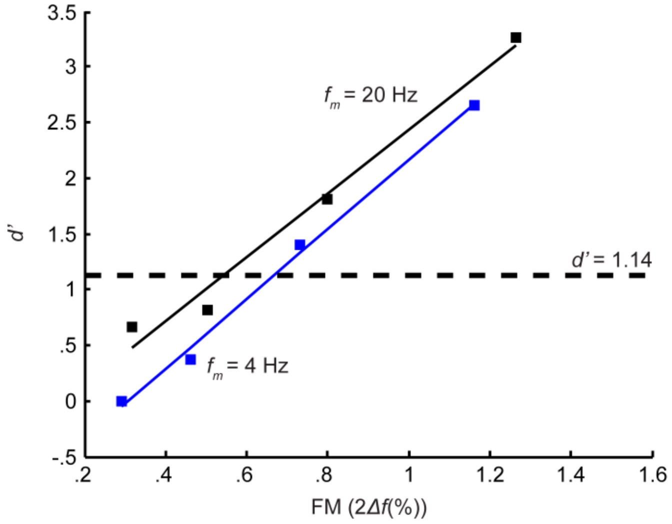

Figure 5.2

Example psychometric function for one-interval FM detection for one subject. The x-axis corresponds to the peak-to-peak frequency change, expressed as a percentage of the carrier frequency. The y-axis corresponds to *d'*, calculated based on the hit and false-alarm rates for each of the modulation indices. Solid lines correspond to the regression line, fitted to the four data points for each modulation rate (blue: f_m = 4 Hz; black: f_m = 20 Hz). Threshold was defined as the peak-to-peak frequency change corresponding to a *d'* of 1.14 (dashed line).

Results from both groups in the two- and one-interval FM and AM tasks are presented in Fig. 5.3 (see Fig. D1 for individual data and text in Appendix Text B2). Mixed-model ANOVAs were conducted on the log-transformed thresholds for both tasks, with modulation rate (4 and 20 Hz) and number of intervals (one vs. two) as the within-subjects factor and group (amusic and control) as the between-subjects factor. Any *p* values less than .05 were considered significant. The ANOVA for FM revealed a main effect of number of intervals [$F(1,22) = 9.34$, $p = .006$, $\eta_p^2 = .298$] and a main effect of group [$F(1,22) = 10.3$, $p = .004$, $\eta_p^2 = .319$], but no significant effect of modulation rate and no interactions (see Table 5.2A). Overall, amusics performed worse than controls at FM detection, and both groups performed better on the one- compared to the two-interval task. Similarly, the ANOVA for AM revealed a main effect of number of intervals [$F(1,22) = 48.5$, $p < .0001$, $\eta_p^2 = .688$] and a main effect of group [$F(1,22) = 14.2$, $p = .001$, $\eta_p^2 = .393$]. There was also a main effect of modulation rate [$F(1,22) = 5.88$, $p = .024$, $\eta_p^2 = .211$], and none of the interactions were significant (see Table 5.2B). The main effect of modulation rate for AM is consistent with previous literature on sinusoidal AM detection with gated carriers (Viemeister, 1979; Sheft and Yost, 1990; Moore and Sek, 1995; Whiteford and Oxenham, 2015), and has been attributed to the increase in the number of cycles for faster rates. Typically, the opposite effect is observed with sinusoidal FM detection (Demany and Semal, 1989; Moore and Sek, 1992, 1994, 1996; Whiteford and Oxenham, 2015), with an increase in performance at slower rates relative to faster rates. Our findings of differential trends of modulation rate for AM (modest effect) versus FM (no effect) are consistent with previous studies.

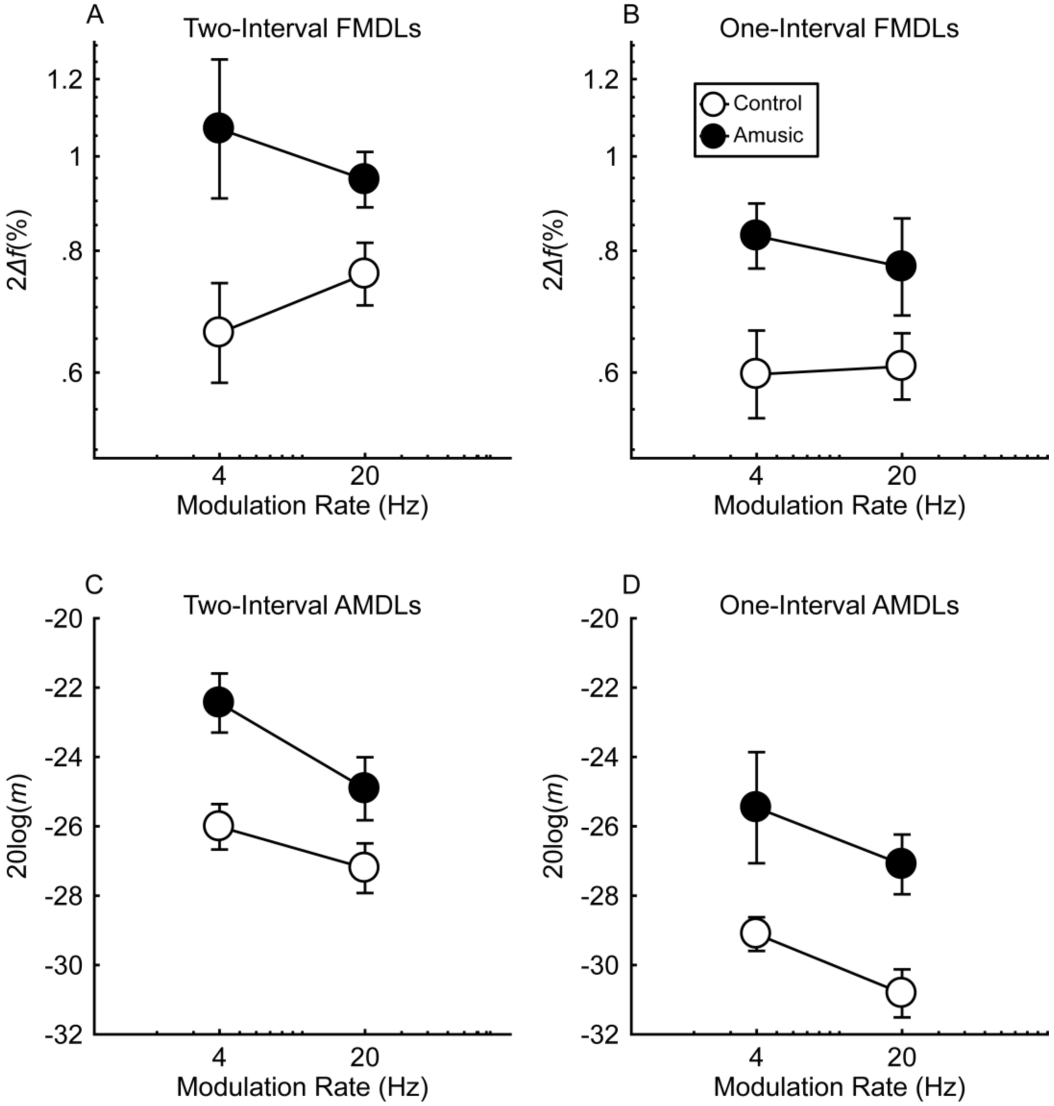

Figure 5.3

Average FM and AM thresholds for the two- versus one-interval tasks at slow (f_m = 4 Hz) and fast (f_m = 20 Hz) modulation rates. Closed circles represent amusics and open circles represent controls. The x-axes corresponds to modulation rate. The y-axis corresponds to the average (A-B) FM threshold, plotted as peak-to-peak frequency change as a percentage of the carrier frequency, and (C-D) AM threshold, plotted as modulation depth in logarithmic units. The left column plots thresholds for the two-interval tasks, while the right column plots the comparable thresholds for the one-interval tasks. Error bars represent ±1 standard error of the geometric mean.

Table 5.2

5.2A. FM Detection					
Effect	DFn	DFD	F	p	η_p^2
Group	1	22	10.3	**.004***	.319
Rate	1	22	.013	.912	.001
Int.	1	22	9.34	**.006***	.298
Group x Rate	1	22	1.26	.273	.054
Group x Int.	1	22	.317	.579	.014
Rate x Int.	1	22	.19	.667	.009
Group x Rate x Int.	1	22	.984	.332	.043
5.2B. AM Detection					
Effect	DFn	DFD	F	p	η_p^2
Group	1	22	14.3	**.001***	.393
Rate	1	22	5.89	**.024***	.211
Int.	1	22	48.5	**< .0001***	.688
Group x Rate	1	22	.173	.681	.008
Group x Int.	1	22	.773	.389	.034
Rate x Int.	1	22	.047	.831	.002
Group x Rate x Int.	1	22	.825	.374	.036

ANOVA results for comparing one- versus two-interval FM (A) and AM (B) detection thresholds. Group refers to amusics versus controls, Rate refers to slow (f_m = 4 Hz) versus fast (f_m = 20 Hz) modulation rates, and Int. to the number of intervals in a given task (one versus two). Significant effects are bolded and marked with *.

Further analyses of response bias (*c*) (Appendix Fig. C1) and overall sensitivity between amusics and controls on one-interval FM and AM detection can be found in Appendix C (Text C1 and C2). Notably, the overall sensitivity to FM and AM was equal between amusics and controls when calculating one *d'* across all modulation indices (see Appendix Fig. C2), as intended by setting the differences in FM and AM depth individually.

5.3.2.2 Comparing one-interval slow FM detection with three-interval frequency discrimination

To further assess the potential effect of memory load on pitch discrimination, performance in the one-interval slow FM-detection task (low memory load) was compared to performance in the three-interval frequency-discrimination task from experiment 1 (high memory load). More specifically, the (log-transformed) peak-to-peak frequency difference at threshold in the one-interval FM task was compared to the (log-transformed) FDL in the discrimination task. Signal detection theory provides a way to compare performance based on the measure of sensitivity, d', which should be independent of task. For the one-interval tasks, the modulation index that corresponded to a d' of 1.28 (the sensitivity equivalent to the 70.7% correct point for a two-interval AFC task) (Macmillan & Creelman, 2005, pg. 429) was estimated for each subject using linear regression. This was necessary to ensure that sensitivity for FM and sensitivity for AM corresponded to the same values of d' for both the one- and three-interval tasks. The slow rather than fast FM task was used because the detection of slow FM ($f_m < \sim 10$ Hz) is believed to rely on the same timing-based peripheral code as frequency discrimination of discrete tones (e.g., Sek and Moore, 1995). The task (FMDL vs. FDL) was the within-subjects factor, and group (amusic vs. control) was the between-subjects factor. All p values less than .05 were considered significant. Results indicated a main effect of task [$F(1,22) = 43.4$, $p < .0001$, $\eta_p^2 = .664$], a main effect of group [$F(1,22) = 12.1$, $p = .002$, $\eta_p^2 = .355$], and a significant interaction [$F(1,22) = 4.54$, $p = .045$, $\eta_p^2 = .171$]. Overall, slow FM thresholds were lower (better) than the FDLs; all subjects had better performance on the one-interval relative to the three-interval task (Fig. 5.4). Post-hoc tests using Bonferroni correction ($\alpha = .025$) indicated that amusics performed more poorly than controls on both measures [FDLs: $t(22) = -2.92$, $p = .004$, one-tailed; FMDLs: $t(22) = -3.06$, $p = .003$, one-tailed]. Therefore, the group by task interaction can be interpreted by viewing the differential trends in average performance for FDLs versus FMDLs in amusics and controls in Fig. 5.4. As is clear from Fig. 5.4, the degree of impairment was greater for the three-interval frequency discrimination task compared to the one-interval FM-detection task in amusics relative to controls.

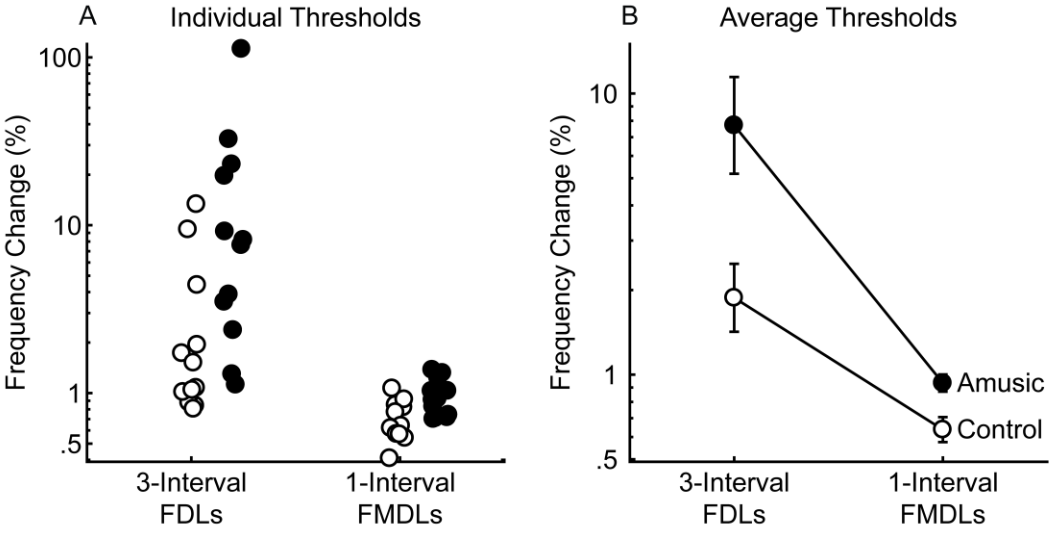

Figure 5.4

Individual (A) and average (B) performance for three-interval FDLs vs. one-interval FMDLs. The x-axis represents task type, with three-interval frequency discrimination (f = 500 Hz; experiment 1) closest to the y-axis and one-interval FM detection (f_c = 500 Hz and f_m = 4 Hz; experiment 2) furthest from the y-axis. The y-axis represents percent frequency change at threshold, where the FMDLs were transformed to peak-to-peak frequency change to be comparable to the FDLs. Error bars represent ±1 standard error of the geometric mean.

5.3.3 Discussion

The ability of amusics to detect FM in a one- or two-interval task was significantly poorer than that of controls at both slow (f_m = 4 Hz) and fast (f_m = 20 Hz) rates. This outcome supports the hypothesis that amusics have an underlying deficit in fine-grained frequency discrimination that extends beyond problems with memory or segmentation. However, the fact that the performance of amusics was equally degraded at FM rates of both 4 and 20 Hz is not, at face value, consistent with the suggestion that amusia results in particularly poor processing of rapid fine-grained pitch information (Albouy et al., 2016). If this were the case, one might expect an interaction between group and modulation rate, with poorer thresholds in amusics at faster modulation rates. It is possible, however, that the task design in the present study is too different from Albouy et al. to expect generalization, perhaps because the tones here are much longer and the changes over time are continuous, rather than discrete. Either poor rapid pitch encoding in amusics only applies to steady tones, and not FM, or the interaction Albouy et al. (2016) observed between group and tone duration reflected ceiling effects in the control group at the longer durations.

Surprisingly, AM detection thresholds were also worse in amusics relative to controls. Despite the fact that the task had low memory load requirements and did not involve any pitch-related changes, the amusics were still at a disadvantage. Tillmann et al. (2016) recently reported impaired reaction times for large intensity changes (20 dB) in amusics relative to controls, but these impairments were only present when gaps between tones were longer (1500 ms) as opposed to shorter (500 ms). These results suggest that the impaired reaction times found in Tillmann et al. could be a result of poorer intensity encoding, perhaps creating a slightly weaker memory trace even for large intensity changes. The present results of poor AM detection, however, was generally not expected, given that amusia is believed to be a pitch-specific deficit.

Trends for FM and AM detection were consistent in the two- and one-interval detection tasks. Overall, amusics performed more poorly at FM and AM detection tasks relative to controls, and their performance was not differentially affected by one- versus two-intervals. The lack of interaction between group and number of intervals for FM (or AM) detection was not unexpected, as the two-interval task could be completed using the same strategy as the one-interval task by, for instance, attending to just the first of the two intervals and determining whether or not it contained a modulated tone. Better performance across all subjects for the one- versus the two-interval tasks was modest for FM and moderate for AM, and could be related to training effects rather than memory, as the two-interval tasks always preceded the one-interval measures. Therefore, it is possible that learning occurred in both groups between the two modulation measures. For example, He et al. (2007) found the opposite effect for FMDLs, with better performance on three- vs. one-interval tasks, and they measured FMDLs in the opposite order as the present study (i.e., one-interval first). The general consensus is that amusia is a lifelong deficit, however, and performance on pitch related tasks cannot improve with training (Hyde and Peretz, 2004). This makes the ordering effect hypothesis unlikely, as there was no interaction between group and number of intervals or task type and number of intervals. It is also possible that one-interval tasks are comparatively easier for our relatively untrained listeners because there are fewer possible strategies that the listener could adopt or switch between, or because of an overall lighter memory load. Further research is needed to clarify this potential discrepancy.

Comparisons of one-interval FMDLs with three-interval FDLs indicated that both amusics and controls were worse in the three-interval discrimination task than would be predicted by performance in the one-interval FM detection task. This difference may be due to the additional memory load required in the three-interval task. Most importantly, however, the detrimental effect of three intervals, relative to one, was markedly greater for the amusics than for the control participants: with the three-interval task, amusics' performance was worse by a factor of 8.26 (mean FDL = 7.71%; mean FMDL = 0.934%), whereas controls were worse by only a factor of 2.94 (mean FDL = 1.88%; mean FMDL = 0.638%). In addition, the effect size of group for the one- and two-interval FM tasks was substantially smaller than that observed for the three-interval task. Thus, it seems that amusia reflects poorer basic coding of frequency (and amplitude), as well as a poorer ability to compare frequencies across time. Experiment 3 explores the latter deficit in more detail.

5.4 Experiment 3: Comparing Discrete and Continuous Changes in Frequency

One potential explanation for why amusics seem more severely affected by having to compare frequencies across time is that their memory trace degrades more rapidly over time than it does for normal controls. Another potential explanation relates to the fact that the changes in pitch in the three-interval task were discrete, or segregated, rather than continuous. Indeed, Foxton et al. (2004) found that for both amusics and controls, the ability to detect a change in pitch was worse for segmented versus gliding tones; however, the segmented tones had 100 ms gaps of silence between them, potentially increasing memory load and/or decreasing pitch salience. Liu et al. (2015) found that Mandarin speaking amusics had deficits in segmented speech processing in speech with a flattened F0 (no pitch changes), suggesting that processing segmented sequences may be impaired in amusia, independent of pitch processing. It is therefore possible that segmentation, rather than memory decay *per* *se*, may be responsible for this deficit. In the general population, the introduction of segmentation tends to elevate (worsen) pitch-discrimination and pitch-change detection thresholds (Sek and Moore, 1999; Lyzenga et al., 2004; Demany et al., 2009), perhaps because comparing pitch changes between two auditory objects is more difficult than tracking changes within one (continuous) object (Demany et al., 2009).

We attempted to distinguish between explanations based on memory decay and segmentation by measuring performance in a two-interval task that involved detecting discrete frequency changes, but using the same time relationships as were used in the two-interval FM detection task of experiment 2 (see Fig. 5.5). If the amusic deficit is due at least in part to difficulties produced by segmentation, then performance in the task with discrete changes should be worse than in the FM-detection task with continuous changes. On the other hand, if the amusic deficit is only due to a more rapid decay in memory regardless of segmentation, then performance in the two tasks should be similar, as they share the same overall duration. Note that a segmentation-related deficit in amusia would not rule out the memory decay hypothesis, as these two interpretations are not necessarily mutually exclusive. Segmentation could, theoretically, put a greater strain on memory resources (i.e., by creating the percept of multiple objects), even when the duration of the stimuli is held constant.

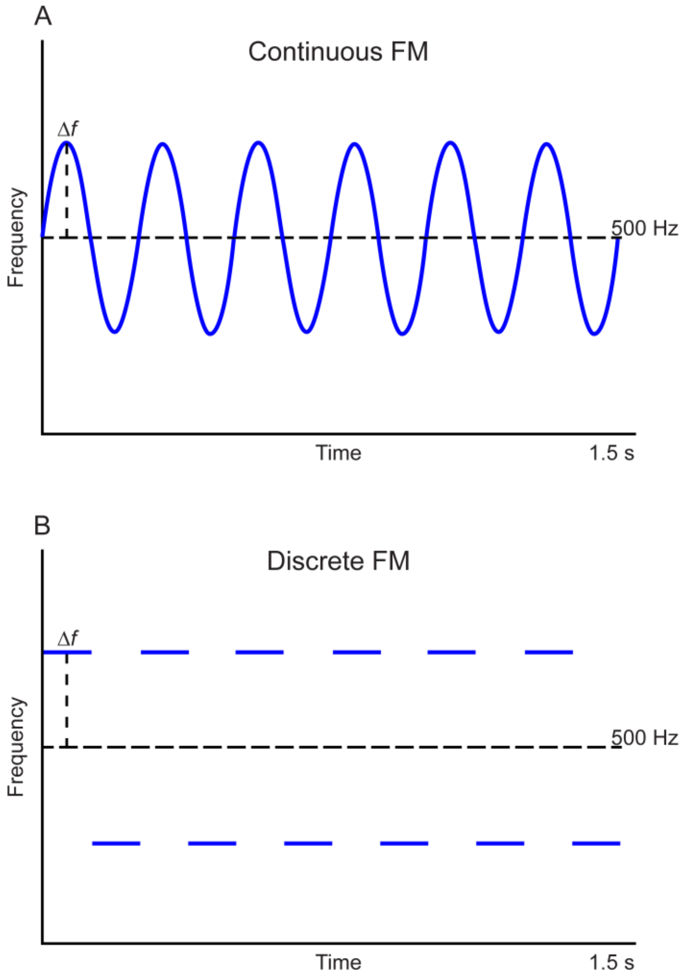

Figure 5.5

Schematic diagram of target stimulus from experiments 2 and 3: Continuous (A) vs. discrete (B) FM. The x-axis corresponds to time, while the y-axis corresponds to frequency. The horizontal dashed line represents the carrier frequency, 500 Hz. The vertical dashed line represents the frequency excursion from the carrier (*Δf*), varied adaptively in the same manner for experiments 2 and 3. Blue lines correspond to the target stimulus. The frequency change contours of the target and reference stimuli were the only differences between experiments 2 and 3.

5.4.1 Methods

5.4.1.1 Participants

Twenty subjects (10 amusics) participated in experiment 3. One of the amusics had not previously participated in experiments 1 or 2, while all other subjects had completed the first two studies. Subjects were matched on the same demographic variables as in experiments 1 and 2 (see Table 5.3). All except one of the amusics performed below the average cutoff of the pitch-based subtasks on the MBEA. Participants provided written informed consent and were compensated with course credit or hourly payment for their time.

Audiometric thresholds were measured at octave frequencies from 250 to 8000 Hz. All but one participant had NH up through 8 kHz, defined as audiometric thresholds no higher than 20 dB HL. One participant from the control group did not meet this criterion but had NH up through 4 kHz and a small loss at 8 kHz (right ear: 25 dB HL; left ear: 35 dB HL). A t-test (two-tailed) revealed no difference in the 8-kHz audiometric thresholds (averaged across ears) between the groups ($t(22) = -.986$, $p = .337$). It should be noted that the controls had a slightly but significantly poorer four-tone pure-tone average (average of audiometric thresholds between 500 and 4000 Hz) (Table 5.3).

Table 5.3

	Group		
	Amusics (n = 10)	**Controls (n = 10)**	***p*-value**
Age (years)	23.6 (10.9)	23.3 (9.6)	.946
Musical Experience (years)	1.55 (1.61)	1.3 (2.26)	.779
Education (years)	14.9 (1.35)	14.9 (1.61)	> .999
IQ	106.7 (10.7)	113.3 (10.5)	.181
Gender	7 females, 3 males	7 females, 3 males	-
Pitch MBEA (%)	65.78 (4.25)	84.67 (4.41)	**< .0001***
Global MBEA (%)	68.06 (4.63)	85.72 (3.66)	**< .0001***
Audiogram- 4 pure-tone average (dB HL)	3.06 (2.29)	6.5 (3.75)	**.024***

Demographic average data for the 10 amusic and 10 control participants from experiment 3. Standard deviations are shown in parentheses. Pitch MBEA scores represent the average percent correct on the pitch based subtasks of the MBEA. Global MBEA scores represent the overall average percent correct across all tasks on the MBEA. All MBEA percentages were transformed to rationalized arcsine units for statistical analyses (Studebaker, 1985). The 4 pure-tone average is the audiometric thresholds between 500-4000 Hz, averaged across frequencies and ears. All statistical tests involved independent-samples t tests. Significant effects ($p < 0.05$) are bolded and marked with an asterisk.

5.4.1.2 Stimuli and procedures

As in experiments 1 and 2, experiment 3 was administered in a sound-attenuating chamber via Sennheiser HD650 headphones at 60 dB SPL. All stimuli were generated in MATLAB using a 24-bit L22 soundcard (LynxStudio, Costa Mesa, CA) at a sampling rate of 48,000 Hz.

Detection of the segmented frequency changes was measured using a 2-interval, 2AFC adaptive procedure. The stimuli and procedures were designed to be analogous to the design from the two-interval, slow (f_m = 4 Hz) FM adaptive procedure in experiment 1. Thus, the target sequence contained frequency changes centered around 500 Hz, but the changes were discrete, rather than continuous. Both the reference and the target sequences were 1.5 s in duration and were comprised of 12 pure tones of 125 ms duration, alternating in frequency between ($500 + \Delta f$) and ($500 - \Delta f$) Hz. The reference sequence consisted of 12 sequential 500-Hz pure tones, also of 125 ms duration. Each individual tone within a sequence (target or reference) had 31.25-ms raised cosine onset and offset ramps, and there were no gaps in between tones within a sequence. The two sequences were separated from each other by a 500-ms gap. Subjects were instructed that they would hear two sequences of tones, one at a time, and their task was to select the sequence that sounded as if it were “changing.” Participants were reminded to attend to the feedback after each trial. All other aspects of the design and procedures were identical to the two-interval FM task from experiment 2. The new participant also completed the two-interval FM detection task for f_m = 4 Hz described in experiment 2.

5.4.2 Results

5.4.2.1 Discrete versus continuous FM detection

Results from the experiment are shown in Fig. 6.6, along with the replotted thresholds from the 2-interval 4-Hz FM detection task of experiment 2. A mixed-model ANOVA was performed on the log-transformed thresholds, with group (amusic or control) as an across-subjects factor and task type (segmented tones from this experiment or two-interval 4-Hz FM detection from experiment 2) as a within-subjects factor. For the purposes of this statistical analysis, the one new subject was treated as the direct replacement for the subject from experiment 1 who did not complete experiment 3. All p values less than .05 were considered significant. Consistent with experiment 2, there was a main effect of group [$F(1,18) = 8.41$, $p = .01$, $\eta_p^2 = .319$], with amusics performing significantly worse than controls. There was also a main effect of task type [$F(1,18) = 24.0$, $p = .0001$, $\eta_p^2 = .571$] and a significant interaction [$F(1,18) = 7.05$, $p = .016$, $\eta_p^2 = .281$]. All participants were worse at detecting frequency changes in segmented sequences compared to continuous changes in frequency (FM). Independent samples t tests indicated that amusics were significantly worse at continuous [$t(18) = -1.94$, $p = .034$, one-tailed] and segmented [$t(18) = -3.06$, $p = .004$, one-tailed] FM detection relative to controls, but the degree of impairment was greater in the segmented task (see Fig. 6.6). Because the segmented FM task was always measured after the continuous FM task, the current results may even underestimate the cost of segmentation for amusics if there was any learning between tasks.

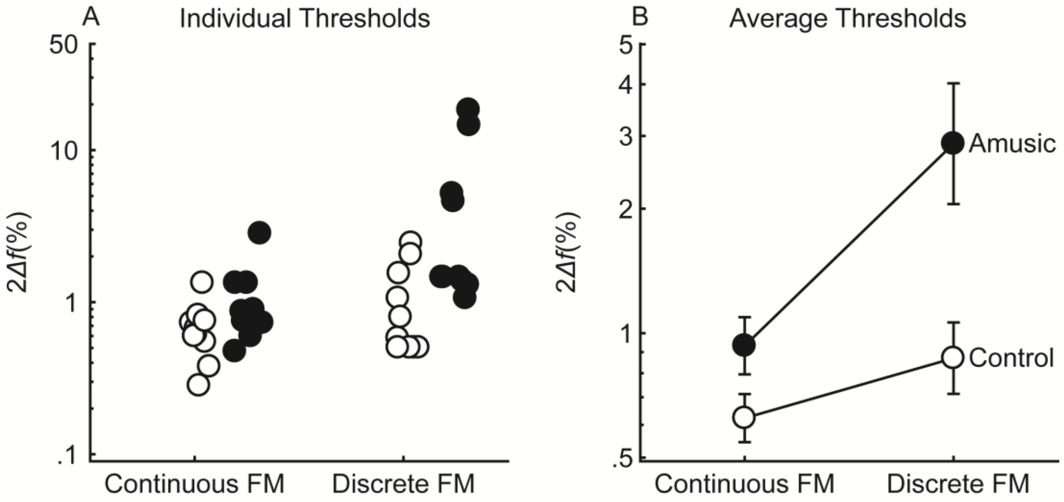

Figure 5.6

Continuous vs. discrete slow (f_m = 4 Hz) FMDLs. Black circles represent amusics and white circles represent controls. The x-axis corresponds to type of modulation (continuous vs. discrete). The y-axis corresponds to the average FM threshold, plotted in peak-to-peak frequency change. Error bars represent ±1 standard error of the geometric mean.

5.4.3 Discussion

The results from experiment 3 indicate that segmentation led to poorer performance for both groups, but the effect was more detrimental to amusics' thresholds than to those of the controls. Importantly, experiment 3 controlled for the number of intervals (2AFC) and tone duration (1.5 s) by keeping these factors constant between the continuous and discrete FM tasks. The results indicate that segmentation contributes to poor pitch perception in amusia, and the effect of segmentation cannot be accounted for by tone duration.

5.5 General Discussion

The purpose of this study was to provide insights into the mechanisms underlying the deficits in pitch perception observed in people with congenital or developmental amusia using well-controlled psychophysical paradigms. A number of previous studies have assessed pitch perception in amusia – according to Vuvan et al. (2015), as of November 2013, 43 unique articles have addressed this topic. However, none have assessed the perception of high frequencies, where both frequency discrimination (Moore, 1973; Micheyl et al., 2012) and melodic perception (Attneave and Olson, 1971; Oxenham et al., 2011) is poor under normal conditions. Furthermore, no studies have assessed how different methodological paradigms using a low or high memory load (e.g., one-interval vs. three-interval AFC) differentially affect pitch perception thresholds in amusics versus controls.

The main findings of this study were: (1) Amusic deficits in pitch discrimination persist even at very high frequencies (8 kHz), despite the lack of musical pitch perception at these high frequencies, even in normal individuals (experiment 1); (2) amusics have deficits in FM perception, suggesting that the deficit involves the representation and coding of frequency, rather than simply memory deficits (experiment 2); (3) the deficits in FM extend to AM as well, suggesting that amusia may involve deficits in fine-grained perception of dimensions other than frequency and pitch (experiment 2); and (4) the impaired ability to compare frequencies over time is not just a function of a more rapid memory decay over time, but appears to reflect a difficulty in comparing information across different auditory “objects” (experiment 3). Note that this effect of segmentation likely affects the nature of how the stimulus memory is stored, so memory decay and segmentation are unlikely to be mutually exclusive.

5.5.1 Poor high-frequency pitch perception

Results showed that amusics had poor pitch perception at low (500 and 2000 Hz) and high (8000 Hz) pure-tone frequencies. Poor perception at musically relevant frequencies coincides with classic findings suggesting that amusia is characterized by an underlying deficit in fine-grained pitch perception (e.g., Foxton et al., 2004; Peretz et al., 2002), with thresholds on the order of a semitone, depending on the particular paradigm used. Our findings suggest that the extent of the pitch perception deficit in part depends on the frequency where pitch perception is assessed, with thresholds significantly higher in both amusics and controls at 8000 Hz relative to the lower frequencies. Poor perception for high pure-tone frequencies in normal-hearing listeners is typically attributed to the use of a less-precise, rate-place code for frequency in the peripheral auditory system (Moore, 1973; Moore and Ernst, 2012). The main effect of frequency and the lack of group by frequency interaction we observed are consistent with this hypothesis. However, the main effect of group, with amusics nearly a factor of 2 worse at all frequencies relative to controls, suggests that central processing can also greatly limit pitch perception. The phenomenon that pitch perception is worse in amusics at 8 kHz, coupled together with recent findings of poor perception for resolved but not unresolved harmonics (Cousineau et al., 2015), supports the hypothesis that amusia is a deficit in central processing specific to spectro-temporal fine structure associated with coding pure tones and resolved harmonics.

The large inter-individual variance (see Fig. 5.1A) demonstrates that not all amusics have poor pitch perception, even when all but one of the amusics from experiment 1 were below cut-off on the pitch-based sub-tasks on the MBEA. This trend has been observed in numerous previous studies that have used adaptive paradigms (e.g., Foxton et al., 2004), reflecting the heterogeneity of the disorder (Vuvan et al., 2015). It is this heterogeneity, and the reliance of short-term or working memory in many pitch perception paradigms, that likely accounts for the possible effects of memory and/or segmentation observed in our sample. Indeed, the largest effect size of group was observed in the three-interval frequency discrimination task, where there were both memory and segmentation demands, while the smallest effect size of group was observed for one-interval continuous FM detection task, where the memory demands were minimized.

5.5.2 Deficits in one-interval, FM detection

One way to control for possible memory confounds in multiple-interval pitch perception paradigms, which require the comparison of pitch information across time, is to use one-interval paradigms. This is because the use of detection paradigms requires simply noticing a change in frequency, rather than comparing information across time or comparing information to a built-in, memory representation (e.g., Jesteadt and Sims, 1975). Even when controlling for possible memory and segmentation confounds, amusics appeared to have impaired pitch perception (measured via slow-rate and fast-rate FM detection) relative to controls. This indicates that amusics have an underlying pitch and frequency coding deficit that cannot be explained entirely by memory load or segmentation, but this effect is much smaller than observed in more traditional frequency discrimination paradigms (e.g., experiment 1). Furthermore, our results could suggest amusia is not a problem with processing rapid, fine-grained pitch information, as proposed by Albouy et al. (2016). This is because amusics were equally impaired at processing slow-rate vs. fast-rate FM. Albouy et al. (2016) found that amusics were impaired at detecting pitch changes in short (100 ms) but not long (350 ms) duration tones relative to controls. While the present study only used long duration FM tones (1500 ms), potentially limiting the comparison to Albouy et al. (2016), a given cycle of FM is much longer for slow-rate FM (250 ms when f_m = 4 Hz) compared to fast-rate FM (50 ms when f_m = 20 Hz). Findings from Albouy et al. (2016) can possibly be attributed to using constant stimuli methods, where all participants were presented with the same Δf (1 or 2 semitones), leading to near ceiling pitch-change detection performance in the controls for the longer-duration tones. However, it is also possible that the mechanism for encoding pitch is different for fast-rate FM compared to slow-rate FM (e.g., Demany and Semal, 1989; Moore and Sek, 1996, 1995, 1994, 1992) and low frequency (< ~4-5 kHz) steady tones (Moore, 1973; Moore and Ernst, 2012), complicating the comparison between the present study and Albouy et al. (2016).

5.5.3 Impairment in one-interval AM detection

Interestingly, amusics also had poorer AM detection thresholds than controls, which has not been previously observed. Cousineau et al. (2012) found that amusics could not perceive a difference between consonant and dissonant intervals, and that this deficit was due to an impairment in the perception of harmonicity (i.e., the frequency spacing of harmonics at integer multiples of F0 in a complex tone) but not amplitude modulated beats (i.e., a perception of “roughness” that occurs when two sinusoidal components fall within the same auditory filter). However, in Cousineau et al.’s experiment, the AM depth was always 100%, which was well above the threshold for amusics. Experiment 2 demonstrated poor AM detection when the changes in AM were more fine-grained, or near threshold with lower detectability – a considerably different process from AM at highly detectable depths, such as that used by Cousineau et al. (2012). A recent study examining slow (f_m = 1 Hz) and fast (f_m = 20 Hz) FM and AM detection at 500 Hz in a large cohort of young, normal-hearing listeners found that detection for near-threshold FM and AM was highly correlated, even across modulation type and modulation rate (Whiteford and Oxenham, 2015). One possible explanation for high multi-collinearity in modulation detection is that near-threshold FM and AM may use a shared, cortical rate-place population code (Micheyl et al., 2013). Micheyl et al. (2013) used mathematical modeling to combine human psychophysical data with animal physiological data and demonstrated that it was possible that the population of cortical neurons that code fine-grained changes in frequency may also code fine-grained changes in intensity. Evidence so far, however comes from frequency and intensity discrimination thresholds, and has not yet been applied to modulation detection tasks. If poor FM and AM in amusia is due to a shared cortical code, then amusics should also exhibit poor intensity discrimination in a paradigm similar to that used in experiment 1.

5.5.4 Auditory object perception

Pitch discrimination was poorer in amusics than in controls, whether the task involved high or low memory load or segregated versus continuous changes in frequency. However, the magnitude of the deficit varied substantially, with amusics performing most poorly in tasks with the highest memory load and segregated tones and performing best in tasks with the lowest memory load and continuous changes in frequency. To illustrate this difference, the average thresholds for amusics were a factor of 8.26 better (lower) in the one-interval FM detection task than in the three-interval frequency-discrimination task. Control participants also had lower thresholds in the one-interval FM task than in the three-interval frequency-discrimination task, but the improvement was only by a factor of 2.94 on average.

Data from amusic participants also illustrated a multiple-interval effect for non-pitch tasks, as illustrated by the one- versus two-interval AM detection, with worse performance on the two-interval task. However, the decrease in performance for the two-interval task relative to the one-interval task was no greater than that observed in the control group. Analogous results were found for the one- versus two-interval FM task. Because the two-interval modulation-detection tasks could be performed using the same strategy as the one-interval task (i.e., by just attending to the second tone in a given trial), it is unclear exactly why performance was elevated across all subjects on the 2-interval task. It could be that participants adopted a poorer listening strategy for 2AFC, imparting memory and segmentation demands when they were not necessary. Alternatively, the relatively better performance of participants in both groups in the one-interval tasks may be result of practice, as the two-interval task was always completed first. However, earlier attempts to train amusics on pitch and melody-related tasks have been unsuccessful (Hyde and Peretz, 2004; Mignault Goulet et al., 2012), casting some doubt on the potential effects of practice, particularly for the amusic group.

Matched controls in the present study and non-amusic individuals in previous studies also exhibit poorer pitch perception in the context of segmented versus non-segmented tones (Sek and Moore, 1999; Lyzenga et al., 2004; Demany et al., 2009). It is not clear what mechanism may be responsible for poor perception of segmented tones, although it is likely separate from pitch processing. Previous studies have suggested the presence of a frequency-change detection mechanism, active only in the presence of continuous frequency changes, but not active during the detection of segmented frequencies (Sek and Moore, 1999; Lyzenga et al., 2004). However, Demany et al. (2009) found that participants were worse at detecting continuous frequency changes that were segmented via changes in amplitude (i.e., the envelope was segmented) compared to the same continuous changes without amplitude segmentation. Their results demonstrate that segmentation effects are not specific to frequency coding, and instead suggest that segmentation could create the percept of multiple auditory objects, which in turn may require greater perceptual or attentional load to process than one continuous, auditory object.

5.5.5 Implications on the nature of amusia

The present study suggests multiple auditory deficits in amusia that extend beyond poor fine-grained pitch perception. However, results from all three experiments show extensive overlap between amusics and controls (Figs. 5.1A, B1, 5.4A, and 5.6A), showing that not all amusics have the same kinds of underlying psychophysical deficits, adding further evidence to the heterogeneity of the disorder (e.g., Vuvan et al., 2015). All amusics do, however, share the commonality of poor melody perception, as assessed with the MBEA. The combined findings suggest underlying perceptual deficits in amusics are typically small and vary across individuals. These issues become further compounded with added demands of memory and segmentation, and hence become quite apparent in more complex tasks, such as musical listening.

CHAPTER 6: LEARNING FOR PITCH AND MELODY DISCRIMINATION IN AMUSIA

Chapter 6 is reprinted from:

Whiteford, K. L., & Oxenham, A. J. (in press). Learning for pitch and melody discrimination in congenital amusia. *Cortex*.

Abstract

Congenital amusia is currently thought to be a life-long neurogenetic disorder in music perception, impervious to training in pitch or melody discrimination. This study provides an explicit test of whether amusic deficits can be reduced with training. Twenty amusics and 20 matched controls participated in four sessions of psychophysical training involving either pure-tone (500 Hz) pitch discrimination or a control task of lateralization (interaural level differences for bandpass white noise). Pure-tone pitch discrimination at low, medium, and high frequencies (500, 2000, and 8000 Hz) was measured before and after training (pretest and posttest) to determine the specificity of learning. Melody discrimination was also assessed before and after training using the full Montreal Battery of Evaluation of Amusia, the most widely used standardized test to diagnose amusia. Amusics performed more poorly than controls in pitch but not localization discrimination, but both groups improved with practice on the trained stimulus. Learning was broad, occurring across all three frequencies and melody discrimination for all groups, including those who trained on the non-pitch control task. Following training, 11 of 20 amusics no longer met the global diagnostic criteria for amusia. A separate group of untrained controls (n=20), who also completed melody discrimination and pretest, improved by an equal amount as trained controls on all measures, suggesting that the bulk of learning for the control group occurred very rapidly from the pretest. Thirty-one trained participants (13 amusics) returned to the lab one year later to assess long-term maintenance of pitch and melody discrimination. On average, there was no change in performance between posttest and one-year follow-up, demonstrating that improvements on pitch- and melody-related tasks in amusics and controls can be maintained. The findings indicate that amusia is not always a life-long deficit when using the current standard diagnostic criteria.

6.1 Introduction

Pitch is a psychological attribute of sound that helps us understand speech prosody and process melodies and harmony. It is well known that musicians tend to have better pitch perception compared to unpracticed non-musicians (Spiegel and Watson, 1984; Kishon-Rabin et al., 2001; Micheyl et al., 2006; Bianchi et al., 2015). While there may be some genetic predispositions relating to across-listener variability in pitch perception (Drayna et al., 2001), differences in pitch discrimination in laboratory settings appear to be highly related to training (Kishon-Rabin et al., 2001; Micheyl et al., 2006; Bianchi et al., 2015). For instance, Micheyl et al. (2006) found that professional, classically trained musicians had lower (better) pure-tone pitch discrimination thresholds than non-musicians by a factor of 6. Despite this initial disadvantage, the non-musicians required only 4-8 hours of laboratory training to perform on a par with professional musicians, providing evidence that pitch discrimination is highly plastic and improves rapidly with training.

Plasticity for pitch discrimination may not apply to a subpopulation of participants with congenital amusia, or “tone deafness”, a neurogenetic disorder in music perception (Peretz, 2016). Impaired music perception in amusia is believed to be at least partially related to an underlying deficit in fine-grained pitch perception (Ayotte et al., 2002; Peretz et al., 2002; Foxton et al., 2004; Tillmann et al., 2009; Liu et al., 2010; Vuvan et al., 2015), which cannot be explained by problems with hearing, peripheral coding, brain damage, or differences in intelligence (Peretz, 2001; Ayotte et al., 2002; Cousineau et al., 2015). Amusics typically report that they are “musically impaired,” and that they have difficulty recognizing familiar tunes without the lyrics, or recognizing when they or others sing out of tune (Peretz et al., 2003, 2008).

Evidence so far suggests that the neural correlates of these behavioral deficits involve cortical malformations in the right frontotemporal network, including the right inferior frontal gyrus and right superior temporal gyrus (for a review, see Peretz, 2016). Several studies have found increased grey matter in right inferior frontal gyrus and right superior temporal gyrus in amusics relative to a group of matched controls (Hyde et al., 2006, 2007), as well as decreased connectivity between these two areas (Loui et al., 2009; Hyde et al., 2011; Albouy et al., 2013a, 2015b; Lévêque et al., 2016). These results support the hypothesis that amusia is related to a dysfunctional frontotemporal network, possibly resulting from abnormal cortical migration early in development (Hyde et al., 2007; Peretz, 2016).

Amusia is regularly described as a life-long deficit, perhaps because the few attempts to train amusics on pitch or music-related tasks have generally been unsuccessful (Hyde and Peretz, 2004; Mignault Goulet et al., 2012; Peretz et al., 2012; Wilbiks et al., 2016; Liu et al., 2017) or have been limited to vocal production (Anderson et al., 2012). However, most studies so far have either used passive listening tasks or have employed very short training schedules. Furthermore, none of the aforementioned studies have specifically trained amusics on a simple pitch-discrimination task.

The purpose of this study was to assess whether amusics can improve their pure-tone pitch discrimination with training and, if so, whether this training generalizes to untrained frequencies and melody discrimination. Participants completed an adaptive psychophysical training paradigm over four separate sessions, with pitch and melody discrimination assessed before and after training. Contrary to the long-held assumptions, both amusics and controls improved their pitch and melody discrimination by similar amounts. Learning was rapid, occurred for both trained and untrained stimuli, and was maintained over a one-year period.

6.2 Materials and Methods

6.2.1 Experiment 1

6.2.1.1 Participants

Forty participants took part in the training; half had congenital amusia, while the other 20 participants were age-matched controls (see Table 6.1). One control participant’s data were excluded from all analyses due to failure to complete the tasks on sessions 3-5; that participant’s demographics are not included in Table 6.1. All participants completed an initial laboratory screening to qualify for the study. The screening included a short questionnaire (adapted from Peretz et al., 2008), an audiometric assessment (measured at octave frequencies between 250 Hz and 8 kHz), the standardized diagnostic test for amusia (the Montreal Battery of Evaluation of Amusia, or MBEA; Peretz et al., 2003), and an IQ test that used the Vocabulary and Matrix Reasoning subtests from the Wechsler Abbreviated Scale of Intelligence –

Second Edition (Wechsler, 2011). Participants were required to have American English as their first language (due to the vocabulary component of the IQ test) and have no reported history of neurological conditions or hearing disorders. To qualify for the amusic group, participants had to perform below both the pitch and global cutoffs from the MBEA (Peretz et al., 2003), while controls had to perform above both of these cutoffs. All participants completed the full MBEA in the lab and were not exposed to any of the MBEA stimuli prior to testing. Those who met the screening criteria were invited to participate in the training portion of the study; the pretest was scheduled at the earliest convenience of the participant, with the constraint that all of the five following sessions (day 1: pretest and first training session; days 2-4: training sessions 2-4; day 5: posttest and second MBEA test) had to be completed within a 2-week period. The median time interval between the initial MBEA screening (test 1) and the pretest is provided in Table 6.1. The first 10 qualifying amusic participants were assigned to the pitch training task, and the second set of 10 qualifying amusic participants were assigned to the interaural level difference (ILD) training task, a localization control task that does not involve exposure to or discrimination of pitch. Controls were assigned to a training task based on the corresponding task of the closest age-matched amusic participant. All participants provided written informed consent and were compensated with hourly payment for their time. The experimental protocols were approved by the Institutional Review Board of the University of Minnesota.

Table 6.1

Average demographics for Pitch-Training (n=20) and ILD-Training (n=19) participants

	Pitch Training			**ILD Training**		
	Amusics (n=10)	**Controls (n=10)**	***p*-value**	**Amusics (n=10)**	**Controls (n=9)**	***p*-value**
Age (years)	24.9 (11.8)	25.7 (13.2)	0.888	34.5 (15.8)	34.2 (16.4)	0.97
Musical Experience (years)	1 (1.56)	1.8 (2.82)	0.443	1.3 (1.34)	0.778 (1.39)	0.416
Education (years)	15.45 (1.04)	15.4 (2.56)	0.954	16.3 (1.64)	16.6 (2.13)	0.771

IQ	105.3 (9.96)	111.7 (8.11)	0.132	112.8 (8.27)	114.6 (7.49)	0.635
Gender	6 females	5 females	-	6 females	6 females	-
Pitch MBEA (%)	61.78 (6.63)	85.78 (3.7)	**< 0.0001***	63.78 (4.89)	83.83 (6.51)	**< 0.0001***
Global MBEA (%)	65.83 (6.03)	86.67 (3.68)	**< 0.0001***	66.56 (7.69)	87.04 (4.53)	**< 0.0001***
Gap (Days)	7.5 (10.35)[a]	8.5 (10.65)[a]	0.909[b]	9.5 (10.65)[a]	4 (9.28)[a]	0.954[b]
Audiogram-Low (dB HL)	6.38 (3.65)	6.13 (2.53)	0.861	5.38 (5.27)	5.42 (4.42)	0.985
Audiogram-Medium (dB HL)	4.38 (3.02)	4.25 (4.42)	0.942	5.13 (6.19)	5.56 (4.81)	0.869
Audiogram-High (dB HL)	5.25 (5.74)	7 (7.1)	0.552	10.1 (12.7)	8.75 (8.97)	0.791

dB = decibel; HL = hearing loss

Values are means with standard deviations in parentheses unless noted otherwise.

P-values = independent-samples t tests, unless noted otherwise.

Gap = Median difference (in days) between initial MBEA screening (test 1) and the pretest.

Pitch MBEA = average percent correct on the three pitch-subtasks on the MBEA.

Global MBEA = average percent correct across all 6 subtasks on the MBEA.

All MBEA percentages were transformed to rationalized arcsine units (RAUs) for statistical analyses (Studebaker, 1985). Audiometric thresholds were grouped into three frequency bands (Low: .25 and .5 kHz; Medium: 1 and 2 kHz; High: 4 and 8 kHz) and averaged across ears.

[a] Mean rank

[b] Wilcoxon rank-sum test

6.2.1.2 MBEA

The MBEA consists of six subtasks (30 trials per subtask, 180 trials total) that evaluate various aspects of melody perception including pitch, rhythm, meter, and memory. The test was presented to participants according to the methods described by Peretz et al. (2003). Four of the subtasks (three pitch-related and one rhythm subtask) involved listening to two short melodies, one at a time, and determining whether they were the same or different. Different melodies were either different by one note (pitch subtasks) or had a rhythmic change (rhythm subtask). The meter subtask required participants to listen to a series of melodies, one at a time, and identify each one as either a waltz or a march. The memory subtask was always presented last, and participants had to indicate whether each of a series of melodies had been presented in the previous tests or not. A full description of the MBEA methods can be found in Peretz et al. (2003).

6.2.1.3 Pretest and posttest stimuli

Pure-tone pitch discrimination was assessed at 500, 2000, and 8000 Hz. Stimuli were 200-ms long, including 50-ms raised-cosine onset and offset ramps, and were presented at an rms level of 60 dB SPL in their steady-state portions.

6.2.1.4 Training task stimuli

Participants in the pitch-training task trained on pure-tone pitch discrimination at 500 Hz only (Fig. 1A). The other half of the participants trained on the discrimination of ILDs (Fig. 1B), a task that is not impaired in amusia (Cousineau et al., 2015) but is also highly susceptible to training in normal listeners (Wright and Fitzgerald, 2001; Wright and Zhang, 2009). The ILD-training task involved detecting an ILD within a white Gaussian noise, bandpass filtered between 20 and 4000 Hz. The ILD task stimulus was a pair of contiguous 100-ms bursts of white noise. Each burst within a pair had 50-ms raised-cosine onset and offset ramps. The rms level of each reference noise burst in each ear was 60 dB SPL at the temporal center of the noise. The target pair always had opposing ILDs imposed: for the first burst, the noise in one ear increased in level by $\Delta L/2$ dB while the noise level in the contralateral ear decreased by $\Delta L/2$ dB, relative to the reference noise burst; for the second burst, the ILD was reversed. The percept of the ILD target was a lateralized stimulus that moved from one ear to the other, while the reference stimuli remained centered. Whether the noise moved from left to right or right to left was randomized on each trial.

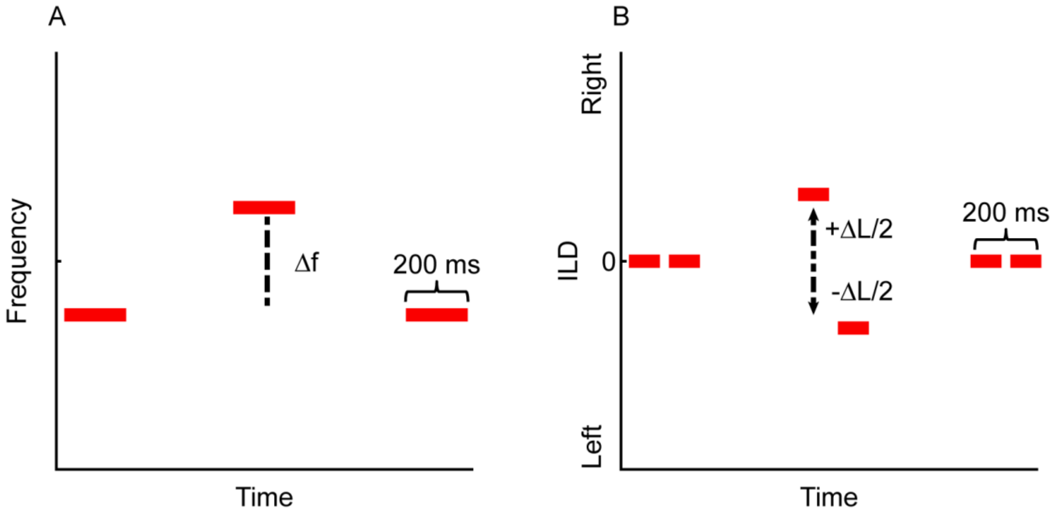

Figure 6.1

Schematic representation of stimuli for pitch training (A) and ILD training (B). Instructions for both training tasks were to pick the interval that was different. The target in the pitch training task was always higher in frequency, while the target in the ILD training task moved from one ear to the contralateral ear.

6.2.1.5 Procedures

All pretest, posttest, and training tasks used two-down one-up adaptive three-interval, three-alternative forced-choice procedures that tracked the 70.7% correct point of the psychometric function (Levitt, 1971). On each trial, three tones (or three pairs of noise bursts) were presented, one at a time, and each tone (or pair of noise bursts) was separated by a 500-ms inter-stimulus interval. The two reference tones (or pairs of noise bursts) were either identical in frequency (for pitch training) or had no ILD (for localization training). The target was always higher in frequency or had an ILD imposed, and the presentation order of the target and reference tones in a given trial was randomly selected. Participants were instructed to pick which of the three intervals was different by clicking the corresponding virtual button on a computer screen, labeled “1”, “2”, and “3”. Trial-by-trial feedback was provided as text on the computer screen (“Correct” or “Incorrect”), and participants were reminded at the beginning of each session to look at the feedback after each trial.

For the pitch tasks, the target and reference frequencies in a given trial were geometrically centered on the test frequency. Each adaptive track began with a frequency difference (Δf) of 20% between the target and reference tones. The Δf changed by a factor of 2 for the first two reversal points, a factor of 1.41 for the third and fourth reversal points, and a factor of 1.12 for the last six reversal points. The frequency difference limen (FDL) for each run was calculated as the geometric mean of the Δf values at the final 6 reversal points.

For the ILD tasks, each adaptive run in the ILD condition began with a ΔL of 3 dB. The ΔL varied by a factor of 1.41 for the first two reversal points, a factor of 1.19 for the following two reversal points, and a factor of 1.06 for the final 6 reversals. The threshold ILD for each run was calculated as the geometric mean value of ΔL at the last six reversals.

The pretest was conducted in a separate session after completing the screening, and consisted of two runs at each of the three frequencies (500, 2000, and 8000 Hz), with the frequency order randomized between subjects. Directly following the pretest, and in the same session, participants completed the first of four training sessions. Each training session consisted of 30 runs of the corresponding training task. Participants were encouraged to look at the feedback on the screen and to take breaks whenever they felt fatigued. Each training session lasted 1-2 hours.

The posttest occurred on a separate, fifth session, using the same method as the pretest. Directly following the posttest and in the same session, participants completed the MBEA for a second time (test 2) to reassess melody discrimination.

6.2.1.6 Equipment

The auditory tasks were administered in a sound-attenuating chamber through Sennheiser HD650 headphones. The MBEA tasks were administered at 70 dB SPL in each ear, consistent with Peretz et al. (2003). Pretest, posttest, and training stimuli were generated in Matlab and presented via a 24-bit L22 soundcard (LynxStudio, Costa Mesa, CA) at a sampling rate of 48 kHz. The same equipment was used in this and all subsequent experiments.

6.2.2 Experiment 2

6.2.2.1 Participants

Twenty additional controls were recruited, matched in age, years of musical experience, IQ, audiometric thresholds, and MBEA performance to the 19 controls from experiment 1 (Appendix Table D1A). These participants were recruited after the conclusion of data collection from experiment 1, in order to assess the amount of learning that occurs from the pretest and MBEA alone. No participants reported a history of neurological conditions. Participants provided written informed consent and were compensated with hourly payment for their time. The experimental protocols were approved by the Institutional Review Board of the University of Minnesota.

6.2.2.2 Stimuli and procedures

The pretest occurred directly after completing the initial screening, so there was no gap between the first MBEA and pretest. Pre and posttest stimuli and procedures were identical to experiment 1, with a minimum of 4 days and a maximum of 14 days between pretest and posttest. No training occurred between pre and posttest and participants were instructed not to partake in any additional psychophysical studies until after completion of the posttest. Directly following the posttest and during the same session, participants completed the MBEA for the second time to re-assess melody discrimination.

6.2.3 Experiment 3

The purpose of experiment 3 was to assess whether or not learning from experiment 1 could be maintained over a period of one year. All trained participants from experiment 1 were invited to return to the lab about one year after posttest to reassess pitch and melody discrimination.

6.2.3.1 Participants

Of the 40 participants from experiment 1, 31 (13 amusics) returned to take part in the follow-up study. Demographic information from these participants is presented in Appendix Table D1B. Participants were provided monetary compensation for their time.

6.2.3.2 Stimuli and procedures

Stimuli and procedures for pitch discrimination and the MBEA were identical to those from Day 5 of experiment 1, with participants first completing pitch discrimination at all three frequencies, followed by the MBEA.

6.2.4 Statistical analysis

One amusic in the ILD group was removed from pretest and posttest pitch discrimination analyses for both experiments 1 and 3 due to hearing loss at 8 kHz (50 dB HL in both ears in experiment 1; 60 dB HL in both ears at one-year follow-up). All analyses were conducted on the log-transformed data [$10\log_{10}(\%\Delta f)$ and $10\log_{10}(\Delta L)$] or rationalized arcsine units (RAUs) to better approximate normality (Studebaker, 1985). Mixed-design ANOVAs with repeated measures were conducted in SPSS Version 22 (IBM) using Type III sums of squares, as this method is not affected by unbalanced designs. Effect sizes were determined using partial-eta-squared (η_p^2) for the ANOVA outcomes and Cohen's d (or d_z) for the post-hoc comparisons (where d_z denotes the effect size for within-subjects comparisons; Lakens, 2013). Greenhouse-Geisser correction was applied where Mauchly's test of sphericity was violated. The homogeneity of variances assumption was tested using Levene's test, as well as by inspecting residuals of the individual data. In cases where this assumption was violated, non-parametric bootstrap analyses were conducted in Matlab 2016b (The Mathworks, Natick, MA) by sampling the control data with replacement for 100,000 iterations using the datasample function. The 95% confidence intervals (CIs) of the bootstrapped control distributions were calculated using the percentile method (Efron and Tibshirani, 1998), and amusic thresholds were compared to the corresponding control CIs. Global MBEA performance consisted of proportion correct across all six subtasks of the MBEA (180 trials total), whereas pitch MBEA performance consisted of proportion correct across the three pitch subtasks only (90 trials total). All MBEA scores were transformed to RAUs for analyses.

6.3 Results

6.3.1 Experiment 1

6.3.1.1 Training

Across-subject average results from the 4-day training sessions are plotted in Fig. 6.2A-B (see Fig. G1 for individual data). Results from the pitch training task were analyzed using a mixed-model ANOVA with repeated measures, with time as a within-subjects factor and group as a between-subjects factor. There was a main effect of time ($F_{1.82,32.8} = 16.5$, $p < 0.0001$, $\eta_p^2 = 0.479$) and a main effect of group ($F_{1,18} = 8.82$, $p = 0.008$, $\eta_p^2 = 0.329$), but no time x group interaction ($F_{1.82,32.8} = 0.104$, $p = 0.885$, $\eta_p^2 = 0.006$). *Post hoc* pairwise comparisons using Bonferroni correction for six multiple comparisons ($\alpha = 0.0083$) showed that performance on Day 1 was significantly worse than all other days ($p \leq 0.0001$ in all cases, $1.11 \leq d_z \leq 1.32$), and that performance did not significantly change after Day 2 ($p > 0.05$ in all cases, $0.12 \leq d_z \leq 0.466$). Training on pitch discrimination improved performance in both amusics and controls, but amusics were still worse than controls across all 4 days of training.

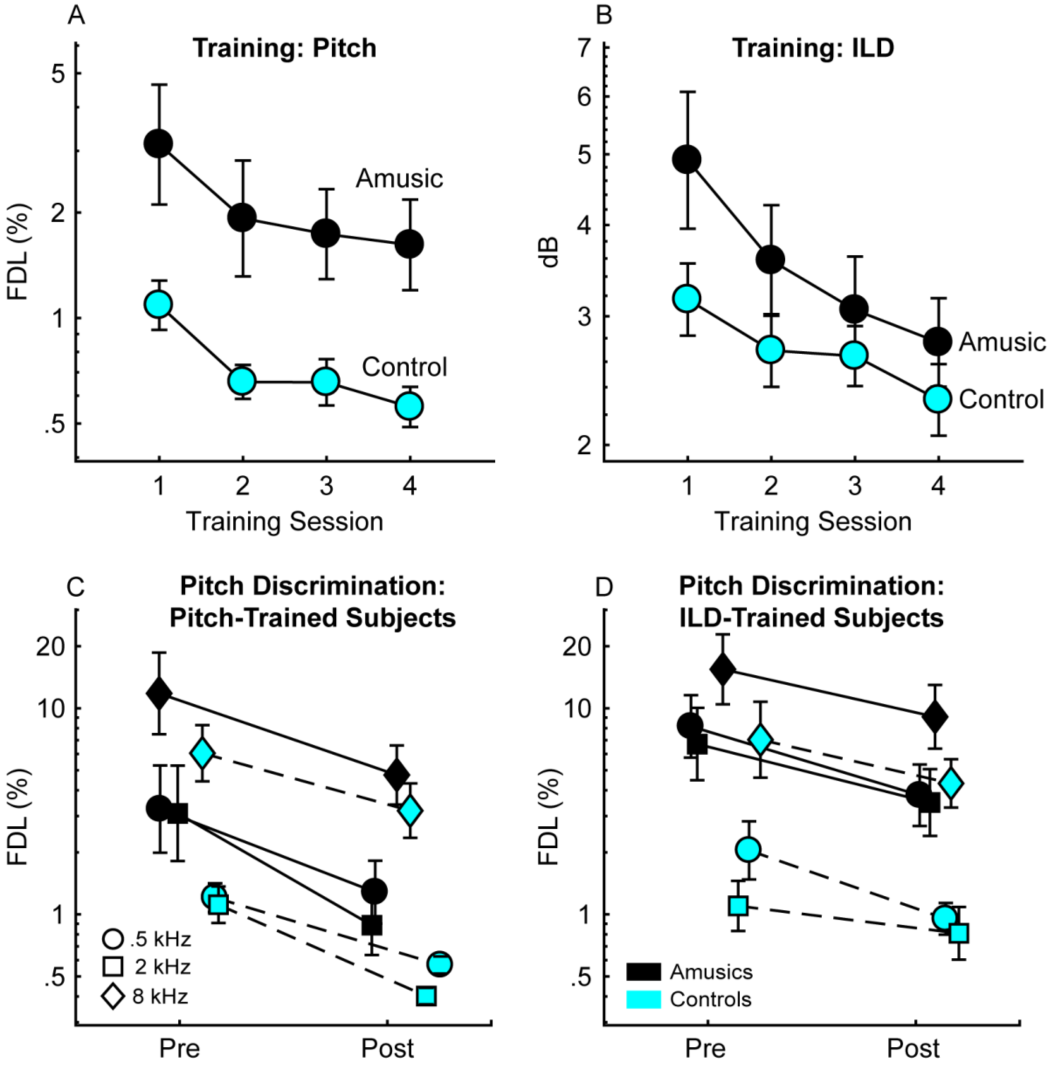

Figure 6.2

Training for 500-Hz pure-tone pitch (A) and white-noise ILD (B) discrimination and pre vs. posttest FDLs for pitch-trained (C) and ILD-trained (D) participants. In panels A and B, circles represent the mean of each group (black: amusics; cyan: controls). Error bars represent ± 1 standard error of the geometric mean. Both amusics and controls improved their pitch (A) and ILD (B) discrimination with practice. In panels C and D, circles correspond to 500 Hz, squares to 2000 Hz, and diamonds to 8000 Hz. Panel C plots average thresholds for participants who trained on 500-Hz pitch discrimination, while panel B plots the averages for those who trained on ILD discrimination. All groups, including amusics, improved their pitch discrimination posttest.

Data from the ILD training task were also analyzed using a mixed-model ANOVA. There was a main effect of time ($F_{3,51} = 22.8$, $p < 0.0001$, $\eta_p^2 = 0.573$) but no main effect of group ($F_{1,17} = 1.68$, $p = 0.212$, $\eta_p^2 = 0.09$). Unlike the pitch task, there was a marginally significant time x group interaction ($F_{3,51} = 2.78$, $p = 0.05$, $\eta_p^2 = 0.141$). To interpret this interaction, post-hoc comparisons were conducted, comparing the performance of amusics and controls on each session of training. There was a trend for amusics to perform worse than controls on Day 1 of training ($p = 0.05$, one-tailed, $d = 0.814$), but this trend was not significant once correcting for four multiple comparisons ($\alpha = 0.0125$). No other trends approached significance ($p > 0.1$, $0.341 \leq d \leq 0.617$).

To further interpret the time x group interaction, a simple effects analysis was conducted comparing performance in each group over time (Bonferroni-corrected $\alpha = 0.0083$). Controls had no significant learning between Days 1 and 2 ($p = 0.044$, $d_z = 0.724$) or between Days 1 and 3 ($p = 0.081$, $d_z = 0.618$), and a significant difference in performance between Days 1 and 4 ($p = 0.008$, $d_z = .997$). There was no difference in thresholds between all other comparisons (all $p \geq 0.028$, $0.092 \leq d_z \leq 0.799$). For amusics, the bulk of learning occurred between Days 1 and 2 ($p = 0.0003$, $d_z = 1.42$). There was a slight trend for improved localization thresholds between Days 2 and 3 ($p = 0.019$, $d_z = 0.818$) and 2 and 4 ($p = 0.003$, $d_z = 1.11$), but performance between Days 3 and 4 did not approach alpha-corrected significance ($p = 0.078$, $d_z = .592$). The time x group interaction can be explained by the generally worse performance of amusics on Day 1 of ILD training (Fig. 6.2B), although this should be interpreted cautiously, given that the effect was non-significant once correcting for multiple comparisons, and a previous study using a similar paradigm (but longer stimuli and no training) found no difference in ILD discrimination between amusics and controls (Cousineau et al., 2015).

6.3.1.2 Pretest versus posttest

Pretest and posttest pitch discrimination thresholds for all groups are plotted in Fig. 2C-D. The individual data (Appendix Fig. D2) illustrate a considerable degree of overlap between amusics and controls at all three frequencies, with amusics generally worse than controls, consistent with previous studies (e.g., Foxton et al., 2004; Vuvan et al., 2015; Whiteford & Oxenham, 2017). Average results (Fig. 2C-D) show that FDLs were generally lower (better) in the posttest than in the pretest, confirm that amusics' thresholds are generally higher than those of controls, and confirm that FDLs at 8 kHz are considerably higher than those at 500 or 2000 Hz (e.g., Moore, 1973; Moore and Ernst, 2012; Whiteford and Oxenham, 2017b). A mixed-model ANOVA with repeated measures was conducted with training task and group as between-subjects factors and time and frequency as within-subjects factors. There were significant main effects of time ($F_{1,34} = 61.5$, $p < 0.0001$, $\eta_p^2 = 0.644$), frequency ($F_{1.6,54.5} = 133$, $p < 0.0001$, $\eta_p^2 = 0.796$), group ($F_{1,34} = 12$, $p = 0.001$, $\eta_p^2 = 0.267$), and training task ($F_{1,34} = 4.2$, $p = 0.048$, $\eta_p^2 = 0.11$). There was also a significant frequency x group interaction ($F_{1.6,54.5} = 5.4$, $p = 0.012$, $\eta_p^2 = 0.137$), likely reflecting the more similar thresholds between amusics and controls at 8 kHz, but no other interactions reached significance (Appendix Table D2). Overall, all groups equally improved their pitch discrimination over time, including amusics and participants trained on ILDs. The ILD-trained participants performed more poorly than the pitch-trained participants, perhaps because the ILD participants were significantly older on average ($p = 0.0497$, two-tailed). Because there was no interaction with training task, the main effect of task can be attributed to group differences unrelated to training.

Because the sample size was relatively small given the number of conditions tested (10 participants per each between-groups condition, except for the ILD-training controls, in which there were 9 participants), and because the homogeneity of variances assumption was violated for several conditions (pretest 2000 Hz: $p = 0.048$; posttest 500 Hz: $p = 0.046$; posttest 2000 Hz: $p = 0.008$), non-parametric bootstrap analyses were conducted to confirm the effect of group (amusic vs. control) at each frequency and the lack of interaction between group and time. Pretest and posttest FDLs were pooled across training task in order to increase power, so that group (amusic vs. control) was the only between-subjects variable. Note that in order to examine effects of training task (pitch vs. ILD) on pitch discrimination in amusics and controls, separate bootstrap analyses would need to be run for pitch- and ILD-trained groups. However, the present sample size is too small to examine both between-subjects (group and training task) factors, so only the primary between-subjects factor of interest (group) was examined for each time point and frequency. Data from the 19 control participants were resampled with replacement for 100,000 iterations. For each iteration, mean FDLs were calculated across the 19 resampled participants for their corresponding thresholds at all three frequencies (500, 2000, and 8000 Hz) at both time points (pretest and posttest). The probability distributions of the mean resampled control FDLs at pretest are plotted in Fig. 3A-C along with 95% CIs (red-dashed lines). The change in FDL between pretest and posttest (FDL ratio) was calculated as the difference between the average pretest and posttest log-transformed FDLs for each of the resampled groups at each of the three frequencies (Fig. 3D-F). As expected, the mean of the actual (not resampled) control data (blue-dashed line) always fell within the 95% CIs of the control distributions for both the pretest and the FDL ratio distributions. For all three frequencies at pretest, including 8 kHz, the mean FDLs across the 19 amusics (black-dashed lines) were outside the 95% CIs of the control distributions. This is consistent with previous findings that amusics have higher (poorer) pure-tone FDLs relative to controls at both low and high frequencies (Whiteford and Oxenham, 2017), although the effect at 8 kHz appears to be smaller than at .5 kHz and 2 kHz. The amount of learning between pretest and posttest, however, was no different between the amusics and the controls for either 500 or 8000 Hz, as the amusic results fell within the 95% CIs of the resampled control distributions. The amusics learned slightly more than controls at 2 kHz, as their FDL ratio was higher than the upper cutoff of the control’s 95% CI. Overall, results from the non-parametric bootstrap analyses were consistent with the parametric analyses, suggesting that amusics performed more poorly than controls at all three frequencies, the difference between groups at 8 kHz seemed smaller than at .5 kHz or 2 kHz, and the amount of learning over time in the amusics was the same or slightly greater than in the controls.

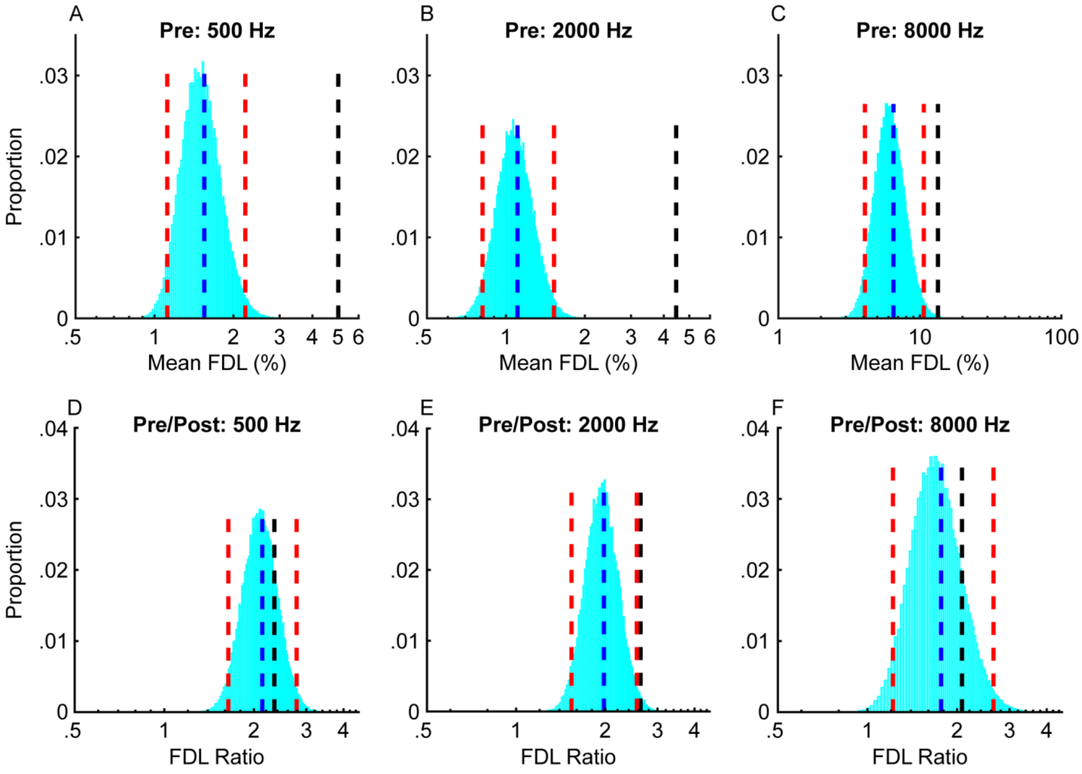

Figure 6.3

Bootstrapped control probability distributions for mean pretest performance (A-C) and the change in mean performance between pretest and posttest (D-F). In panels A-C, the x-axis represents the mean FDLs (%Δf) of the resampled data. In panels D-F, the x-axis represents the difference between the log-transformed mean pretest and posttest FDLs (FDL ratio), transformed back to a linear scale; 1 corresponds to no learning, values greater than 1 to improved performance at posttest, and values less than 1 to worse performance at posttest. For all panels, the y-axis corresponds to proportion. Blue-dashed lines correspond to the actual control data, black-dashed lines to the amusic data, and red-dashed lines to 95% CIs.

6.3.1.3 MBEA

Average and individual MBEA scores for both the pretest and posttest are plotted in Fig. 4. A mixed-model ANOVA for global MBEA scores (converted to RAUs) was conducted with time as a within-subjects factor and group (amusic vs. control) and task (pitch vs. ILD training type) as between-subjects factors. There was a main effect of time ($F_{1,35} = 100.9$, $p < 0.0001$, $\eta_p^2 = 0.742$), a main effect of group ($F_{1,35} =$

92.2, $p < 0.0001$, $\eta_p^2 = 0.725$), and a time x group interaction ($F_{1,35} = 5.91$, $p = 0.02$, $\eta_p^2 = 0.145$). No other main effects or interactions reached significance (Table D3). Both controls and amusics significantly improved their MBEA scores after training (controls: $t_{18} = -5.18$, $p < 0.0001$, $d_z = -1.22$; amusics: $t_{19} = -9.33$, $p < 0.0001$, $d_z = -2.14$; two-tailed). The time x group interaction can be explained by a greater improvement in MBEA scores in amusics relative to controls, perhaps because the control group was closer to ceiling performance initially. The same trends were found when limiting analyses to only the pitch subtasks from the MBEA (i.e., same/different melody discrimination where different trials involve a change in pitch), although the time x group interaction did not reach significance ($F_{1,35} = 2.74$, $p = 0.107$, $\eta_p^2 = 0.073$) (Fig. 6.4B). Surprisingly, 11 of 20 amusics no longer met the global criterion for amusia after training, while 14 of 20 amusics no longer met the criterion based just on the pitch subtasks. Analogous repeated-measures ANOVAs on MBEA performance using sensitivity (d') and response bias (c) demonstrated that improvements in all groups were due to increased sensitivity but not a change in response bias (see Appendix D1 Text and Appendix Figs. D3-D4).

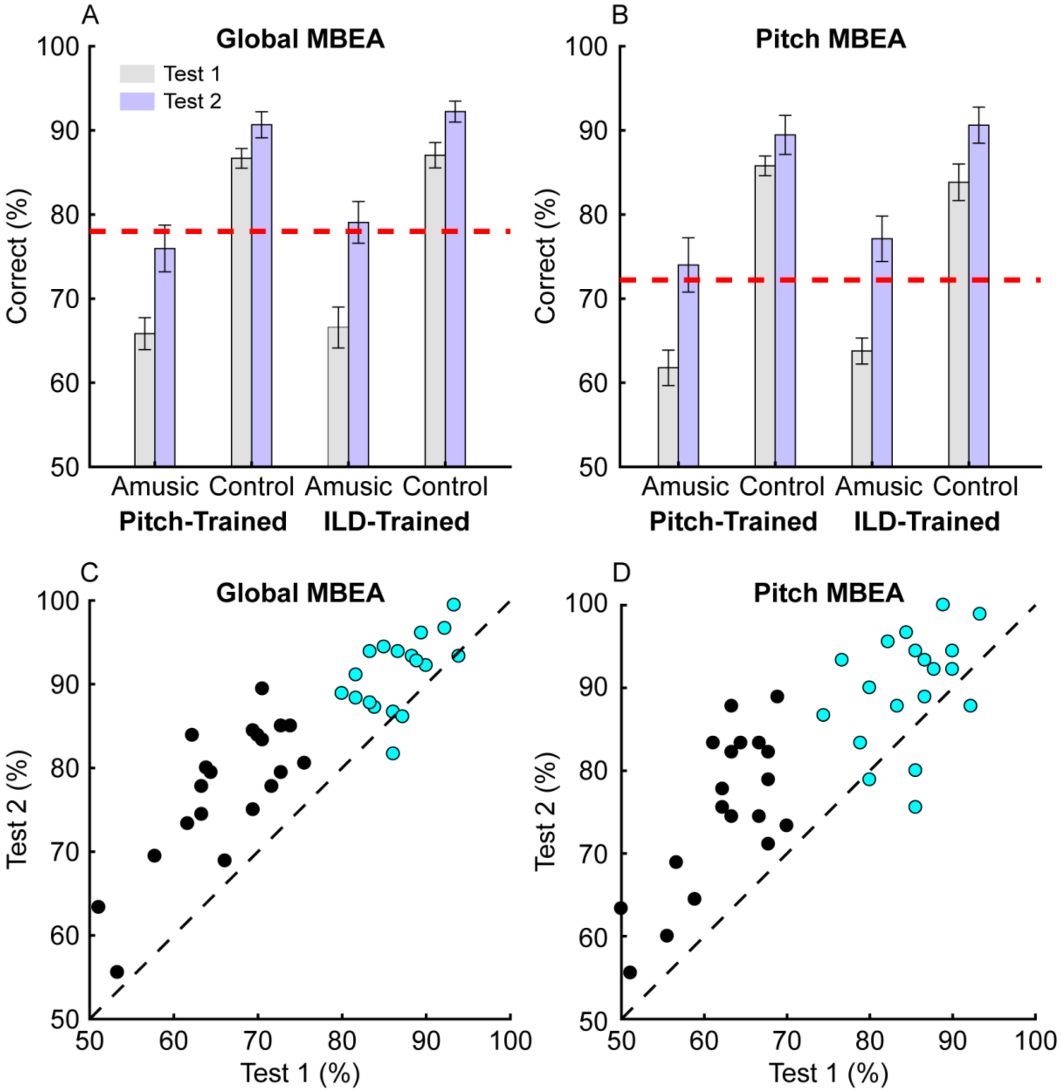

Figure 6.4

Melody discrimination scores before vs. after training. Average (A-B) and individual (C-D) test 1 (Panels A-B: grey bars; Panels C-D: x-axis) and test 2 (Panels A-B: purple bars; Panels C-D: y-axis) global (A and C) and pitch-subtask (B and D) MBEA scores. The y-axis of panels A-B plots MBEA performance in percent correct (note that RAUs were used for statistical analyses). The two sets of bars closest to the y-axis correspond to the pitch-trained participants, while the two sets furthest from the y-axis correspond to the ILD-trained participants. The red-dashed line corresponds to the cutoff for amusia based on Peretz et al. (2003). Error bars represent ± 1 standard error of the mean.

6.3.2 Experiment 2: Trained vs. untrained controls

6.3.2.1 Pitch discrimination

Pretest and posttest pitch discrimination thresholds for untrained controls (blue) are plotted in Fig. 6.5A along with thresholds from the trained controls (cyan) from experiment 1 (see Appendix Fig. D5 for individual data). Analyses were conducted on the log-transformed thresholds. A mixed-model ANOVA was conducted with time (pre vs. post) and frequency (500, 2000, and 8000 Hz) as within-subjects factors and task (trained vs. untrained) as a between-subjects factor. Thresholds at posttest were better than pretest, as indicated by a main effect of time ($F_{1,37}$ =, $p <$ 0.0001, $\eta_p^2 = 0.617$). There was also a main effect of frequency ($F_{1.41,52.2} = 139.1$, $p <$ 0.0001, $\eta_p^2 = 0.79$) but no main effect of task ($F_{1,37} = 0.823$, $p = 0.37$, $\eta_p^2 = 0.022$), and no interactions (time x frequency: $F_{1.67,61.6} = 0.436$, $p = 0.612$, $\eta_p^2 = 0.012$; time x task: $F_{1,37} = 1.1$, $p = 0.3$, $\eta_p^2 = 0.029$; frequency x task: $F_{1.41,52.2} = 0.349$, $p = 0.632$, η_p^2 = 0.009; time x frequency x task: $F_{1.67,61.6} = 0.267$, $p = 0.726$, $\eta_p^2 = 0.007$). Pairwise comparisons for frequency using Bonferroni correction ($\alpha = 0.0125$) showed significant differences between all comparisons (all ps < 0.0125, $.601 \leq d_z \leq 1.89$), with performance best at 2000 Hz and worst at 8000 Hz. The lack of time by task interaction suggests that similar improvements on pitch discrimination between pretest and posttest, regardless of whether participants completed psychophysical training.

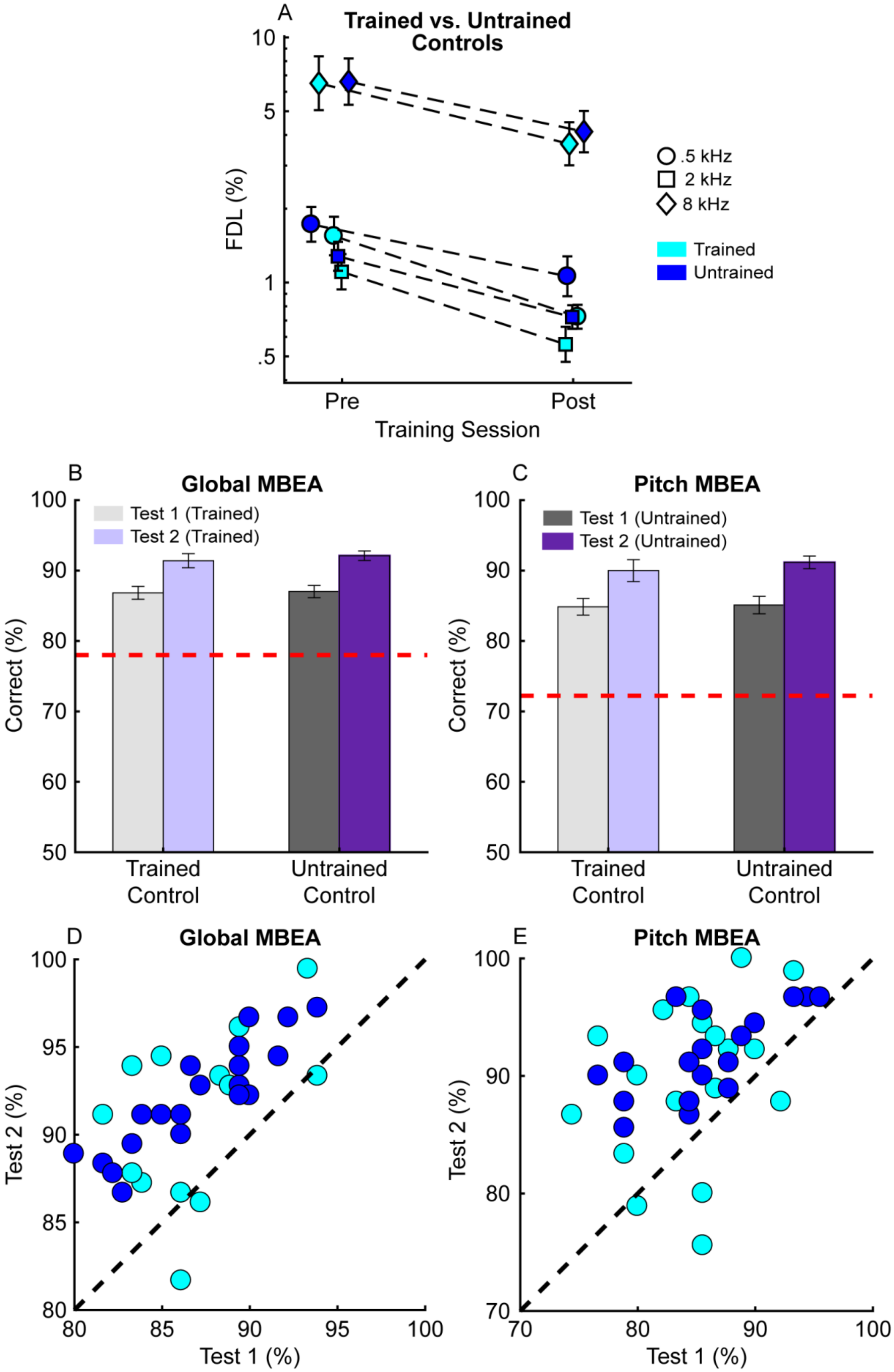
A
Trained vs. Untrained Controls
FDL (%)
10
5
1
.5
Pre
Post
Training Session
○ .5 kHz
□ 2 kHz
◇ 8 kHz
Trained
Untrained
B
Global MBEA
Test 1 (Trained)
Test 2 (Trained)
Correct (%)
100
90
80
70
60
50
Trained Control
Untrained Control
C
Pitch MBEA
Test 1 (Untrained)
Test 2 (Untrained)
Correct (%)
100
90
80
70
60
50
Trained Control
Untrained Control
D
Global MBEA
Test 2 (%)
100
95
90
85
80
80
85
90
95
100
Test 1 (%)
E
Pitch MBEA
Test 2 (%)
100
90
80
70
70
80
90
100
Test 1 (%)

Figure 6.5

Pitch discrimination (A) and MBEA (B-E) performance for trained (light) and untrained (dark) controls. Panel A shows pitch discrimination thresholds pre and posttest at 500 Hz (circles), 2000 Hz (squares), and 8000 Hz (diamonds). Panels B-C show test 1 (light/dark grey) and test 2 (light/dark purple) global (B and D) and pitch-related subtasks (C and E) of the MBEA. Average (B and C) and individual (D and E) MBEA results. The red-dashed line corresponds to the cutoff for amusia based on Peretz et al. (2003). Error bars represent ± 1 standard error of the mean. In panels D and E, the black-dashed line references no change in performance across sessions, so that data points above this line represent improvement in performance over time, while data points below this line represent poorer performance over time.

6.3.2.2 MBEA

Global MBEA scores (Fig. 6.5B and 6.5D) were converted to RAUs and entered into a mixed-model ANOVA, with time as the within-subjects factor and task (trained vs. untrained) as the between-subjects factor. There was a main effect of time ($F_{1,37} = 103.3$, $p < 0.0001$, $\eta_p^2 = 0.736$) but no main effect of task ($F_{1,37} = 0.073$, $p = 0.789$, $\eta_p^2 = 0.002$) or task by time interaction ($F_{1,37} = 0.12$, $p = 0.731$, $\eta_p^2 = 0.003$), indicating that controls' improvement on MBEA performance appeared to be unrelated to training. Conducting the same analysis but restricting MBEA performance to just the 3 pitch-related subtasks shows the same trends as those with the global MBEA performance, with a main effect of time ($F_{1,37} = 42.2$, $p < 0.0001$, $\eta_p^2 = 0.533$), and no main effect of task ($F_{1,37} = 0.073$, $p = 0.789$, $\eta_p^2 = 0.002$) or task by time interaction ($F_{1,37} = 0.008$, $p = 0.929$, $\eta_p^2 = 0.0002$). Thus, trained and untrained controls seemed to equally improve their MBEA performance over time, even when restricting analyses to only the pitch subtasks.

6.3.3 Experiment 3: One-year follow-up

6.3.3.1 Pitch discrimination

Average posttest vs. follow-up FDLs are plotted in Fig. 6A, with pretest FDLs plotted for reference. Of the 13 returning amusics, one amusic in the ILD group was removed from pitch discrimination analyses due to hearing loss at 8 kHz (60 dB HL in both ears). Levene's test of equality of error variances as well as inspection of the residuals (Appendix Fig. D6) indicated heterogeneous between-groups variances for multiple conditions (posttest 500 Hz: $p = 0.003$; posttest 2000 Hz: $p = 0.014$; follow-up

500 Hz: *p* = 0.03; follow-up 2000 Hz: 0.017). Because the homogeneity of variances assumption was violated and the sample size was markedly different between groups, non-parametric bootstrap analyses were conducted instead of a mixed-model ANOVA. The 18 returning control participants were resampled with replacement for 100,000 iterations. The mean posttest and follow-up FDLs were calculated across the 18 resampled subjects at each frequency (500, 2000, and 8000 Hz) for each iteration, to create probability distributions of the posttest FDLs (Figs. 6.6B-D) and the change in FDLs between posttest and follow-up (FDL ratio; Figs. 6E-G). The FDL ratio was calculated as the difference between the average log-transformed posttest and follow-up FDLs for each of the 100,000 resampled groups at each of the three frequencies. The mean of the non-resampled control data (blue-dashed lines) was always centered in the 95% CIs (red-dashed lines). As with the full sample of participants from experiment 1, the mean FDLs of the subset of 12 amusics at posttest (black-dashed lines) were always outside of the 95% CIs of the control distributions. However, the change in performance from posttest to follow-up was no different between amusics and controls, with the amusic thresholds falling within the control 95% CIs. The one exception was found at 2 kHz, where the FDL ratio was slightly below the low tail of the 95% control CI.

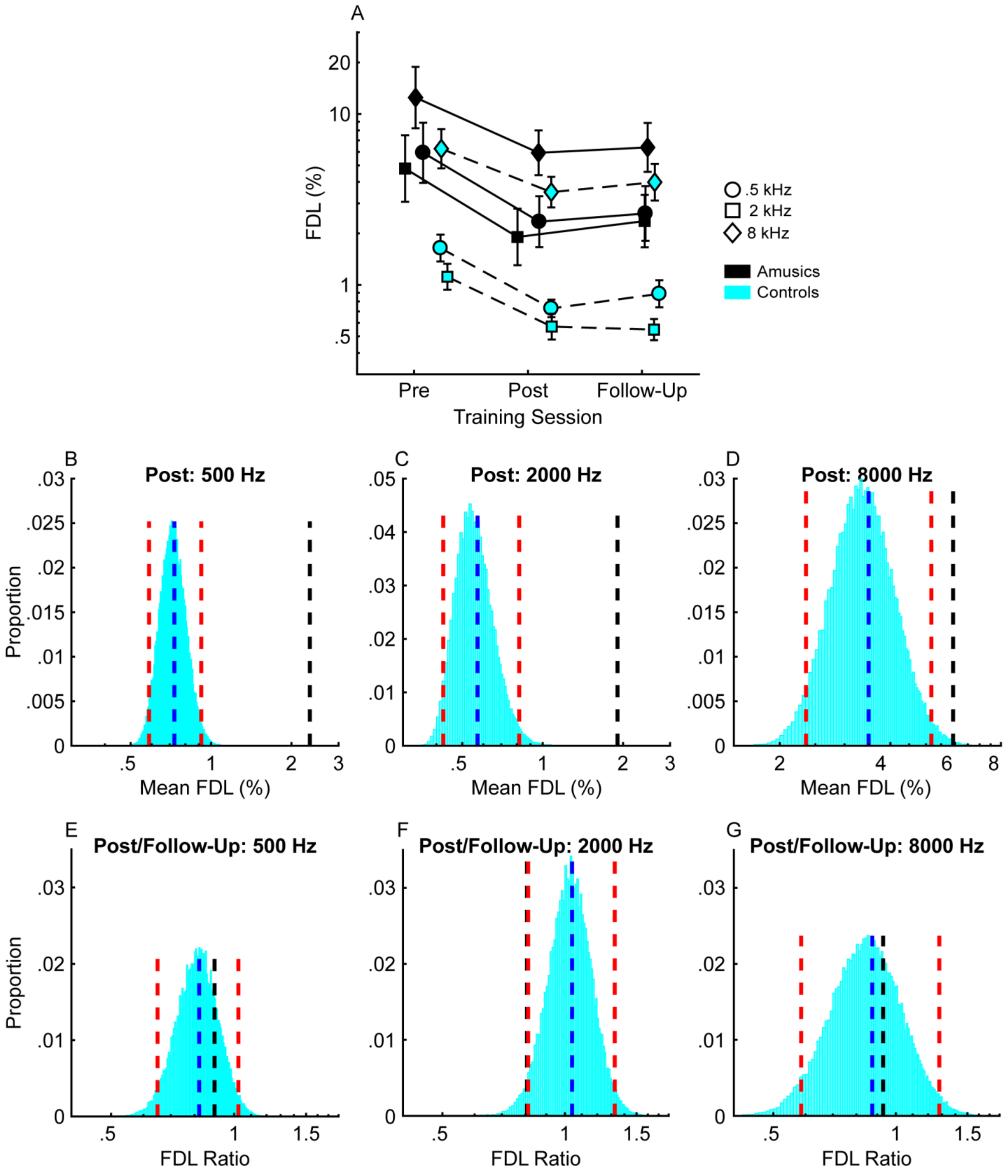

Figure 6.6

Pitch discrimination thresholds at three frequencies for 12 returning amusics (black) and 18 returning controls (cyan). Average pretest, posttest, and follow-up FDLs (A). Error bars represent ± 1 standard error of the mean. Bootstrapped control probability distributions for mean posttest performance (B-D) and the change in mean performance between posttest and one-year follow-up (E-G). In panels B-D, the x-axis represents the mean FDLs (%Δf) of the resampled data. In panels E-G, the x-axis represents the difference between the log-transformed mean posttest and follow-up FDLs (FDL ratio), transformed back to a linear scale; 1 corresponds to no learning, values greater than 1 to improved performance at follow-up, and values less than 1 to worse performance at posttest. For all panels, the y-axis corresponds to proportion. Blue-dashed lines correspond to the actual control data, black-dashed lines to the amusic data, and red-dashed lines to 95% CIs.

6.3.3.2 MBEA

A mixed-model ANOVA was conducted on the global MBEA performance (Fig. 6.7A), with time as a within-subjects factor (posttest vs. follow-up) and group as a between-subjects factor (amusic vs. control). Results indicated a main effect of group ($F_{1,29} = 42.9$, $p < 0.0001$, $\eta_p^2 = 0.597$), with amusics performing significantly worse than controls, but no main effect of time ($F_{1,29} = 0.19$, $p = 0.667$, $\eta_p^2 = 0.006$) or time by group interaction ($F_{1,29} = 0.304$, $p = 0.585$, $\eta_p^2 = 0.01$). These same trends were observed when limiting analyses to the pitch-subtasks only (group: $F_{1,29} = 25.2$, $p < 0.0001$, $\eta_p^2 = 0.465$; time: $F_{1,29} = 2.12$, $p = 0.156$, $\eta_p^2 = .068$; time x group: $F_{1,29} = 0.234$, $p = 0.632$, $\eta_p^2 = 0.008$). The lack of change in performance between test 2 and the one-year follow-up demonstrates that melodic-related learning was maintained in both amusics and controls for up to one year after training. The same analyses performed on *d'* and *c* showed no change in sensitivity or bias between posttest and follow-up (see Appendix D2 Text and Appendix Figs. D7-D8). Of the 13 returning amusics, only 5 still met the global diagnostic criterion for amusia, while only 4 met the pitch diagnostic criterion.

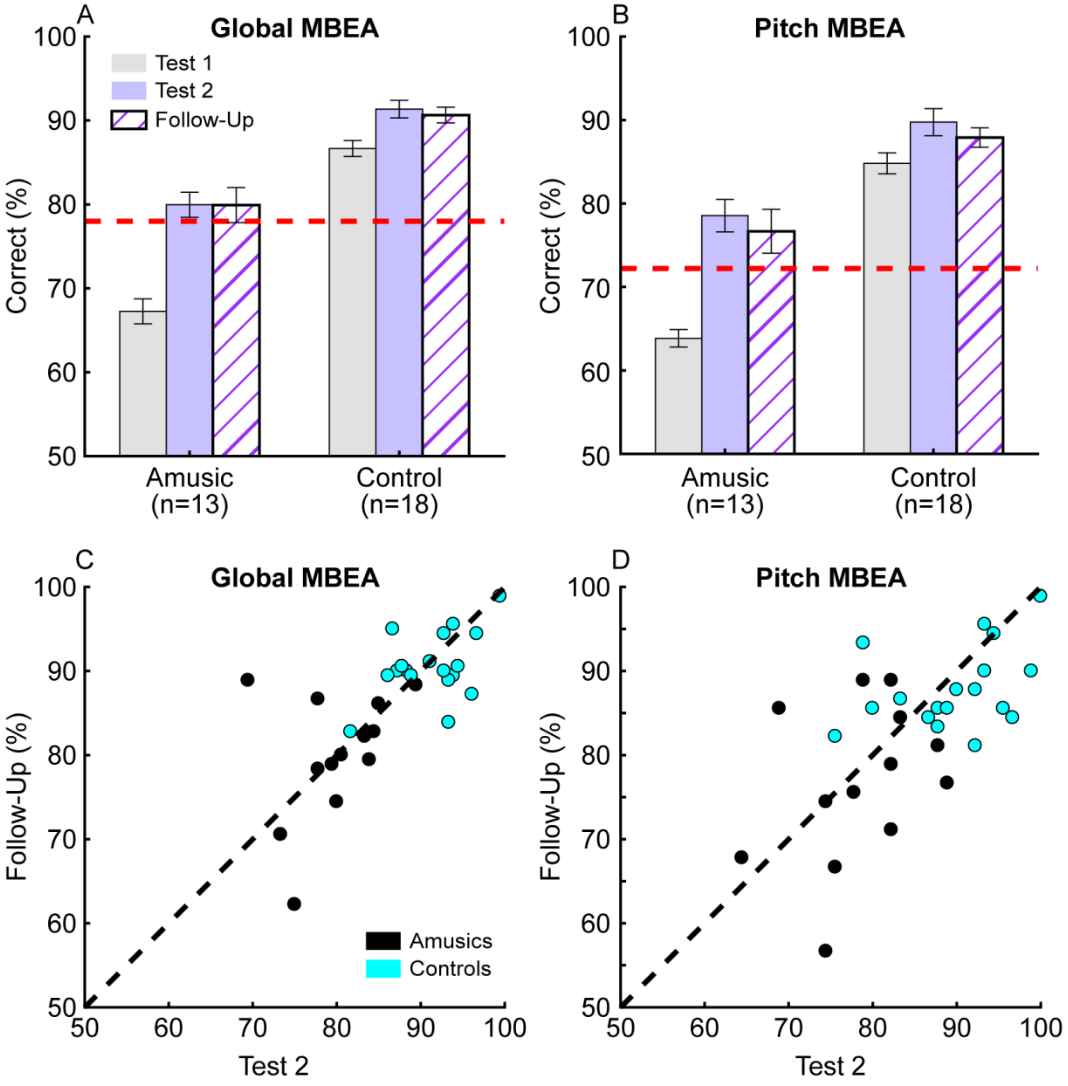

Figure 6.7

One-year follow-up MBEA performance for trained amusics (black) and controls (cyan). Test 1 (grey), test 2 (purple), and follow-up (striped) global (A and C) and pitch-related subtasks (B and D) of the MBEA. Average (A and B) and individual (C and D) MBEA results. The red-dashed line corresponds to the cutoff for amusia based on Peretz et al. (2003). Error bars represent ± 1 standard error of the mean. In panels C and D, the black-dashed line references no change in performance across sessions, so that data points above this line represent improvement in performance over time, while data points below this line represent poorer performance over time.

6.4 Discussion

6.4.1 Summary of findings

The results demonstrate several important findings: (1) Amusics and controls can improve pitch and ILD discrimination with training; (2) learning occurs for pitch discrimination across untrained frequencies and also across tasks; (3) sufficient learning may occur even during the brief pretest and MBEA to account for most of the learning observed in this study; and (4) learning is maintained over the period of at least a year.

6.4.2 Plasticity over 4-Day training

Learning on 500-Hz pure-tone pitch discrimination was rapid, with the bulk of learning occurring between sessions 1 and 2, consistent with the data from non-musicians (not amusics) from Micheyl et al. (2006). The rapidity of this learning effect seems at odds with prior studies that have found no learning by amusics on either pitch change detection (Hyde and Peretz, 2004) or pitch matching (Anderson et al., 2012). Both these earlier studies did not provide participants with feedback, and the sample size of 5 amusics tested by Anderson et al. (2012) was unlikely to have provided sufficient statistical power to detect learning on their pitch matching task. The present findings, however, are consistent with improved pitch-direction identification in amusics (n=10) with training (Liu et al., 2017), and expand on those findings by demonstrating that amusics are capable of learning to discriminate fine-grained pitch changes, beyond labeling the pitch direction. Amusics were also capable of learning ILD discrimination with training.

Even though amusics improved over time, they remained worse than controls at pitch but not ILD discrimination. This could mean that pitch-discrimination deficits in amusics remain after practice, due to an underlying neurogenetic abnormality (e.g., Peretz, 2013), or simply that more training is needed for amusics to reach their best possible level of performance.

6.4.3 Broad Generalization or a Retest Effect?

Despite the specificity of the trained stimulus and task, improvements were observed across untrained frequencies and melody discrimination for both amusics and controls, even for participants trained on the non-pitch control task. One explanation is that learning from the 4-session training generalized to untrained stimuli. Some perceptual learning paradigms in normal listeners have found pure-tone pitch discrimination learning to be frequency-specific (Demany and Semal, 2002), while others have found partial across-frequency generalization (Demany, 1985; Irvine et al., 2000; Delhommeau et al., 2005; Wright and Fitzgerald, 2005) or broad, across-frequency learning that may or may not be a result of generalization (Grimault et al., 2003; Amitay et al., 2005). The similar amount of learning in the untrained control group on pitch and melody discrimination relative to the trained controls suggests that the bulk of learning, at least for the control group, was a result of taking the tasks twice (i.e., a retest effect, or the learning that occurs from the pretest stage) and not necessarily generalization from the intervening training sessions. Retest effects can be specific to the procedure, task, or stimulus (e.g., Hawkey, Amitay, & Moore, 2004; Mossbridge, Fitzgerald, O'Connor, & Wright, 2006; Ortiz & Wright, 2009; Wright & Fitzgerald, 2001), and results from the present study cannot differentiate between these three possibilities. This does not detract from the primary finding that both amusics and controls learned by a roughly equal amount, regardless of training task. It is possible that learning in amusics was also primarily a retest effect, although the lack of change between test-retest in pitch-subtask MBEA performance in a separate group of 10 untrained amusics (Liu et al., 2017), and the lack of any significant change in pitch-subtask MBEA performance even for the 10 amusics trained on Liu et al's pitch contour identification task, cast some doubt on this interpretation. The amusics from Liu et al. had a test-retest gap of 2 weeks, very comparable to the gap used in the present study. Furthermore, their amusics had similar pretest pitch MBEA scores (untrained amusics: 59.78%; trained amusics: 58.22%) to the pretest scores of the amusics who underwent training in the present study (pitch-trained amusics: 61.78%; ILD-trained amusics: 63.78%). It is therefore possible that the psychophysical training may have had some influence on the improved melody and/or pitch discrimination performance in our pitch and ILD-trained amusics. Given that the present study did not have a group of 20 untrained amusics matched to the 20 trained amusics, and amusics form a heterogeneous population (Vuvan et al., 2015), it is not possible to discern whether the learning in amusics was entirely a retest effect, learning caused from the psychophysical training, or some combination of the two. Future research should examine test-retest effects on same/different melody discrimination in trained and untrained amusics and controls. In addition, using a set of new posttest melodies that are perceptually distinct from pretest but equated for difficulty would help differentiate whether or not the learning that occurs in amusia is stimulus specific or more general. While the present findings are limited, in that they cannot discern the type of learning that occurs in amusia, they demonstrate clear and large effects of pitch and melody discrimination learning in amusics, which was previously thought not to be possible (e.g., Hyde & Peretz, 2004).

6.4.4 Maintenance and Implications for Amusia

Perhaps the most surprising finding was that both amusics and controls improved on the MBEA, and for the subset of 31 participants that returned at follow-up, this learning was, on average, maintained for at least one year. More than half of the amusics no longer met the diagnostic criteria for amusia at posttest, and 8/13 no longer met the criteria at follow-up. This outcome is inconsistent with the notion that amusia, as defined by performance in the MBEA test, is a life-long disorder. The results of the present study demonstrate rapid and substantial improvements in performance on pitch and melody-related tasks in congenital amusia, which opens many further questions regarding the mechanisms responsible for this learning, and whether this learning affects the neurological correlates of amusia (e.g., Albouy et al., 2013; Hyde et al., 2007; Peretz, 2016). The broad improvements observed here, for example, could reflect improvements related to auditory short-term or working memory for pitch, rather than specific changes to pitch encoding. The psychophysical paradigm involved geometrically centering the target and reference frequencies from trial to trial, forcing participants to make on-line comparisons of the tones across time. It is possible that the nature of the specific tasks amusics were tested on resulted in more general learning, although further work is needed to confirm this speculation. The implications of the present results are contingent on several factors that make our categorization of amusia relevant to the amusia literature. First, our amusics performed below the global cutoff on the MBEA, the most commonly accepted standard for amusia (Peretz et al., 2003). Second, in order to qualify to participate, amusics also had to perform below the pitch-subtask cutoff from Peretz et al. (2003).

This is a more conservative criterion than used in most other studies, as described in the recent Montreal Protocol for Identification of Amusia (MPIA; see Table 1 from Vuvan et al., 2017). Lastly, inclusion in the study and/or analyses was also in line with recent suggestions from the MPIA, including an audiometric and cognitive assessment for each participant. This means that worse performance in the amusic group is unlikely to be driven by extraneous factors, such as hearing loss or cognitive ability, as all participants were given an audiometric assessment, performed within normal limits on the IQ test, and the controls were matched in IQ to the amusics. It is therefore not possible that the amusics could have simply been miscategorized based on currently accepted definitions. It is possible, however, that the currently accepted criteria for amusia do not fully capture all aspects of the disorder. If within-experiment learning, even on non-pitch tasks, is capable of dramatically affecting melody discrimination, then this poses methodological issues with using the same sample of amusic participants across multiple experiments.

It is possible that some labs already periodically retest amusics on all or parts of the MBEA. As of now, however, retest procedures are either minimally discussed or not mentioned in published studies. Future studies should examine whether or not their prescreening processes curtail some of the melody-discrimination learning effects that can occur in amusics, as well as consider retesting amusics periodically in a laboratory setting to determine whether their scores have changed after undergoing extensive laboratory testing.

6.5 Conclusions

Our results demonstrate that pitch and melody discrimination can be improved rapidly with training in people with congenital amusia to the extent that many would no longer be defined as amusic under current standard diagnostic procedures using the MBEA. Although amusic participants improved their pitch discrimination with training, their discrimination abilities remained worse than controls. This suggests either that the asymptotic limits for pitch discrimination in amusic participants are higher (worse) than in controls, or that amusic participants require more prolonged or intensive training to further improve their discrimination abilities.

CHAPTER 7: CONCLUSIONS AND FUTURE DIRECTIONS

The goal of this dissertation work was to better understand how the auditory system represents pitch. Pitch discrimination varies widely across individuals. We utilized these individual differences to better understand the peripheral and central mechanisms for coding pitch. Together, our findings have several important implications and open up new directions for future research, as described below.

7.1 Variability across NH Listeners is Probably not Peripheral

7.1.1 Multicollinearity between FMDLs and AMDLs

The high correlations we observed between FM and AM tasks in NH listeners, even across tasks originally thought to use different mechanisms (e.g., slow dichotic FM and slow dichotic AM), suggest that the bulk of normal-hearing variability in our tasks was not driven by differences in peripheral coding. This conclusion was further confirmed by the general lack of relationship between slow or fast FMDLs and the steepness of the auditory filter slopes at the test frequency.

What, then, can explain the variability in FM and AM detection? FM detection is a relatively simple task, and the chapter on FM detection in amusics (Chapter 5) showed that FM induces a lower working memory load than three-interval frequency discrimination. Furthermore, there are only minimal learning effects in FMDLs, suggesting they are a more accurate estimation of the system’s sensitivity to frequency compared to frequency difference limens (Moore, 1976). The high multicollinearity amongst the FM and AM tasks as well as the impairment we found in amusics for both FM and AM point towards a possible shared cortical mechanism. Micheyl et al. (2013) demonstrated that it is theoretically plausible for the same rate-place cortical code to account for human behavioral performance in frequency and intensity discrimination. But, without current modeling applied to FM and AMDLs, this remains an open question for future research.

7.1.2 What does poor TFS coding look like?

Perhaps most concerning was the conclusion that the substantial variability between listeners over a wide age range observed in the slow dichotic FM task, which was meant to be an index of TFS coding, was unlikely to reflect variability in TFS coding. This is because slow dichotic FM was highly correlated with slow dichotic AM (Fig. 3.8), our measure for ILD coding. Binaural IPD and ITD tasks, like the slow dichotic FM task we used, have been viewed as ideal measures for an individual's ability to use time coding for TFS. It is well established that precise time coding must be utilized to perform well on these tasks (Hafter et al., 1979). The problem is that poor performance on binaural ITDs does not necessitate the listener to have poor peripheral time coding. There may be many other reasons why someone might have difficulty with ITDs and IPDs, from poor binaural coding to poor processing efficiency to poor executive functioning. The inability to differentiate generally poor binaural performers from listeners with a specific issue in time coding to TFS is highly problematic for the auditory literature, as these are currently the best measures we have available. Slow-rate, low-carrier FM was thought to be a useful monaural index for TFS coding (e.g., Moore and Sek, 1995; Strelcyk and Dau, 2009), but our findings from SNHL listeners (Chapter 4) now show that slow FM is strongly affected by place coding constraints. It would be useful for basic science, particularly in pitch research, to have a non-invasive but objective physiological measure for TFS coding in humans.

7.2 FM is Represented by a Place or Combined Place-Time Code

Based on the bulk of previous literature supporting the use of a time code for slow-rate, low-carrier FM but a place code for faster rates, we had predicted that SNHL, which leads to poorer place coding, should adversely affect fast but not slow FMDLs. Specifically, fast FMDLs were expected to correlate with the steepness of the cochlear filter slope at the test frequency, but slow FMDLs were not. We were surprised to find that both slow and fast FMDLs correlated with the steepness of the low slope, consistent with Zwicker's (1952) predictions over half a century ago. The lack of any clear role for frequency selectivity in NH listeners makes for a relatively clean comparison (Chapters 2 and 3; note that NH listeners were tested at 500 Hz while SNHL listeners at 1 kHz), as they had less variability in frequency selectivity, presumably because they were all NH. The fact that the correlation between frequency selectivity and FMDLs was only clearly present on the low-frequency side of the excitation pattern, as opposed to the low and high side, further suggests this relation is real. Contrary to the widely accepted understanding that slow FM uses time coding, our results show that either a place or combined place-time code is utilized.

7.3 Congenital Amusia is not necessarily a Life-Long Disorder

Chapter 6 presented the first study to demonstrate rapid learning for pitch and melody discrimination in congenital amusia. From the perspective of the pitch training literature, this conclusion may seem obvious. Non-musicians, for example, only appear to need 4-8 hours of pitch discrimination training to perform on par with professional musicians, despite the fact that the musicians began with a pitch-discrimination advantage greater than a factor of 6 (Micheyl et al., 2006). But for quite some time, congenital amusia has been described as a life-long disorder that is impervious to training (Hyde and Peretz, 2004; Mignault Goulet et al., 2012; Peretz et al., 2012; Wilbiks et al., 2016; Liu et al., 2017). Using an adaptive task with simple, highly controlled stimuli, we demonstrated learning effects that were so large that over half the amusics no longer met the standardized diagnostic criterion for amusia after the training. In fact, the learning was so rapid that it is possible the training may not have even been necessary to see the improvements.

Our findings have important implications for the amusia literature, where the sample sizes are small and the same amusics are often used in multiple experiments. This means experimenters risk the chance of “training out” their between-group effects, which could potentially cause replication failures in the literature, particularly for any between-group effects that are small. Our findings perhaps generate more questions than they answer: What does it mean to be amusic if one can train out the disorder in a matter of days? Are the recovered amusics truly normal or do they only perform in the normal range for task-specific stimuli? Do the well-established neural malformations in amusics change as a function of pitch or melody-related learning? What were the amusics learning? Do these findings have the potential to generalize to the acquired amusia literature, where acquired amusia is a quite common side effect of right-hemispheric strokes (Sihvonen et al., 2016b)? The topic of pitch learning, and particularly pitch learning in amusia, has many open questions for future research.

7.4 Conclusion

Overall, results demonstrate that both peripheral and central mechanisms can be responsible for individual differences in pitch perception. When all the participants have NH, the variance appears to be dominated by more central factors or even possibly cognitive factors such as working memory, depending on the particular task. The benefit of central noise affecting pitch perception is that it is highly susceptible to auditory training interventions, even in amusic participants who have a disorder in fine-grained pitch perception. A great deal of further work is needed to better understand the central mechanisms for coding pitch as well as the nature of the mechanisms involved in pitch learning paradigms. For listeners with SNHL, FM detection is directly affected by the fidelity of place (tonotopic) coding in the cochlea. Such peripheral effects are unlikely to improve simply through practice. Targeted medical and/or prosthetic interventions, aimed at improving the tonotopic frequency-to-place mapping within the cochlea or auditory nerve, have the potential to improve pitch coding for people with hearing loss.

BIBLIOGRAPHY

Albouy, P., Cousineau, M., Caclin, A., Tillmann, B., and Peretz, I. (**2016**). “Impaired encoding of rapid pitch information underlies perception and memory deficits in congenital amusia,” Sci. Rep. **6**, 18861.

Albouy, P., Lévêque, Y., Hyde, K. L., Bouchet, P., Tillmann, B., and Caclin, A. (**2015**). “Boosting pitch encoding with audiovisual interactions in congenital amusia,” Neuropsychologia **67**, 111–120.

Albouy, P., Mattout, J., Bouet, R., Maby, E., Sanchez, G., Aguera, P. E., Daligault, S., et al. (**2013**). “Impaired pitch perception and memory in congenital amusia: The deficit starts in the auditory cortex,” Brain **136**, 1639–1661.

Albouy, P., Mattout, J., Sanchez, G., Tillmann, B., and Caclin, A. (**2015**). “Altered retrieval of melodic information in congenital amusia: insights from dynamic causal modeling of MEG data,” Front. Hum. Neurosci. **9**, 1–13.

Albouy, P., Schulze, K., Caclin, A., and Tillmann, B. (**2013**). “Does tonality boost short-term memory in congenital amusia?,” Brain Res. **1537**, 224–232.

Altman, D. G., and Bland, J. M. (**1983**). “Measurement in medicine: The analysis of method comparison studies,” Statistician **32**, 307–317.

Amitay, S., Hawkey, D. J. C., and Moore, D. R. (**2005**). “Auditory frequency discrimination learning is affected by stimulus variability,” Percept. Psychophys. **67**, 691–698.

Amitay, S., Irwin, A., and Moore, D. R. (**2006**). “Discrimination learning induced by training with identical stimuli,” Nat. Neurosci. **9**, 1446–1448.

Anderson, S., Himonides, E., Wise, K., Welch, G., and Stewart, L. (**2012**). “Is there potential for learning in amusia? A study of the effect of singing intervention in congenital amusia,” Ann. N. Y. Acad. Sci. **1252**, 345–353.

Arehart, K. H. (**1994**). “Effects of harmonic content on complex-tone fundamental-frequency discrimination in hearing-impaired listeners,” J. Acoust. Soc. Am. **95**, 3574–3585.

Attias, H., and Schreiner, C. E. (**1997**). “Temporal low-order statistics of natural sounds,” Adv. Neural Inf. Process. Syst. **9**, 27–33. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.53.201&rep=rep1&type=pdf

Attneave, F., and Olson, R. K. (**1971**). “Pitch as a medium: A new approach to psychophysical scaling,” Am. J. Psychol. **84**, 147–166.

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

Baker, R. J., and Rosen, S. (**2002**). “Auditory filter nonlinearity in mild/moderate hearing impairment,” J. Acoust. Soc. Am. **111**, 1330–1339.

Baker, R. J., and Rosen, S. (**2006**). “Auditory filter nonlinearity across frequency using simultaneous notched-noise masking,” J. Acoust. Soc. Am. **119**, 454–462.

Bendor, D. (**2012**). “Does a pitch center exist in auditory cortex?,” J. Neurophysiol. **107**, 743–746.

Bendor, D., and Wang, X. (**2005**). “The neuronal representation of pitch in primate auditory cortex,” Nature **436**, 1161–1165.

Bendor, D., and Wang, X. (**2006**). “Cortical representations of pitch in monkeys and humans,” Curr. Opin. Neurobiol. **16**, 391–399.

Bendor, D., and Wang, X. (**2010**). “Neural coding of periodicity in marmoset auditory cortex,” J. Neurophysiol. **103**, 1809–1822.

Bernstein, J. G., and Oxenham, A. J. (**2003**). “Pitch discrimination of diotic and dichotic tone complexes: Harmonic resolvability or harmonic number?,” J. Acoust. Soc. Am. **113**, 3323–3334.

Bernstein, J. G. W., and Oxenham, A. J. (**2006**). “The relationship between frequency selectivity and pitch discrimination: sensorineural hearing loss,” J. Acoust. Soc. Am. **120**, 3929–3945.

Bernstein, J. G. W., and Oxenham, A. J. (**2006**). “The relationship between frequency selectivity and pitch discrimination : Effects of stimulus level,” J. Acoust. Soc. Am. , doi: 10.1121/1.2372451.

Bharadwaj, X. H. M., Masud, S., Mehraei, G., Verhulst, S., and Shinn-cunningham, B. G. (**2015**). “Individual differences reveal correlates of hidden hearing deficits,” J. Neurosci. **35**, 2161–2172.

Bianchi, F., Santurette, S., Wendt, D., and Dau, T. (**2015**). “Pitch discrimination in musicians and non-musicians: Effects of harmonic resolvability and processing effort,” J. Assoc. Res. Otolaryngol. **79**, 69–79.

Blauert, J. (**1972**). “On the lag of lateralization caused by interaural time and intensity differences," Int. J. Audiol. **11**, 265–270.

Bregman, A. S. (**1990**). *Auditory Scene Analysis: The Perceptual Organization of Sound*, Bradford Books, MIT Press, Cambridge, MA.

Burns, E. M., and Viemeister, N. F. (**1976**). "Nonspectral pitch," J. Acoust. Soc. Am. **60**, 863–869.

Burns, E. M., and Viemeister, N. F. (**1981**). "Played-again SAM: Further observations on the pitch of amplitude-modulated noise," J. Acoust. Soc. Am. **70**, 1655–1660.

Buss, E., Hall, J. W., and Grose, J. H. (**2004**). "Temporal fine-structure cues to speech and pure tone modulation in observers with sensorineural hearing loss," Ear Hear. **25**, 242–250.

Carcagno, S., and Plack, C. J. (**2011**). "Subcortical plasticity following perceptual learning in a pitch discrimination task," J. Assoc. Res. Otolaryngol. **12**, 89–100.

Carlyon, R. P., and Shackleton, T. M. (**1994**). "Comparing the fundamental frequencies of resolved and unresolved harmonics: Evidence for two pitch mechanisms?," J. Acoust. Soc. Am. **95**, 3541–3554.

Chen, H., and Zeng, F.-G. (**2004**). "Frequency modulation detection in cochlear implant subjects," J. Acoust. Soc. Am. **116**, 2269–2277.

Chen, J. L., Kumar, S., Williamson, V. J., Scholz, J., Griffiths, T. D., and Stewart, L. (**2015**). "Detection of the arcuate fasciculus in congenital amusia depends on the tractography algorithm," Front. Psychol. **6**, 1–11.

de Cheveigné, A. (**2005**). "Pitch perception models," In C. J. Plack, A. J. Oxenham, R. R. Fay, and A. N. Popper (Eds.), Pitch Neural Coding Percept. Springer, pp. 169–233.

Coninx, F. (**1978**). "The detection of combined differences in frequency and intensity," Acustica **39**, 137–150.

Coninx, F. (**1978**). "The perception of combined frequency and amplitude modulations with clearly audible modulation depths," Acustica **39**, 151–154.

Conzelmann, K., and Süß, H. M. (**2015**). "Auditory intelligence: Theoretical considerations and empirical findings," Learn. Individ. Differ. **40**, 27–40.

Cousineau, M., Oxenham, A. J., and Peretz, I. (**2015**). "Congenital amusia: A cognitive disorder limited to resolved harmonics and with no peripheral basis," Neuropsychologia **66**, 293–301.

Coutinho, E., and Dibben, N. (**2013**). “Psychoacoustic cues to emotion in speech prosody and music,” Cogn. Emot. **27**, 658–684.

Dai, H., Nguyen, Q. T., and Green, D. M. (**1995**). “A two-filter model for frequency discrimination,” Hear. Res. **85**, 109–114.

Deal, J. A., Betz, J., Yaffe, K., Harris, T., Purchase-Helzner, E., Satterfield, S., Pratt, S., et al. (**2017**). “Hearing impairment and incident dementia and cognitive decline in older adults: The Health ABC Study,” Journals Gerontol. Ser. A Biol. Sci. Med. Sci. **72**, 703–709.

Delhommeau, K., Micheyl, C., and Jouvent, R. (**2005**). “Generalization of frequency discrimination learning across frequencies and ears: Implications for underlying neural mechanisms in humans,” J. Assoc. Res. Otolaryngol. **6**, 171–179.

Demany, L. (**1985**). “Perceptual learning in frequency discrimination,” J. Acoust. Soc. Am. **78**, 1118–1120.

Demany, L., Carlyon, R. P., and Semal, C. (**2009**). “Continuous versus discrete frequency changes: different detection mechanisms?,” J. Acoust. Soc. Am. **125**, 1082–1090.

Demany, L., and Semal, C. (**1986**). “On the detection of amplitude modulation and frequency modulation at low modulation frequencies,” Acustica **61**, 243–255.

Demany, L., and Semal, C. (**1989**). “Detection thresholds for sinusoidal frequency modulation,” J. Acoust. Soc. Am. **85**, 1295–1301.

Demany, L., and Semal, C. (**2002**). “Learning to perceive pitch differences,” J. Acoust. Soc. Am. **111**, 1377–1388.

Drayna, D., Manichaikul, A., de Lange, M., Snieder, H., and Spector, T. (**2001**). “Genetic correlates of musical pitch recognition in humans,” Science (80-.). **291**, 1969–1972.

Durlach, N. I., and Braida, L. D. (**1969**). “Intensity perception I Preliminary theory of intensity resolution,” J. Acoust. Soc. Am. **46**, 372–383.

Edwards, B. W., and Viemeister, N. F. (**1994**). “Modulation detection and discrimination with three-component signals,” J. Acoust. Soc. Am. **95**, 2202–2212.

Edwards, B. W., and Viemeister, N. F. (**1994**). “Frequency modulation versus amplitude modulation discrimination: Evidence for a second frequency modulation encoding mechanism," J. Acoust. Soc. Am. **96**, 733–740.

Efron, B., and Tibshirani, R. J. (**1998**). "Confidence intervals based on bootstrap percentiles," An Introd. to Bootstrap Chapman & Hall/CRC, New York, pp. 168–177.

Elliott, D. N., Riach, W. D., Sheposh, J. P., and Trahiotis, C. (**1966**). "Discrimination performance of high school sophomores on a battery of auditory tests," Acta Otolaryngol. **216**, 1–58.

Ernst, S. M. A., and Moore, B. C. J. (**2012**). "The role of time and place cues in the detection of frequency modulation by hearing-impaired listeners," J. Acoust. Soc. Am. **131**, 4722–4731.

Ernst, S. M. a, and Moore, B. C. J. (**2010**). "Mechanisms underlying the detection of frequency modulation," J. Acoust. Soc. Am. **128**, 3642–3648.

Feng, L., and Wang, X. (**2017**). "Harmonic template neurons in primate auditory cortex underlying complex sound processing," Proc. Natl. Acad. Sci. **114**, E840–E848.

Festen, J. M., and Plomp, R. (**1981**). "Relations between auditory functions in normal hearing," J. Acoust. Soc. Am. **70**, 356–369.

Festen, J. M., and Plomp, R. (**1983**). "Relations between auditory functions in impaired hearing," J. Acoust. Soc. Am. **73**, 652–662.

Feth, L. L. (**1972**). "Combinations of amplitude and frequency differences in auditory discrimination," Acustica **26**, 67–77.

Flanagan, J. L., and Guttman, N. (**1960**). "On the pitch of periodic pulses," J. Acoust. Soc. Am. **32**, 1308–1319.

Foxton, J. M., Dean, J. L., Gee, R., Peretz, I., and Griffiths, T. D. (**2004**). "Characterization of deficits in pitch perception underlying 'tone deafness,'" Brain **127**, 801–810.

Füllgrabe, C. (**2013**). "Age-dependent changes in temporal-fine-structure processing in the absence of peripheral hearing loss," Am. J. Audiol. **22**, 313.

Füllgrabe, C., and Moore, B. C. J. (**2017**). "Evaluation of a method for determining binaural sensitivity to temporal fine structure (TFS-AF Test) for older listeners with normal and impaired low-frequency hearing," Trends Hear. **21**, 233121651773723.

Füllgrabe, C., Moore, B. C. J., and Stone, M. A. (**2015**). “Age-group differences in speech identification despite matched audiometrically normal hearing: Contributions from auditory temporal processing and cognition,” Front. Aging Neurosci. **7**, 1–25.

Gallun, F. J., McMillan, G. P., Molis, M. R., Kampel, S. D., Dann, S. M., and Konrad-Martin, D. L. (**2014**). “Relating age and hearing loss to monaural, bilateral, and binaural temporal sensitivity,” Front. Neurosci. **8**, 1–14.

Gaser, C., and Schlaug, G. (**2003**). “Brain structures differ between musicians and non-musicians,” J. Neurosci. **23**, 9240–9245.

Glasberg, B. R., and Moore, B. C. J. (**1986**). “Auditory filter shapes in subjects with unilateral and bilateral cochlear impairments,” J. Acoust. Soc. Am. **79**, 1020–1033.

Glasberg, B. R., and Moore, B. C. J. (**2000**). “Frequency selectivity as a function of level and frequency measured with uniformly exciting notched noise,” J. Acoust. Soc. Am. **108**, 2318–2328.

Goldstein, J. L. (**1973**). “An optimum processor theory for the central formation of the pitch of complex tones,” J. Acoust. Soc. Am. **54**, 1496–1516.

Gosselin, N., Jolicœur, P., and Peretz, I. (**2009**). “Impaired memory for pitch in congenital amusia,” Ann. N. Y. Acad. Sci. **1169**, 270–272.

Grant, K. W. (**1987**). “Frequency modulation setection by normally hearing and profoundly hearing-impaired listeners,” J. Speech, Lang. Hear. Res. **30**, 558.

Grimault, N., Micheyl, C., Carlyon, R. P., Bacon, S. P., and Collet, L. (**2003**). “Learning in discrimination of frequency or modulation rate: Generalization to fundamental frequency discrimination,” Hear. Res. **184**, 41–50.

Grose, J. H., and Mamo, S. K. (**2010**). “Processing of temporal fine structure as a function of age,” Ear Hear. **31**, 755–760.

Grose, J. H., and Mamo, S. K. (**2012**). “Frequency modulation detection as a measure of temporal processing: Age-related monaural and binaural effects,” Hear. Res. **294**, 49–54.

Hafter, E. R., Dye, R. H., and Gilkey, R. H. (**1979**). “Lateralization of tonal signals which have neither onsets nor offsets,” J. Acoust. Soc. Am. **65**, 471–477.

Hall, D. A., Edmondson-Jones, A. M., and Fridriksson, J. (**2006**). “Periodicity and frequency coding in human auditory cortex," Eur. J. Neurosci. **24**, 3601–3610.

Hall, D. A., and Plack, C. J. (**2009**). "Pitch processing sites in the human auditory brain," Cereb. Cortex **19**, 576–585.

Harrison, R. V, and Evans, E. F. (**1979**). "Some aspects of temporal coding by single cochlear fibres from regions of cochlear hair cell degeneration in the guinea pig," Arch. Otorhinolaryngol. **224**, 71–78.

Hartmann, W. M., and Hnath, G. M. (**1982**). "Detection of mixed modulation," Acustica **50**, 297–312.

Hawkey, D. J. C., Amitay, S., and Moore, D. R. (**2004**). "Early and rapid perceptual learning," Nat. Neurosci. **7**, 1055–1056.

He, N., Mills, J. H., Ahlstrom, J. B., and Dubno, J. R. (**2008**). "Age-related differences in the temporal modulation transfer function with pure-tone carriers," J. Acoust. Soc. Am. **124**, 3841–3849.

He, N., Mills, J. H., and Dubno, J. R. (**2007**). "Frequency modulation detection: Effects of age, psychophysical method, and modulation waveform," J. Acoust. Soc. Am. **122**, 467–477.

Henry, K. S., and Heinz, M. G. (**2012**). "Diminished temporal coding with sensorineural hearing loss emerges in background noise," Nat. Neurosci. **15**, 1362–1364.

Henry, M. J., and McAuley, J. D. (**2013**). "Failure to apply signal detection theory to the Montreal Battery of Evaluation of Amusia may misdiagnose amusia," Music Percept. An Interdiscip. J. **30**, 480–496.

Hopkins, K., and Moore, B. C. J. (**2007**). "Moderate cochlear hearing loss leads to a reduced ability to use temporal fine structure information," J. Acoust. Soc. Am. **122**, 1055–1068.

Hopkins, K., and Moore, B. C. J. (**2011**). "The effects of age and cochlear hearing loss on temporal fine structure sensitivity, frequency selectivity, and speech reception in noise," J. Acoust. Soc. Am. **130**, 334–349.

Houtsma, A. J. M., and Smurzynski, J. (**1990**). "Pitch identification and discrimination for complex tones with many harmonics," J. Acoust. Soc. Am. **87**, 304.

Humes, L. E., Busey, T. A., Craig, J., and Kewley-Port, D. (**2013**). "Are age-related changes in cognitive function driven by age-related changes in sensory processing?," Attention, Perception, Psychophys. **75**, 508–524.

Hyde, K. L., Lerch, J. P., Zatorre, R. J., Griffiths, T. D., Evans, A. C., and Peretz, I. (**2007**). "Cortical thickness in congenital amusia: When less is better than more," J. Neurosci. **27**, 13028–13032.

Hyde, K. L., and Peretz, I. (**2004**). "Brains that are out of tune but in time," Psychol. Sci. **15**, 356–360.

Hyde, K. L., Zatorre, R. J., Griffiths, T. D., Lerch, J. P., and Peretz, I. (**2006**). "Morphometry of the amusic brain: A two-site study," Brain **129**, 2562–2570.

Hyde, K. L., Zatorre, R. J., and Peretz, I. (**2011**). "Functional MRI evidence of an abnormal neural network for pitch processing in congenital amusia," Cereb. Cortex **21**, 292–299.

Irvine, D. R., Martin, R. L., Klimkeit, E., and Smith, R. (**2000**). "Specificity of perceptual learning in a frequency discrimination task," J. Acoust. Soc. Am. **108**, 2964–2968.

Ives, D. T., Calcus, A., Kalluri, S., Strelcyk, O., Sheft, S., and Lorenzi, C. (**2013**). "Effects of noise reduction on AM and FM perception," J. Assoc. Res. Otolaryngol. **14**, 149–157.

Jesteadt, W., and Sims, S. L. S. (**1975**). "Decision processes in frequency discrimination," J. Acoust. Soc. Am. **57**, 1161–1168.

Jiang, C., Lim, V. K., Wang, H., and Hamm, J. P. (**2013**). "Difficulties with pitch discrimination influences pitch memory performance: Evidence from congenital amusia," (C. R. Larson, Ed.) PLoS One **8**, e79216.

Johnson, D. H. (**1980**). "The relationship between spike rate and synchrony in responses of auditory-nerve fibers to single tones," **68**, 1115–1122.

Johnson, D. M., Watson, C. S., and Jensen, J. K. (**1987**). "Individual differences in auditory capabilities I," J. Acoust. Soc. Am. **81**, 427–438.

Joris, P. X., Schreiner, C. E., and Rees, A. (**2004**). "Neural processing of amplitude-modulated sounds," Physiol. Rev. **84**, 541–577.

Joris, P. X., and Verschooten, E. (**2013**). "On the limit of neural phase locking to fine structure in humans," Basic Asp. Hear. Physiol. Percept. Springer, New York, pp. 101–108.

Karlin, J. E. (**1941**). "Music ability," Psychometrika **6**, 61–65.

Karlin, J. E. (**1942**). “A factorial study of auditory function,” Psychometrika **7**, 251–279.

Khanna, S. M., and Teich, M. C. (**1989**). “Spectral characteristics of the responses of primary auditory-nerve fibers to frequency-modulated signals,” Hear. Res. **39**, 159–176.

Kidd, G. R., Watson, C. S., and Gygi, B. (**2007**). “Individual differences in auditory abilities,” J. Acoust. Soc. Am. **122**, 418–435.

King, A., Hopkins, K., and Plack, C. J. (**2014**). “The effects of age and hearing loss on interaural phase difference discrimination,” J Acoust Soc Am **135**, 342–351.

Kishon-Rabin, L., Amir, O., Vexler, Y., and Zaltz, Y. (**2001**). “Pitch discrimination: Are professional musicians better than non-musicians?,” J. Basic Clin. Physiol. Pharmacol. **12**, 125–143.

Kollmeier, B., and Gilkey, R. H. (**1990**). “Binaural forward and backward masking: Evidence for sluggishness in binaural detection,” J. Acoust. Soc. Am. **87**, 1709–1719.

Lacher-Fougère, S., and Demany, L. (**1998**). “Modulation detection by normal and hearing-impaired listeners,” Audiology **37**, 109–121.

Lakens, D. (**2013**). “Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs,” Front. Psychol. **4**, 1–12.

Lévêque, Y., Fauvel, B., Groussard, M., Caclin, A., Albouy, P., Platel, H., and Tillmann, B. (**2016**). “Altered intrinsic connectivity of the auditory cortex in congenital amusia,” J. Neurophysiol. **116**, 88–97.

Levitt, H. (**1971**). “Transformed up-down methods in psychoacoustics,” J. Acoust. Soc. Am. **49**, 467–477.

Licklider, J. C. R. (**1956**). “Auditory frequency analysis,” In C. Cherry (Ed.), Inf. Theory Academic Press, New York, pp. 253–268.

Lin, F. R., and Albert, M. (**2014**). “Hearing loss and dementia – who is listening?,” Aging Ment. Health **18**, 671–673.

Lin, F. R., Metter, E. J., O’Brien, R. J., Resnick, S. M., Zonderman, A. B., and Ferrucci, L. (**2011**). “Hearing loss and incident dementia,” Arch. Neurol. **68**, 214–220.

Lin, F. R., Yaffe, K., Xia, J., Xue, Q.-L., Harris, T. B., Purchase-Helzner, E., Satterfield, S., et al. (**2013**). “Hearing loss and cognitive decline in older adults,” JAMA Intern. Med. **173**, 293.

Liu, F., Jiang, C., Francart, T., Chan, A. H. D., and Wong, P. C. M. (**2017**). “Perceptual learning of pitch direction in congenital amusia,” Music Percept. **34**, 335–351.

Liu, F., Patel, A. D., Fourcin, A., and Stewart, L. (**2010**). “Intonation processing in congenital amusia: Discrimination, identification and imitation,” Brain **133**, 1682–1693.

Liu, L.-F., Palmer, A. R., and Wallace, M. N. (**2006**). “Phase-locked responses to pure tones in the inferior colliculus,” J. Neurophysiol. **95**, 1926–1935.

Loeb, G. E. (**2005**). “Are cochlear implant patients suffering from perceptual dissonance?,” Ear Hear. **26**, 435–450.

Loeb, G. E., White, M. W., and Merzenich, M. M. (**1983**). “Spatial cross-correlation,” Biol. Cybern. **47**, 149–163.

Lorenzi, C., Debruille, L., Garnier, S., Fleuriot, P., and Moore, B. C. J. (**2009**). “Abnormal processing of temporal fine structure in speech for frequencies where absolute thresholds are normal,” J. Acoust. Soc. Am. **125**, 27–30.

Lorenzi, C., Gilbert, G., Carn, H., Garnier, S., and Moore, B. C. J. (**2006**). “Speech perception problems of the hearing impaired reflect inability to use temporal fine structure,” Proc. Natl. Acad. Sci. **103**, 18866–18869.

Loui, P., Alsop, D., and Schlaug, G. (**2009**). “Tone deafness: A new disconnection syndrome?,” J. Neurosci. **29**, 10215–10220.

Lu, T., and Wang, X. (**2000**). “Temporal discharge patterns evoked by rapid sequences of wide- and narrowband clicks in the primary auditory cortex of cat,” J. Neurophysiol. **84**, 236–246.

Lyzenga, J., Carlyon, R. P., and Moore, B. C. J. (**2004**). “The effects of real and illusory glides on pure-tone frequency discrimination,” J. Acoust. Soc. Am. **116**, 491–501.

Macmillan, N. A., and Creelman, C. D. (**2005**). *Detection Theory: A User’s Guide*, Lawrence Erlbaum Associates, Inc., 2nd ed.

Maiwald, D. (**1967**). “Ein Funktionsschema des Gehörs zur Beschreibung der Erkennbarkeit kleiner Frequenz-und Amplitudenänderungen,” Acustica **18**, 81–

92.

Maiwald, D. (**1967**). “Die Berechnung von Modulationsschwellen mit Hilfe eines Funktionsschemas,” Acustica **18**, 193–207.

Martin, J. A. M., and Martin, D. (**1973**). “Auditory Perception,” Br. Med. J. **2**, 459–461.

McDermott, J. H., Lehr, A. J., and Oxenham, A. J. (**2010**). “Individual differences reveal the basis of consonance,” Curr. Biol. **20**, 1035–1041.

Mcleish, J. (**1950**). “The validation of Seashore’s measures of musical talent by factorial methods,” Br. J. Stat. Psychol. **3**, 129–140.

Micheyl, C., Delhommeau, K., Perrot, X., and Oxenham, A. J. (**2006**). “Influence of musical and psychoacoustical training on pitch discrimination,” Hear. Res. **219**, 36–47.

Micheyl, C., Schrater, P. R., and Oxenham, A. J. (**2013**). “Auditory frequency and intensity discrimination explained using a cortical population rate code,” (T. D. Griffiths, Ed.) PLoS Comput. Biol. **9**, e1003336.

Micheyl, C., Xiao, L., and Oxenham, A. J. (**2012**). “Characterizing the dependence of pure-tone frequency difference limens on frequency, duration, and level,” Hear. Res. **292**, 1–13.

Mignault Goulet, G., Moreau, P., Robitaille, N., and Peretz, I. (**2012**). “Congenital amusia persists in the developing brain after daily music listening,” PLoS One , doi: 10.1371/journal.pone.0036860.

Miller, R. L., Schilling, J. R., Franck, K. R., and Young, E. D. (**1997**). “Effects of acoustic trauma on the representation of the vowel /ɛ/ in cat auditory nerve fibers,” J. Acoust. Soc. Am. **101**, 3602–3616.

Moerel, M., De Martino, F., and Formisano, E. (**2014**). “An anatomical and functional topography of human auditory cortical areas,” Front. Neurosci. **8**, 1–14.

Moore, B. C. J. (**1973**). “Frequency difference limens for short-duration tones,” J. Acoust. Soc. Am. **54**, 610–619.

Moore, B. C. J. (**1976**). “Comparison of frequency DL’s for pulsed tones and modulated tones,” Br. J. Audiol. **10**, 17–20.

Moore, B. C. J. (**2007**). *Cochlear Hearing Loss: Physiological, Psychological and Technical Issues*, Wiley, Chichester, 2nd ed.

Moore, B. C. J. (**2008**). “The role of temporal fine structure processing in pitch perception, masking, and speech perception for normal-hearing and hearing-impaired people," JARO - J. Assoc. Res. Otolaryngol. **9**, 399–406.

Moore, B. C. J. (**2014**). *Auditory Processing of Temporal Fine Structure: Effects of Age and Hearing Loss*,.

Moore, B. C. J., and Ernst, S. M. A. (**2012**). "Frequency difference limens at high frequencies: Evidence for a transition from a temporal to a place code," J. Acoust. Soc. Am. **132**, 1542–1547.

Moore, B. C. J., and Glasberg, B. R. (**1986**). "The Relationship Between Frequency Selectivity and Frequency Discrimination for Subjects with Unilateral and Bilateral Cochlear Impairments," In B. C. J. Moore and R. D. Patterson (Eds.), Audit. Freq. Sel. Plenum, New York.

Moore, B. C. J., and Glasberg, B. R. (**1987**). "Formulae describing frequency selectivity as a function of frequency and level, and their use in calculating excitation patterns," Hear. Res. **28**, 209–225.

Moore, B. C. J., and Glasberg, B. R. (**1989**). "Mechanisms underlying the frequency discrimination of pulsed tones and the detection of frequency modulation," J. Acoust. Soc. Am. **86**, 1722–1732.

Moore, B. C. J., Glasberg, B. R., and Hopkins, K. (**2006**). "Frequency discrimination of complex tones by hearing-impaired subjects: Evidence for loss of ability to use temporal fine structure," Hear. Res. **222**, 16–27.

Moore, B. C. J., Glasberg, B. R., Stoev, M., Füllgrabe, C., and Hopkins, K. (**2012**). "The influence of age and high-frequency hearing loss on sensitivity to temporal fine structure at low frequencies (L)," J. Acoust. Soc. Am. **131**, 1003.

Moore, B. C. J., and Moore, G. a. (**2003**). "Discrimination of the fundamental frequency of complex tones with fixed and shifting spectral envelopes by normally hearing and hearing-impaired subjects," Hear. Res. **182**, 153–163.

Moore, B. C. J., and Peters, R. W. (**1992**). "Pitch discrimination and phase sensitivity in young and elderly subjects and its relationship to frequency selectivity," J. Acoust. Soc. Am. **91**, 2881–2893.

Moore, B. C. J., and Sek, A. (**1992**). "Detection of combined frequency and amplitude modulation," J. Acoust. Soc. Am. **92**, 3119–3131.

Moore, B. C. J., and Sek, A. (**1994**). "Effects of carrier frequency and background noise on the detection of mixed modulation," J. Acoust. Soc. Am. **96**, 741–751.

Moore, B. C. J., and Sek, A. (**1995**). "Effects of carrier frequency, modulation rate, and modulation waveform on the detection of modulation and the discrimination of modulation type (amplitude modulation versus frequency modulation)," J. Acoust. Soc. Am. **97**, 2468–2478.

Moore, B. C. J., and Sek, A. (**1996**). "Detection of frequency modulation at low modulation rates: Evidence for a mechanism based on phase locking," J. Acoust. Soc. Am. **100**, 2320–2331.

Moore, B. C. J., and Skrodzka, E. (**2002**). "Detection of frequency modulation by hearing-impaired listeners: Effects of carrier frequency, modulation rate, and added amplitude modulation," J. Acoust. Soc. Am. **111**, 327–335.

Moore, B. C. J., Vickers, D. a., and Mehta, A. (**2012**). "The effects of age on temporal fine structure sensitivity in monaural and binaural conditions," Int. J. Audiol. **51**, 715–721.

Moore, B. C. J., Vickers, D. A., Plack, C. J., and Oxenham, A. J. (**1999**). "Inter-relationship between different psychoacoustic measures assumed to be related to the cochlear active mechanism," J. Acoust. Soc. Am. **106**, 2761–2778.

Mossbridge, J. A., Fitzgerald, M. B., O'Connor, E. S., and Wright, B. A. (**2006**). "Perceptual-learning evidence for separate processing of asynchrony and order tasks," J. Neurosci. **26**, 12708–12716.

Neff, D. L. (**1986**). "Confusion effects with sinusoidal and narrow-band noise forward maskers," J. Acoust. Soc. Am. **79**, 1519–1529.

Nelken, I., Rotman, Y., and Yosef, O. B. (**1999**). "Responses of auditory-cortex neurons to structural features of natural sounds," Nature **397**, 154–157.

Norman-Haignere, S., Kanwisher, N., and McDermott, J. H. (**2013**). "Cortical pitch regions in humans respond primarily to resolved harmonics and are located in specific tonotopic regions of anterior auditory cortex," J. Neurosci. **33**, 19451–19469.

Norman-Haignere, S. V., Albouy, P., Caclin, A., McDermott, J. H., Kanwisher, N. G., and Tillmann, B. (**2016**). "Pitch-responsive cortical regions in congenital amusia," J. Neurosci. **36**, 2986–2994.

Ochi, A., Yamasoba, T., and Furukawa, S. (**2014**). "Factors that account for inter-individual variability of lateralization performance revealed by correlations of performance among multiple psychoacoustical tasks," Front. Neurosci. **8**, 1–10.

Ortiz, J. A., and Wright, B. A. (**2009**). "Contributions of procedure and stimulus learning to early, rapid perceptual improvements," J. Exp. Psychol. Hum. Percept. Perform. **35**, 188–194.

Otsuka, S., Furukawa, S., Yamagishi, S., Hirota, K., and Kashino, M. (**2014**). "Interindividual variation of sensitivity to frequency modulation: Its relation with click-evoked and distortion product otoacoustic emissions," J. Assoc. Res. Otolaryngol. **15**, 175–186.

Otsuka, S., Furukawa, S., Yamagishi, S., Hirota, K., and Kashino, M. (**2016**). "Relation between cochlear mechanics and performance of temporal fine structure-based tasks," J. Assoc. Res. Otolaryngol. **17**, 541–557.

Oxenham, A. J. (**2012**). "Pitch perception," J. Neurosci. **32**, 13335–13338.

Oxenham, A. J. (**2013**). "Revisiting place and temporal theories of pitch," Acoust. Sci. Technol. **34**, 388–396.

Oxenham, A. J. (**2016**). "Predicting the perceptual consequences of hidden hearing loss," Trends Hear. **20**, 1–6.

Oxenham, A. J., Bernstein, J. G. W., and Penagos, H. (**2004**). "Correct tonotopic representation is necessary for complex pitch perception," Proc. Natl. Acad. Sci. U. S. A. **101**, 1421–1425.

Oxenham, A. J., Micheyl, C., and Keebler, M. V (**2009**). "Can temporal fine structure represent the fundamental frequency of unresolved harmonics?," J. Acoust. Soc. Am. **125**, 2189–2199.

Oxenham, A. J., Micheyl, C., Keebler, M. V, Loper, A., and Santurette, S. (**2011**). "Pitch perception beyond the traditional existence region of pitch," Proc. Natl. Acad. Sci. U. S. A. **108**, 7629–7634.

Palmer, A. R., and Russell, I. J. (**1986**). "Phase-locking in the cochlear nerve of the guinea-pig and its relation to the receptor potential of inner hair-cells," Hear. Res. **24**, 1–15.

Paraouty, N., Ewert, S. D., Wallaert, N., and Lorenzi, C. (**2016**). "Interactions between amplitude modulation and frequency modulation processing: Effects of age and hearing loss," J. Acoust. Soc. Am. **140**, 121–131.

Paraouty, N., and Lorenzi, C. (**2017**). “Using individual differences to assess modulation-processing mechanisms and age effects,” Hear. Res. **344**, 38–49.

Patterson, R. D. (**1976**). “Auditory filter shapes derived with noise stimuli,” J. Acoust. Soc. Am. **59**, 640–654.

Patterson, R. D., Uppenkamp, S., Johnsrude, I. S., and Griffiths, T. D. (**2002**). “The processing of temporal pitch and melody information in auditory cortex,” Neuron **36**, 767–776.

Penagos, H., Melcher, J. R., and Oxenham, A. J. (**2004**). “A neural representation of pitch salience in nonprimary human auditory cortex revealed with functional magnetic resonance imaging,” J. Neurosci. **24**, 6810–6815.

Peretz, I. (**2001**). “Brain specialization for music: New evidence from congenital amusia,” Neuroscientist **930**, 153–165.

Peretz, I. (**2013**). “The biological foundations of music: Insights from congenital amusia,” Psychol. Music Elsevier Inc., Third Edit., pp. 551–564.

Peretz, I. (**2016**). “Neurobiology of congenital amusia,” Trends Cogn. Sci. **20**, 857–867.

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

Peretz, I., Champod, A. S., and Hyde, K. (**2003**). “Varieties of musical disorders: The Montreal Battery of Evaluation of Amusia,” Ann. N. Y. Acad. Sci. **999**, 58–75.

Peretz, I., Cummings, S., and Dubé, M.-P. (**2007**). “The genetics of congenital amusia (tone deafness): a family-aggregation study,” Am. J. Hum. Genet. **81**, 582–588.

Peretz, I., Gosselin, N., Tillmann, B., Cuddy, L. L., Gagnon, B., Trimmer, C. G., Paquette, S., et al. (**2008**). “On-line identification of congenital amusia,” Music Percept. **25**, 331–344.

Peretz, I., Kolinsky, R., Tramo, M., Labrecque, R., Hublet, C., Demeurisse, G., and Belleville, S. (**1994**). “Functional Dissociations following Bilateral Lesions of Auditory Cortex,” Brain **117**, 1283–1301.

Peretz, I., Saffran, J., Schön, D., and Gosselin, N. (**2012**). “Statistical learning of speech, not music, in congenital amusia,” Ann. N. Y. Acad. Sci. **1252**, 361–366.

Pfeifer, J., and Hamann, S. (**2015**). “Revising the diagnosis of congenital amusia with the Montreal Battery of Evaluation of Amusia," Front. Hum. Neurosci. **9**, 1–15.

Plack, C. J., Barker, D., and Hall, D. a. (**2014**). "Pitch coding and pitch processing in the human brain," Hear. Res. **307**, 53–64.

Plack, C. J., and Oxenham, A. J. (**2005**). "The psychophysics of pitch," In C. J. Plack, A. J. Oxenham, R. R. Fay, and A. N. Popper (Eds.), Pitch Neural Coding Percept. Springer, New York, pp. 7–55.

Plack, C. J., Oxenham, A. J., Fay, R. R., and Popper, A. N. (Eds.) (**2005**). *Pitch*, Springer Handbook of Auditory Research, Springer New York, New York, NY, Vol. 24.

Plomp, R. (**1964**). "The ear as a frequency analyzer," J. Acoust. Soc. Am. **36**, 1628–1636.

Rance, G., and Starr, A. (**2015**). "Pathophysiological mechanisms and functional hearing consequences of auditory neuropathy," Brain , doi: 10.1093/brain/awv270.

Ritsma, R. J. (**1962**). "Existence region of the tonal residue I," J. Acoust. Soc. Am. **34**, 1224–1229.

Rose, J. E., Brugge, J. F., Anderson, D. J., and Hind, J. E. (**1967**). "Phase-locked response to low-frequency tones in single auditory nerve fibers of the squirrel monkey," J. Neurophysiol. **30**, 769–793.

Ross, B., Fujioka, T., Tremblay, K. L., and Picton, T. W. (**2007**). "Aging in binaural hearing begins in mid-life: Evidence from cortical auditory-evoked responses to changes in interaural phase," J. Neurosci. **27**, 11172–11178.

Ruggero, M. A. (**2013**). "Cochlear delays and the traveling waves: Comments on 'experimental look at cochlear mechanics,'" Audiology **33**, 131–142.

Ruggles, D., Bharadwaj, H., and Shinn-Cunningham, B. G. (**2011**). "Normal hearing is not enough to guarantee robust encoding of suprathreshold features important in everyday communication," Proc. Natl. Acad. Sci. U. S. A. **108**, 15516–15521.

Ruggles, D., Bharadwaj, H., and Shinn-Cunningham, B. G. (**2012**). "Why middle-aged listeners have trouble hearing in everyday settings," Curr. Biol. **22**, 1417–1422.

Saberi, K., and Hafter, E. R. (**1995**). "A common neural code for frequency- and amplitude-modulated sounds," Nature **374**, 537–539.

Schellenberg, G., Krysciak, A. M., and Campbell, R. J. (**2000**). "Perceiving Emotion in

Melody: Interactive Effects of Pitch and Rhythm Author," Music Percept. **18**, 155–171.

Schneider, P., Scherg, M., Dosch, H. G., Specht, H. J., Gutschalk, A., and Rupp, A. (**2002**). "Morphology of Heschl's gyrus reflects enhanced activation in the auditory cortex of musicians," Nat. Neurosci. **5**, 688–694.

Schreiner, C. E., and Langner, G. (**1988**). "Coding of temporal patterns in the central auditory nervous system," In G. M. Edelman, W. E. Gall, and W. M. Cowan (Eds.), Audit. Funct. Neurobiol. Basis Hear. John Wiley & Sons, New York, pp. 337–361.

Sek, A., and Moore, B. C. J. (**1995**). "Frequency discrimination as a function of frequency, measured in several ways," J. Acoust. Soc. Am. **97**, 2479–2486.

Sek, A., and Moore, B. C. J. (**1999**). "Discrimination of frequency steps linked by glides of various durations," J. Acoust. Soc. Am. **106**, 351–359.

Sergeyenko, Y., Lall, K., Liberman, M. C., and Kujawa, S. G. (**2013**). "Age-related cochlear synaptopathy : An early-onset contributor to auditory functional decline," J. Neurosci. **33**, 13686–13694.

Shackleton, T. M., and Carlyon, R. P. (**1994**). "The role of resolved and unresolved harmonics in pitch perception and frequency modulation discrimination," J. Acoust. Soc. Am. **95**, 3529–3540.

Shamma, S. A. (**1985**). "Speech processing in the auditory system II: Lateral inhibition and the central processing of speech evoked activity in the auditory nerve," J. Acoust. Soc. Am. **78**, 1622–1632.

Shamma, S., and Klein, D. (**2000**). "The case of the missing pitch templates: How harmonic templates emerge in the early auditory system," J. Acoust. Soc. Am. **107**, 2631–2644.

Shannon, R. V, Zeng, F.-G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). "Speech recognition with primarily temporal cues," Science (80-.). **270**, 303–304.

Sheft, S., Shafiro, V., Lorenzi, C., McMullen, R., and Farrell, C. (**2012**). "Effects of age and hearing loss on the relationship between discrimination of stochastic frequency modulation and speech perception," Ear Hear. **33**, 709–720.

Sheft, S., and Yost, W. A. (**1990**). "Temporal integration in amplitude modulation detection," J. Acoust. Soc. Am. **88**, 796–805.

Shera, C. A., Guinan, J. J., and Oxenham, A. J. (**2002**). "Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements," Proc. Natl. Acad. Sci. U. S. A. **99**, 3318–3323.

Sihvonen, A. J., Ripolles, P., Leo, V., Rodriguez-Fornells, A., Soinila, S., and Sarkamo, T. (**2016**). "Neural basis of acquired amusia and its recovery after stroke," J. Neurosci. **36**, 8872–8881.

Sihvonen, A. J., Ripolles, P., Leo, V., Rodriguez-Fornells, A., Soinila, S., and Sarkamo, T. (**2016**). "Neural basis of acquired amusia and its recovery after stroke," J. Neurosci. **36**, 8872–8881.

Smith, Z. M., Delgutte, B., and Oxenham, A. J. (**2002**). "Chimaeric sounds reveal dichotomies in auditory perception," Nature **416**, 87–90.

Spiegel, M. F., and Watson, C. S. (**1984**). "Performance on frequency-discrimination tasks by musicians and nonmusicians," J. Acoust. Soc. Am. **76**, 1690–1695.

Srulovicz, P., and Goldstein, J. L. (**1983**). "A central spectrum model: A synthesis of auditory-nerve timing and place cues in monaural communication of frequency spectrum," J. Acoust. Soc. Am. **73**, 1266–1276.

Stankov, L., and Horn, J. L. (**1980**). "Human abilities revealed through auditory tests," J. Educ. Psychol. **72**, 21–44.

Steiger, J. H. (**1980**). "Tests for comparing elements of a correlation matrix," Psychol. Bull.

Stelmachowicz, P. G., Lewis, D. E., Larson, L. L., and Jesteadt, W. (**1987**). "Growth of masking as a measure of response growth in hearing-impaired listeners," J. Acoust. Soc. Am. **81**, 1881–1887.

Strelcyk, O., and Dau, T. (**2009**). "Relations between frequency selectivity, temporal fine-structure processing, and speech reception in impaired hearing," J. Acoust. Soc. Am. **125**, 3328–3345.

Studebaker, G. A. (**1985**). "A 'rationalized' arcsine tranform," J. Speech Hear. Res. **28**, 455–462.

Taberner, A., and Liberman, M. (**2005**). "Response properties of single auditory nerve fibers in the mouse," J. Neurophysiol. **93**, 557–569.

Terhardt, E. (**1974**). "Pitch , consonance , and harmony," J. Acoust. Soc. Am. **55**, 1061–1069.

Thomson, R. S., Auduong, P., Miller, A. T., and Gurgel, R. K. (**2017**). “Hearing loss as a risk factor for dementia: A systematic review,” Laryngoscope Investig. Otolaryngol. **2**, 69–79.

Tillmann, B., Lévêque, Y., Fornoni, L., Albouy, P., and Caclin, A. (**2016**). “Impaired short-term memory for pitch in congenital amusia,” Brain Res. **1640**, 251–263.

Tillmann, B., Schulze, K., and Foxton, J. M. (**2009**). “Congenital amusia: A short-term memory deficit for non-verbal, but not verbal sounds,” Brain Cogn. **71**, 259–264.

Tyler, R. S., Wood, E. J., and Fernandes, M. (**1983**). “Frequency resolution and discrimination of constant and dynamic tones in normal and hearing-impaired listeners,” J. Acoust. Soc. Am. **74**, 1190–1199.

Viemeister, N. F. (**1979**). “Temporal modulation transfer functions based upon modulation thresholds,” J. Acoust. Soc. Am. **66**, 1364–1380.

Vuvan, D. T., Nunes-Silva, M., and Peretz, I. (**2015**). “Meta-analytic evidence for the non-modularity of pitch processing in congenital amusia,” Cortex **69**, 186–200.

Vuvan, D. T., Paquette, S., Mignault Goulet, G., Royal, I., Felezeu, M., and Peretz, I. (**2017**). “The Montreal Protocol for Identification of Amusia,” Behav. Res. Methods , doi: 10.3758/s13428-017-0892-8.

Wallaert, N., Moore, B. C. J., Ewert, S. D., and Lorenzi, C. (**2017**). “Sensorineural hearing loss enhances auditory sensitivity and temporal integration for amplitude modulation,” J. Acoust. Soc. Am. **141**, 971–980.

Wallaert, N., Moore, B. C. J., and Lorenzi, C. (**2016**). “Comparing the effects of age on amplitude modulation and frequency modulation detection,” J. Acoust. Soc. Am. **139**, 3088–3096.

Watson, C. S., Qiu, W. W., Chamberlain, M. M., and Li, X. (**1996**). “Auditory and visual speech perception: Confirmation of a modality-independent source of individual differences in speech recognition,” J. Acoust. Soc. Am. **100**, 1153–1162.

Wechsler, D. (**2011**). *Wechsler Abbreviated Scale of Intelligence - Second Edition*, Pearson, Bloomington, MN.

Whiteford, K. L., Kreft, H. A., and Oxenham, A. J. (**2017**). “Assessing the role of place and timing cues in coding frequency and amplitude modulation as a function of age," J. Assoc. Res. Otolaryngol. **18**, 619–633.

Whiteford, K. L., and Oxenham, A. J. (**2015**). "Using individual differences to test the role of temporal and place cues in coding frequency modulation," J. Acoust. Soc. Am. **138**, 3093–3104.

Whiteford, K. L., and Oxenham, A. J. (**2017**). "Auditory deficits in amusia extend beyond poor pitch perception," Neuropsychologia **99**, 213–224.

Wightman, F. L. (**1973**). "The pattern-transformation model of pitch," J. Acoust. Soc. Am. **54**, 407–416.

Wilbiks, J. M. P., Vuvan, D. T., Girard, P.-Y., Peretz, I., and Russo, F. A. (**2016**). "Effects of vocal training in a musicophile with congenital amusia," Neurocase **22**, 526–537.

Williamson, V. J., McDonald, C., Deutsch, D., Griffiths, T. D., and Stewart, L. (**2010**). "Faster decline of pitch memory over time in congenital amusia," Adv. Cogn. Psychol. **6**, 15–22.

Williamson, V. J., and Stewart, L. (**2010**). "Memory for pitch in congenital amusia: beyond a fine-grained pitch discrimination problem," Memory **18**, 657–669.

Woolf, N. K., Ryan, A. F., and Bone, R. C. (**1981**). "Neural phase-locking properties in the absence of cochlear outer hair cells," Hear. Res. **4**, 335–346.

Wright, B. A., and Fitzgerald, M. B. (**2001**). "Different patterns of human discrimination learning for two interaural cues to sound-source location," Proc. Natl. Acad. Sci. **98**, 12307–12312.

Wright, B. A., and Fitzgerald, M. B. (**2005**). "Learning and generalization of five auditory discrimination tasks as assessed by threshold changes," In D. Pressnitzer, A. de Cheveigné, S. McAdams, and L. Collet (Eds.), Audit. signal Process. Physiol. psychoacoustics, Model. Springer, New York, NY, pp. 509–515.

Wright, B. A., and Zhang, Y. (**2009**). "A review of the generalization of auditory learning," Philos. Trans. R. Soc. B Biol. Sci. **364**, 301–311.

Yost, W. A. (**1974**). "Discriminations of interaural phase differences," J. Acoust. Soc. Am. **55**, 1299–1303.

Zeng, F.-G., Nie, K., Stickney, G. S., Kong, Y.-Y., Vongphoe, M., Bhargave, A., Wei, C., et al. (**2005**). "Speech recognition with amplitude and frequency modulations,” Proc. Natl. Acad. Sci. **102**, 2293–2298.

Zhang, Y., Moore, D. R., Guiraud, J., Molloy, K., Yan, T.-T., and Amitay, S. (**2016**). “Auditory discrimination learning: Role of working memory,” (P. Allen, Ed.) PLoS One **11**, e0147320.

Zurek, P. M., and Formby, C. (**1981**). “Frequency-discrimination ability of hearing-impaired listeners,” J. Speech Lang. Hear. Res. **24**, 108.

Zwicker, E. (**1952**). “Die Grenzen der Horbarkeit der Amplituden-modulation und der Frequenzmodulation eins Tones,” Acustica **3**, 126–133.

Zwicker, E. (**1956**). “Die elementaren Grundlagen zur Bestimmung der Informationskapazität des Gehörs,” Acustica **6**, 365–381.

Zwicker, E. (**1970**). “Masking and psychological excitation as consequences of the ear’s frequency analysis,” In R. Plomp and G. F. Smoorenburg (Eds.), Freq. Anal. Period. Detect. Hear. Leiden: Sijthoff, pp. 376–394.

Appendices

Appendix A

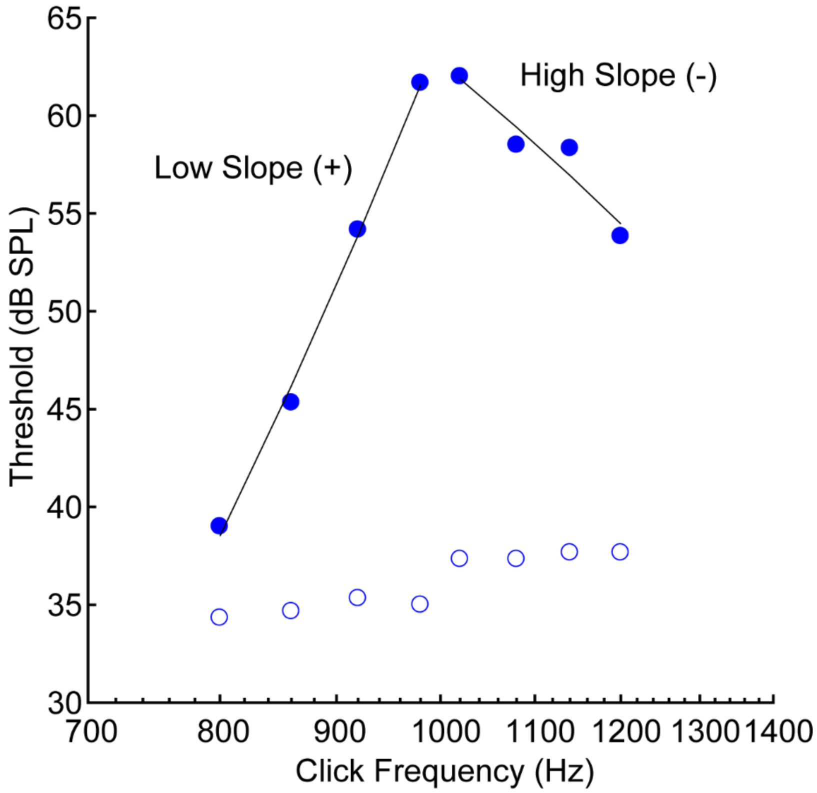

Figure A1

Example forward masking pattern. Absolute thresholds for the 20-ms tone in quiet (unfilled circles) and when preceded by a 500-ms, 1-kHz pure-tone forward masker (filled circles). The level of the tone must be much higher to be perceived when the tone is very close in frequency to the masker as opposed to when it is farther away. The slopes were calculated by conducting two linear regressions: one between the thresholds of the four lowest (low slope) and one between the four highest (high slope) signal frequencies.

Text A1 Quantifying Contributions from Absolute Thresholds, Age, and Sensitivity to AM

The correlations between the low slope and FM detection demonstrate a significant role for frequency selectivity, even after accounting for sensitivity to AM, age, and absolute thresholds. But this begs the question: How much do each of these factors directly contribute to FM sensitivity? Unlike correlations, which are bidirectional, multiple linear regression (MLR) is a conservative, directional approach to examining the amount of variance accounted for by each variable. Because many of the variables are correlated (e.g., low slope and absolute thresholds: $r = -.595$), the order the variables are entered into the model will affect the percentage of variance explained by each variable. We took the most conservative approach by entering the low and high slopes last, after all of the other variables. Factors known or believed to contribute to FM sensitivity (absolute thresholds in the measured ear, age, sensitivity to AM at the corresponding rate, low slope, and high slope, entered in this order) were entered into the model, fitted using the Ordinary Least Squares method. The full models, with all variables entered, explained 64.5% ($p < .0001$) and 55.2% ($p < .0001$) of the variance in slow and fast FM. When sequentially entering each variable, absolute thresholds accounted for 48.2% ($p = .014$) and 24.9% ($p = .165$, n.s.) of the variance for slow and fast FM, respectively (note that all of the p values here correspond to the significance of the variable in the full model). Because age is known to impair FM detection (He et al., 2007; Strelcyk and Dau, 2009; Grose and Mamo, 2012; Paraouty et al., 2016; Wallaert et al., 2016; Paraouty and Lorenzi, 2017; Whiteford et al., 2017), age was entered into the model second, accounting for an additional 4.5% (slow FM: $p = .031$) and 3.8% (fast FM: $p = .151$, n.s.) of the variance, while AM, entered third accounted for 2.1% (slow FM: $p = .212$) and 16.3% (fast FM: $p =.0001$). The low (slow FM: 9.7%, $p = .001$; fast FM: 10.1%, $p =.003$) but not the high slope (slow FM: 0%, $p = .757$; fast FM: .1%, $p =.810$) significantly contributed to the variance in sensitivity to FM at both rates, consistent with the partial correlation analysis (Fig. 4). Note that entering the slopes first, instead of last, into the regression means that the variance explained is the same as the squared correlations plotted in the left-most two panels of Fig. 3 (e.g., slow FM and low slope: 41.7%; fast FM and low slope: 25%), but the total variance explained in the full models is unaffected. Entering just the low slope and AM at the corresponding rate into the MLRs, which would be consistent with Zwicker's (1956) place model, accounts for 46.5% and 47% of the variance for slow and fast FM, with significant contributions from both the low slope (slow FM: $p < .0001$; fast FM: $p < .0001$) and AM (slow FM: $p = .048$; fast FM: $p < .0001$).

Appendix B

Text B1 Experimental Protocol

On average, all three experiments (including MBEA and audiograms) took approximately 4-5 hours in total, spread over 2-3 separate sessions, each of no more than 2 hours duration. Experiment 1 and the first part of experiment 2 (the two-interval task) were run on the first day, with the two-interval task run before experiment 1. The second part of experiment 2 was run on the second day. Experiment 3 was always run last. Most participants completed experiments 1-3 within the same week, although some took longer to complete all three experiments. Two controls and three amusics did not return to complete experiment 3.

Text B2 Individual Continuous FMDLs and AMDLs

There was considerable overlap between individual thresholds in amusics and controls for one- and two-interval FMDLs and AMDLs at slow (f_m = 4 Hz) and fast (f_m = 20 Hz) modulation rates (Fig. B1).

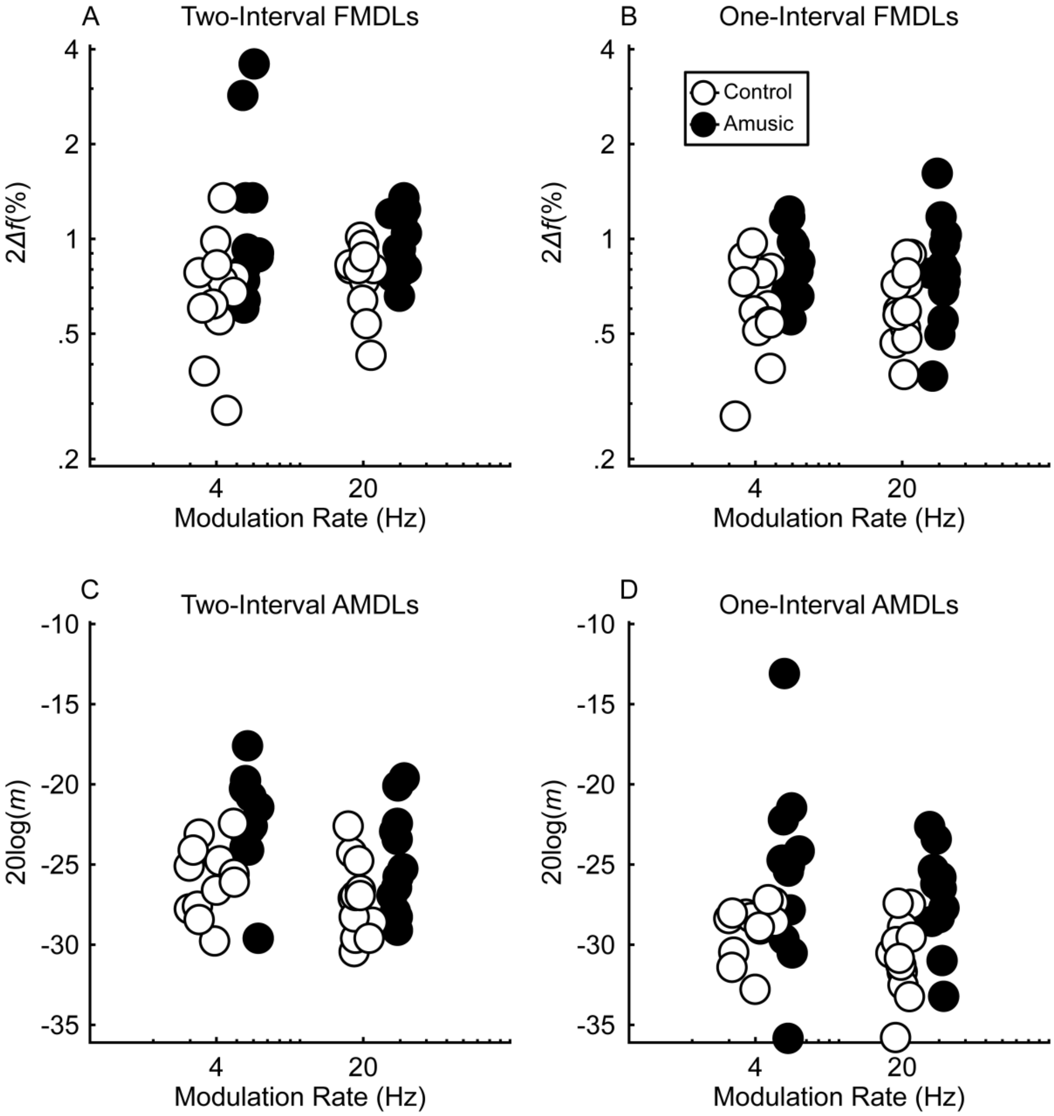

Figure B1

Individual FM and AM thresholds for slow (f_m = 4 Hz) and fast (f_m = 20 Hz) modulation rates in the two- (A and C) and one-interval (B and D) tasks from Experiment 2. FMDLs are plotted in panels A and B, while AMDLs are shown C and D. Open circles correspond to controls and filled circles correspond to amusics.

Appendix C

Text C1 Response Bias for One-Interval FM and AM Detection

The response bias (criterion), *c*, reflects the underlying tendency of participants to respond one way or the other (Macmillan and Creelman, 2005). The value for *c* is calculated via the following formula:

$$c = -\frac{1}{2}[z(H) + z(F)]$$

A positive *c* value indicates a bias towards responding "no," whereas a negative *c* value indicates a bias towards responding "yes." In order to calculate one response bias measure per subject, *H* was calculated as the total hit rate across all four signal conditions. Notably, because all of the signal and noise conditions were randomized within a block, only one false-alarm rate was estimated.

Both the amusics and the controls had a positive average response bias for all AM and FM conditions, indicating a tendency for each group towards responding "no" (Appendix Fig. C1). Interestingly, a 2 x 2 x 2 (Group x Modulation Type x Modulation Rate) ANOVA revealed a main effect of group [$F(1,22) = 10.7$, $p = .003$, $\eta_p^2 = .328$], with the amusics responding "yes" more often than controls. None of the other main effects or interactions approached significance (see Table C1).

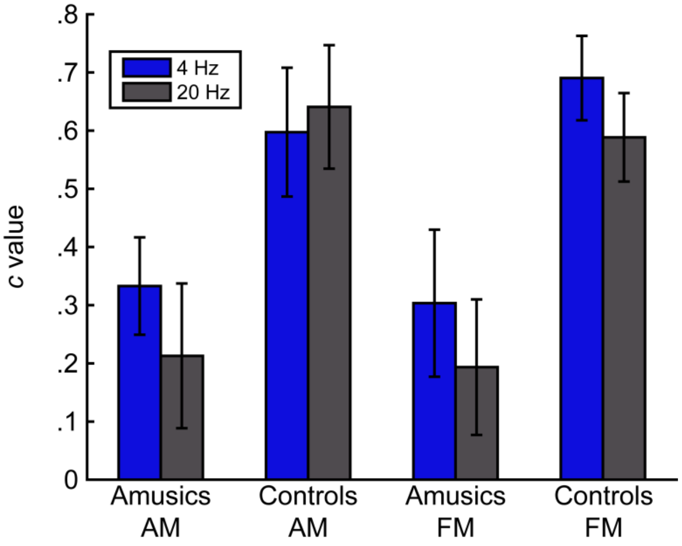

Figure C1

Average response bias for one-interval AM and FM tasks. Blue bars correspond to f_m = 4 Hz, while grey bars correspond to f_m = 20 Hz. The y-axis represents the *c* value, where greater *c* values correspond to more "no" responses. A *c* value of 0 means there was an equal proportion of "yes" and "no" responses. Error bars are ± 1 standard error of the mean.

Table C1

Results from the mixed-model ANOVA comparing average response bias (*c*) with Group (amusics vs. controls) as the between-subjects factor and Type (FM vs. AM) and Rate (slow vs. fast) as within-subjects factors. Significant effects are bolded and marked with *.

Effect	**DFn**	**DFD**	***F***	***p***	η_p^2
Group	1	22	10.7	**.003***	.328
Type	1	22	.001	.973	< .0001
Rate	1	22	1.844	.188	.077
Group x Type	1	22	.14	.712	.006
Group x Rate	1	22	.651	.428	.029
Type x Rate	1	22	.44	.514	.02
Group x Type x Rate	1	22	.58	.455	.026

Why might the amusics respond "yes" more often than the controls? The one-interval FM and AM tasks were individualized, meaning that the modulation indices were set so that sensitivity was equivalent across subjects. Further analysis of the overall *d'* (i.e., where the hit rate now corresponds to the overall hit rate across all of the signal conditions) indicated no main effect of group, modulation type, modulation rate, or any interaction (see Appendix Fig. C2 for group averages and Table C2 for statistics). This means that, overall, the changes in frequency and amplitude were, in fact, equally difficult to detect for both amusics and controls. Because task difficulty cannot explain the differences in response bias across groups, one possible explanation is that amusics utilized a different listening strategy as a direct result of the task instructions. It could be that amusics, knowing that they have difficulty with auditory tasks, set a looser criterion than controls when instructed to "listen carefully, as the modulation will be very subtle." Task instructions are well-known to influence response bias (Macmillan and Creelman, 2005). However, another possibility is that amusics always have a propensity to say "yes" more often than the controls when the tones are set to equal detectability, regardless of task instructions. Further experimentation is needed to determine whether amusics always have a tendency to respond "yes" to near-threshold signals that are equated in sensitivity relative to controls or whether this difference was due to group differences in the interpretation of task instructions.

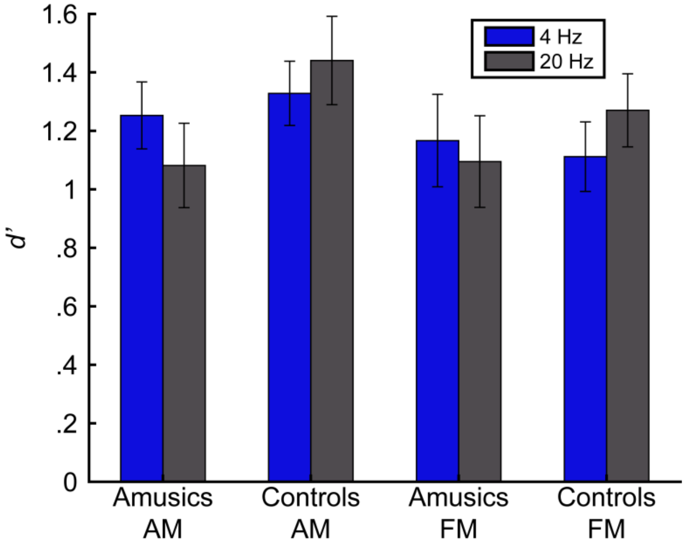

Figure C2

Average overall *d'*s for one-interval AM and FM tasks from Experiment 2. Blue bars correspond to f_m = 4 Hz, while grey bars correspond to f_m = 20 Hz. The y-axis represents the overall *d'* value, where greater *d'*s indicate better sensitivity. There was no significant difference in overall *d'* for task type (AM vs. FM) or group (amusics vs. controls) and no interaction. Error bars are ± 1 standard error of the mean.

Table C2

Results from the mixed-model ANOVA comparing overall sensitivity (*d'*) for the one-interval FM and AM tasks from Experiment 2. None of the main effects or interactions were significant, indicating that difficulty was equivalent across Group (amusics vs. controls) and was not differentially effected by Type (FM vs. AM) or Rate (slow vs. fast).

Effect	**DFn**	**DFD**	***F***	***p***	**η_p^2**
Group	1	22	1.61	.218	.068
Type	1	22	1.3	.267	.056

Rate	1	22	.006	.938	< .001
Group x Type	1	22	.604	.445	.047
Group x Rate	1	22	2.08	.163	.086
Type x Rate	1	22	.191	.667	.009
Group x Type x Rate	1	22	.025	.875	.001

Text C2 Discussion

While both amusics and controls had a tendency to respond “no” more often than “yes” in the one-interval FM and AM tasks, amusics responded “no” significantly less often than controls. The lower *c* but similar overall *d’* means that both amusics’ false alarm rate (tendency to respond “yes” when no modulation is present) and their hit rate was higher than controls in the one-interval tasks. The opposite trend has been found in previous *d’* analyses in amusia, with amusics tending to miss the signal more often than controls (Tillmann et al., 2009; Albouy et al., 2013b, 2015a; Henry and McAuley, 2013; Pfeifer and Hamann, 2015). The previous studies using measures of signal detection theory to calculate *c*, however, do not control for individual sensitivity to the signal. A higher miss rate is a perceptually relevant bias when the stimuli are designed to be subthreshold for amusics but not controls. In our experiment, the overall *d'* (i.e., where *H* corresponds to the overall hit rate across all of the signal conditions) indicated no main effect of group, modulation type, modulation rate, or any interaction. This means that, overall, the changes in frequency and amplitude were, in fact, equally difficult to detect for both amusics and controls. Once controlling for sensitivity, *d’*, we found the opposite trend in response bias from that reported in earlier studies, with a higher false alarm rate in amusics relative to controls. The shift in criterion in the amusic group may have been specifically related to instructional priming or may reflect a general tendency for amusics to respond “yes” more often than controls when sensitivity is equated between the two groups. In other words, participants in the amusic group believe that they perform poorly at auditory tasks, so when uncertain if a given trial is a signal or noise, they assume the trial is more likely to be a signal, perhaps due to being instructed that the modulation would be subtle.

Appendix D

Text D1

It is possible that changes in MBEA performance over time may be related to changes in response bias rather than changes in underlying sensitivity, as two previous studies have found that amusics have a greater tendency to respond “same” on the same/different subtasks of the MBEA (Henry and McAuley, 2013; Pfeifer and Hamann, 2015). Therefore, test 1 vs. test 2 MBEA performance was analyzed in terms of sensitivity (*d′*) and response bias (*c*) using signal detection theory (Macmillan and Creelman, 2005). As recommended by Macmillan and Creelman (2005, pg. 8), proportions of 1 were transformed to $1 - 1/(2N)$, where N corresponds to the total number of trials. Hits and false alarms were defined in the same manner as described by Pfeifer and Hamann (2015), so that higher response biases on the pitch subtasks indicates a greater propensity to respond “same” versus “different”. Average and individual *d′* and *c* are presented in Figs. D3 and D4.

A mixed-model ANOVA for global MBEA, using *d′* as the dependent variable, was conducted with time as a within-subjects factor and group and task as between-subjects factors. The results mirrored those from the when the analyses were conducted on RAU-transformed proportion correct: there was a significant main effect of time ($F_{1,35} = 58.7$, $p < 0.0001$, $\eta_p^2 = 0.627$) and a significant main effect of group ($F_{1,35} = 59.3$, $p < 0.0001$, $\eta_p^2 = 0.629$). The main effect of task was not significant ($F_{1,35} = 0.065$, $p = 0.8$, $\eta_p^2 = 0.002$), nor was the task by group ($F_{1,35} = 0.31$, $p = 0.581$, $\eta_p^2 = 0.009$), task by time ($F_{1,35} = 0.014$, $p = 0.907$, $\eta_p^2 = 0.0004$), time by group ($F_{1,35} = 2.99$, $p = 0.092$, $\eta_p^2 = 0.079$), or the three way interaction ($F_{1,35} = 0.666$, $p = 0.42$, $\eta_p^2 = 0.019$), indicating learning was not affected by training type or group. Analyzing just the *d′* performance of the pitch subtasks from the MBEA showed the same trends as the global MBEA analysis. There was a main effect of group ($F_{1,35} = 48.8$, $p < 0.0001$, $\eta_p^2 = 0.582$) and a main effect of time ($F_{1,35} = 27.8$, $p < 0.0001$, $\eta_p^2 = 0.443$) but no significant main effect of task ($F_{1,35} = 0.312$, $p = 0.58$, $\eta_p^2 = 0.009$) or any interactions (time x task: $F_{1,35} = 0.636$, $p = 0.43$, $\eta_p^2 = 0.018$; time x group: $F_{1,35} = 1.44$, $p = 0.238$, $\eta_p^2 = 0.04$; task x group: $F_{1,35} = 0.775$, $p = 0.385$, $\eta_p^2 = 0.022$; time x task x group: $F_{1,35} = 0.223$, $p = 0.640$, $\eta_p^2 = 0.006$). For both pitch and global MBEA performance, there were significant improvements in sensitivity for melody discrimination over time, and these improvements were not related to the training task or group.

Similarly, response bias (*c*) was entered into a mixed-model ANOVA for global MBEA, with time as the within-subjects factor and group and task as between-subjects factors. There was no significant main effect of group ($F_{1,35} = 0.357$, $p = 0.554$, $\eta_p^2 = 0.01$), time ($F_{1,35} = 0.075$, $p = 0.786$, $\eta_p^2 = 0.002$), and no interactions (time x task: $F_{1,35} = 0.378$, $p = 0.543$, $\eta_p^2 = 0.011$; time x group: $F_{1,35} = 2.91$, $p = 0.097$, $\eta_p^2 = 0.077$; task x group: $F_{1,35} = 0.162$, $p = 0.689$, $\eta_p^2 = 0.005$; time x task x group: $F_{1,35} = 0.738$, $p = 0.396$, $\eta_p^2 = 0.021$), indicating no difference in response bias between amusics and controls and no change in response bias with learning. Restricting analyses to just the pitch subtasks showed the same, non-significant trends across the board, with no main effect of group ($F_{1,35} = 0.05$, $p = 0.824$, $\eta_p^2 = 0.001$), time ($F_{1,35} = 0.361$, $p = 0.552$, $\eta_p^2 = 0.01$) or any interactions (time x task: $F_{1,35} = 0.219$, $p = 0.643$, $\eta_p^2 = 0.006$; time x group: $F_{1,35} = 2.93$, $p = 0.096$, $\eta_p^2 = 0.077$; task x group: $F_{1,35} = 0.032$, $p = 0.859$, $\eta_p^2 = 0.001$; time x task x group: $F_{1,35} = 0.092$, $p = 0.763$, $\eta_p^2 = 0.003$). Unlike previous studies that used much smaller samples of amusics (Henry and McAuley, 2013; Pfeifer and Hamann, 2015), there was no propensity for amusics to respond one way over another. The learning observed in the percent correct MBEA scores (Fig. 6.4) was therefore driven by increased sensitivity and not simply a change in response bias.

Text D2

Test 2 and follow-up MBEA scores were calculated in terms of *d′* and *c*, as described in Appendix Text D1 above. Average and individual sensitivity and bias at test 2 and follow-up are plotted in Figs. D7 and D8, along with test 1 scores for reference. The *d′* MBEA scores were entered into a mixed-model ANOVA, with time (test 2 vs. follow-up) as a within-subjects factor and group (amusics vs. controls) as a between-subjects factor. An analogous mixed-model ANOVA was run for response bias. Results for the sensitivity and response bias ANOVAs are presented in Table D4. Just as when results were analyzed on RAU-transformed proportion correct (main text), there was no change in MBEA *d′* performance between test 2 and one-year follow-up for either amusics or controls. Similarly, there was no significant change in response bias over time, and no significant difference in response bias between amusics and controls. Learning that occurred between test 1 and test 2 was thus maintained over a one year period for the subset of amusics and controls who participated in experiment 3 and was unrelated to response bias.

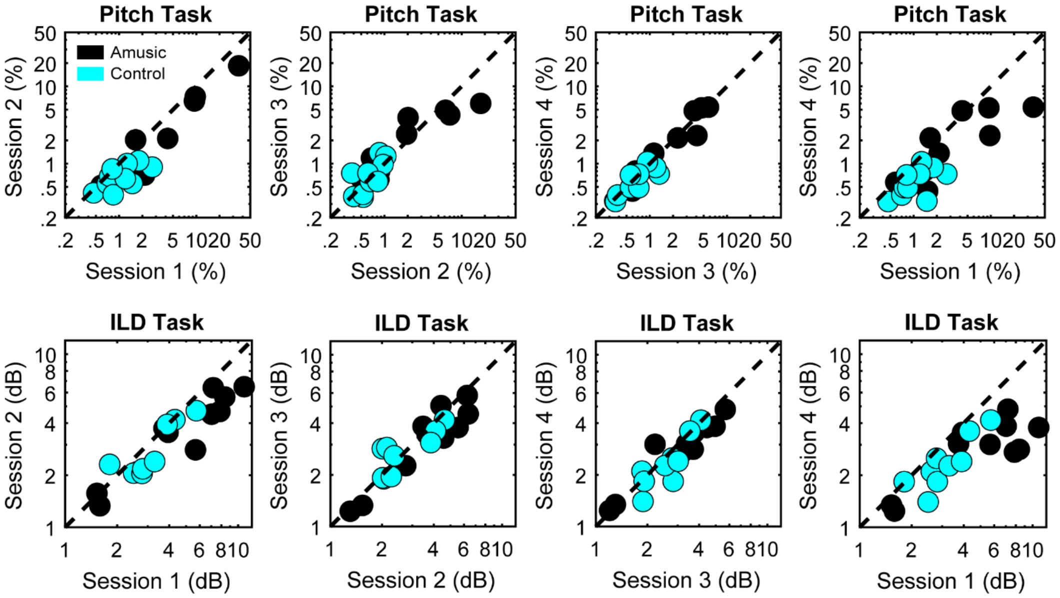

Figure D1

Individual training thresholds. Data points (black: amusics; cyan: controls) in the top row correspond to individual FDLs (%) at 500 Hz, while data points in the bottom row correspond to individual ILD thresholds (dB). The black-dashed line references no change in performance across sessions, so that data points below this line represent improvement in performance over time, while data points above this line represent poorer performance over time. Columns 1-3 plot performance for two consecutive training sessions, with the earlier session on the x-axis. Column 4 shows performance for the first (x-axis) and last (y-axis) training sessions.

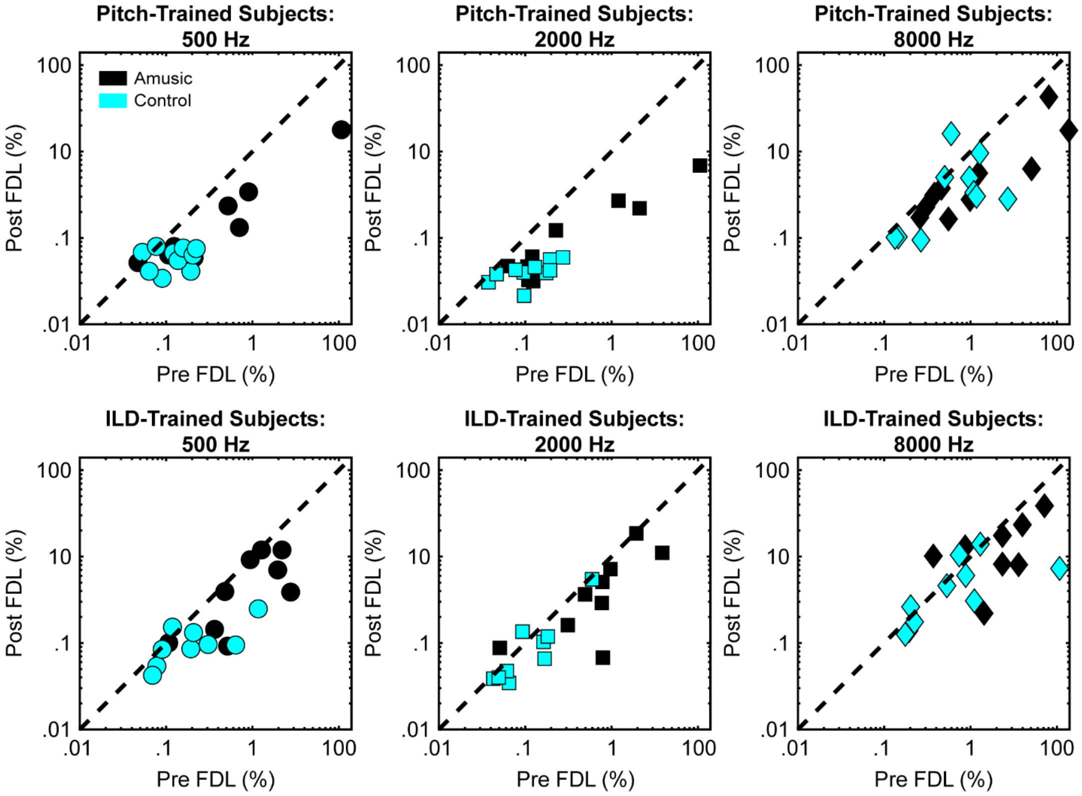

Figure D2

Individual pre vs. posttest pitch discrimination thresholds. Pretest (x-axis) versus posttest (y-axis) FDLs for amusics (black) and controls (cyan) who trained on pitch discrimination at 500 Hz (top row) versus ILD discrimination for white noise (bottom row). The black-dashed line references no change in performance across sessions, so that data points below this line represent improvement in performance over time, while data points above this line represent poorer performance over time (circles: 500 Hz; squares: 2000 Hz; diamonds: 8000 Hz).

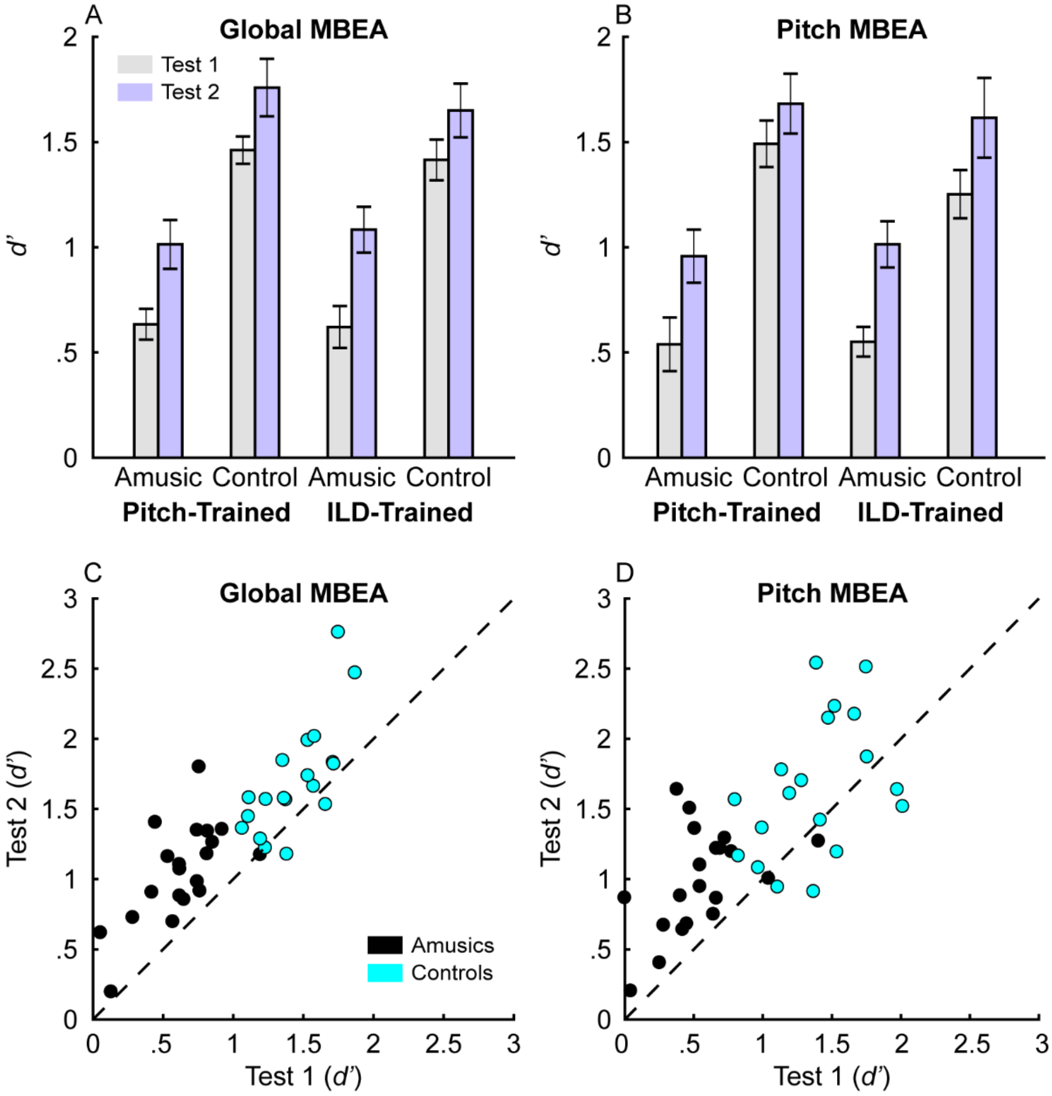

Figure D3

Average (A-B) and individual (C-D) test 1 (Panels A-B: grey bars; Panels C-D: x-axis) and test 2 (Panels A-B: purple bars; Panels C-D: y-axis) global (A and C) and pitch-subtask (B and D) MBEA scores, calculated in *d'*. The two sets of bars closest to the y-axis in Panels A and B correspond to the pitch-trained participants, while the two sets furthest from the y-axis correspond to the ILD-trained participants. Error bars represent ± 1 standard error of the mean.

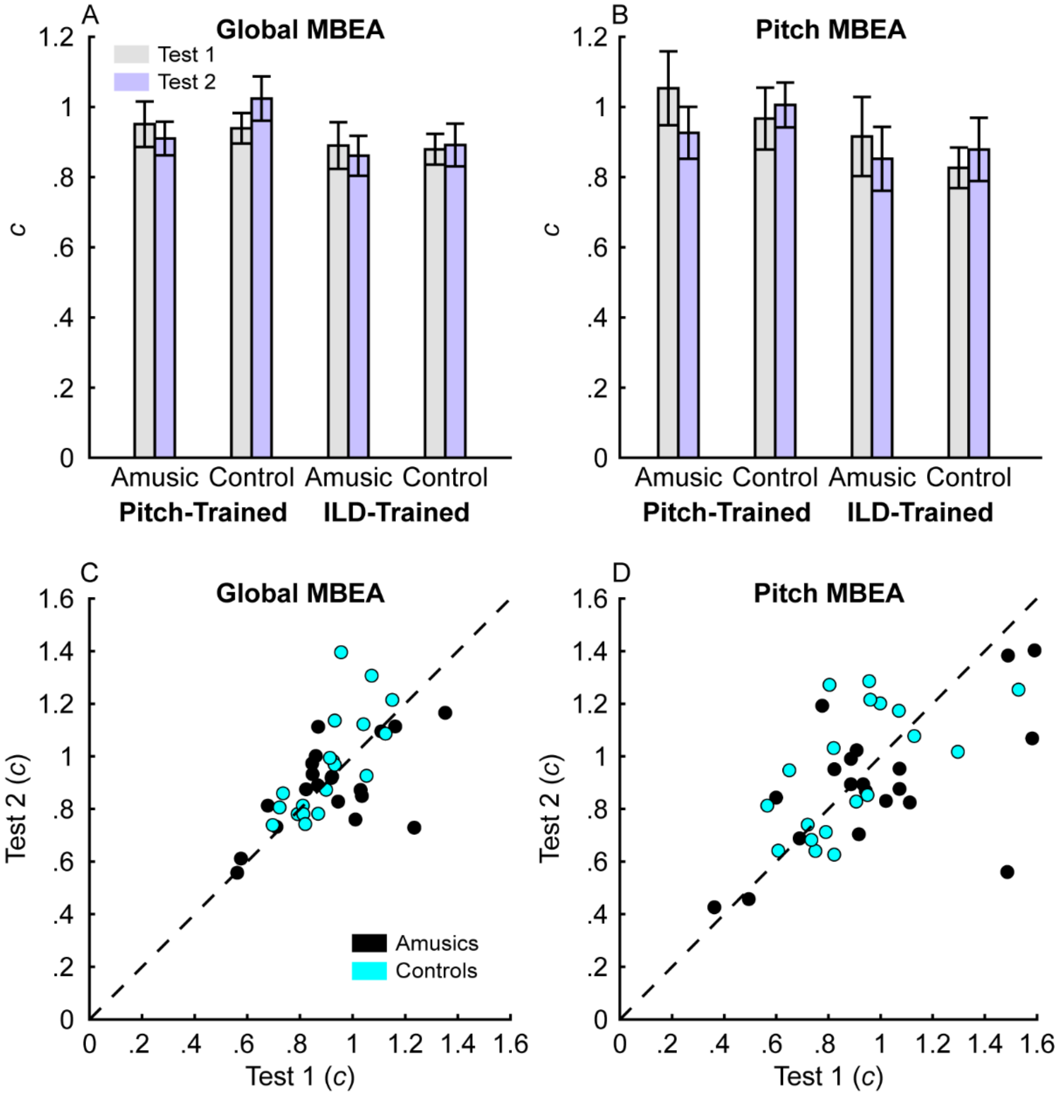

Figure D4

Average (A-B) and individual (C-D) test 1 (Panels A-B: grey bars; Panels C-D: x-axis) and test 2 (Panels A-B: purple bars; Panels C-D: y-axis) global (A and C) and pitch-subtask (B and D) MBEA response bias (*c*). The two sets of bars closest to the y-axis in Panels A and B correspond to the pitch-trained participants, while the two sets furthest from the y-axis correspond to the ILD-trained participants. Error bars represent ± 1 standard error of the mean.

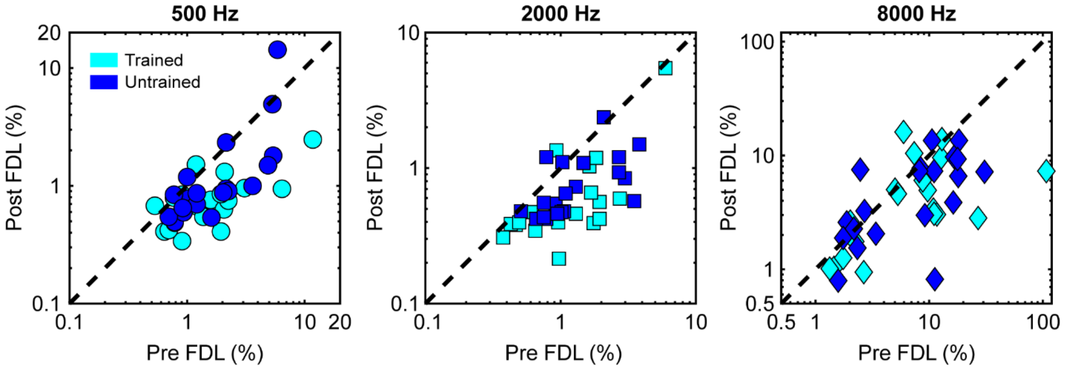

Figure D5

Individual pre (x-axis) and posttest (y-axis) pitch discrimination thresholds for trained (cyan) and untrained (blue) controls. The black-dashed line references no change in performance across sessions, so that data points below this line represent improvement in performance over time, while data points above this line represent poorer performance over time (circles: 500 Hz; squares: 2000 Hz; diamonds: 8000 Hz).

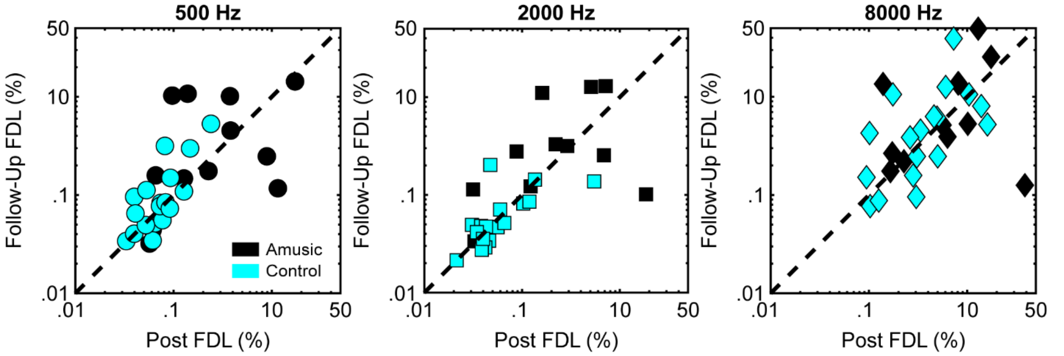

Figure D6

Individual post (x-axis) and follow-up (y-axis) pitch discrimination thresholds for trained controls (cyan) and amusics (black). The black-dashed line references no change in performance across sessions, so that data points below this line represent improvement in performance one year later, while data points above this line represent poorer performance one year later (circles: 500 Hz; squares: 2000 Hz; diamonds: 8000 Hz).

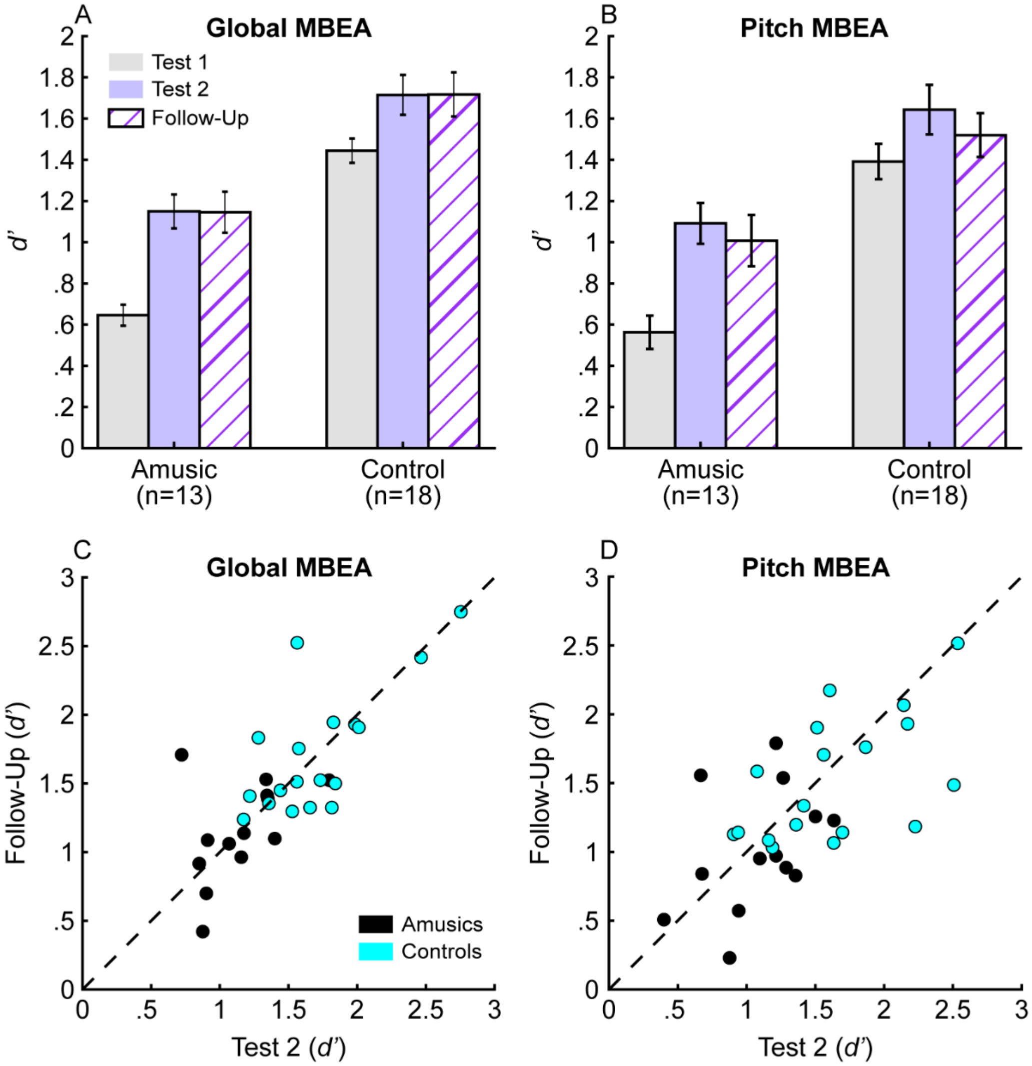

Figure D7

Test 1 (grey), test 2 (purple), and follow-up (striped) *d′* global (A and C) and pitch-related subtasks (B and D) of the MBEA. Average (A and B) and individual (C and D) MBEA results. Error bars represent ± 1 standard error of the mean. In panels C and D, the black-dashed line references no change in performance across sessions.

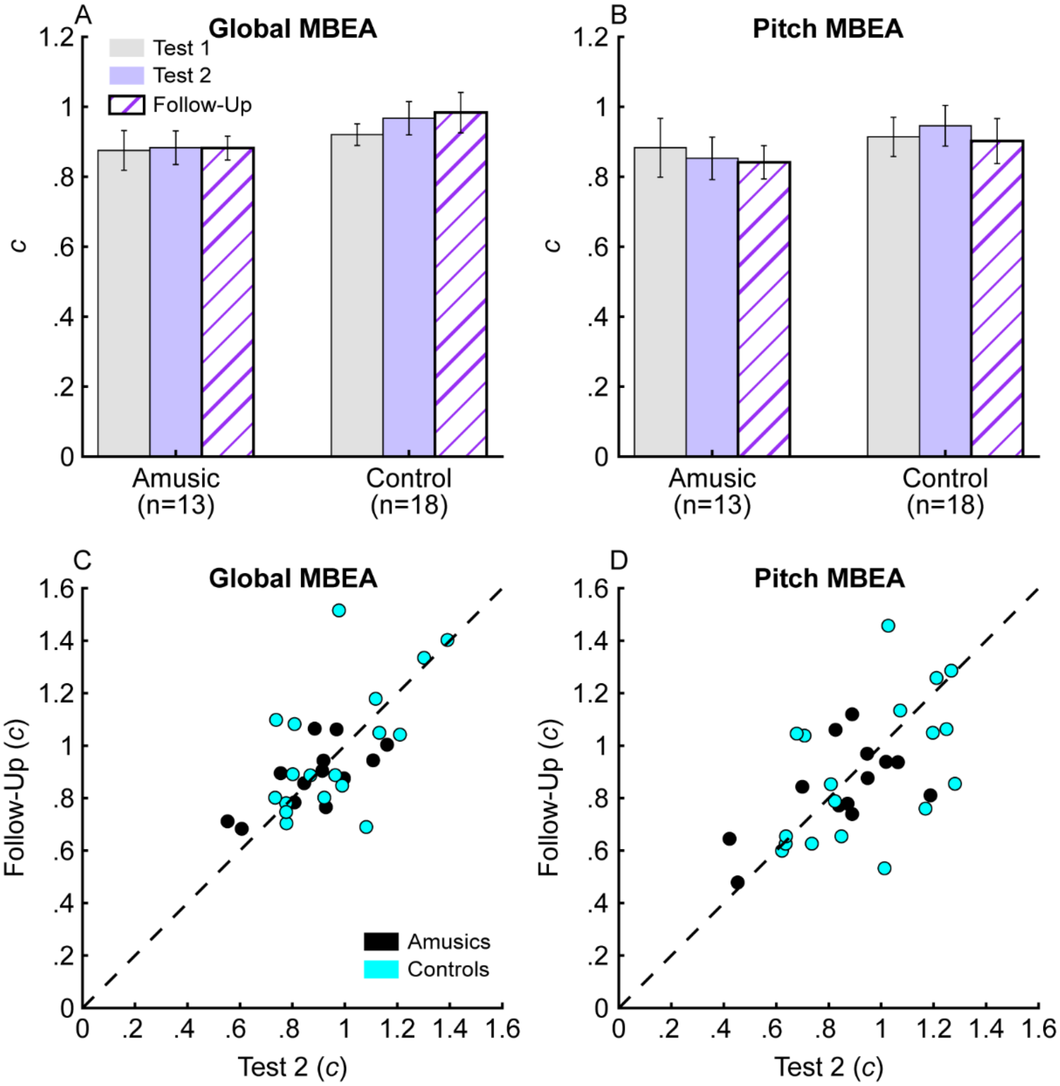

Figure D8

Test 1 (grey), test 2 (purple), and follow-up (striped) response bias for global (A and C) and pitch-related subtasks (B and D) of the MBEA. Average (A and B) and individual (C and D) response bias. Error bars represent ± 1 standard error of the mean. In panels C and D, the black-dashed line references no change in performance across sessions.

Table D1

Average demographics for experiment 2 (A) and experiment 3 (B).

1A. Experiment 2			
	Trained Controls (n=19)	**Untrained Controls (n=20)**	***p*-values**
Age (years)	29.7 (15)	25.1 (8.81)	0.244
Musical Experience (years)	1.32 (2.26)	2.13 (2.21)	0.266
Education (years)	16 (2.37)	15.7 (1.84)	0.717
IQ	113.1 (7.74)	112.5 (9.13)	0.84
Gender	11 females	15 females	-
Pitch MBEA (%)	84.85 (5.16)	85.11 (5.55)	0.882
Global MBEA (%)	86.84 (3.99)	87.03 (3.86)	0.884
Audiogram- Low (dB HL)	5.79 (3.47)	5.38 (4.35)	0.745
Audiogram- Medium (dB HL)	4.87 (4.52)	3.81 (4.13)	0.451
Audiogram- High (dB HL)	7.83 (7.86)	4 (3.73)	0.058

1B. Experiment 3			
	Returning Amusics (n=13)	**Returning Controls (n=18)**	***p*-values**
Age (years)	27 (14.1)	30.1 (15.5)	0.504
Musical Experience (years)	1.25 (1.22)	1.39 (2.3)	0.85
Education (years)	15.6 (1.29)	16.1 (2.33)	0.482
IQ	111.5 (9.28)	114.1 (6.4)	0.368
Gender	6 females	10 females	-
Pitch MBEA (%)	63.89 (3.92)	84.82 (5.31)	**< 0.0001***
Global MBEA (%)	67.08 (3.92)	86.67 (4.03)	**< 0.0001***
Audiogram- Low (dB HL)	5 (3.15)	5.29 (4.02)	0.829
Audiogram- Medium (dB HL)	4.04 (4.18)	5.15 (5.15)	0.533

Audiogram- High (dB HL)	4.9 (5.27)	7.72 (8.19)	0.29

MBEA = Montreal Battery of Evaluation of Amusia; dB = decibel; HL = hearing loss

Standard deviation is in parentheses.

p-values = independent-samples t tests.

Pitch MBEA = average percent correct on the three pitch-subtasks on the MBEA.

Global MBEA = average percent correct across all 6 subtasks on the MBEA.

All MBEA percentages were transformed to rationalized arcsine units (RAUs) for statistical analyses (Studebaker, 1985). Audiometric thresholds were grouped into three frequency bands (Low: .25 and .5 kHz; Medium: 1 and 2 kHz; High: 4 and 8 kHz) and averaged across ears.

Table D2

Pretest vs. posttest results for experiment 1 from a four-way mixed-design ANOVA with repeated measures, with between-subjects variables Group (trained amusics vs. trained controls) and Task (pitch training vs. ILD training).

Effect	**DF_{effect}**	**DF_{error}**	***F***	***p***	**η_p^2**
Time	1	34	61.5	**< 0.0001***	0.644
Frequency	1	54.5	133	**< 0.0001***	0.796
Group	1	34	12	**0.001***	0.261
Task	1	34	4.21	**0.048***	0.11
Time*Frequency	1.57	53.2	1.23	0.293	0.035
Time*Group	1	34	.864	0.359	0.025
Time*Task	1	34	3	0.092	0.081
Frequency*Group	1.6	54.5	5.4	**0.012***	0.137
Frequency*Task	1.6	54.5	2.6	0.094	0.071
Time*Frequency*Group	1.57	53.2	0.359	0.648	0.01
Time*Frequency*Task	1.57	53.2	3.07	0.067	0.083
Time*Task*Group	1	34	0.057	0.812	0.002
Frequency*Task*Group	1.6	54.5	0.773	0.44	0.022
Time*Frequency*Task*Group	1.57	53.2	0.3	0.688	0.009

Significant effects are bolded and marked with an asterisk.

Table D3

Experiment 1 MBEA results

3A. Global MBEA

Effect	**DF$_{\text{effect}}$**	**DF$_{\text{error}}$**	***F***	***p***	**η_p^2**
Time	1	35	100.9	**< 0.0001***	0.742
Group	1	35	92.2	**< 0.0001***	0.018
Task	1	35	0.641	0.429	0.018
Time*Group	1	35	5.91	**0.02***	0.145
Time*Task	1	35	1.13	0.295	0.031
Time*Group*Task	1	35	0.125	0.726	0.004

3B. Pitch MBEA

Effect	**DF$_{\text{effect}}$**	**DF$_{\text{error}}$**	***F***	***p***	**η_p^2**
Time	1	35	56.3	**< 0.0001***	0.617
Group	1	35	80.1	**< 0.0001***	0.696
Task	1	35	0.212	0.648	0.006
Time*Group	1	35	2.74	0.107	0.073
Time*Task	1	35	0.679	0.416	0.019
Time*Group*Task	1	35	0.124	0.727	0.004

Significant effects are bolded and marked with an asterisk.

Table D4

Experiment 3 *d'* (A-B) and *c* (C-D) MBEA results.

4A. Global MBEA (*d'*)

Effect	**DF$_{\text{effect}}$**	**DF$_{\text{error}}$**	***F***	***p***	**η_p^2**
Time	1	29	< 0.0001	0.992	< 0.0001
Group	1	29	19.4	**0.0001***	0.401
Time*Group	1	29	0.003	0.957	0.0001

4B. Pitch MBEA (*d'*)

Effect	**DF$_{\text{effect}}$**	**DF$_{\text{error}}$**	***F***	***p***	**η_p^2**
Time	1	29	1.58	0.218	0.052
Group	1	29	14	**0.001***	0.326
Time*Group	1	29	0.058	0.812	0.002

4C. Global MBEA (*c*)					
Effect	**DF$_{effect}$**	**DF$_{error}$**	***F***	***p***	**η_p^2**
Time	1	29	0.054	0.818	0.002
Group	1	29	2.1	0.158	0.068
Time*Group	1	29	0.072	0.719	0.002

4D. Pitch MBEA (*c*)					
Effect	**DF$_{effect}$**	**DF$_{error}$**	***F***	***p***	**η_p^2**
Time	1	29	0.444	0.511	0.015
Group	1	29	1.05	0.315	0.035
Time*Group	1	29	0.156	0.697	0.005

Significant effects are bolded and marked with an asterisk.