The Dissertation Committee for Kirsten Elisabeth Smayda Certifies that this is the approved version of the following dissertation:

Enhancing Older Adult Speech Perception in Challenging Listening Environments: Contextual Cues and Music Training

Committee:

Bharath Chandrasekaran, Supervisor

Alison Preston

David Schnyer

Karen Fingerman

Enhancing Older Adult Speech Perception in Challenging Listening Environments: Contextual Cues and Music Training

by

Kirsten Elisabeth Smayda

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

December 2017

Dedication

This work is dedicated to my parents, Beth N. and Greg J. Smayda. I hope you'll see this work as a foundational step in the direction of what I intend to accomplish during my time on this planet. Thanks for letting "me do."

Acknowledgements

I would like to acknowledge my funding sources, The National Institute on Aging pre-doctoral NRSA grant 5F31AG052308-02 to myself, National Institutes of Health grant 1R01DC013315 to Dr. Bharath Chandrasekaran, and National Institute on Aging grant 26167870 to Dr. David Schnyer and Dr. Darrell Worthy.

Enhancing Older Adult Speech Perception in Informational Noise: Contextual Cues and Music Training

Kirsten Elisabeth Smayda, Ph.D.

The University of Texas at Austin, 2017

Supervisor: Bharath Chandrasekaran

Normal aging is associated with difficulty understanding speech in adverse listening conditions and can lead to problems for the elderly such as social isolation, anxiety, depression, and diminished quality of life. A large literature suggests at least two types of noise can negatively interfere with speech intelligibility: energetic and informational noise. Energetic masking results when the noise spectro-temporally overlaps with the speech signal (e.g., near a construction site). Informational masking results when the noise contains information beyond spectro-temporal overlap with the speech signal (e.g., the 'cocktail' party situation). Cognitive processes have been implicated in mediating individual differences in speech-in-noise (SPIN) perception such as auditory working memory, attention, and processing speed; as well as perceptual processes such as temporal processing and gap detection. Importantly, the cognitive and perceptual subprocesses involved in accurate speech-in-noise perception also decline as we age. An expansive literature suggests that music training is positively associated with enhancements in not only SPIN processing, but also the perceptual and cognitive abilities supporting SPIN perception. Importantly, the causal effect of music training on older adult SPIN perception is poorly understood. The overarching goal of this thesis is to characterize the contextual and listener features that can improve older adult speech-in-

noise perception. The first paper in this dissertation explores the extent to which contextual cues, such as visual and semantic information, can aid in older adult speech-in-noise processing. In Paper 2, we examine the source of a musician advantage in learning novel speech categories. Using computational modeling we show that the musician advantage is due to both cognitive and perceptual processes. Paper 3 tests the extent to which age of onset of music training improves decision-making later in life. The broader implications of Papers 1 through 3 are explored in the General Discussion, which includes a proof-of-concept training study experimentally testing the effect of ten weeks of group piano lessons on older adult speech-in-noise processing. Preliminary results suggest that music training confers larger SPIN improvements relative to no training, and participants in the music training condition were more motivated to complete their training relative to those in the active control group.

Table of Contents

List of Tables

List of Figures

INTRODUCTION

Engaging in conversations in noisy environments is an unavoidable and difficult task we oftentimes find ourselves in. Often we reallocate attentional resources towards the speaker, read lips, or ask for the speaker to repeat him or herself when noise interferes with our conversation. Perhaps unsurprisingly, our ability to understand a conversation in noisy environments decreases with age. Noise creates a greater burden on the auditory processing systems by masking the speech signal, and the brain compensates for the reduced speech information by recruiting cognitive resources (Pichora-Fuller et al., 2016). As we age, however, our cognitive resources also decline (Park & Reuter-Lorenz, 2009), further complicating the already difficult task of understanding speech-in-noise (SPIN).

The implications of the age-related decline associated with hearing amongst a noisy background are diffuse and have dramatic downstream repercussions for older adults' wellbeing. An inability to hear speech-in-noise (SPIN) can lead to problems such as social isolation, depression, and lowered quality of life (Heine & Browning, 2002). Older adults are less likely to leave home and attend events if they know they will not be able to comprehend what is said to them. Hearing becomes a circular issue then, as older adults choose to avoid noisy environments by staying home, they are also reducing the amount and variety of speech signals entering the brain, therefore propagating the auditory system's decline. As the baby-boomer generation reaches older adulthood, they will have to live with hearing loss for longer than ever before. This is especially important because, as alluded to earlier, hearing loss is intimately related to quality of life (QoL; Ciorba, Bianchini, Pelucchi, & Pastore, 2012; Dalton et al., 2003; Raina, Wong, & Massfeller, 2004). In a meta analysis of articles published between 2000 and 2011,

Ciorba et al. (2012) found that presbycusis affects emotional reaction such as loneliness, isolation, dependence, anxiety, depression, and embarrassment; behavioral reactions including withdrawing; and cognitive reactions such as difficulty focusing, distracting thoughts, communication disorders, and confusion (Ciorba et al., 2012). Importantly, of the population with hearing loss, 39% perceive themselves to have an excellent QoL or very good physical health, whereas a staggering 68% of those without hearing loss perceive their QoL as excellent.

In a longitudinal population study, Dalton and colleagues tested the effect of hearing loss on communication difficulties, health, activities of daily life (ADLs) and instrumental ADLs in individuals aged 52-97. Findings suggested that participants ages 60 and older with moderate to severe hearing loss (> 40 dB pure tone HL in each ear) indicated impairment with both ADLs and instrumental ADLs significantly more than individuals without hearing impairment even after age, sex, education, arthritis, and visual acuity, and chronic diseases were controlled for (Dalton et al., 2003). It is worth noting that a larger proportion of females experience ADL impairment associated with sensory loss than males (Raina et al., 2004). ADLs include activities such as eating, walking, bathing, dressing oneself, etc., while preparing meals, taking prescriptions, using the telephone are examples of instrumental ADLs. Impairment to either ADLs or instrumental ADLs can necessitate the recruitment of extra resources, either from family or a care-taking service, potentially placing additional emotional and financial strains on a senior's entire support network. Because of the individual and societal burden hearing loss confers, it is imperative that we test ecologically valid and engaging training tools to limit the age-related decline in hearing, particularly understanding speech-in-noise. New methods to improve speech perception across the lifespan can have a dramatic effect on

not only the individual but society as well, as it can reduce the financial burden placed on the nation's healthcare system in managing older adult's declining hearing.

The types of noise in which listeners must resolve a masked speech-in-noise signal fall broadly under two types: *energetic masking*, which occurs when the noise spectro-temporally overlaps with the speech signal at the level of the cochlea (such as near a construction site), and *informational masking*, which occurs when the noise contains information beyond spectro-temporal overlap with the speech signal at the level of the brainstem, auditory cortex and cognitive regions (e.g., a loud talker near your conversation, (Kidd, Mason, Richards, Gallun, & Durlach, 2008). Prior research implicates cognitive processes such as auditory working memory (Akeroyd, 2008; Anderson, White-Schwoch, Parbery-Clark, & Kraus, 2013; Parbery-Clark, Strait, Anderson, Hittner, & Kraus, 2011) auditory attention (Strait & Kraus, 2011; Strait, Kraus, Parbery-Clark, & Ashley, 2010), and processing speed (Akeroyd, 2008; Pichora-Fuller, 2003); and perceptual processes such as temporal processing, gap detection, and frequency discrimination (Anderson et al., 2013; Jin, Liu, & Sladen, 2014; Pichora-Fuller, 2003) to be involved in mediating individual differences in SPIN processing. Our lab and others have provided converging evidence that working memory, such as the ability to hold onto information over a short duration and despite interefering information, is implicated more in informational than energetic noise conditions (Koelewijn, Zekveld, Festen, Rönnberg, & Kramer, 2012; Wong et al., 2009; Xie, Maddox, Knopik, McGeary, & Chandrasekaran, 2015; Zekveld, Rudner, Johnsrude, Heslenfeld, & Rönnberg, 2012). In addition, based on a functional magnetic resonance imaging study, older adults rely more on executive function-related brain regions, such as the prefrontal cortex and the fronto-parietal network and less on the primary auditory cortex relative to young adults during SPIN processing (Wong et al., 2009). Adults with

and without hearing loss also recruit the cingulo-opercular network during successful speech-in-noise recognition (Vaden, Kuchinsky, Ahlstrom, Dubno, & Eckert, 2015; Vaden et al., 2016). These findings can be considered within the framework of the decline-compensation hypothesis that suggests as primary sensory mechanisms in the brain decline in functioning, the older adult brain recruits a wider network of brain regions, including regions supporting general cognition, such as the prefrontal cortex as compensatory mechanisms (Cabeza, Anderson, Locantore, & McIntosh, 2002). As such, when the central auditory system declines, older adults may recruit cognitive processes to compensate during speech-in-noise processing (Kim & Oh, 2013)

In order to understand how aging affects speech-in-noise perception, it is important to understand the ways in which aging affects the supporting sensory system, namely the peripheral and central auditory systems. In the peripheral system, when the speech sound waves are transduced through the staples into the inner ear, the vibrations distort the basilar membrane in the region sensitive to the vibrating frequency. The hair cells on the basilar membrane convert the vibrations (a mechanical process) to an electrical signal, which is then conducted through the auditory nerve to the brain. The aging process results in several changes the aforementioned processes. Specifically, aging relates to a decrease in basilar membrane thickness (Bhatt, Liberman & Nadol, 2001; Howarth& Shane 2006), decreased outer hair cell density and inner hair cell loss (Wright, Davis, Bredberg, Ulehlova, & Spencer; Perez & Bao, 2011; Raphael, 2002), and such changes can be exacerbated by auditory trauma early in age (Kujawa & Liberman, 2006; however see Corwin, 1988 regarding sensory hair cell regeneration).

Aging also affects the central auditory system (see Humes et al., 2012). Age-related hearing loss is accompanied by a reduction in white matter fractional anisotropy, which is a measurement of “strength of directionality”, in the auditory pathway (Chang et

al., 2004; Lin et al., 2008) and an increase in fractional anisotropy in the inferior colliculus (Lutz et al., 2007). Other researchers have found differing accounts of aging's affect on fractional anisotropy in the central auditory system (Husain et al., 2011; Profant et al., 2014). Additionally, aging confers chemical (see Ouda, Profant, & Syka, 2015 for a review), and functional changes in the central auditory system. Temporal processing (Humes, Busey, Craig, & Kewley-Port, 2009), frequency discrimination (He, Dubno, & Mills, 1998), and inhibition (Caspary, 2005) all decline with age. As per the decline-compensation hypothesis (Cabeza et al., 2002), with age-related sensory loss will be accompanied by increased recruitment of cognitive resources (Kim & Oh, 2013; Wong, Ettlinger, Sheppard, Gunasekera, & Dhar, 2010). Given the profound decline in the central auditory system, improving older adult speech perception begs the question of whether it can be accomplished by training the cognitive processes sub-serving accurate speech-in-noise perception (however see Wayne, Hamilton, Jones Huyck, & Johnsrude, 2016).

Although older adults will recruit cognitive resources during speech-in-noise processing to support a declining sensory system, older adults' cognitive abilities also decline. Excitingly, however, research suggests training can improve cognition in older adults (Ball et al., 2002; Basak, Boot, Voss, & Kramer, 2008; Borella, Carretti, Riboldi, & De Beni, 2010; Stepankova et al., 2014; Willis et al., 2006). A recently developed theoretical framework provides a basis to examine the extent to which different training activities may influence cognitive abilities in older adults. Park et al. suggest that *productive engagement* in an activity that results in learning a novel skill and does not rely on previous knowledge, will produce greater cognitive gains than *receptive engagement* in an activity that builds upon existing skills or knowledge (Park et al., 2014; Park, Gutchess, Meade, & Stine-Morrow, 2007). One example of productive engagement

is a non-musician learning to play an instrument. Music-making is a rich, multimodal experience that requires cognitive processes such as remembering the notes you just played, and planning for the next measure of music to achieve an intended sound. All of these processes interact and ultimately result in a rewarding process of skill acquisition.

Long-term music training has been associated with cognitive benefits (see Benz, Sellaro, Hommel, & Colzato, 2016 for a review) such as enhanced auditory working memory (Bergman Nutley, Darki, & Klingberg, 2014; Pallesen et al., 2010; Parbery-Clark, Skoe, Lam, & Kraus, 2009; Parbery-Clark et al., 2011; Talamini, Carretti, & Grassi, 2016; Vandervert, 2015), attention (Ouimet, Foster, & Hyde, 2012; Rodrigues, Loureiro, & Caramelli, 2013; Strait et al., 2010), and perceptual benefits such as robustness of the neural encoding of speech sounds (Chandrasekaran, Krishnan, & Gandour, 2009; Kraus & Chandrasekaran, 2010; Wong, Skoe, Russo, Dees, & Kraus, 2007) and non-native speech discrimination (Gottfried, Staby, & Ziemer, 2004; Marie, Delogu, Lampis, Belardinelli, & Besson, 2011; Smayda et al., 2015). Musicians also outperform non-musicians in SPIN perception tasks across the lifespan (Parbery-Clark, Anderson, Hittner, & Kraus, 2012; Parbery-Clark et al., 2011; Strait & Kraus, 2011). These results have led researchers to suggest that music training may act as neural protection for age-related decline in not only cognitive ability, but also perceptual ability. Taken together, we posit that music training provides a productively engaging experience that will broadly enhance both the cognitive and perceptual processes associated with successful SPIN processing, thereby enhancing SPIN processing across both informational and energetic noise types.

Beyond the supportive cognitive and perceptual processes involved in speech-in-noise processing ability, music training has also been related to improved speech-in-noise perception ability both experimentally, and non-experimentally. Experimental research

suggests that two years of weekly music engagement including ensembles and lessons can lead to improved speech-in-noise perception in three- to five-year olds (Slater et al., 2015). Correlational research suggests that older adults with music training show preservative neural and behavioral effects against the typical age-related decline in speech-in-noise perception (Parbery-Clark et al., 2012, 2011). Importantly, this effect has not been studied experimentally in older adults (however see Zendel, West, Belleville, & Peretz, preprint). There could be a myriad of reasons throughout the lifespan that account for a musician-advantage in speech perception in older adults. For instance, an already acute hearing ability in childhood, parental suggestion, or an innate musical talent beyond hearing status may lead to a lifetime of music playing. Therefore, to limit pre-existing explanations for music training's observed relationship with speech-in-noise hearing ability, an experimental intervention study must be conducted in the aging population, where music training is the manipulated variable.

Overview of Dissertation Papers

Paper 1: Audio-Visual and Meaningful Semantic Context Enhancements in Older and Younger Adults

In the first paper (Smayda, Van Engen, Maddox, & Chandrasekaran, 2016), we examine the extent to which older adults and younger adults differ in their use of visual and semantic cues to understand speech amongst a noisy background. Visual cues, such as speaking in person rather than over the phone, can aid in the listener's ability to understand what is being said by allowing for lip-reading, and constraining the possibilities of the perceived auditory signal by matching to corresponding motor movements. For instance, within a myriad of distracting noises, the onset of a sentence can be more easily perceived when you notice the speakers' lips begin to move.

Meaningful semantic cues can also help listeners of all ages comprehend a speech signal by providing a context and expectations for what will be said. Although these two cues have been studied independently in the speech perception and aging literature, a recent study by Van Engen et al. (2014) suggests that a meaningful semantic context may increase the visual benefit received during speech perception in young adults (Van Engen, Phelps, Smiljanic, & Chandrasekaran, 2014). The present paper extends this research into the aging population to test the extent to which visual and semantic information interacts in younger and older listeners. Participants were presented with semantically meaningful and anomalous sentences at in the presence or lack of visual cues across a range of signal-to-noise ratios (SNR). Findings suggest that younger adults outperform older adults in the proportion of keywords identified across all combinations of visual and semantic contexts except for the most supportive – sentences with visual cues and semantically meaningful sentences in the easiest signal-to-noise ratios. When either the visual cues or the semantic cues were removed, older adults showed a significant decrease in keyword identification accuracy. We also replicated and extended Van Engen et al. (2014) by showing that both older and younger listeners receive more visual benefit in semantically meaningful relative to semantically anomalous contexts, and greater meaningful benefit in audio-visual relative to auditory-only contexts. Interestingly, older adults received the same amount of semantic and visual benefit as younger adults. These results highlight the importance of understanding the mechanism behind the age-related decline in speech perception because of the wide range of contexts in which older adults show difficulty.

PAPER 2: ENHANCED COGNITIVE AND PERCEPTUAL PROCESSING: A COMPUTATIONAL BASIS FOR THE MUSICIAN ADVANTAGE IN SPEECH LEARNING

In the first paper (Smayda et al. 2016), we show an age-related speech perception deficit across many listening contexts. In the second paper, we explore the mechanisms that have given rise to a musician enhancement in novel speech learning. This paper switches focus from speech intelligibility to speech sound learning; however, there are very similar processes involved. For instance, in the speech sound learning process, listeners must discriminate between novel speech sounds to form categorical representations of the sounds. Similarly, in speech perception, listeners must discriminate between sounds and categorize sounds into semantically and grammatically relevant language units such as words or sentences.

In the second paper (Smayda et al., 2015) fifteen musicians and fifteen non-musicians learned to categorize non-native speech sounds (i.e. Mandarin tones). Accuracy rates across learning were obtained for each group, and computational modeling of participant responses highlighted potential perceptual and cognitive mechanisms for group accuracy differences.

PAPER 3: BETTER LATE THAN NEVER (OR EARLY): MUSIC TRAINING IN LATE CHILDHOOD IS ASSOCIATED WITH ENHANCED DECISION-MAKING

In the third paper, we expand upon the music-related cognitive enhancements discussed in Paper 2 by testing the extent to which music training begun at different critical periods in childhood can improve later decision-making ability. Given that brain regions supporting decision-making begin maturing late in childhood, we hypothesized that adults who began learning to play an instrument later in childhood would show a decision-making advantage relative to adults who began playing an instrument early in childhood or not at all. In this study, young adults who began playing music early in

childhood, late in childhood, and not at all (early-trained musicians, ET; late-trained musicians, LT; non-musicians, NM respectively) performed a decision-making task. Late-trained musicians significantly outperformed both early-trained musicians and non-musicians. To better understand the mechanism of the LT advantage, we conducted computational modeling on participant responses and found that LT were less biased by recent outcomes and incorporated longer strings of outcomes when deciding among the choice options. These results tentatively suggest that music training may confer decision-making enhancements, and carry strong implications for the utility of music training in childhood.

Paper 1: Audio-Visual and Meaningful Semantic Context Enhancements in Older and Younger Adults[1]

INTRODUCTION

Perhaps the most critical feature of the complex auditory soundscape in which we live is the speech of the people around us. Although background noise often degrades the speech signal, noise rarely disrupts ongoing conversations. This is partly due to various contextual cues listeners use to mitigate the deleterious effects of background noise. Of particular importance in such situations are visual (Helfer, 1997, 1998; Sommers, Tye-Murray, & Spehar, 2005; Sumby & Pollack, 1954; Van Engen et al., 2014) and semantic (Bradlow & Alexander, 2007; Pichora-Fuller, Schneider, & Daneman, 1995; Smiljanic & Sladen, 2013; Van Engen et al., 2014) cues. While there are other cues listeners may use to understand speech in challenging situations, the modulatory influences of visual cues and semantic context are well-studied in young adults and known to be important for speech perception in noise (Bradlow & Alexander, 2007; Sumby & Pollack, 1954; Van Engen et al., 2014). Prior research in young adults suggests that individual differences in speech perception in noise can be partly explained by differential use of visual cues and semantic context. In the present study, we investigate the interaction of these two cues in enhancing speech perception across age groups.

As we age, our ability to hear declines. Lin and colleagues found that 63.1% of adults aged 70 years and older in the United States experience some form of hearing loss, as defined by the World Health Organization (Lin, Thorpe, Gordon-Salant, & Ferrucci, 2011). Even so, older adults, for the most part, maintain their ability to understand

---

[1] This paper has been previously published as Smayda, K. E., Van Engen, K. J., Maddox, W. T., & Chandrasekaran, B. (2016). Audio-Visual and Meaningful Semantic Context Enhancements in Older and Younger Adults. *PLOS ONE*, *11*(3), e0152773. https://doi.org/10.1371/journal.pone.0152773. KES performed the experiments, analyzed the data, partly contributed materials/analysis tools, and wrote the paper.

speech. This is partly due to increased reliance on non-auditory cues. A large body of research suggests that older adults receive the same amount of visual benefit as younger adults during speech perception (Helfer, 1997, 1998; Sommers et al., 2005; Stevenson et al., 2015). For instance, older adults and younger adults showed similar amounts of visual benefit when asked to repeat phonemes in multi-talker babble across a wide range of signal-to-noise ratios (SNRs; Power(signal)-Power(noise) in decibels; (Stevenson et al., 2015)). Older adults also showed similar amounts of visual benefit to young adults across SNRs during a word identification task, except at the most difficult SNR (-16 dB), where older adults received less visual benefit than young adults (Stevenson et al., 2015). Older adults also received the same amount of visual benefit as middle-aged adults when asked to repeat sentences presented audio-visually in speech envelope noise (Walden, Busacco, & Montgomery, 1993). Finally, Sommers, Tye-Murray and Spehar (2005), found that older adults and young adults used visual cues to the same extent when asked to identify consonants, words, and semantically meaningful sentences using audio-visual presentation in 20-talker babble noise when accuracy during the audio-only presentation was controlled for between age groups (Sommers et al., 2005).

Relative to visual cues, results are mixed as to whether or not older adults make greater use of semantic context compared to younger adults. Some studies have found that young and nor- mal-hearing older adults receive the same amount of performance boost in identifying final words of sentences in highly predictive semantic contexts relative to low-predictability con- texts. This has been found in both multi-talker babble (Frisina & Frisina, 1997) and spectrally-shaped noise (Dubno, Ahlstrom, & Horwitz, 2000). Pichora-Fuller and colleagues, on the other hand, found that older adults (with normal hearing or with hearing loss) benefitted more than younger adults from a high degree of semantic con- text when identifying sentence-final words in multi-talker babble

(Pichora-Fuller et al., 1995). Older adults' use of semantic cues to aid in speech perception also appears to make them more susceptible than young adults to false hearing: when a sentence is highly predictive of a particular final word but a non-predicted word is presented, older adults will report high confidence in their incorrect (but semantically predicted) responses (Rogers, Jacoby, & Sommers, 2012).

In the previous sections we discussed the role of visual and semantic cues in independently modulating speech intelligibility. A recent study conducted on younger adults shows that these cues interact in enhancing speech intelligibility. Van Engen and colleagues (2014) found that young adults received greater visual benefit when there was a strong semantic context relative to a weak semantic context during a speech-in-noise perception task (Van Engen et al., 2014). One possible explanation for their findings is that visual cues can be used more effectively when the set of possible words has already been significantly constrained by semantic context. Critically, the extent to which the semantic context may increase the visual benefit garnered during speech perception has never been studied in older adults. Given the literature suggesting that older adults may rely more strongly on semantic context than younger adults, it is important to examine whether semantic context modulates older adults' ability to use visual cues during speech perception.

To explore these issues, we presented older adults (ages 60–90) and young adults (ages 18– 35) with sentences that were either semantically meaningful or anomalous in both audio-only and audio-visual modalities in speech-shaped noise (SSN) across a range of signal-to-noise ratios. Our analyses were split into two sections: 1) keyword identification accuracy, and 2) relative visual and meaningful benefit. In our first set of analyses, we compared accuracy of key- word identification across all SNRs (0, -4, -8, -12, -16) and semantic contexts (meaningful and anomalous) in audio-only and audio-

visual modalities separately. In keeping with previous studies, we predicted that young adults would achieve higher accuracy than older adults across SNRs, that audio-visual presentation would enhance accuracy relative to audio-only presentation, and that meaningful semantic context would support higher accuracy relative to an anomalous semantic context.

In our second set of analyses, we used the keyword identification accuracy results to com- pare the relative benefit received from visual cues and semantic cues across age groups. For these analyses, we restricted our analyses to the -8 and -12 SNRs on the basis of Ross et al. (2006), which found maximal AV benefit at -12 SNR relative to higher or lower SNRs (Ross, Saint-Amour, Leavitt, Javitt, & Foxe, 2006). Using intermediate SNRs (-8 and -12 SNR) circumvents ceiling or floor constraints in measuring visual benefit. We hypothesized that young adults would benefit more from visual cues in meaningful contexts relative to anomalous contexts, as was found in Van Engen, Phelps, Smiljanic, and Chandrasekaran (2014; Van Engen et al., 2014). Similarly, we predicted that young adults would use semantic cues to a greater extent in audio-visual conditions as was suggested by Van Engen et al. (2014) in their speech-shaped-noise conditions. In older adults, there are two possible outcomes for the extent to which they use visual cues in semantically meaningful relative to semantically anomalous contexts. It is possible that because older adults rely heavily on semantic context during speech perception, they will rely on visual cues more in the anomalous con- text relative to the meaningful context, which would be indicated by higher visual benefit scores in the anomalous context relative to the meaningful context, or no difference in visual benefit between the contexts at all. Conversely, like younger adults, older adults might use visual cues more when a semantically meaningful context is available to constrain the set of possible auditory candidates for the listener to perceive. Lastly, we expect that older adults, similar to

young adults, will use semantic cues more in audio-visual conditions relative to audio-only conditions given the benefit older adults receive from audio-visual and semantically meaningful presentation, when studied independently.

MATERIALS AND METHODS

Participants

Forty-five young adults (ages 18–35, average age = 21.8) and thirty-three older adults (ages 60– 90, average age = 66.8) were recruited from the University of Texas at Austin community. All participants passed a hearing threshold test (PTA < 40 dB over 500, 1000, 2000, and 4000 Hz) and had no known psychiatric disorder. The Institutional Review Board at The University of Texas approved the experiment and materials, and written consent was obtained from each participant.

Neuropsychological Testing

Older adults were administered a series of standard neuropsychological tests in order to quantify their cognitive abilities. These included measures of memory (California Verbal Learning Test, CVLT; (Delis, Kramer, Kaplan, & Ober, 1987)), attention (Wechsler Adult Intelligence Scale, Third Edition (WAIS-III), Digit Span and Vocabulary (Wechsler, D., 1997)), and executive functioning (Trail Making Test A & B (Lezak, 1995); FAS and Wisconsin Card Sorting Task (WCST) (Heaton, Chelune, Talley, Kay, & Curtiss, 1993); Stroop Interference (Stroop, 1935)). Young adults were tested with a shortened series of neuropsychological tests including the WAIS-III Digit Span, Vocabulary, and Stroop Interference.

Testing occurred during a prescreening session, and all scores were converted to age-normalized Z-scores based on each test's standards. To ensure that our participants were within the "normal" range of neuropsychological ability, participants who scored

more than two standard deviations below the mean for at least one test in each measurement group (memory, attention, and executive functioning) were not included in the study. Z-scores and demographic information are presented in Table 1 for the older adults and in Table 2 for the young adults.

Table 1. Neuropsychological test results and demographic information for older adults.

Measure	Mean (*SD*)	Range
Neuropsychological Test		
WAIS Vocabulary	0.86 (0.82)	-0.7 to 2.3
Digit Span	0.17 (0.79)	-1.3 to 1.7
CVLT Immediate Recall (Free)	0.83 (0.95)	-1.0 to 2.5
CVLT Delayed Recall (Free)	0.85 (0.85)	-1.0 to 2.0
CVLT Immediate Recall (Cued)	0.56 (0.86)	-1.5 to 1.5
CVLT Delayed Recall (Cued)	0.70 (0.91)	-1.0 to 2.0
CVLT Recognition False Positives	-0.50 (0.65)	-1.0 to 1.5
CVLT Recognition True Positives	-0.08 (0.99)	-3.0 to 1.0
FAS	0.13 (1.04)	-2.0 to 2.8
TMT-A	-0.44 (0.87)	-1.7 to 1.9
TMT-B	-0.42 (0.49)	-1.1 to 1.0
WCST Errors	0.44 (0.93)*	-1.4 to 2.5*
WCST Perseveration	0.50 (0.76)*	-0.9 to 2.5*
Stroop Interference	0.37 (0.88)	-1.0 to 3.0
Demographic Information		
Age (years)	66.82 (4.22)	60 to 75
Years of Education	15.97 (2.86)	10 to 25

* Missing one participant's data

Table 2. Neuropsychological test results and demographic information for younger adults.

Measure	Mean (*SD*)	Range
Neuropsychological Test		
WAIS Vocabulary	0.84 (1.03)	-1.3 to 3.0
Digit Span	0.04 (0.92)	-1.3 to 2.3
Stroop Interference	0.76 (0.86)	-1.2 to 2.6
Demographic Information		
Age (years)	21.84 (3.64)	18 to 32
Years of Education	14.76 (2.29)	12 to 22

Stimuli

Target Sentences

One 22-year-old native English-speaking male was video-recorded producing both semantically meaningful and semantically anomalous sentences in a sound-attenuated stage at The University of Texas at Austin. The speaker, who read the sentences from a teleprompter, was instructed to speak in a conversational manner. He was recorded using a Sony PMW-EX3 studio camera. The anomalous sentences were derived from the Syntactically Normal Sentence Test (Nye & Gaitenby, 1974), and the meaningful sentences were adapted from the Basic English Lexicon (Calandruccio & Smiljanic, 2012; Van Engen et al., 2014). Both sets of sentences contained four keywords per sentence.

The video recording was processed through a Ross crosspoint video switcher and recorded on an AJA Pro video camera. The audio was recorded using an Audio Technica AT835b microphone at a sampling rate of 48000 Hz. The video and audio were segmented and separated using Final Cut Pro, and the audio tracks were equalized for RMS amplitude using Praat (Boersma & Weenink, 2009). For the audio-visual stimuli, the leveled audio was reattached to the video using Final Cut Pro. **Figure 1** provides a schematic of the experimental variables. The individual pictured in **Figure 1** of this manuscript has given written informed consent (as outlined in PLOS ONE consent form) to publish these details.

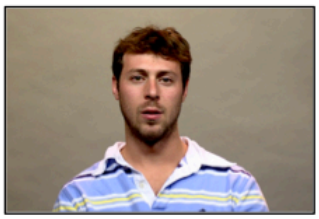
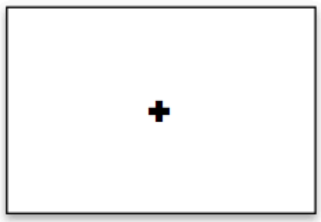
Context	Modality
Meaningful: The **hungry dog ate** the **food**.	Audio-visual 
Anomalous: The **good tree set** the **hair**.	Audio-only 
Bold words indicate keyword.	

Figure 1. Schematic of stimuli presentation and experimental variables.

Mixing noise and target sentences

The speech-shaped noise (SSN) was created by filtering white noise to match the long-term average spectrum of the 80 sentences spoken by the speaker. The RMS amplitude of the noise was scaled using Praat (Boersma & Weenink, 2009) to produce SNRs of 0, -4, -8, -12, and -16 dB when mixed with the target speech. The level of the target speech, therefore, was consistent across the experiment. In each trial, 500 ms. of noise preceded the target sentence and 500 ms. of noise followed the target sentence.

Procedure

The speech-in-noise task was run using E-prime 2.0 in a quiet testing room. Participants viewed stimuli on a 1280 x 1024 pixel computer screen and listened to stimuli at a comfortable level through Sennheiser HD 280 Pro headphones. Participants were instructed that they would hear each target sentence preceded by one half second of noise, and that they should type the target sentence using the keyboard provided. Each trial was participant-initiated. During audio-only trials, a black fixation cross against a white background was presented on the computer monitor. During audio-visual trials, the video clip covered the entire computer screen.

Participants were presented with a total of 80 sentences across two blocks. Each block contained either meaningful sentences or anomalous sentences, and the order of these blocks was counter-balanced across participants. Within each block, participants were presented with 20 audio-only (AO) sentences and 20 audio-visual (AV) sentences with 5 levels of SNR (0, -4, -8, -12, -16). Therefore, there were 4 sentences for each context, SNR, and modality combination. Each sentence was trial-unique, that is, if a sentence were presented in the audio-only modality, it would not be presented again in the audio-visual modality. Modality and SNR were randomly presented as well. Therefore, while the presentation order of the semantic context blocks was

counterbalanced between subjects, within each block every participant experienced a different order and presentation modality of a given target sentence.

Each sentence's four keywords were hand-scored as correct or incorrect. Misspellings of the target words were acceptable as long as they did not produce another English word or change the tense or pluralization. For instance, if the target word was "plate," "late" or "plates" would be counted as incorrect, but "platee" would be counted as correct. Because both groups' accuracy reached floor performance (accuracy rate less than .15 correct) at -16 SNR, we excluded that condition from our analyses.

Analyses

Results from the current study are presented in two analyses: keyword identification accuracy and relative benefit.

Keyword Identification Analyses

Keyword identification accuracy for every participant's response to each sentence was analyzed using a generalized linear mixed-effects logistic model for audio-only and audio-visual modalities separately. For both analyses, accuracy on each of the four keywords per sentence served as the dependent variable in our models. Fixed effects included group (older or younger adults), context (meaningful or anomalous), SNR, and their interactions. Group and context were treated as categorical variables, and SNR, which was mean-centered, was treated as a continuous variable. The model included by-subject and by-sentence random intercepts (Barr, Levy, Scheepers, & Tily, 2013). Lastly, pure-tone averages (PTA) of each participant were added as a covariate to the model:

$$Response \sim PTA + Group \times Context \times SNR + (1|Subject) + (1|Sentence)$$

Analyses for keyword identification were performed using the glmer function, which fits a generalized linear mixed-effects model on independent variables that are

binomial (1 or 0), and analyses for relative benefit were performed using the lmer function, which fits a linear mixed-effect model on independent variables that are not binomial. Both lmer and glmer are from the lme4 package in R (Bates, Maechler, & Bolker, 2012). In addition, we utilized the contr.sdif contrast-setting function from the MASS package in R in order to test levels of each factor against one another (Venables & Ripley, 2002). To reduce the risk of over-fitting the data, we systematically removed non-significant fixed effects and their interactions, and compared the progressively simpler model to the more complex model using the likelihood ratio (Baayen, Davidson, & Bates, 2008). Only the fixed effect estimates (β), standard errors of the estimates (SE), and estimates of significance (Z and p values) from the simplest, best-fitting model are reported for the keyword identification analyses.

Relative Benefit Analyses

To measure the benefit each participant received from either visual cues or semantic cues, we conducted our analysis at both -12 and -8 SNR. We began at -12 SNR because it has been identified as a "sweet spot" for audiovisual integration (Ross et al., 2006). Therefore, any differences in visual benefit due to context or group designation will likely show at this SNR. Additionally, we conducted the same analysis at -8 to test the generalizability of any pattern at an easier SNR. Similarly, we restricted our analysis of relative meaningful benefit to -12 SNR and -8 SNR.

To obtain each participant's relative visual benefit (RVB) score for both meaningful and anomalous contexts, we used the formula:

$$RVB = \frac{AV - AO}{1 - AO}$$

where AV represents the participant's average accuracy during audio-visual modality presentation and AO represents the participant's average accuracy during audio-

only modality presentation. To investigate the effect of context and group designation on relative visual benefit score, we ran a linear mixed-effects model with relative visual benefit as the dependent variable of interest, and group and context as categorical fixed effects. We also included a by-subject random intercept.

Each participant's relative meaningful benefit (RMB) score was derived using a similar equation to relative visual benefit:

$$RMB = \frac{MEAN - ANOM}{1 - ANOM}$$

where MEAN is the participant's average accuracy during semantically meaningful trials, and ANOM is the participant's average accuracy during semantically anomalous trials. A mixed- effects model was carried out on RMB with group and modality as categorical fixed effects and a by-subject random intercept.

Using the relative benefit score formulas described above has the advantage of interpretation as it reflects the extent to which visual and semantic cues improve performance across a range of unimodal performance levels. The relative benefit scores also allow us to use data from a participant with floor performance in the audio-only conditions or anomalous contexts.

RESULTS

Keyword Identification

Audio-Only Modality

As shown in Table 3, the probability of identifying a keyword correctly in the audio-only modality was higher for the meaningful context relative to anomalous context ($p < 0.0001$). In addition, as the SNR became easier, the probability of identifying a keyword correctly increased ($p < 0.0001$). There is no significant effect of group or PTA. Separate models for older adults and younger adults that include audio-only accuracy as

the dependent variable and PTA as the only independent variable suggest PTA significantly predicts audio-only accuracy for older adults, but not young adults. There is a significant interaction of group and SNR ($p < 0.05$), suggesting that as the SNR becomes easier, the difference between older and younger adults increases. Lastly, there is a significant interaction between context and SNR ($p < 0.0001$), suggesting that as the SNR becomes easier, the difference between meaningful and anomalous becomes larger. These results are also represented in **Figure 2**.

Table 3. Log odds of producing a correct response across group, context, and SNR in the audio-only modality.

Fixed Effect	***β***	***SE***	***Z value***	***p***
Intercept	0.61	0.31	1.95	0.05
PTA	-0.02	0.01	-1.80	0.07
$\text{Group}_{\text{young adults – older adults}}$	0.14	0.22	0.64	0.52
$\text{Context}_{\text{meaningful-anomalous}}$	1.39	0.07	19.28	< 0.0001
SNR	0.33	0.01	45.32	< 0.0001
$\text{Group}_{\text{young adults-older adults}}$: SNR	0.03	0.01	2.26	< 0.05
$\text{Context}_{\text{meaningful-anomalous}}$: SNR	0.01	0.01	9.07	< 0.0001

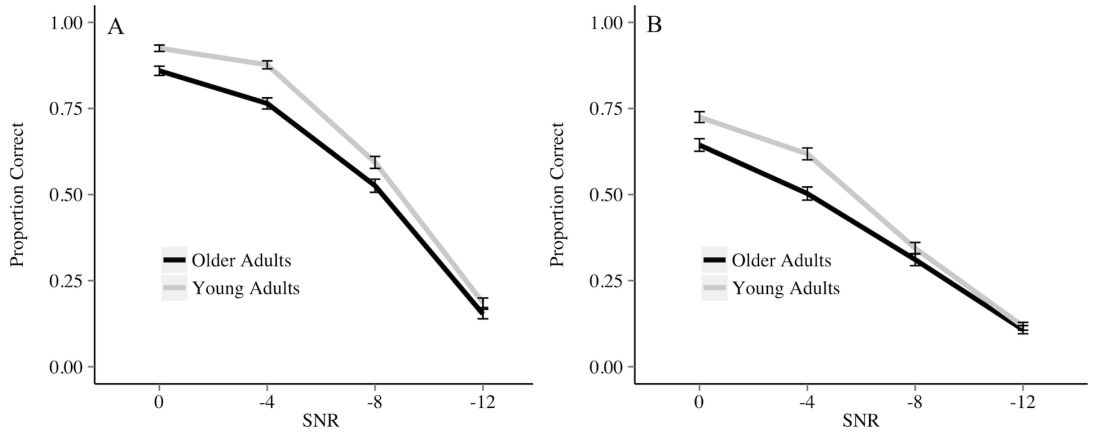


Figure 2. Proportion of keywords correctly identified by older and younger adults in the audio-only meaningful (A) and anomalous (B) semantic contexts. Bars represent standard error.

Audio-Visual Modality

The simplest, best fitting generalized linear mixed-effects model for the audio-visual presentation modality results included all fixed effects and their interactions. The results of the analysis are displayed in Table 4. In the audio-visual presentation modality, as pure tone averages decreased, the probability of a correct response increased($p < 0.01$). As in the audio-only condition, separate models for older adults and younger adults that include audio-visual accuracy as the dependent variable, and PTA as the only independent variable suggest PTA significantly predicts accuracy for older adults, but not young adults. The effect of group was not significant ($p = 0.36$). The effect of context was significant ($p < 0.0001$), suggesting that a meaningful context will lead to a higher probability of keyword correctly identified. The effect of SNR was also significant ($p < 0.0001$), suggesting that as the SNR became easier, there was an increase in the probability of correctly identifying a keyword. In addition, there was a significant interaction of context and SNR ($p < 0.0001$), suggesting that as the SNR became easier, the difference between correctly identifying a keyword in the meaningful and anomalous contexts increased. Lastly there was a significant three-way interaction of group, SNR, and context suggesting that in the anomalous context but not meaningful con- text, as the SNR becomes easier, the difference between older adults and young adults becomes larger (anomalous: $p < 0.05$; meaningful: $p = 0.46$). These results are represented in **Figure 3**.

Table 4. Log odds of producing a correct response across group, context, and SNR in the audio-visual modality.

Fixed Effect	***β***	***SE***	***Z value***	***p***
Intercept	2.04	0.35	5.75	< 0.0001
PTA	-0.05	0.02	-3.09	< 0.01
$\text{Group}_{\text{young adults – older adults}}$	-0.23	0.25	-0.92	0.36
$\text{Context}_{\text{meaningful - anomalous}}$	1.55	0.08	20.21	< 0.0001
SNR	0.26	0.01	36.50	< 0.0001
$\text{Group}_{\text{young adults - older adults}}$: $\text{Context}_{\text{meaningful - anomalous}}$	-0.23	0.12	-1.93	0.05
$\text{Context}_{\text{meaningful - anomalous}}$: SNR	0.09	0.01	6.34	< 0.0001
$\text{Group}_{\text{young adults - older adults}}$: $\text{Context}_{\text{meaningful}}$:SNR	-0.02	0.02	-0.74	0.46
$\text{Group}_{\text{young adults - older adults}}$: $\text{Context}_{\text{anomalous}}$:SNR	0.04	0.02	2.23	< 0.05

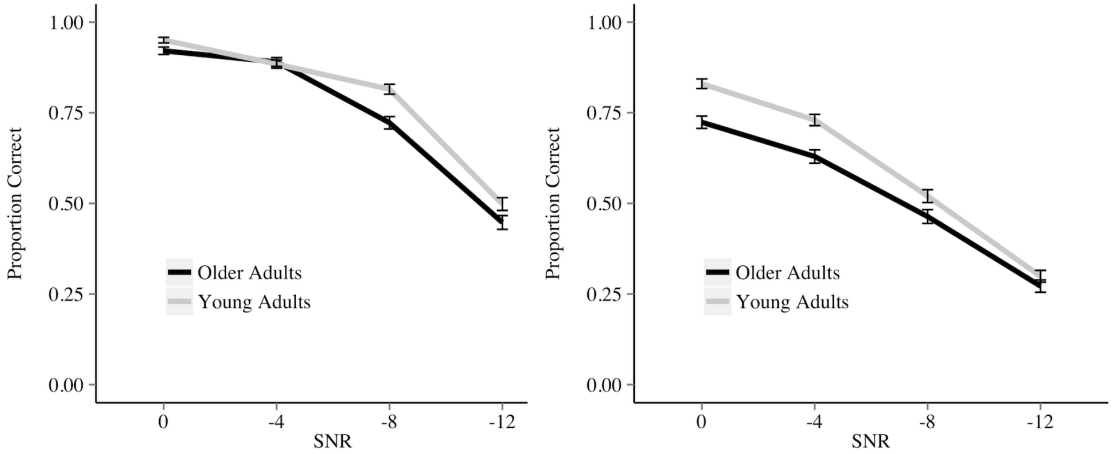


Figure 3. Proportion of keywords correctly identified by older and younger adults in the audio-visual meaningful (A) and anomalous (B) semantic contexts. Bars represent standard error.

Relative Benefit Analyses

Relative Visual Benefit

The results of the linear mixed-effects model on the relative visual benefit in meaningful and anomalous contexts for both older and younger adults suggest no difference between older and younger adults ($\beta = 0.004$, $p = 0.92$), but a difference in context ($\beta = 0.16$, $p < 0.0001$) at -12 SNR. The positive estimate for the context fixed effect suggests that the average relative visual benefit in a meaningful context is higher than in an anomalous context. The results of the simplest, best-fitting linear mixed-effects model are presented in Table 5. **Figure 4** graphically displays these results as well. Results of the linear mixed-effects model on the relative benefit in meaningful and anomalous contexts across age groups at -8 SNR suggests a similar pattern to what was found at -12 SNR. There is no difference between older and younger adults ($\beta = 0.15$, $p = 0.11$), but a difference in context ($\beta = 0.17$, $p < 0.05$) at -8 SNR. The positive estimate for the context fixed effect suggests that the average relative visual benefit in a meaningful context is higher than in an anomalous context. Lastly, we do not present relative visual benefit data from -4 and 0 SNR because we excluded numerous data points ($n = 16$ and $n = 37$, respectively) due to ceiling effects. Participants often scored perfect accuracy in the audio-only context, which placed a zero in the denominator of the RVB formula. We would also like to note that using a simple difference score (AV-AO) to measure benefit produces the same pattern of results.

Table 5. Results of the effect of group and semantic context on relative visual benefit at -12 SNR.

Fixed Effect	***β***	***SE***	***t value***	***p***
Intercept	0.28	0.02	13.00	< 0.0001
$\text{Group}_{\text{young adults – older adults}}$	-0.004	0.04	-0.11	0.92
$\text{Context}_{\text{meaningful – anomalous}}$	0.16	0.04	4.66	< 0.0001

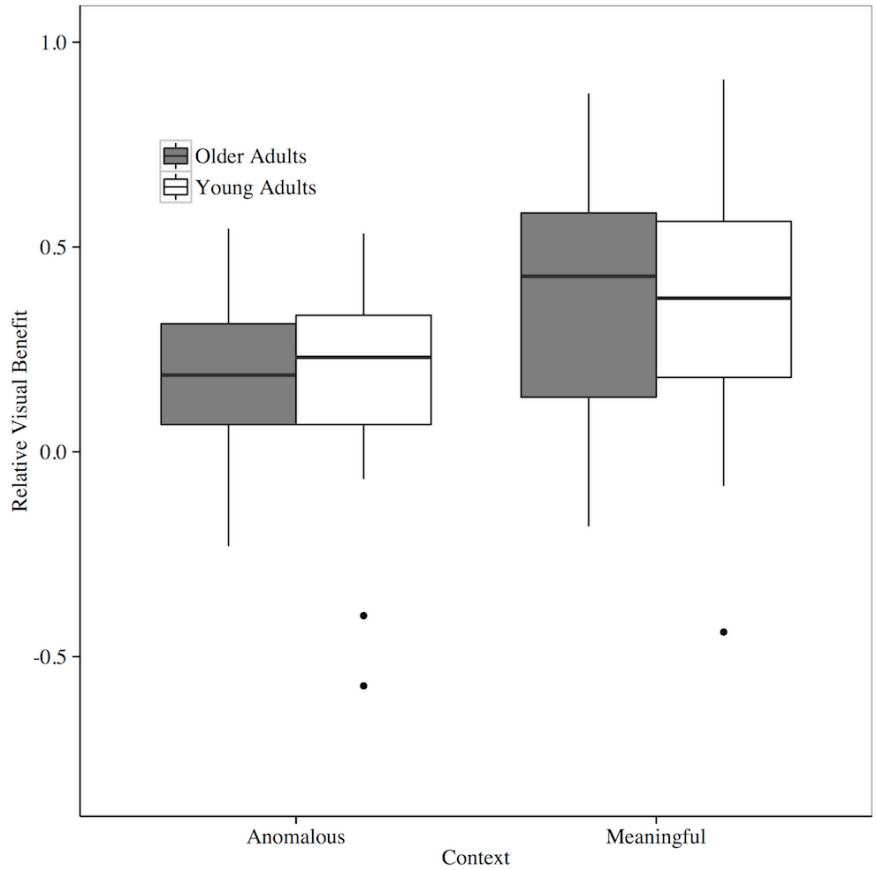


Figure 4. Relative Visual Benefit In Meaningful and Anomalous Semantic Contexts at -12 SNR. Bars represent upper and lower limits of the sample data, box limits represent upper and lower quartile, and the heavy line represents the sample median.

Relative Meaningful Benefit

Results suggest that group designation is not a significant predictor of relative meaningful benefit ($\beta = 0.005$, $p = 0.92$), but presentation modality is ($\beta = 0.20$, $p < 0.0001$) at -12 SNR. The positive estimate for the modality fixed effect suggests that the average relative meaningful benefit in the audio-visual presentation modality is higher than in the audio-only presentation modality. The results of the simplest, best-fitting model are dis- played in Table 6. **Figure 5** displays a graphical representation of the results as well. Results at -8 SNR suggest a similar pattern of relative meaningful benefit across age groups and presentation modality as was found at -12 SNR. Group designation is not a significant predictor of relative meaningful benefit ($\beta = 0.11$, $p = 0.07$), but presentation modality is ($\beta = 0.23$, $p < 0.0001$) at -8 SNR. Lastly, we do not present relative meaningful benefit data from -4 and 0 SNR because of the high rate of a ceiling effect: 50 instances where RMB = 1.0 across presentation modalities at -4 SNR, and 81 instances where RMB = 1.0 at 0 SNR across presentation modalities.

Table 6. Results of the effect of group and presentation modality on relative meaningful benefit at -12 SNR.

Fixed Effect	***β***	***SE***	***t value***	***p***
Intercept	0.14	0.02	6.45	< 0.0001
$\text{Group}_{\text{young adults-older adults}}$	0.005	0.04	0.11	0.92
$\text{Modality}_{\text{audio-visual-audio only}}$	0.20	0.04	4.91	< 0.0001

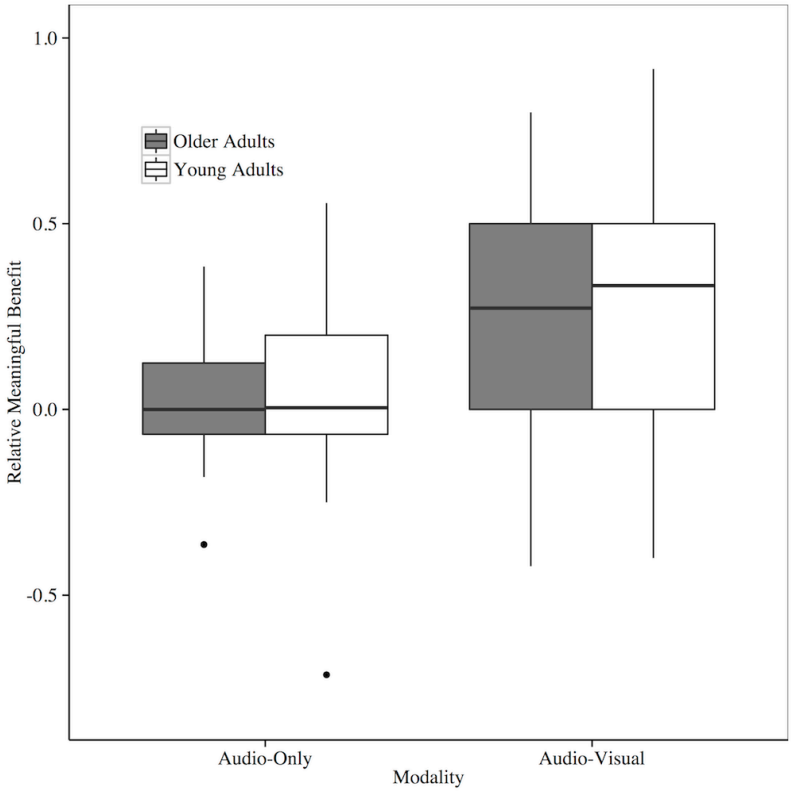


Figure 5. Relative Meaningful Benefit in Audio-Only and Audio-Visual Modalities at -12 SNR. Bars represent upper and lower limits of the sample data, box limits represent upper and lower quartile, and the heavy line represents the sample median. Circles indicate outliers.

DISCUSSION

The present study investigated the role of age on the ability to use visual and semantic cues to understand speech in a noisy environment. In both audio-only and audio-visual presentation modalities, as the SNR became easier, older and younger adults responded correctly more often, and the difference between the age groups in responding correctly increased, with young adults scoring higher than older adults. These results replicate prior findings suggesting that young adults have a general accuracy advantage in speech perception in speech-shaped noise relative to older adults (Helfer & Freyman, 2007). Similarly, as the SNR became easier, the difference in accuracy between meaningful and anomalous contexts became larger, with meaningful contexts leading to a higher accuracy rate at easier SNRs. This suggests that decreasing the amount of noise present can help listeners not only by giving them more access to the speech signal, but also by allowing them to make greater use of semantic context. Only in the audio-visual presentation modality did we find a significant three-way interaction of group, context, and SNR: in the anomalous context, as the SNR became easier, the accuracy difference between young adults and older adults became larger, with young adults achieving a higher accuracy at easier SNRs. In the meaningful context, there was no change in the accuracy difference between older adults and young adults as the SNR became easier. This interaction suggests that the older adult accuracy deficit can be ameliorated to the accuracy level of young adults by the presence of both semantic context and visual cues.

One interesting pattern in the current data is that young adults outperform older adults only in the easier SNRs and in certain conditions: an anomalous context presented in the audio-visual and audio-only modalities, and a meaningful context presented in the audio-only modality. As the SNR becomes more challenging in the aforementioned conditions, the difference between groups reduces. In addition, it is worth noting that

older adults maintain comparable accuracy rates to young adults in the two easiest SNRs (0 and -4) in the most supportive and commonly occurring context–audio-visual meaningful contexts. It is not until -8 SNR that older adults diverge from the young adults in terms of their speech perception ability in the audio-visual and meaningful conditions. This suggests that both visual cues and a semantic context are very important for older adults, and that with them, older adults show no deficit in speech perception ability relative to young adults.

Turning to the visual and meaningful benefit analyses, we found that older adults received the same amount of visual benefit and meaningful benefit as young adults at -12 dB and -8 dB SNRs. In addition, we found that young adults received more visual benefit during meaningful contexts relative to anomalous contexts, replicating Van Engen et al., (2014). Similarly, older adults received more visual benefit in meaningful contexts relative to anomalous contexts. In addition, young adults and older adults did not differ in the amount of meaningful benefit they received in both audio-only and audio-visual modalities. They received more meaningful benefit during audio-visual presentation relative to audio-only presentation.

To our knowledge, this study represents the first investigation of how visual and semantic cues are used by older and younger adults within the framework of a single study. Our results replicate some prior work in older and younger adults showing similar levels of visual and semantic context enhancement (Dubno et al., 2000; Frisina & Frisina, 1997; Pichora-Fuller et al., 1995; Sommers et al., 2005; Stevenson et al., 2015). These results suggest that the extent to which we use visual and semantic cues remains fairly consistent as we age. It is also worth acknowledging that our older adult sample may be particularly high-functioning, and therefore may not be representative of the general population. In addition, we limited our sample to participants who passed our hearing

screening of < 40 dB at several frequencies, and were above two standard deviations below the age-normalized mean of tests across memory, attention, and executive functioning. It is possible with a more heterogeneous sample in terms of hearing and neuropsychological abilities may yield larger age-related effects.

There are several limitations in the current study that should be addressed in future research. For instance, it will be important to examine visual and meaningful benefits at easier SNRs. A paradigm that equates group performance in one modality (e.g., audio-only or visual- only), for example, will circumvent the ceiling effects in accuracy (see Sommers et al., 2005). We also acknowledge that the constrains we place on our experimental parameters, such as the pre-recorded stimuli read off of a teleprompter, may limit the generalizability of our results to real-life speech- in-noise scenarios. For instance, Gilbert et al., (2014) showed that speech spoken without the presence of noise carries different characteristics than speech spoken in the presence of noise (Gilbert, Chandrasekaran, & Smiljanic, 2014). Therefore, since we recorded the sentences in quiet, they may differ from how they would be produced in real life. In addition, although we instructed the speaker of the stimuli to speak in a conversational manner, to mimic the temporal processing of speech in normal speech conversations, it is possible that there was still a slowing of speech given the artificial environment in which the speaker was recorded. Because older adults have difficulty with the temporal processing of speech (Schneider & Pichora-Fuller, 2001), it is imperative that future work use stimuli that capture the temporal aspects of everyday speech. In addition, a further investigation of individual differences will be important to our understanding of why some young and older adults are better than others at perceiving speech, especially in noise. Anderson and colleagues (2013) provide a model for understanding how individual differences such as cognition and perceptual ability relate to speech perception in older

adults; however, important cognitive measures implicated in speech perception such as processing speed (Schneider & Pichora-Fuller, 2001) were not included in their analysis (Anderson et al., 2013). Therefore, future models of speech perception in the aging population should include processing speed as a predictor of speech perception in noise.

Investigating how speech production affects older adults' use of visual and semantic cues (as in Van Engen et al., 2014) will also elucidate the extent to which older adults use other important cues in speech perception. Interestingly, Helfer (1998) found that older adults receive more visual benefit during conversational speech relative to clear speech, but the extent to which speaking style interacts with semantic benefit in older adults has not been studied. Future work should also focus on training methods that seek to enhance speech perception in older adults when visual and semantic cues are not present. Cognitive training methods, such as the one used in Anderson, White-Schwoch, Choi, and Kraus (2014) show promise of enhancing older adult speech perception under difficult listening conditions (Anderson, White-Schwoch, Choi, & Kraus, 2014). Examining the extent to which cognitive training can improve speech perception in a variety of noise contexts carries the potential to help older adults communicate better.

In conclusion, the present study found that when both visual cues and a meaningful semantic context were present in easy SNRs, older adults showed no accuracy difference in their ability to perceive speech in noise relative to young adults. In addition, older and younger adults received the same amount of visual and meaningful benefit during speech perception, suggesting a consistency in using these cues as we age. Finally, more visual benefit was found in meaningful relative to anomalous semantic contexts and more meaningful benefit was found in the audio-visual presentation modality relative to the audio-only presentation modality. These results highlight an

important facet of older adults' speech perception: with both visual and semantic cues, older adults can excel at speech perception.

Paper 2: Enhanced Cognitive and Perceptual Processing: A Computational Basis for the Musician Advantage in Speech Learning[2]

Music training is a rich, multimodal experience that has been found to modify the brain in many positive ways. For instance, long-term music training is associated with enhanced processing of musical information such as pitch discrimination and perception (Bidelman, Krishnan, & Gandour, 2011; Magne, Schön, & Besson, 2006; Schön, Magne, & Besson, 2004; Tervaniemi, Just, Koelsch, Widmann, & Schröger, 2004; Zarate, Ritson, & Poeppel, 2012) rhythm production (Bailey, Zatorre, & Penhune, 2014; Chen, Penhune, & Zatorre, 2008), beat perception (Grahn & Rowe, 2012), and timbre discrimination (Crummer, Walton, Wayman, Hantz, & Frisina, 1994). Processing of musical information has also been studied in non- human primates. For instance, extensive pitch discrimination training has been used to characterize the plastic nature of the non-human auditory cortex (Brosch, Selezneva, Bucks, & Scheich, 2004; Brosch, Selezneva, & Scheich, 2005). In addition to musical information processing advantages, recent studies have also found that long-term music training is associated with advantages that extend beyond the musical domain, such as speech processing. For example, musicians show more robust neural encoding of speech sounds relative to non-musicians (Wong et al., 2007; Chandrasekaran et al., 2009; Kraus and Chandrasekaran, 2010) and outperform non-musicians in recognizing speech embedded in noise (Parbery-Clark et al., 2009; Strait and Kraus, 2011). Musicians also show superior non-native speech discrimination (Gottfried et al., 2004; Marie et al., 2011) and learning (Gottfried & Riester, 2000; Alexander et al., 2005; Wong and Perrachione, 2007; Lee and Hung, 2008) compared to

[2] This paper was previously published as Smayda, K. E., Chandrasekaran, B., & Maddox, W. T. (2015). Enhanced cognitive and perceptual processing: a computational basis for the musician advantage in speech learning. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00682

non-musicians. While the musician advantage for learning non-native speech sounds is robust, the underlying mechanisms giving rise to this advantage are poorly understood.

Recently, a framework was developed to explore the mechanisms underlying the cross-domain auditory plasticity induced by long-term music training. The OPERA hypothesis posits that music training will affect the neural encoding of speech because: there is **O**verlap between the networks used to process both music and language; there is a greater **P**recision of processing of music relative to language; music elicits strong **E**motional experiences; **R**epetition is integral to music learning; and musical engagement requires sustained **A**ttention (Patel, 2011). The OPERA hypothesis was later updated to clarify the "precision" aspect of the hypothesis (Patel, 2012). More recently it was expanded to include the cognitive benefits of non- vocal music training on speech processing, motivation for using animal models, and preliminary data from a study investigating music training's impact on speech perception in cochlear-implant users (Patel, 2014). Per this framework, music and speech share similarities in acoustics, such as pitch, timbre, and timing (Kempe, Bublitz, & Brooks, 2013), as well as higher-level cognitive processes such as working memory and attention (Besson, Chobert, & Marie, 2011; Nina Kraus, Strait, & Parbery-Clark, 2012), suggesting that the musician advantage in learning non-native speech could arise from enhanced perceptual processing, cognitive processing, or both. To date, the evidence in support of these hypotheses comes from studies that target domain-general cognitive or perceptual processes with unique tasks. For instance, musicians show enhanced cognitive abilities compared to non-musicians in areas such as executive function (Bialystok & DePape, 2009), working memory (George & Coch, 2011; Nina Kraus et al., 2012; Pallesen et al., 2010; Parbery-Clark et al., 2009; D. L. Strait, O'Connell, Parbery-Clark, & Kraus, 2013), and switching (Hanna-Pladdy & MacKay, 2011), while a separate literature shows perceptual enhancements in speech

processing (Parbery-Clark, Strait, & Kraus, 2011; Parbery-Clark et al., 2012, 2011; White-Schwoch, Carr, Anderson, Strait, & Kraus, 2013; Zendel & Alain, 2012). To date, the cognitive and perceptual processes mediating the musician advantage in non-native speech learning has never been investigated within a single task. The current study addresses this shortcoming by examining non-native speech learning in musicians and non-musicians using traditional measures of performance (e.g., accuracy), and computational models that allow us to independently estimate the perceptual and cognitive processing.

We examine perceptual and cognitive processing within the specific theoretical framework of multidimensional (MD) signal detection theory(F. Gregory Ashby & Townsend, 1986; Maddox & Ashby, 1993). Within this framework, repeated presentations of the same physical stimulus yield unique perceptual effects that result in a multivariate normal distribution of perceptual effects (F. Gregory Ashby & Townsend, 1986; Green & Swets, 1967). Changes in the perceptual variances are associated with perceptual selectivity. To explore changes in perceptual processing as a function of musical training, we separately estimate a measure of perceptual selectivity (also referred to as perceptual variance or noise) along the pitch height and pitch direction dimensions. In addition, we can look at decision processes that involve constructing decision bounds (defined in detail later) that divide the perceptual space into separate response regions. Critically, perceptual and decisional processes are theoretically independent, and have unique, identifiable parameters (Gregory Ashby & Townsend, 1986; Green & Swets, 1967; Maddox & Ashby, 1993).

In the current study, we examine the extent to which long- term music training impacts learning to categorize Mandarin lexical pitch patterns. Mandarin Chinese is a tone language, wherein changes in pitch patterns within a syllable result in changes to

word meaning. Learning to categorize the four pitch patterns in Mandarin is a challenging task for monolingual American adults (Wang, Spence, Jongman, & Sereno, 1999), and therefore provides an excellent paradigm for studying the perceptual and cognitive mechanisms underlying learning. The four Mandarin Chinese tone categories and their descriptions are: T1, "high-level," T2, "mid-rising," T3, "low-dipping," and T4, "high-falling" (Howie, 1976). Pitch height (how high or low a tone is) and pitch direction (average movement of a pitch) have been found to be the most prominent dimensions used in categorizing lexical tones such as in Mandarin (Gandour, 1983; Gandour & Harshman, 1978).

Native English speakers exhibit differential sensitivity to the dimensions underlying tone perception relative to native Mandarin Chinese speakers. MD scaling analyses of native English speakers and Mandarin speakers found that while English speakers weight the pitch height dimension equally to that of tone language speakers, they weight the pitch direction dimension less than tone language speakers (Chandrasekaran, Gandour, & Krishnan, 2007; Gandour & Harshman, 1978). This is likely due to the fact that pitch direction is not as salient a cue in English as it is in Mandarin, where it is required to distinguish pitch patterns that vary dynamically within the syllable. Although native English speakers and Mandarin speakers attend to the pitch height dimension to a similar degree, this dimension is highly influenced by variability in talkers (different talkers have different average pitches). In previous studies using the same computational modeling methods utilized in the current report, we have shown that the optimal decision strategy is one in which the participant attends to and utilizes both pitch height and pitch direction in making categorization judgments (Bharath Chandrasekaran, Yi, & Maddox, 2013; Maddox & Chandrasekaran, 2014; Maddox, Chandrasekaran, Smayda, & Yi, 2013; Yi, Maddox, Mumford, & Chandrasekaran, 2014).

This is referred to as a MD decision strategy and is contrasted with a unidimensional (UD) strategy in which the participant bases their decision solely on one stimulus dimension (usually pitch height). In the present study, we applied rigorous computational models to each participant's response pattern on a block-by-block basis. We included one model that instantiates a MD strategy, two that instantiate UD strategies, and one that instantiates a random responder (RR) strategy. Computational models are necessary to make this distinction because the same accuracy rate can be obtained using qualitatively different strategies.

In addition to providing critical insights into the decisional strategies used by musicians and non-musicians, the computational models also allow us to explore perceptual processes independent of decisional processes. To explore changes in perceptual processing as a function of musical training, we separately estimate a measure of perceptual selectivity (also referred to as perceptual variance or noise) along the pitch height and pitch direction dimensions. Since pitch height is highly salient in English we make no strong predictions regarding the effects of musical training on perceptual selectivity along the pitch height dimension. However, although pitch direction is not as salient a feature in English as it is in Mandarin, musicians train for many hours a week to become sensitive to pitch direction (i.e., melodies), thus capitalizing on the narrow frequency tuning capabilities of the human primary auditory cortex (Bitterman, Mukamel, Malach, Fried, & Nelken, 2008). Therefore it is likely that musicians will show enhanced perceptual selectivity (i.e., reduced perceptual noise) along the pitch direction dimension compared to non-musicians. Detailed descriptions of the computational models can be found below in Section "Computational Modeling Descriptions."

To summarize, we predict a musician advantage in non-native speech learning. Our goal is to go beyond accuracy measures and to provide mechanistic explanations for the musician advantage. We predict that this advantage is due to an increased use of optimal MD decision strategies, as well as enhanced perceptual selectivity along the pitch direction dimension.

MATERIALS AND METHODS

Stimulus Characteristics

Training stimuli consisted of the four Mandarin tones, tone 1 (T1), tone 2 (T2), tone 3 (T3), and tone 4 (T4) in the context of five syllables found in both Mandarin Chinese and English ("bu," "di," "lu," "ma," "mi") by one male talker and one female talker (40 stimuli total). Both speakers are originally from Beijing, and stimuli were RMS amplitude and duration normalized (70 dB, 0.4 s) using the software Praat (Francis & Nusbaum, 2002; Perrachione, Lee, Ha, & Wong, 2011). Five native speakers of Mandarin were asked to identify the tone categories (they were given four choices) and rate their quality and naturalness. High identification (>95%) was achieved across all five native speakers and speakers rated these stimuli as highly natural. We can represent these stimuli in a two-dimensional space with pitch height (how high or low a tone is) on the *x*-axis and pitch direction (average movement of the tone) on the *y*-axis (**Figure 6**). These two dimensions have been found to be especially relevant dimensions when categorizing the Mandarin tones (Francis, Ciocca, Ma, & Fenn, 2008).

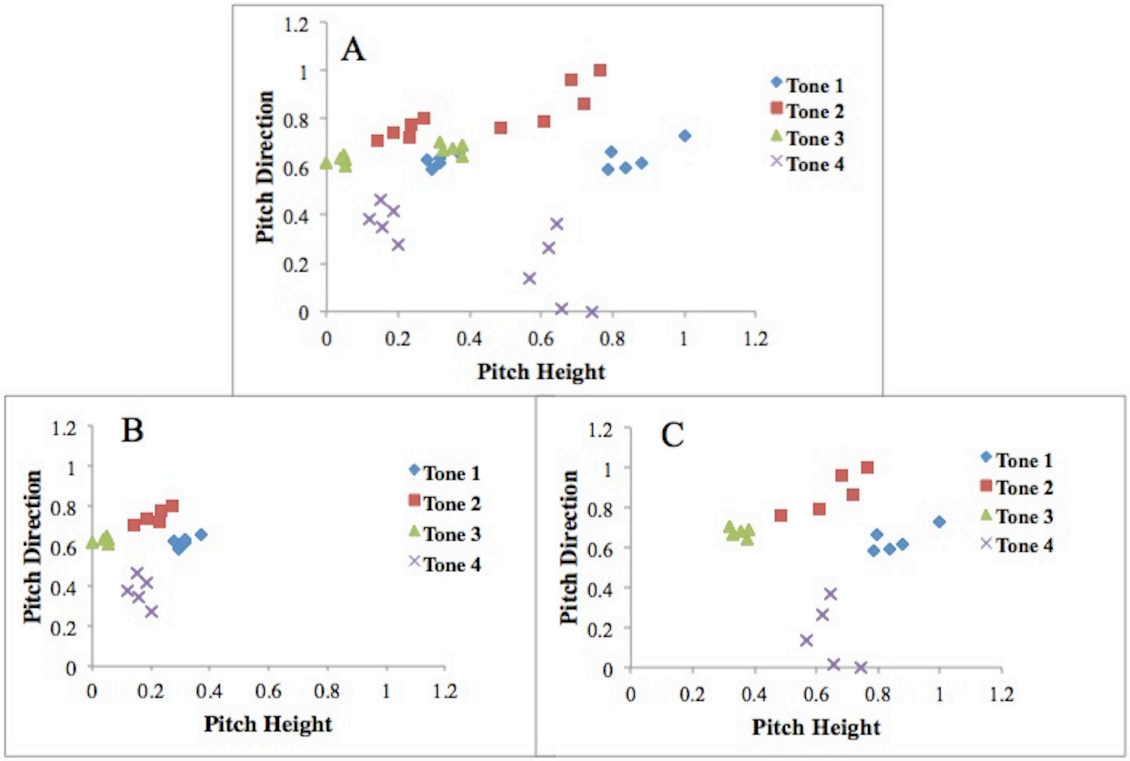


Figure 6. Scatterplot of all stimuli (A). Scatterplot of male-talker stimuli (B). Scatterplot of female-talker stimuli (C). Stimuli dimensions (Pitch Height and Pitch Direction) were normalized between 0 and 1. Pitch height refers to how high or low the pitch is, and pitch direction refers to (end pitch – start pitch)/duration.

Participants

Fifteen musicians (7 female; mean age = 25 years, SD = 5.29) and fifteen non-musicians (12 female; mean age = 23 years, SD = 3.96) from The University of Texas at Austin and greater Austin, Texas community were paid $10 per hour for their participation. The University of Texas at Austin Institutional Review Board approved the study protocol, and informed consent was obtained from all participants. Exact thresholds were recorded for over half of the participants (8 of the 15 non-musicians; 9 of the 15 musicians). We conducted a mixed model ANOVA on the effect of ear (within subjects: left/right), frequency (within subjects: 500, 100, 2000 Hz), and group (between subjects: musician/non-musician) on pure tone audiometry thresholds. "Participant" was treated as a random variable. We found no difference between groups with respect to left and right ear thresholds [$F(1,14) = 0.72$, $p = 0.41$, partial $\eta 2 = 0.05$] or pure tone averages (500, 1000, 2000 Hz) [$F(2,29) = 2.10$, $p = 0.14$, partial $\eta 2 = 0.13$]. In addition participants reported no significant issues related to hearing. Musicians had at least 10 years of group or private instrumental lessons, and currently play or sing at least 3 h a week (instruments included organ, piano, flute, guitar, viola, and voice). Non- musicians had 3 years or less of group or private music lessons, and do not currently play an instrument. Participants' musical history can be found in Table 7. Stimuli were presented at comfortable supra-threshold listening levels through Sennheiser HD 280 Pro headphones.

Table 7. Participants' Music History

	Years of Training	Age of Onset, yr.	Hours Play Per Week	Instrument
Musician				
1	15	7	20	Flute
2	15	8	28	Flute
3	11	5	6	Guitar
4	15	7	36	Organ
5	15	6	3	Piano
6	16	4	11	Piano
7	11	12	8.5	Piano
8	11	9	12	Piano
9	17	5	11	Piano
10	21	5	4	Piano
11	20	6	33	Piano
12	30	7	10	Viola
13	16	6	27	Viola
14	14	10	26	Voice
15	12	9	7	Voice
Mean	15.93	7.07	16.17	
Non-Musician				
16	2	7	0	Flute
17	1	12	0	Flute
18	1	13	0	Guitar
19	1	9	0	Piano
20	2	8	0	Piano
21	3	8	0	Piano
22	0.5	10	0	Recorder
23	3	12	0	Saxophone
24	2	11	0	Trumpet
25	1	11	0	Violin
26	2	NA*	0	Violin
27	0	NA	0	NA
28	0	NA	0	NA
29	0	NA	0	NA
30	0	NA	0	NA
Mean	1.23	10.10	0	

**Subject 26 did not report age of onset*

Procedure

On each trial, participants were presented with a single exemplar from one of four Mandarin tone categories (T1, T2, T3, or T4) and instructed to categorize the stimulus into one of four equally likely categories. During the training blocks, participants were given feedback after each trial and exposed to multiple talkers that were randomized within a block. Participants listened to 40 stimuli per block (4 tone categories × 5 syllables × 2 talkers). Each participant completed five 40-trial blocks of training and was instructed that high accuracy levels were possible. Participants generated a response by pressing one of four number button keys on the left side of the computer keyboard, labeled "1," "2," "3," or "4." Corrective feedback was provided for 1 s on the screen immediately following the button press and consisted of the word "Correct" or "No" followed by the label of the tone that was actually presented. For example, on a correct T1 trial the feedback display was as follows: "Correct, that was a category 1." On an incorrect response trial where T4 was the correct response the feedback display was as follows: "No, that was a category 4." A 1-s ITI followed the feedback.

After participants completed five 40-trial blocks, they completed one 20-trial generalization block. For the generalization block, all four tones and five syllables were presented, but were spoken by a different male speaker from the five training blocks. This resulted in 20 tokens (4 tones × 5 syllables × 1 new talker), and therefore 20 trials. In addition, feedback was not given. The generalization block was modeled separately from the five training blocks. The entire task lasted about 35 min.

Surveys and Neuropsychological Tests

All participants completed a demographics survey, and a music and language experience survey. In addition, all participants completed WAIS-III Digit Span task to assess working memory capacity (Wechsler, D., 1997), and no difference was found

between the two groups' composite working memory sore (backward score + forward score) [$t(28) = 1.51$, $p = 0.14$]. Participants were matched on age and education (musicians: mean = 16.77 years, SD = 1.76; non-musicians: mean = 16.07, SD = 2.15).

Computational Modeling Descriptions

Decisional Processing Assumptions

Accuracy rates provide an excellent source of information regarding how well an individual is performing in a task. Although a good starting point, one weakness of accuracy-based measures is that the same accuracy rate can often be achieved with qualitatively different strategies (e.g., UD or MD strategies). Within the domain of category learning, computational models can be utilized to address this shortcoming and can provide important insights into the nature of the strategy an individual is applying in a given task. In this study we apply a series of decision-bound models originally developed for application in the visual domain (Ashby & Maddox, 1993; Maddox & Ashby, 1993) and recently extended to the auditory domain by Maddox and Chandrasekaran (2014; Bharath Chandrasekaran et al., 2013; Maddox et al., 2013; Yi et al., 2014) on a block-by-block basis at the individual participant level because of problems with interpreting fits to aggregate data (Ashby, Maddox, & Lee, 1994; Estes, 1956; Maddox, 1999). We assume that the two-dimensional space (pitch height vs. pitch direction) displayed in **Figure 6A** accurately describes the perceptual representation of the stimuli, and based on the results from our earlier work (Maddox & Chandrasekaran, 2014), we also assume that participants applied category learning strategies separately to the male- and female- talker perceptual spaces (**Figures 6B, C**, respectively). Each model assumes that decision bounds (or category boundaries created by the participant as they

learn the categories) were used to classify stimuli into each of the four Mandarin tone categories (T1, T2, T3, or T4).

To explore the types of strategies that participants used, we applied three types of models: UD, MD, and RR. **Figure 7** displays stimuli and response regions for the four tone categories generated from a hypothetical participant using strategies implicated by one version of the UD_Height model (**Figure 7A**), one version of the UD_Direction model (**Figure 7B**), and the MD model (**Figure 7C**). Each UD model assumed that the participant set three criteria along a given dimension, which effectively partitioned the perceptual space into four response regions. For example, the UD_Height model assumes that the participant sets three criteria along the pitch height dimension, which are used to separate the stimuli into those that are low, medium–low, medium–high, or high pitch height. Importantly, this model ignores the pitch direction dimension. The eight most reasonable variants of the model were examined and differ only in the assignment of the category labels (T1, T2, T3, T4) to response regions (low, medium-low, medium–high and high, respectively). Therefore, the eight most reasonable variants were: 3214, 3412, 3241 (shown in **Figure 7A**), 3421, 2314, 4312, 2341, and 4321. For example, a participant who carved up the space using the 3241 variant of the model would categorize a low tone as category 3, a medium–low tone as category 2, a medium– high tone as category 4, and a high tone as category 1. The UD_Direction model assumes that the participant sets three criteria along the pitch direction dimension. The model assumes that the three criteria along the pitch direction dimension are used to separate the stimuli into those that have a low slope, medium–low slope, medium–high slope, or high slope. Importantly, this model ignores the pitch height dimension. The two most reasonable variants of the model were examined and differ only in the assignment of category labels (T1, T2, T3, T4) to response regions (low, medium–low, medium–high, and high slopes).

These were: 4312 and 4132 (shown in **Figure 7B**). Each UD model contains three free decision parameters—three criteria along the relevant dimension.

The MD model that we used also partitioned the space into four separate response regions, but unlike the UD models, the MD model focused on both the pitch height and pitch direction dimensions. In addition, whereas the UD model decision bounds were vertically oriented (in the UD_Height model) or were horizontally oriented (in the UD_Direction model), in the MD model the decision bound orientations were not constrained. A model of this sort can be instantiated in a number of ways. In line with some of our previous work (Maddox & Chandrasekaran, 2014; Maddox et al., 2013; Yi et al., 2014), we used a simple-prototype model framework in which each category is represented by a single prototype and each exemplar is classified into the category with the most similar prototype. Because the location of one of the prototypes can be fixed, and since a uniform expansion or contraction of the space will not affect the location of the resulting response region partitions, the MD model contains five free decision parameters that determine the location of the prototypes, and a single free parameter that represents noise in their placement. **Figure 7C** displays a scatterplot of the stimuli and response regions for the four tone categories generated from a hypothetical participant using one version of the MD model. A key feature of this model is that it assumes the participant is integrating information from both pitch height and pitch direction dimensions in their classification of Mandarin tones, making this a model that implicates a MD strategy. Importantly, we introduce the decisional models we present here, and the perceptual models we present in Section "Perceptual Processing Assumptions" as "cognitive" and "perceptual" models within a specific theoretical framework – multiple signal detection theory (Gregory Ashby & Townsend, 1986; Maddox & Ashby, 1993). These models are referred to as "cognitive" models because working memory, attention,

and executive functioning are relevant to the distinction between UD and MD strategies. We explore working memory capacities of UD and MD strategy users in section Working memory and cognitive strategies.

The third model is a RR model that assumes that the participant guesses on each trial.

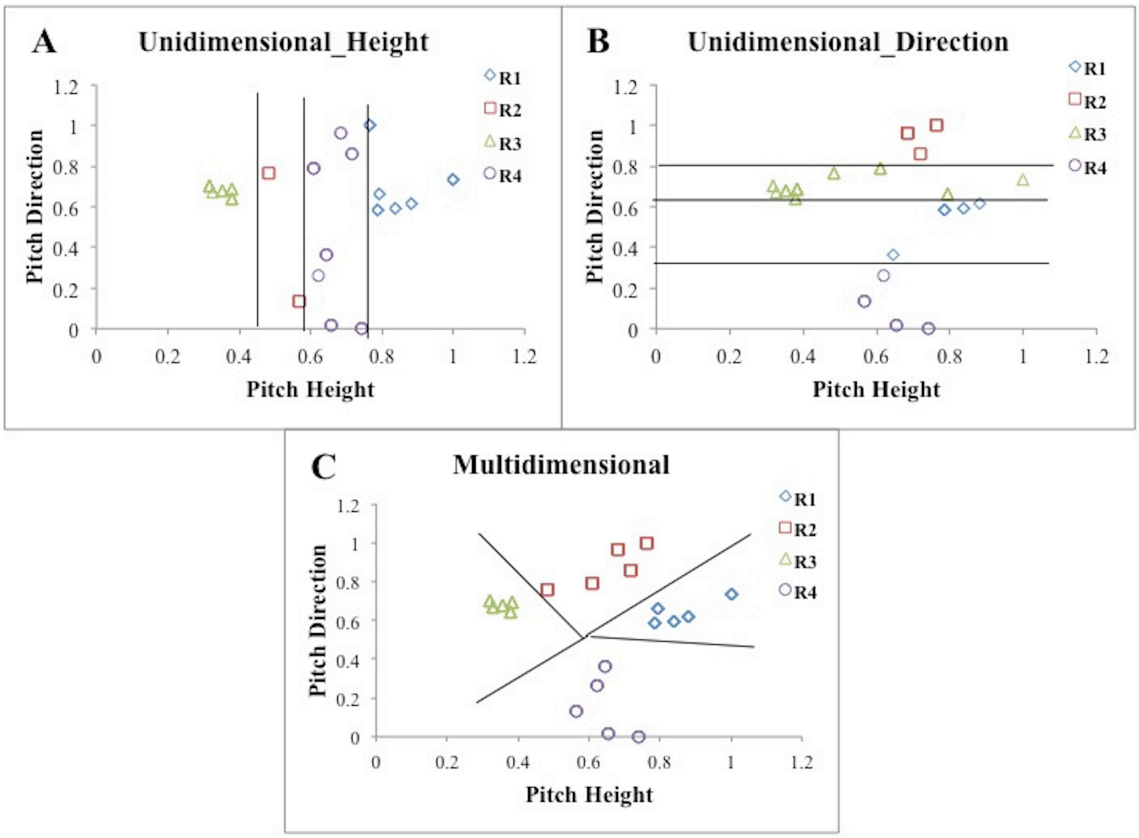


Figure 7. Scatterplots of the responses along with the decision boundaries that separate response regions from a hypothetical participant using a version of the Unidimensional (UD)_Height (A), UD_Direction (B) and Multidimensional (MD; C) models as applied to the female talker stimuli shown in Figure 6C.

Perceptual Processing Assumptions

Whereas **Figures 6A–C** denotes the mean perceptual effects of the stimuli, variability in the trial-by-trial perceptual effects is estimated from the data. We assume that the perceptual variance along the pitch height dimension is identical across all 40 stimuli and that the perceptual variance along the pitch direction dimension is identical across all 40 stimuli (referred to as a stimulus invariant perceptual representation; Ashby and Maddox, 1992; Maddox, 2001, 2002; Maddox and Dodd, 2003), but that the perceptual variance along the pitch height and pitch direction dimensions are uncorrelated (referred to as perceptual independence; (F. Gregory Ashby, 1988; F. Gregory Ashby & Townsend, 1986). In other words, while we estimate the perceptual variability along the pitch height dimension separately from that along the pitch direction dimension, we assume those variability estimates are constant across stimuli (stimulus invariance), and that the perceptual covariance between pitch height and pitch direction is zero (perceptual independence). A smaller perceptual variance is associated with a more veridical percept. The decisional processes introduced above, and the perceptual processes introduced in this section are independent of one another (Green & Swets, 1967; Maddox & Ashby, 1993).

Model Fitting Procedure

In this section, we elaborate on the procedures used to fit models to behavioral data. On each trial, the participant is presented with a single stimulus and emits one categorization response. Thus for each stimulus the observed probability of responding T1–T4 is either 1 or 0 with three of these responses having an observed probability of 0 and one of 1. For example, if the participant generated a T1 response on trial 1, then the observed probability of responding T1, T2, T3, and T4 would be 1, 0, 0, and 0, respectively. The same holds for each of the 40 trials in a block. For a given model and a

fixed set of parameters, the model generates a set of predicted response probabilities for each of the 40 trials. The observed and predicted values are combined using maximum likelihood, and are used to produce an Akaike information criterion (AIC; Akaike, 1974) value:

$$AICi = -2lnLi + 2Vi \quad (1)$$

where Li is the maximum likelihood for model i, and Vi is the number of free parameters in the model. The model parameters are adjusted until the smallest AIC value is identified, and this is defined as the best fitting version of that model for a given set of data. This process is repeated for all of the models and the model with the smallest AIC value is defined as the best fitting model for that data set. Notice that AIC penalizes models with more free parameters. Thus, if two models provide equivalent maximum likelihood fits to a set of data, but one has more free parameters, the model with more free parameters will be rejected in favor of the model with fewer free parameters.

Data Analysis

Several of our results derive from an examination of the effects of music training on performance across blocks of trials, such as accuracy, and perceptual selectivity measures from the computational models. In these cases, we conducted a 2 between group (musician vs. non-musician) × 5 within group (block: 1–5, repeated measure) mixed design ANOVA with "participant" as a random variable. Other results derive from simple comparisons between musician and non-musicians. These include the first block of trials best fit by a MD strategy model, total number of blocks fit by a MD strategy model, working memory comparisons between MD and UD users, and measures of accuracy and perceptual variance in the generalization block. For these analyses, we used

t-tests to compare measures between groups. All analyses were carried out using R version 3.0.3 (R Core Team, 2014).

RESULTS

We first present accuracy analyses comparing block-by-block training and generalization performance between musicians and non-musicians. Then we present model-based analyses to explore the types of decision strategies that participants use to learn during the task, working memory comparisons of different strategy users, and the magnitude of perceptual noise along the pitch height and pitch direction dimensions.

Accuracy Results

Learning curves for the musicians and non-musicians are presented in **Figure 8**. We begin with a 2 group (between subjects: musician vs. non-musician) × 5 training block (within subjects: blocks 1–5) mixed design ANOVA on the accuracy data with "participant" as a random variable. The main effect of participant group was significant [$F(1,28) = 11.07$, $p = 0.0018$, partial $\eta 2 = 0.3$] and suggests a performance advantage for musicians (average accuracy = 0.74) over non- musicians (average accuracy = 0.50). The main effect of block was also significant [$F(4,112) = 47.60$, $p < 0.001$, partial $\eta 2 = 0.063$]. Finally, the interaction between participant group and block was significant [$F(4,112) = 5.911$, $p < 0.001$, partial $\eta 2 = 0.174$]. *Post hoc* pairwise comparisons of the groups at each block suggest that the musician advantage held in all blocks except block 1 (all p's < 0.01). In addition, we tested the difference in learning trajectories between the two groups by conducting polynomial contrast tests on accuracy between the musician and non-musician groups across blocks. Results revealed a significant linear relationship of the group × block interaction [$F(1,112) = 14.01$, $p < 0.001$, partial $\eta 2 = 0.111$], a significant quadratic trend of the interaction [$F(1,112) = 4.25$, $p < 0.05$, partial $\eta 2 =$

0.037], and a significant cubic trend of the interaction [$F(1,112) = 4.59$, $p < 0.05$, partial $\eta = 0.039$]. Contrast analyses using the linear, quadratic, and cubic scores for each participant indicated that the linear trend was significantly different for the musician and non-musician groups. The average linear increase in accuracy for the musician group ($M = 0.49$, SD = 0.41) is significantly larger than the average linear increase in accuracy for the non-musician group [$M = 0.89$, SD = 0.42; $t(148) = 5.93$, $p < 0.001$]. The quadratic trend also differed significantly for the musician and non-musician groups across blocks and was significantly greater for the non-musician group ($M = -0.17$, SD = 0.27) than for the musician group ($M = -0.43$, SD = 0.29) [$t(148) = 5.93$, $p < 0.001$]. Lastly, the cubic trend was significantly different for musicians and non-musicians across blocks. The cubic trend from the musicians was significantly larger for musicians

($M = 0.20$, SD = 0.24), than non-musicians [$M = -0.04$, SD = 0.21) [$t(148) = 6.34$, $p < 0.001$]. These results suggest different learning trajectories for musicians and non-musicians, where across blocks, musicians show a significantly stronger linear and cubic trend relative to non-musicians, who show a significantly stronger quadratic trend. As suggested by an examination of **Figure 8**, generalization performance for musicians was superior to that for non-musicians [$t(28) = 3.48$, $p < 0.005$].

To determine whether more training trials might result in a different pattern of accuracy rates for musicians and non- musicians, we compared accuracies in block 4 and 5 for musicians and non-musicians separately. Using two one-way repeated measures ANOVA's, results reveal that accuracy rates for both musicians and non-musicians did not significantly change from block 4 to 5 [musicians: $F(1,14) = 2.88$, $p = 0.11$; non-musicians: $F(1,14) = 0.01$, $p = 0.91$].

Taken together, these data suggest that musicians show better Mandarin tone category learning and generalization than non- musicians. These findings replicate a large

body of work in showing an accuracy advantage in learning non-native speech categories for musicians relative to non-musicians (Alexander, Wong, & Bradlow, 2005; T.L. Gottfried & Riester, 2000; Lee & Hung, 2008; P. C. M. Wong & Perrachione, 2007). Next we explore computational modeling of participants' responses to better understand the locus of the musician performance advantage.

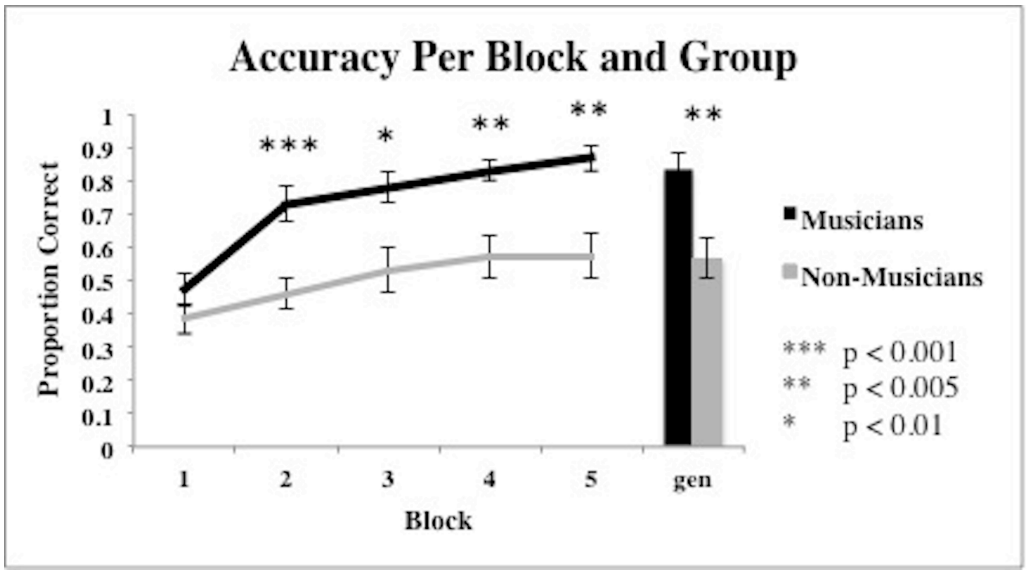


Figure 8. Average proportion correct for the five training blocks and generalization block for musicians and non-musicians. Bars represent SEM.

Computational Modeling Results

The accuracy-based analyses suggest that musicians showed a learning and generalization advantage over non-musicians when asked to categorize Mandarin tones. Accuracy measures are informative, but they do not provide a mechanistic explanation for this performance advantage – for instance, whether this advantage is due to cognitive and/or perceptual processing advantages in musicians. It is possible that non-musicians are using the same strategies as musicians, just sub-optimally, or they could be using different strategies altogether. In addition, musicians and non-musicians may show similarities or differences in perceptual selectivity along each dimension. Model-based analyses allow us to address these important questions.

Cognitive Strategies and Accuracy Rates Across Blocks

In this section, we compare the cognitive strategies used by musicians and non-musicians during Mandarin tone category learning. Specifically, we compare the use of a MD, UD, and RR strategies across musicians and non-musicians. A breakdown of strategies per block and group can be found in **Figure 9**.

To investigate the use of strategies over the course of the five training blocks, we examined three aspects of the data between musicians and non-musicians. First we determined the first block of trials for which the MD model provided the best account of the data and compared these values for musicians and non- musicians. Second, we determined the total number of blocks of trials for each participant for whom the MD model provided the best account of the data and compared these values for musicians and non-musicians. Finally, we examined the learning curves for musicians and non-musicians whose final block of data was best fit by either a MD or a UD strategy. To determine the first block of trials for which musicians (as a group) and non-musicians (as a group) used a MD strategy, we identified the first block of trials for each participant for

which the MD model provided the best account of the data. We then computed the average of these blocks for musicians and non-musicians separately. For instance, if the first block of trials for which a MD strategy best fit the data from musicians 1–3 were blocks 3, 4, and 4, then the average block when they first used a MD strategy would be block 3.67. We found that the first use of a MD strategy occurred significantly earlier for musicians (average 1.87 blocks) than for non-musicians (average 3.20 blocks) [$t(28) = 2.24$, $p < 0.05$]. Next, we examined the number of blocks of trials for which a MD strategy provided the best fit to the data for musicians and non-musicians. We found that the number of blocks of trials best fit by a MD model was larger for musicians (average 4.07 blocks) than non-musicians (average 2.13 blocks) [$t(28) = 3.24$, $p < 0.01$].

Finally, we examined the learning curves associated the best fitting model during the final training block. We classified participants as UD-Musician, UD-Non-Musician (UD groups also included those best fit by RRs), MD-Musician, and MD- Non-Musician based upon the best fitting model from block five. As suggested by an examination of **Figure 9**, none of the 15 musicians' data was best fit by a UD model in block 5. Thus, we cannot generate a learning curve for this group. The goal of this analysis was to determine how the strategy used in the final block of trials might affect the course of learning. **Figure 10** shows the learning curves for each group based on this classification. A 3 group (between subjects: musicians using MD, non-musicians using MD, non-musicians using UD, or RR strategies) × 5 training block (within subjects) mixed design ANOVA conducted on proportion correct (accuracy) revealed a significant main effect of group [$F(2,27) = 23.69$, $p < 0.0001$, partial $\eta 2 = 0.64$], a significant main effect of block [$F(4,108) = 52.99$, $p < 0.0001$, partial $\eta 2 = 0.66$], and a significant interaction2between block and group [$F(8,108) = 5.38$, $p < 0.0001$, partial $\eta = 0.28$]. *Post hoc* pair-wise comparisons with Bonferroni correction of the group main effect revealed that both

musicians and non-musicians using MD strategies were significantly more accurate than non-musicians using UD strategies in all blocks (all p's < 0.01). The comparison of musicians and non-musicians who used MD strategies did not reach significance ($p > 0.38$). Thus, although musicians are more likely to utilize MD strategies than non-musicians, those musicians and non-musicians who use MD strategies do so with nearly the same accuracy. This is an important finding as it suggests a critical mechanism (MD strategy use) associated with enhanced speech learning (Chandrasekaran et al., 2013; Maddox & Chandrasekaran, 2014; Maddox et al., 2013; Yi et al., 2014).

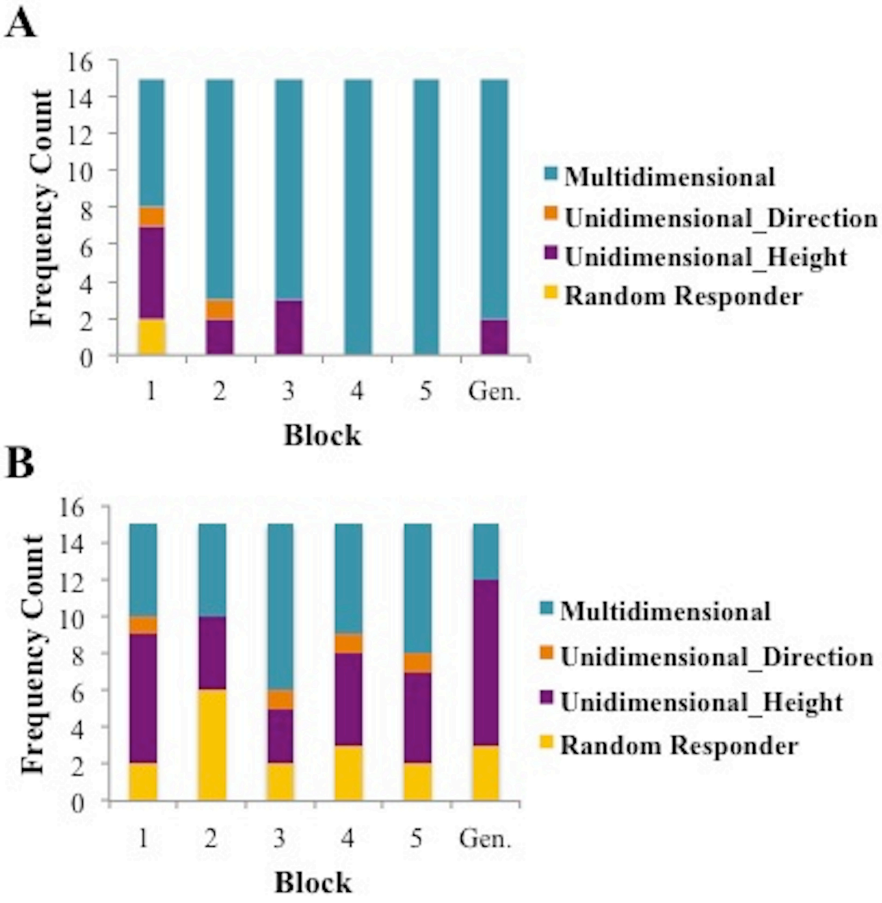


Figure 9. Strategy use counts per block for musicians (A) and non-musicians (B).

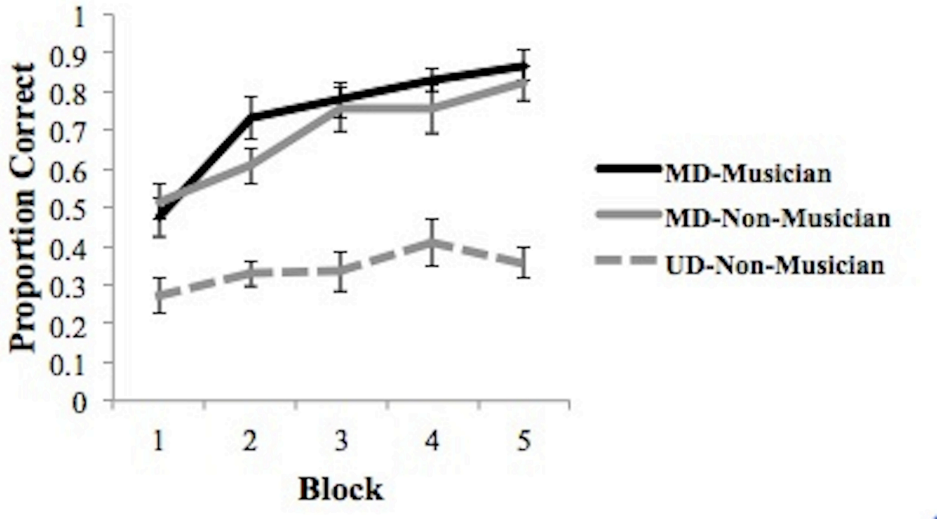


Figure 10. Average proportion correct across all training blocks for MD musicians, MD and UD non-musicians based on final block strategy. Bars represent SEM.

Working Memory and Cognitive Strategies

We also investigated any working memory differences between participants who arrived at a UD strategy versus participants who arrived at a MD strategy in block 5. Importantly, we did not find any working memory difference between our musician group and non-musician group [$t(28) = 1.51$, $p = 0.14$]. While this does not replicate previous work (Parbery-Clark et al., 2009; Pallesen et al., 2010; George and Coch, 2011; Kraus et al., 2012; Strait et al., 2013), our computational modeling can give us insight into why this may be.

Executive function is critical for MD strategy use as it is a complex decisional process requiring the maintenance of multiple pieces of auditory information in order to make a categorical judgment. Thus, we predict that participants who use MD strategies will have a higher working memory capacity. To test this, we conducted a one-way ANOVA of group (between subjects: musician, non-musician) and block 5 strategy [between subjects: MD, non-MD (UD and RR)] on composite working memory scores (forward score + backward score). The ANOVA revealed a significant main effect of strategy [$F(1,27) = 7.28$, $p < 0.01$], but no significant main effect of group [$F(1,27) = 2.80$, $p = 0.11$] on composite working memory score. *Post hoc* t-tests between groups suggest that block 5 MD users have a significantly higher working memory composite score than block 5 non-MD users [$t(28) = 3.21$, $p < 0.005$]. Within just non-musicians, block 5 MD users have a significantly higher working memory composite score relative to block 5 non-MD users [$t(13) = 2.55$, $p < 0.05$]. In addition, there is no difference in working memory composite scores between non-musician block 5 MD users and musician block 5 MD users [$t(20) = 0.27$, $p = 0.79$]. Because there were no UD or RR musicians, we could not compare their working memory scores to those of MD

musicians. These results suggest that working memory abilities may partially explain who uses a MD strategy by the end of training, regardless of music training.

Strategies and Accuracy Rates in Generalization Block

Turning to the generalization block, a Fisher exact test reveals that there were significantly more musicians using a MD strategy relative to non-musicians using a MD strategy ($p < 0.001$). Next, we explored the accuracy rates associated with musicians and non-musicians who were either MD strategy users or UD strategy users in the generalization block (strategy counts in **Figure 9**) and found that non-musicians using MD strategies obtained marginally higher accuracy rates than non-musicians using UD strategies [$t(10) = 2.03$, $p = 0.07$]. Likewise, musicians using MD strategies obtained significantly higher accuracy rates than musicians using UD strategies [$t(13) = 2.43$, $p < 0.05$] whereas musicians using MD strategies were no more accurate than non- musicians using MD strategies [$t(14) = 0.59$, $p = 0.56$]. Just as in the training blocks, these results suggest that employing a MD strategy, regardless of music experience, enhances accuracy. However, these results should be interpreted with caution due to the small sample sizes.

Computational Modeling Results of Perceptual Representation Across Blocks

In this section, we examine the effects of musical training on perceptual selectivity along the pitch height and pitch direction dimensions. A number of studies in the literature (Goldstone, 1994; Maddox, 2001; Maddox & Bogdanov, 2000; Maddox & Dodd, 2003) suggest that perceptual forms of selectivity often follow when decisional forms of selectivity are operative, but not always (Filoteo & Maddox, 1999). Given that English speakers naturally weight pitch height, due to its relevance in English, it is reasonable to suppose that musicians and non-musicians will not show any differences in

perceptual selectivity along the pitch height dimension before training (however, see (Perrachione, Fedorenko, Vinke, Gibson, & Dilley, 2013 regarding the influence of music experience on perceptual selectivity at the sentence-level). It is likely, however, that musical training leads to enhanced perceptual selectivity along the pitch direction dimension and thus musicians will show smaller estimates of perceptual noise. Because we focus on the perceptual variability estimates, we wanted to use the model that best accounted for the data. This, by definition, is the most general MD model.

First, we examined the effects of musical training on perceptual selectivity along the pitch height dimension. We conducted a 2 group (between subjects) × 5 block (within subjects) mixed design ANOVA, with "participant" as a random variable. We found a main effect of group [$F(1,28) = 4.16$, $p = 0.051$, partial $\eta 2 = 0.129$], and a main effect of block [$F(4,112) = 23.59$, $p < 0.001$, partial $\eta 2 = 0.457$], but no interaction [$F(4,112) = 1.55$, $p = 0.194$, partial $\eta 2 = 0.052$]. Musicians showed better perceptual selectivity in the form of smaller perceptual variance (mean = 0.17) compared to non- musicians (mean = 0.29). In addition, perceptual variance across groups decreased with learning (mean of block 1 = 0.43; mean of block 5 = 0.12). These results are displayed in **Figure 11**.

Second, we examined the effects of musical training on perceptual selectivity along the pitch direction dimension. We conducted a 2 group (between subjects) × 5 block (within subjects) mixed design ANOVA, with "participant" as a random variable. We found a significant interaction [$F(4,112) = 2.87$, $p < 0.05$, partial $\eta 2 = 0.093$], along with a significant main effect of group [$F(1,28) = 11.38$, $p < 0.005$, partial $\eta 2 = 0.289$], and a significant main effect of block [$F(4,112) = 3.62$, $p < 0.01$, partial $\eta 2 = 0.115$]. To identify the locus of the significant interaction, we conducted two analyses. First, we ran *t*-tests comparing musicians and non-musicians at each block. We found significant smaller perceptual variance estimates for musicians in all blocks except the first [block 1:

$t(28) = 0.42$, $p = 0.68$; block 2: $t(28) = 4.33$, $p < 0.0005$; block 3: $t(28) = 2.13$, $p < 0.05$; block 4: $t(28) = 2.92$, $p < 0.01$; block 5: $t(28) = 3.01$, $p < 0.01$]. Next, we conducted separate one-way repeated measures ANOVA's within each group and found musicians' perceptual variance estimates along the pitch direction dimension declined significantly across blocks [$F(4,56) = 15.24$, $p < 0.0001$, partial $\eta 2 = 0.521$] whereas non-musicians' did not [$F(4,56) = 0.57$, $p = 0.69$, partial $\eta 2 = 0.039$].

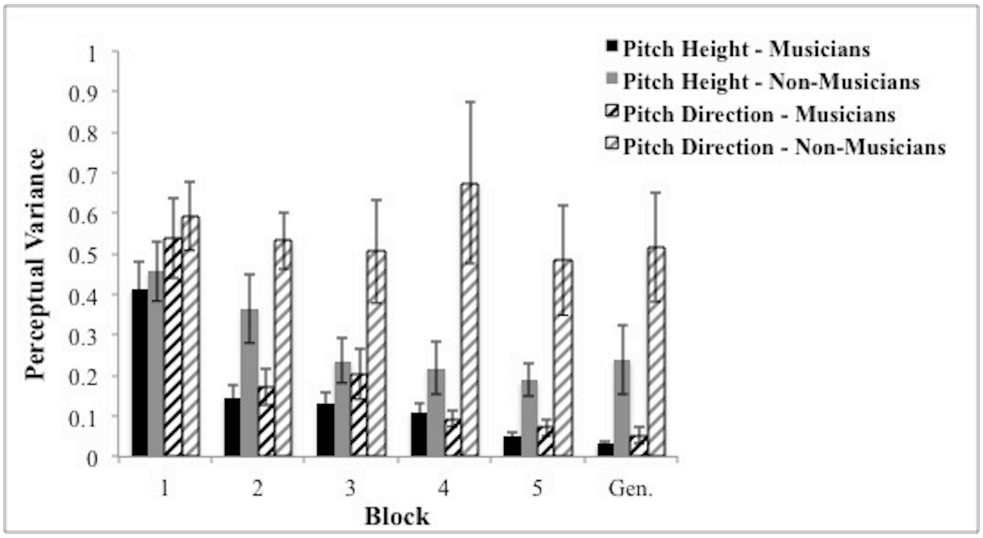


Figure 11. Average perceptual variance of groups across training blocks and generalization block in the pitch height and pitch direction dimensions. Bars represent SEM.

Computational Modeling of Perceptual Representation in Generalization Block

Here we examine the perceptual variance estimates in the generalization block. These analyses allow us to determine how perceptual variability along the pitch height and pitch direction dimensions changes in the context of a new talker and no feedback. Perceptual variance estimates were smaller for musicians relative to non-musicians along both the pitch height [$t(28) = 2.42$, $p < 0.05$], and pitch direction dimensions [$t(28) = 3.39$, $p < 0.005$]. These results are depicted in **Figure 11**. We also compared the perceptual variance estimates in the final training block to those in the generalization block. Interestingly, the pitch height and pitch direction perceptual variance estimates were numerically smaller in the generalization block than in the final training block for musicians, but were numerically larger for non-musicians. Even so, the only one of these comparisons to reach statistical significance was for musicians along the pitch height dimension [$t(14) = 2.21$, $p < 0.05$].

DISCUSSION

We examined the effects of long-term musical training on non-native speech learning in native English speakers, none of who had prior experience with Mandarin tones. Our results show a musician advantage (average accuracy = 0.74) relative to non-musicians (average accuracy = 0.50) in learning to categorize naturally produced Mandarin tones. Our results are consistent with previous studies that have identified a musician advantage in learning speech categories (Gottfried and Riester, 2000; Alexander et al., 2005; Wong and Perrachione, 2007; Lee and Hung, 2008). While accuracy differences help identify a cross-domain (between music and speech learning) advantage for musicians, they do not provide information on the specific mechanistic underpinnings of the advantage. To this end, we employed computational modeling analyses to examine the locus of the musician advantage. Specifically, our models specified decisional

strategies used by musicians and non-musicians, as well as perceptual processes that are independent of the decisional processes. The computational modeling results revealed that musicians used the optimal, MD strategy faster, and more frequently than non-musicians. This suggests musicians have enhanced cognitive processing supporting categorical decisional judgments relative to non-musicians as a group. Importantly, the model-based analyses allow us to examine decision processes in each individual. Although musicians used MD strategies faster and more frequently than non-musicians, when compared to non-musicians who used MD strategies by block 5, there were no differences in accuracy rates. In addition, across participant groups, participants who used MD strategies in the final training block had a significantly higher working memory composite score than those who used UD strategies. Specifically, musicians and non-musicians who used MD strategies in block 5 were no different in their composite working memory scores. In addition, non-musicians who used MD strategies in block 5 had a significantly higher working memory score than non-musicians who had did not use a MD strategy in block 5. These are critical findings as they suggest a mechanism for the musician advantage; namely, an increased use of MD strategies, since musicians and non-musicians who used MD strategies by the end of the training were very similar with respect to accuracy and working memory capacity.

Increased use of MD strategies leads to enhanced speech learning, but changes in perceptual processing may also explain better performance. Importantly, these parameters are theoretically independent from the decision parameters (Green & Swets, 1967; Maddox & Ashby, 1993) and in some cases are empirically independent (Filoteo & Maddox, 1999). The current results suggest that both musicians and non-musicians show increased perceptual selectivity (i.e., reduced perceptual variance or noise) along the pitch height dimension with learning. However, only musicians show increased

perceptual selectivity (or more veridical perception) along the under- weighted pitch direction dimension. Together, this suggests that the performance advantage in learning non-native speech sounds for musicians relative to non-musicians is due not only to cognitive processes, but also perceptual processes and is consistent with enhanced perceptual representation of dynamic pitch changes in musicians, relative to non-musicians (Wong et al., 2007). Why would long-term music training promote cross-domain auditory plasticity? Studies examining plasticity related to music training have examined basic perceptual encoding of auditory signals as well as higher-level linguistic and cognitive processes. For example, musicians show enhanced encoding of linguistic pitch patterns at the level of the midbrain/brainstem relative to non-musicians (Wong et al., 2007). Such perceptual encoding advantages could lead to faster speech learning in musicians by relaying a more faithful representation of the speech signal to the cortex than non-musicians. A general cognitive/decisional advantage could drive enhanced speech learning as well. In fact, a previous proposal posits a reciprocal process where cognitive advantages drive perceptual advantages in a top– down manner (Strait et al., 2010). The OPERA hypothesis suggests that music training places a significantly greater demand on the perceptual and/or cognitive circuitry that is shared between music and speech (Patel, 2014). In addition, recent findings suggest common mechanisms underlying music aptitude and speech-sound processing (Kempe et al., 2013). Thus, long-term training could alter cognitive and perceptual processes that are common to music and speech, resulting in enhanced learning of components shared between the two domains.

In the current study we examined the extent to which music training enhanced learning of non-native, linguistic pitch patterns. Pitch is a basic element in music and speech, and both domains use pitch patterns to convey information extensively. In

English speech, pitch patterns can convey information related to prosody and emotion. Pitch patterns are also used in some languages (e.g., Mandarin Chinese) within a syllable to change word meaning. Native English-speaking adults struggle in learning Mandarin pitch patterns and often confuse one tone category with another (Bharath Chandrasekaran, Sampath, & Wong, 2010; Wang et al., 1999). Our results show that music training can enhance the ability to categorize non-native linguistic pitch patterns. Computational modeling helps pinpoint the locus of this advantage by showing that musicians use the optimal MD strategy sooner and more often than non-musicians. In addition, musicians shower greater perceptual selectivity of the stimuli along the pitch direction dimension relative to non-musicians.

Lexical tones are well characterized by a MD space with two dimensions related to pitch (pitch height and direction) that can help disambiguate tone categories. The relative weighting between dimensions is language-dependent, where native English speakers tend to weight pitch direction less than native Mandarin speakers, reflecting the relative difference in the use of this dimension in their language (Chandrasekaran et al., 2007; Gandour & Harshman, 1978). Thus, native English speakers focus predominantly on pitch height to disambiguate tone categories. In previous studies using computational models we found that relying on only one of the two dimensions during learning (a UD decision strategy) is a sub-optimal strategy (Maddox et al., 2013; Yi et al., 2014). For example, an over-reliance on pitch height (is it high or low?) is not optimal because it leads to confusions between the rising and the falling tones (which have similar average heights but differ considerably in direction).

Pitch height is also highly talker-dependent; for example, it is a critical cue in differentiating male and female talkers. Thus, an over-reliance on this dimension may lead to category confusions across talkers. The computational modeling results of the

current study show that relative to non-musicians, musicians were faster and more frequent users of MD strategies, which incorporate both pitch height and pitch direction information- an advantageous strategy that promotes greater differentiation between tone categories.

While learning is important, generalization of the learned material is also important, especially in the case of speech as it rarely occurs in the same context. Different talkers with variable speaking characteristics such as rate of speech, average pitch, etc., all create unique contexts in which speech must be understood. Therefore, in addition to during the five blocks of learning, we examined accuracies, strategies, and perceptual selectivity during a generalization block in which participants were required to categorize the four Mandarin tones in the context of a single, new speaker and received no feedback. Musicians showed an accuracy advantage that was supported by enhancements in both decisional strategies (larger number of MD users) and perceptual selectivity (smaller perceptual variance along pitch height and pitch direction dimensions). A large literature suggests that non- native speech sound training which implements highly variable training stimuli is more conducive than low variable training stimuli to successfully generalizing learned speech sounds to new contexts (see Bradlow, 2008; Perrachione et al., 2011). Importantly, prior research has manipulated the training paradigm in order to produce successful generalization. The current results build off of this literature and suggest there may also be individual differences (such as music training) involved in how successful a participant is in generalizing learned non-native speech sounds to novel contexts. Future research should investigate how and which individual differences lead to successful generalization of learned non-native speech sounds.

The burgeoning literature on the cross-domain plasticity induced by long-term music training has led several researchers to propose music training as a clinical training tool. Our current findings hold promise for using long-term music training as a method to help clinical populations that demonstrate greater auditory–perceptual variability (Hornickel & Kraus, 2013) and learning-related difficulties. However, on a cautionary note, several questions and criticisms should be addressed before pursuing more clinical goals. For example, first, it is unclear whether the cognitive and perceptual advantages reflect an effect of long-term music training, or a general predisposition that drives individuals toward music training. A recent longitudinal study suggests the former (Kraus et al., 2014). Using a longitudinal design, children from the Los Angeles area were randomly assigned to either defer music involvement for a year and receive only 1 year of music lessons, or begin music instruction immediately and receive a total of 2 years of music training. By the end of the 2-year training, the second group, which had received 2 years of music training, showed stronger neurophysiological distinctions of /ba/ versus /ga/ sounds, while the first group did not. In addition, within the second group, number of hours spent practicing over the 2-year training period positively correlated with improvement in neural differentiation (Kraus et al., 2014). However, there were several limitations that prevent strong inferences from being drawn. For instance, an active control group against which they could compare the gains in the 2-year music group was not included. In addition, there were several issues regarding analyses of the data, and no behavioral data were presented (Evans et al., 2014). Next, we need to evaluate the specificity of the musician advantage. Pitch and changes in pitch are clearly important attributes of music. Whether cognitive and perceptual advantages percolate to other attributes of sound such as loudness and duration needs to be addressed in future studies. Lastly, in the current study we use a definition of ‘musician’ that is derived from the

larger existing literature; however, this definition is admittedly narrow (see Levitin, 2012 for example), as is the definition of a 'non-musician.' In addition, a larger sample size, allowing the examination of music training to be a continuous variable, and a well-established performance-based measure would prove useful.

Future Directions

There are many available future directions. One is to more broadly explore the extent of the observed musician cognitive advantage in speech learning. For instance, cognitive tasks that show musician advantages are frontally mediated cognitive tasks that test executive function (Bialystok & DePape, 2009), working memory (George & Coch, 2011; Nina Kraus et al., 2012; Pallesen et al., 2010; Alexandra Parbery-Clark et al., 2009; Strait et al., 2013), and switching (Hanna-Pladdy & MacKay, 2011). Musicians also show increased gray matter volume in the dorsolateral prefrontal cortex (Bermudez, Lerch, Evans, & Zatorre, 2009). Given that musicians show frontally mediated advantages, it is possible these complex frontally mediated rule-based strategies drive cross-domain auditory plasticity, especially given the task-dependent nature of activation in the human auditory cortex (Ohl & Scheich, 2005). Notably, when construed within a dual-learning systems perspective, a rule-based learning advantage may not transfer to all learning conditions. Within the dual-learning systems framework, a *reflective* system, which uses executive attention and working memory, is *competitive* with the *reflexive* system, which relies on dopamine-mediated reward signals in the striatum (Ashby & Maddox, 2005, 2011; Maddox & Chandrasekaran, 2014; Maddox et al., 2013). Since these two systems are competitive, if the musician advantages in cross-domain plasticity are driven purely by the frontally mediated cognitive advantages, musicians should perform worse on auditory tasks that require the reflexive, striatally mediated, system than on auditory tasks

that require the reflective system. Thus a robust theoretical framework may help establish the limits of neuroplasticity related to music training.

A second future direction is to investigate whether different music-training environments provide different cognitive or perceptual benefits related to non-native speech learning. In the present study, we used musicians who have at least 10 years of formal group or private training. It is possible that musicians with less training, those who play instruments from different instrument families, those who are self-taught, or those who play instruments that use non-Western tonality will show different learning patterns compared to the musicians in this study. For instance, many non-Western styles of music use tonalities distinguish between smaller differences in pitch than Western music. This may result in greater demands on the perceptual system, and consequently lead to a non-Western trained musician advantage over Western-trained musicians in learning non-native speech sounds due to the increased sensitivity to smaller pitch differences. Lastly, research suggests that non-human species are capable of differentiating between different types of pitch movements – a skill trained during music learning and used in non-native speech learning (Brosch et al., 2004; Ohl, Scheich, & Freeman, 2001). As suggested by Patel (2014), animal models may provide valuable insight into how specific aspects of music training (i.e., pitch movements) may influence species-specific language components such as vocalizations, and thus clarify how music training may affect speech learning.

CONCLUSION

Using rigorous computational modeling, we extended prior research by showing that the musician accuracy advantage relative to non-musicians observed in prior studies can be attributed to both cognitive advantages, as evidenced by earlier and more frequent

use of the optimal MD strategy; and perceptual advantages, as evidenced by smaller perceptual noise along both the pitch height and pitch direction dimensions. In addition, musicians and non-musicians who used MD strategies by the end of training showed no differences in accuracy and working memory scores. Contrastingly, participants who used MD strategies by the end of training showed higher accuracy rates and working memory scores than those who used UD or RR strategies. These results suggest a cognitive mechanism for the musician accuracy advantage. Specifically, the use of MD strategies faster and more often relative to non-musicians. In the generalization block, where stimuli were presented by a new talker, and no feedback was given, more musicians used the optimal strategy and obtained a higher accuracy relative to non-musicians. At the perceptual level, our modeling revealed that musicians' perception of the stimuli is more veridical, especially along the normally underweighted pitch direction dimension. This pattern extended to novel stimuli during a generalization phase. These results provide further evidence for cross-domain auditory plasticity due to music training.

Paper 3: Better Late Than Never (or Early): Music Training in Late Childhood is Associated with Enhanced Decision-Making[3]

INTRODUCTION

Decision-making is a critical skill required for everyday functioning. We use it to make choices, guide our actions, and complete tasks throughout the day. We can choose options that are riskier or safer and those decisions may lead to positive or negative outcomes immediately or after some delay. Prior work has shown that decision-making relies heavily on the prefrontal cortex across species (Barraclough, Conroy, & Lee, 2004; Bechara, Damasio, Tranel, & Anderson, 1998; Broche-Pérez, Herrera Jiménez, & Omar-Martínez, 2016; Hare, O'Doherty, Camerer, Schultz, & Rangel, 2008; Li, Lu, D'Argembeau, Ng, & Bechara, 2009). More specifically, Hare and colleagues found that activity in the medial orbitofrontal cortex is associated with the predicted amount of reward each option will confer (i.e. goal values), while the central orbitofrontal region is associated with measuring the net reward of choosing an option (i.e., decision values; Hare et al., 2008). Goals and decision values are critical components of the decision-making process that guide the participant to actions that result in the largest net benefit. Additionally, animal studies suggest converging evidence implicating the orbitofrontal region in updating expected outcomes during a decision-making task (Sul, Kim, Huh, Lee, & Jung, 2010). A recent study also implicates the medial prefrontal cortex in tracking differences between a received reward and the expected reward on a trial-by-trial basis (Samanez-Larkin, Worthy, Mata, McClure, & Knutson, 2014). Lastly, prior studies

[3] This paper was previously published as Smayda, K. E., Worthy, D. A., & Chandrasekaran, B. (2017). Better late than never (or early): Music training in late childhood is associated with enhanced decision-making. *Psychology of Music*, 305735617723721. KES partially conceived and designed the experiment, performed the experiment, computational modeling, analyses of the data, and wrote the paper. KES partly conceived and designed the experiment, analyzed the data, contributed materials/analysis tools, and wrote the paper.

have found that damage to the ventromedial prefrontal cortex (Bechara, Damasio, Damasio, & Anderson, 1994) dorsolateral and dorsomedial prefrontal cortex (Bechara et al., 1998; Manes et al., 2002) results in impaired decision-making ability. Together, these results suggest that while different subregions of the prefrontal cortex may be responsible for specific reward-based processes during a decision-making task, the prefrontal cortex as a whole is a critical brain region for successful decision-making.

The prefrontal cortex demonstrates protracted development through childhood (Gogtay et al., 2004), suggesting that as children age, and their prefrontal cortices develop, they should show enhanced decision-making abilities. Supporting this notion, prior work suggests that in a classic decision-making task, the Iowa Gambling Task (IGT), in which the participant must choose between decks that lead to net gains or net losses, the tendency to avoid disadvantageous deck choices increases from childhood through adolescence to adulthood (Hooper, Luciana, Conklin, & Yarger, 2004) and overall performance also increases throughout childhood (Crone & van der Molen, 2004). The IGT has also been used as a measure of incentive-based decision-making across a variety of populations including healthy adolescents (Hooper et al., 2004), healthy older adults (Wood, Busemeyer, Koling, Cox, & Davis, 2005), and clinical populations such as patients with schizophrenia (Shurman, Horan, & Nuechterlein, 2005), bipolar disorder (Ono et al., 2015), and damage to the ventromedial prefrontal cortex (Bechara et al., 1994).

Music training across the lifespan confers advantages in domains well beyond music (Herholz & Zatorre, 2012; Kraus & Chandrasekaran, 2010) such as speech perception (Costa-Giomi, 1999; Kraus & Chandrasekaran, 2010; Sylvain Moreno et al., 2011) and cognitive abilities (Costa-Giomi, 1999; Sylvain Moreno et al., 2011; Schellenberg, 2005) and frontally-mediated cognitive abilities such as working memory (

Nutley et al., 2014; George & Coch, 2011; Kraus et al., 2012), processing speed (Bugos, 2010; Bugos & Mostafa, 2011), and cognitive control (Bialystok & DePape, 2009; Pallesen et al., 2010). Music training beginning before the age of seven relative to later in life, can enhance sensorimotor abilities that persist beyond childhood (Steele, Bailey, Zatorre, & Penhune, 2013). These findings support the concept of a "critical period," which suggests that introducing a novel skill that utilizes brain regions that are undergoing significant development will confer long-lasting cross-domain benefits (Hensch, 2005; Steele et al., 2013). Given that regions implicated in decision-making, namely the prefrontal cortex, do not develop until late childhood, we posit that beginning to play music in late childhood may provide an enduring enhancement in decision-making relative to beginning to play music earlier in childhood or never at all.

The goal of the current study is to examine the extent to which music training impacts decision-making, as measured by the Iowa Gambling Task. We hypothesize that musicians will show a performance advantage in decision-making relative to non-musicians, supported by the literature showing associations between music training and enhanced cognitive ability. We also hypothesize that if there is an advantage, it may only be present in musicians who began music training later in childhood, given that the brain regions supporting decision-making begin maturing later in childhood (Gogtay et al., 2004).

We examined decision-making performance using behavioral and computational modeling approaches in three groups: Adults who began playing music at or before the age of 8, classified as "early-trained" musicians (ET), adults who began playing music after the age of 8 classified as a "late-trained" (LT) musicians. Adults who never played an instrument were considered non-musicians (NM). We predicted that late-trained musicians would perform better than both early-trained musicians and non-musicians.

Although accuracy data is informative for showing group differences, we fit computational models to response patterns to identify the mechanism behind any performance differences between groups. Lastly, we included data from a "many labs" collaboration as a Super Control group in our behavioral analyses. The purpose of including the Super Control group is to compare any group effect to a large, established well-studied sample that is representative of healthy young adult performers on the Iowa Gambling Task (Steingroever et al., 2015).

METHOD

Participants

A total of 69 participants (aged 18–35 years) were recruited from The University of Texas at Austin community and were part of a larger multi-day study, which was occurring at the same time as testing for the present study. Written consent was obtained from all participants and the Institutional Review Board of The University of Texas at Austin approved the experimental procedure. Participants were classified as non-musicians (NM), early-trained musicians (ET), or late-trained musicians (LT). Operationally, non-musicians had less than 2 years of music training and were not currently playing any instrument. ET had at least 8 years of music training beginning at or before age 8, while LT had at least 8 years of music training beginning after the age of 8. Several intervention studies examining neural indices of cross-domain enhancements from music training have used 8 years as the starting point of music training (Chobert, Francois, Velay, & Besson, 2014; Moreno et al., 2009). These studies consistently find evidence of music training induced plasticity even past 8 years of age. We used 8 years as the cut-off to be consistent with prior literature examining early versus late training as well as accommodating for the work on brain plasticity induced by music training past

the “so-called” sensorimotor critical period (Bailey & Penhune, 2012; Steele et al., 2013). Both ET and LT actively practiced instruments at the time of testing and did not significantly differ in the number of hours they currently played, $t(44) = 0.44$, $p = .66$, the number of instruments they played, $t(44) = 0.85$, $p = .40$, or their age, $t(44) = 1.11$, $p = .27$. Participant demographic details can be found in Table 8. Reported participant ages for the Super Control group from the “many labs” collaboration were not complete, but from the five studies that published age information, the mean age of participants is 25.6 years (datasets from Horstmann, 2012; Kjome et al., 2010; Premkumar et al., 2008; Steingroever, Šmíra, Lee, & Pachur, n.d.; Wood et al., 2005), and for the two studies which did not report ages, the samples included only undergraduates (Maia & McClelland, 2004; Worthy, Pang, & Byrne, 2013). No music experience metric was reported for any of the studies included in the “many labs” collaboration. Several participants across both early-trained and late-trained groups in the present study played multiple instruments, as represented in the instrument count in Table 9.

Table 8. Demographic information of participants.

	n	Age Mean (SD)	Years Playing Music Mean (SD)	Age Began Playing Mean (SD)	Hours Currently Play Mean (SD)
Non-Musicians	23	23.91 (4.47)	0.00 (0.00)	NA	0.00 (0.00)
Early-Trained Musicians	23	23.70 (5.34)	16.91 (4.82)	5.61 (1.85)	7.42 (5.82)
Late-Trained Musicians	23	25.17 (3.50)	13.65 (3.72)	11.43 (2.11)	6.46 (8.78)

Table 9. Instrument counts across musician groups.

	Strings	Brass	Woodwind	Keys	Percussion	Voice	Other
Early-Trained Musicians	11	1	6	18	7	10	3
Late-Trained Musicians	17	2	4	12	3	7	2

Procedure

The Iowa Gambling Task was run on PCs using Matlab with Psychtoolbox (version 2.5). Participants were seated at a computer and instructed that on each trial, they were to choose a card from one of four decks (A, B, C, D) by pressing "Z," "W," "P" or "/" on the keyboard. They were told that after each card choice, they would be shown the number of points gained or lost, and that their goal was to maximize the number of points gained. The participants were unaware that two of the decks (C and D) led to a net gain of points (and are therefore considered "good decks") while the other two led to a net loss of points (A and B; considered "bad decks").

In addition, the decks varied by magnitude and frequency of points won or lost such that choosing decks B or D resulted in large but infrequent losses, while choosing decks A or C led to more frequent and smaller-magnitude losses. The task consisted of five 20-trial blocks and was identical to the task in Worthy, Pang, et al. (2013).

Neuropsychological testing measure descriptions

Participants completed several neuropsychological tests. The Stroop Test was administered to assess selective attention and inhibitory control ability (Stroop, 1935). A "relative interference score" was calculated as the number of words read in the "incongruent" condition normalized by the average of the number of words read in the "color" and "word" conditions. Operation Span was used to assess complex working memory ability. At a computer, participants were asked to remember a string of letters while completing a secondary math problem between each letter presented in the sequence. After each sequence of letters was presented, participants were asked to recall the sequence of letters in the order they were presented. Additionally, participants were told to maintain an accuracy of 85% or higher on the math problems (Unsworth, Heitz, Schrock, & Engle, 2005; see also Xie et al., 2015). Lastly, the WAIS-III Digit Span was

administered to assess verbal working memory. For this task, participants were required to repeat strings of numbers read aloud by the researcher; first exactly how the researcher read them, and then backwards (Wechsler, 1997).

RESULTS

Iowa Gambling Task behavioral results

Overall performance on the IGT was our primary measure and was calculated as the proportion of good minus bad deck selections for each subject in five 20-trial blocks (C + D – A – B; Worthy, Pang, et al., 2013). A mixed ANOVA with group as a between-subjects variable (non-musician, early-trained, late-trained) and block as a within-subjects variable (1 through 5), revealed significant main effects of group, $F(2, 66) = 5.10$, $p < .01$, partial $\eta 2 = 0.13$, block, $F(4, 264) = 8.19$, $p < .001$, partial $\eta 2 = 0.11$, and an interaction between group and block, $F(8, 264) = 2.56$, $p < .05$, partial $\eta 2 = 0.07$. To decompose the significant interaction, we ran a series of one-way ANOVAs at each block comparing the effect of group on overall performance. We found a significant effect of group in blocks 4 and 5. Pairwise comparisons of overall performance between groups in blocks 4 and 5 suggest an LT advantage relative to NM in block 4, $t(44) = 3.29$, $p < .005$, and ET and NM in block 5, $t(44) = 4.61$, $p < .001$; $t(44) = 2.08$, $p < .05$, respectively. These results are displayed in Figure 1(a).

We also compared LT to a 441-participant "Super Control" group of young adults who performed the 100-trial IGT (shown in gray in Figure 1(a); Steingroever et al., 2015). To determine if the late-trained musicians performed significantly different from the Super Control group, we ran a mixed model ANOVA with group as a between-subjects variable (Super Control group, LT musician) and block as a within-subjects variable (1 through 5). We found significant main effects of group, $F(1, 462) = 6.96$, $p <$

.009, partial $\eta 2 = 0.01$, and block, $F(4, 1848) = 83.67$, $p < .0001$, partial $\eta 2 = 0.15$, and interaction between group and block, $F(4, 1848) = 2.39$, $p = .05$, partial $\eta 2 = 0.005$. These results suggest that late-trained musicians perform significantly better than the Super Control group.

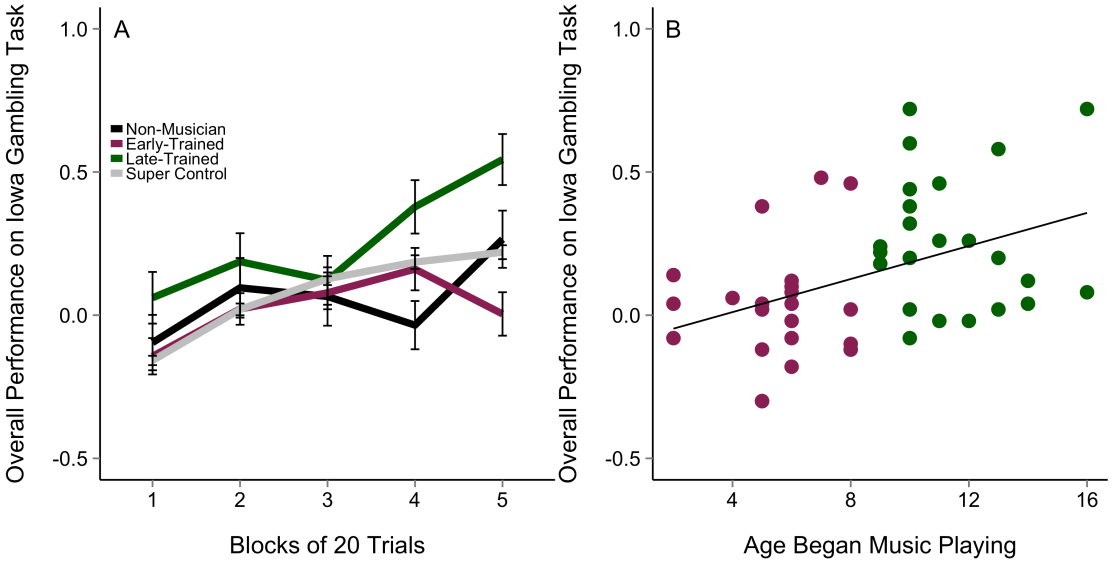


Figure 12. Overall performance on the IGT. Error bars represent standard error of the mean (a). Overall performance of musicians on the IGT plotted with the age they began playing music (b).

Within just the early- and late-trained musician groups, a linear regression of age of onset of music training as a predictor of overall performance suggests that the later an individual begins playing music in childhood, the greater the benefit to their decision-making skills as young adults, $F(1, 44) = 8.96$, $p < .01$, $R^2 = 0.17$, $b = 0.03$; displayed in Figure 1(b). Interestingly, although the two musician groups differed significantly in their years of playing music, LT: $M = 13.65$, ET: $M = 16.91$; $t(44) = 2.57$, $p < .05$, years of playing music was not a significant predictor of performance, $F(1, 44) = 1.90$, $p = .17$, $R^2 = 0.04$; $b = -0.01$.

Deck choice results

We also compared differences in deck choices between our four participant groups (late-trained musicians, early-trained musicians, non-musicians, and the Super Control group) in two steps. First, we compared deck choices between non-musicians, late-trained musicians, and early-trained musicians; and second, we compared deck choices between the late-trained musicians and the Super Control group. For our first analysis, we performed a two-way ANOVA on proportion of deck chosen with group (non-musicians, late-trained musicians, early-trained musicians) as a between-subjects variable and deck (A, B, C, D) as a within-subjects variable. We did not find a significant main effect of group, $F(2, 66) = 1.24$, $p = .300$, but we did find a significant effect of deck, $F(3, 198) = 11.14$, $p < .001$, partial $\eta 2 = 0.14$. There was no significant interaction of group and deck, $F(6, 198) = 2.14$, $p = 0.051$. Results are displayed in Figure 2 and suggest that Deck D was chosen most often ($M = 0.32$), then Decks B ($M = 0.26$) and C ($M = 0.24$), and Deck A was chosen the least often ($M = 0.19$). Post-hoc pairwise comparisons (with Holm correction) of selections in each deck suggest a group difference in only Deck B selection. Late-trained musicians selected Deck B significantly less ($M =$

0.21) than both early-trained musicians, $M = 0.28$; $t(44) = 2.97$, $p < .05$, and non-musicians, $M = 0.28$; $t(44) = 2.38$, $p < .05$. To compare deck choices between late-trained musicians and the Super Control group, we ran a two-way ANOVA on proportion of deck chosen with group and deck (A, B, C, D) and found significant main effects of deck, $F(3, 1386) = 98.43$, $p < .001$, partial $\eta 2 = 0.18$, and group, $F(1, 462) = 14.91$, $p < .0002$, partial $\eta 2 = 0.03$, and a significant interaction of deck and group, $F(3, 1386) = 4.36$, $p < .006$, partial $\eta 2 = 0.5$. Post-hoc analyses at each deck suggest late-trained musicians and the Super Control group differ in the proportion they chose a deck in only Deck B, $F(1, 462) = 10.95$, $p < .002$, partial $\eta 2 = 0.02$. Results are displayed in Figure 13.

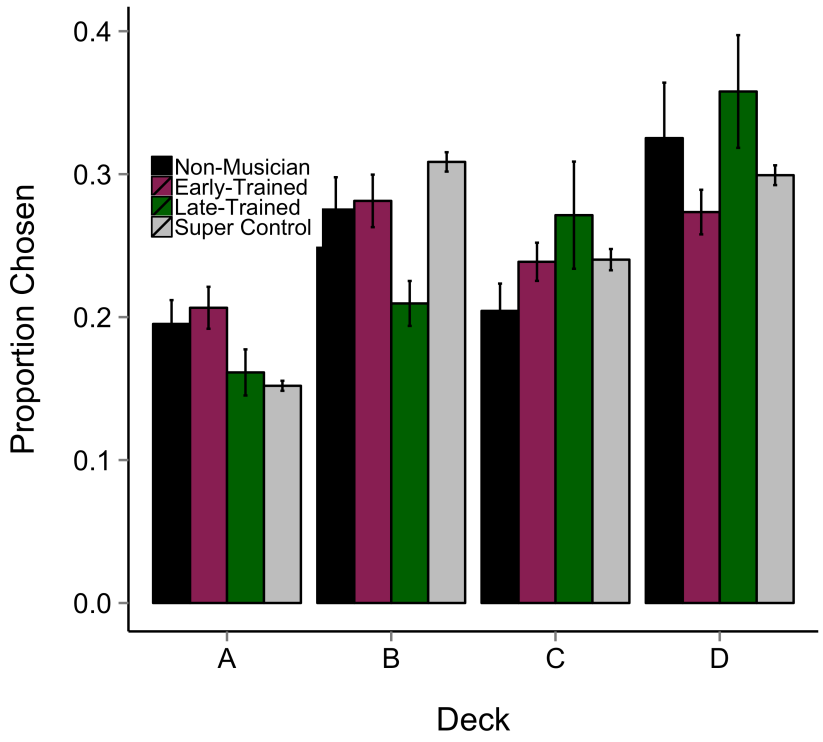


Figure 13. Proportion of decks chosen by non-musicians, early-trained musicians, late-trained musicians, and the Super Control group. Error bars represent standard error of the mean.

Neuropsychological testing results

To determine if there are any other differences between our participant groups that may mediate the late-musician advantage in the Iowa Gambling Task, we ran a series of planned pairwise t tests with Holm correction between the neuropsychological test scores of our three groups. Our results suggest there are no differences between groups with respect to Stroop Interference score [ET and LT: $t(44) = 0.50$, $p = 1.00$; ET and NM: $t(44) = 0.65$, $p = 1.00$; LT and NM: $t(44) = 0.25$, $p = 1.00$], Digit Span Total score [ET and LT: $t(44) = 0.50$, $p = .93$; ET and NM: $t(44) = 0.65$, $p = .91$; LT and NM: $t(44) = 1.08$, $p = 0.87$], and Operation Span score [ET and LT: $t(38) = 0.88$, $p = 1.00$; ET and NM: $t(41) = 0.88$, $p = 1.00$; LT and NM: $t(39) = 0.71$, $p = 1.00$].

Computational modeling results

To understand the mechanism involved in the LT performance advantage, we fit a series of computational models to the data (Worthy, Pang, et al., 2013). The full details of the models are provided in the online Supplementary Material. We fit a total of four reinforcement learning (RL) models: The Prospect Valence Learning Delta and Decay models (PVL2-Delta & PVL2-Decay), Valence-Plus-Perseveration (VPP) model, and the Expectancy-Frequency-Perseveration (EFP) model. These RL models all assume that participants track and update expected values for each action and then make decisions based on a relative comparison of the value of each deck. They differ in the ways that they account for specific processes like recency, loss aversion, frequency of gains versus losses, and perseveration. We also fit a win-stay-lose-shift model (WSLS; (Worthy, Hawthorne, & Otto, 2013). This model fundamentally differs from the RL models because it makes no assumption about the expected value of alternatives, but instead assumes that participants stay with the same option or switch to a different option depending upon whether the outcome on the previous trial was a net gain or loss.

Critically, in the IGT, the WSLS model assumes that participants will switch randomly between the three alternative decks on switch trials, rather than decks with higher expected values. Another critical assumption of the WSLS model is that it relies only on the outcome of the previous trial, and is thus heavily recency-biased. We also fit a Baseline or null model. Readers who are interested in a comprehensive account of each model beyond the basic mechanisms are directed to the online Supplementary Material.

We next used BIC or Schwarz weights to compare the relative fit of each model (Wagenmakers & Farrell, 2004). BIC weights are calculated from BIC to obtain a continuous measure of goodness-of-fit (see online Supplementary Material, section 1.2). These weights can be simply interpreted as the probability that the model is the best model given the data set and the set of candidate models (Wagenmakers & Farrell, 2004). We computed the BIC weights for each model for each participant. The average weights for participants in each group are shown in Table 10. Average weights for the WSLS model were nearly twice as high for participants in the NM and ET groups (both M = 0.50) compared to participants in the LT group (M = 0.27). This suggests that participants in the LT group relied on a simple WSLS strategy much less than participants in the other groups and that they may have incorporated a more complex RL strategy by comparing the expected values of each action.

Table 10. BIC (Schwarz) weights for each model. *Standard deviations are listed in parentheses*.

Model	PVL2-Delta	PVL2-Decay	VPP	EFP	WSLS	Baseline
Non-Musician	0.06 (0.20)	0.37 (0.41)	0.00 (0.00)	0.05 (0.20)	0.50 (0.42)	0.02 (0.06)
Early-Trained Musician	0.06 (0.16)	0.22 (0.32)	0.10 (0.29)	0.09 (0.20)	0.50 (0.42)	0.03 (0.07)
Late-Trained Musician	0.18 (0.26)	0.22 (0.31)	0.14 (0.34)	0.15 (0.26)	0.27 (0.38)	0.04 (0.17)

We also examined the best-fitting parameter estimates for each model across groups (Table 11) to compare specific decision-making sub-processes assumed by the models. For instance, the best-fitting recency or learning rate parameter (Φ) was lower for participants in the LT group than for NM and ET participants for both the PVL2-Delta and EFP models. Because the recency parameter represents how strongly someone weights recent trial outcomes in updating expected values, a lower recency parameter suggests that LT may have valued options based on a longer series of recent outcomes while NM and ET may have based decisions on only the most recent outcomes. Greater reliance on recent actions has been associated with WSLS strategy use in prior work as well (Otto, Taylor, & Markman, 2011; Worthy, Otto, & Maddox, 2012). Enhanced memory of more distant outcomes could account for LT participants' ability to avoid selecting Deck B, which yields very large but infrequent losses, less than NM and ET participants as well as the Super Control group.

Table 11. Average best-fitting parameter estimates for each model and group. Standard deviations are listed in parentheses. **Significant difference between Non-Musician and Late groups at p<.05, †Significant difference between Early and Late groups at p<.05, ‡ Significant difference between Non-Musician and Early Groups at p<.05. All tests were two-tailed.*

	Non-Musician	Early-Trained Musician	Late-Trained Musician
IGT			
PVL2-Delta			
α	0.40 (0.34)	0.42 (0.32)	0.53 (0.35)
λ	3.00 (2.15)	2.16 (2.16)	3.09 (2.01)
Φ	0.56 (0.43)*	0.48 (0.43)†	0.21 (0.32)*†
c	0.24 (0.27)	0.21 (0.32)	0.32 (0.35)
p	25.10 (37.5)*†	-7.00 (36.0)‡	-8.70 (38.40)*
PVL2-Decay			
α	0.30 (0.35)	0.41 (0.28)	0.30 (0.32)
λ	1.34 (1.73)‡	2.93 (2.18)†‡	1.69 (2.00)†
Φ	0.43 (0.25)	0.55 (0.34)	0.58 (0.32)
c	0.56 (0.80)	0.23 (0.40)	0.46 (0.50)
VPP			
α	0.45 (0.29)	0.40 (0.37)	0.42 (0.35)
λ	2.40 (2.43)	1.95 (2.46)	2.68 (2.05)
Φ	0.30 (0.39)	0.32 (0.33)	0.18 (0.27)
c	1.84 (0.98)	2.08 (0.90)	2.51 (1.06)
w	0.53 (0.42)	0.63 (0.39)	0.69 (0.36)
k	0.42 (0.35)	0.52 (0.38)	0.43 (0.34)
ε_{pos}	0.30 (0.53)‡	-0.12 (0.58)‡	0.06 (0.57)
ε_{neg}	0.10 (0.83)‡	-0.41 (0.62)†‡	0.07 (0.78)†
EFP			
λ	1.36 (2.04)*	1.71 (2.19)	2.88 (2.30)*
Φ	0.52 (0.40)*	0.57 (0.44)†	0.28 (0.33)*†
c	1.02 (1.13)	1.12 (1.02)	0.72 (0.52)
p	2.7 (6.40)	2.90 (15.9)	-2.30 (22.6)
ω	0.48 (0.36)	0.48 (0.38)	0.56 (0.41)
WSLS			
$P(stay\|win)$	0.55 (0.20)‡	0.33 (0.24)‡	0.51 (0.32)
$P(shift\|loss)$	0.59 (0.34)‡	0.80 (0.22)†‡	0.57 (0.34)†

DISCUSSION

We examined the extent to which music training confers long-lasting decision-making benefits. Our results suggest that music training impacts decisional processes differently depending on the age at which the musician began his or her training. In particular, we showed that late-trained musicians (who began after age 8) show a performance advantage in the Iowa Gambling Task, a well-studied decision-making task, relative to early-trained musicians (who began playing at or before age 8) and non-musicians. The performance advantage is observed in the last two blocks of the task when the gains and losses associated with each deck have already been learned. Critically, "age of onset" but not "years playing" was a significant predictor of performance in the Iowa Gambling Task in musician groups. One possible explanation for these results is that music training beginning late in childhood capitalizes on the period of significant maturation in the prefrontal cortex, a region of the brain implicated in optimal performance on the IGT (Bechara et al., 1994; Hare et al., 2008; Samanez-Larkin et al., 2014). Per a critical period theory, taxing the neural regions implicated in decision-making during a period of significant growth can lead to long-lasting functional benefits in decision-making ability (Hensch, 2005). Because the prefrontral cortex does not begin a stage of rapid development until late childhood, introducing a novel skill that relies on the prefrontal cortex during this stage of rapid development can have a profound and positive impact on other functions that the area subserves.

Importantly, our results suggest that the age at which novel skill acquisition begins, and not necessarily the ages during which the skills are developed beyond initial training, may be an important metric for characterizing long-lasting impact on the brain regions the skill utilizes. This is because the process of learning a novel procedural skill can be decomposed into at least two stages (Karni et al., 1998; Miyachi, Hikosaka, & Lu,

2002; Yin et al., 2009). The first stage is characterized by rapid skill attainment at the onset of the learning process. Imaging research suggests that the initial fast learning of a procedural skill in the brain can characterized by first a reduction in activation in the primary motor cortex, followed by an increase in the area of activation in the primary motor cortex (Karni et al., 1998). Subcortically, initial fast learning is associated with increased activation in the association regions of the basal ganglia (Miyachi et al., 2002). The second stage is considered to be slower with neural changes occurring within the context of the template set up during the initial, fast stage of learning. The early learning changes to the motor cortex during procedural skill learning are thought to "set up" the motor cortex for a task-specific process and serve as a template for ongoing modification of the slower learning process (Karni et al., 1998; Miyachi et al., 2002).

The literature on the neural mechanisms of learning a procedural skill supports the results of the current study. As would be expected from the literature, we found that the best predictor of performance on the Iowa Gambling Task was when our participants began playing music, which corresponds to the initial (or fast learning) stage of procedural skill acquisition. Although the early-trained musicians played their instruments during late childhood, they were by that point beyond the initial stage of skill acquisition, and therefore may have received less cognitive benefit from taxing the decision-making regions of the brain with music-learning.

The findings of the current study are corroborated by a large literature supporting the notion that beginning music training at an age at when the underlying brain regions are undergoing significant development may grant long-term performance enhancements in other skills those brain regions serve. For instance, research suggests that beginning music training at an early age (before age 7 years) can produce significant structural and behavioral enhancements in perceptual-motor tasks and the associated regions because

major neural development in perceptual-motor regions occur before the age of seven (Bailey, Zatorre, & Penhune, 2014b; Bailey & Penhune, 2012; Steele et al., 2013).

Our computational modeling results suggest that late-trained musicians more successfully compared reward expectations of each option in order to perform well, the process of which has been associated with the medial prefrontal cortex (Samanez-Larkin et al., 2014). In addition, NM and ET participants were more likely to rely on the most recent outcomes and on recency-based WSLS strategies more than LT participants who may have avoided the disadvantageous Deck B more successfully by better retention of the infrequent, but very large, losses (–1,250 points) it provided. It is possible that late music training enhances individuals' ability to incorporate a longer series of events when engaging in complex cognitive processes like decision-making. Interestingly, the results of our linear regression also support this notion, as years of music-playing, was not a significant predictor of performance in the Iowa Gambling Task, but age of onset of playing was.

Our study design is similar to previous research that has examined the advantage of early music training, relative to late music training on brain plasticity (Bailey & Penhune, 2012; Steele et al., 2013). As in previous studies, a limitation is that our data are correlational in nature. We compared non-musicians and two types of musicians in a classic decision-making task, but this method does not allow us to draw a causal inference regarding the effects of age of onset of music-playing and decision-making later in life. Although additional analyses have ruled out a number of obvious alternative explanations, there may be other factors (e.g. personality; Corrigall, Schellenberg, & Misura, 2013) that influence when a person begins playing music that may mediate the intriguing relationship we find in the present study. Future work should seek to further

refine this relationship using a longitudinal methodology to more directly test the effect of beginning music training earlier or later in life on decision-making ability.

The results of the current study have real-world implications: in addition to the rich sensorimotor benefits of early music playing may have on the developing brain, music training may also confer long-lasting benefits in complex cognitive functions like decision-making. To take a real-world example, the music classes offered during many children's elementary and high school education in America (which correspond to the age of music-training onset in the present "late-trained musician" group) may result in improved decision-making ability as an adult. Understanding tools that can serve individuals who show profound decision-making deficits as young as 8 years old could have a dramatic effect on the rate of risky behaviors later in life. Specifically, our results lead to an exciting and testable hypothesis: learning to play music during elementary and high school may lead to dramatic enhancements on late-maturing brain regions that mediate decision-making abilities.

The results of the current study also suggest that music training should be added as a measure of individual differences for future studies in decision-making. The Super Control group data which came from a "many labs" collaboration is composed of young adults with a mean age of 25.6 years. Music experience was not reported for any of the studies included in the Super Control group; therefore, we cannot exclude the possibility that music training was a factor in any results presented. Our results suggest that music-training metrics may be easy and worthwhile measurements to obtain in future decision-making studies. Therefore, we recommend that future decision-making studies include measurements of music training to help elucidate any effects that might be otherwise obscured in a musically heterogeneous sample.

We also cannot completely rule out alternative explanations for our results. For instance, it can be argued that very few aspects about a child's environment throughout childhood are under the child's control. This means that the age at which a child begins playing music, or any other engaging activity, is most likely determined by the parents. It is possible that children who were introduced to music at an early age may have different motivations for pursuing music training than the children who began playing music later in childhood. It is reasonable to assume that parents may have had a larger influence in the decision to play music at an earlier age than later, suggesting that early-trained musicians would have a different motivation to play relative to the late-trained musician. Research also suggests that several factors such as self-regulation, deliberate practice strategies, and self-perception of competence all significantly influence musical achievement. These factors will likely differ in their relationship to musical achievement depending on what age the musician begins practicing (Bonneville-Roussy & Bouffard, 2015). For instance, a younger musician may practice a certain number of hours per week because of parental involvement, which shapes their practice schedule and influences how they self-regulate their practicing later in life. Although the current study was not equipped to measure the existence of such motivational differences, we cannot rule out that a difference in the primary contributor of motivation (i.e., parent or self) may have resulted in the current results.

A limitation of the current study is that our results are based on a group of young adults who had access to music education during childhood and may not generalize to the rest of the population. Factors such as socioeconomic status and race are important factors that can affect access to music education (Palmer, 2011). Because the participants in this study were among a cohort who had access to music education in schools or as an extracurricular activity and were later recruited in a university setting, our results are

limited to such a cohort. Future studies should endeavor to include participants from diverse socioeconomic statuses, and races in order to better understand the ways in which music training may affect the developing brain.

Future directions should also include a test for the existence of an upper age limit for beginning music training by which receiving music training after such an age would no longer provide the cognitive benefits explored in this paper. In our current sample, late-trained musicians began playing music between the ages of 9 and 16 years old." Because the neural regions supporting decision-making continue to develop into adulthood, it is plausible that beginning music training at any point between late childhood and adulthood might confer decision-making benefits (the correlational results displayed in Figure 1(b) suggests this may be the case). However, a study similar in recruitment methodology to Savion-Lemieux, et al. including several age groups binned within a late-trained musician group, and an early-trained control group would be better suited to test such a phenomenon (Savion-Lemieux, Bailey, & Penhune, 2009). Lastly, we encourage researchers to replicate the current study and publish both significant and non-significant findings in order to better understand the impact of music training in late childhood on decision-making abilities.

In conclusion, the goal of the present study was to test the extent to which the music training affects decision-making. Using the Iowa Gambling Task, we found that musicians who began training at age 8 or later showed improved decision-making ability relative to musicians who began training before the age of 8 or non-musicians. Using computational modeling, we found that the late-trained musician performance advantage was a result of using longer strings of trial outcomes to inform their current trial's strategy. Our results have important implications for science-based approaches to incorporating music education in school systems, cognitive development, and the

development of ecologically valid tools to enhance individual deficits in decision-making.

GENERAL DISCUSSION

In a series of experiments, I investigated the extent to which contextual cues and music training may provide support to the brain systems supporting accurate speech-in-noise perception, especially those in which informational masking is present. As we age, our ability understand a conversation amongst a noisy backgrounds decreases, which can lead to societal health concerns such as depression and lowered quality of life (Heine & Browning, 2002). In tandem, advanced age negatively affects the cognitive and perceptual processes responsible for accurate speech-in-noise perception. Importantly, long-term music training has been associated with enhanced cognitive and perceptual processes required for accurate speech-in-noise perception across the lifespan. However, a critical gap in the literature exists in the assessment of music training as a tool for improving speech perception in the aging population. The papers included in this thesis span a wide range of topics in order to provide a well-rounded and theoretically driven approach to advancing our knowledge of a complex and important phenomenon: speech-in-noise perception deficits in the aging brain. What follows is an outline of the three published papers of my thesis, which provide the motivation to conduct a fully randomized control trial directly testing the effects of music training on the brain. Before such a large project can be undertaken, there are several logistical concerns that must be addressed such as "how much training is enough, and in what interval?", "what is an appropriate active control group?", and "which musical instrument is most appropriate for the training?" The proof-of-concept study presented below aims to answer such questions.

In Paper 1, we characterized a speech perception deficit in the aging population (Paper 1; Smayda, Van Engen, Maddox, & Chandrasekaran, 2016). The presence of

either semantic cues (Bradlow, 2008; Pichora-Fuller et al., 1995; Smiljanic & Sladen, 2013; Van Engen et al., 2014) or visual cues (Helfer, 1997, 1998; Sommers et al., 2005; Sumby & Pollack, 1954; Van Engen et al., 2014) have independently been shown to aid in a listener's perception of the conversation. Recently, Van Engen et al. (2016) provided the first test of the integrative effect of semantic and visual cues in younger adult speech-in-noise perception. The current paper extends the literature into the older adult population and tests the extent to which semantic cues and visual cues aid in speech perception in both younger adults (ages 18-35) and older adults (ages 60-90). Participants listened to sentences produced in speech-shaped noise (SSN) by a single speaker across a range of signal-to-noise ratios (SNRs). On any given trial, the sentences were presented in the audio-visual or audio-only modality, and in either an anomalous or meaningful semantic context. Results were broken down into keyword identification accuracy and relative benefit analyses. With respect to the keyword identification results in the audio-only context, older and younger adults perform more similarly in the easier SNRs relative to tougher SNRs across both meaningful and anomalous contexts. In the audio-visual modality, a different pattern of results emerged. We found a significant three-way interaction between SNR, age group, and semantic context such that in only the anomalous semantic context, as the SNR becomes easier, the difference in accuracies between age groups increases with younger adults performing better than older adults. However, the effect was present only in the anomalous condition, suggesting that in the meaningful context, which requires less working-memory resources, the added visual information may bootstrap the older adults' performance.

We next calculated relative benefit scores representing the amount of performance gain received from adding a visual component to the speech stream (relative to audio-only modality context) from adding a meaningful semantic context (relative to an

anomalous semantic context). Results suggest that listeners benefit from visual and semantic information to a consistent degree across the lifespan. In addition we found that for both age groups, visual information provided more benefit in the meaningful semantic context relative to the anomalous context. Reciprocally, semantic information provided a greater benefit for speech perception in audio-visual relative to audio-only contexts. These results replicate the young adult findings of Van Engen et al. (2014), and extend the literature to the aging population. Importantly, the results of Paper 1 suggest that in all listening conditions other than the most supportive (audio-visual and meaningful semantic information present in speech stimulus), older adults show an accuracy deficit in repeating a spoken speech stream relative to younger adults. The findings of this paper provide support for the importance of testing the interactive effect of multiple real-world contextual components such as visual and semantic information in speech perception ability across the lifespan. To our knowledge, this paper is the first to concurrently test the effect of semantic and visual information in the aging population, finding that the benefit received from visual information is dependent on the level of semantic listening support, and the benefit received from a meaningful semantic context is dependent on the level of visual support across the lifespan.

It is worth noting that in Paper 1 when the most supportive listening context in was available to older adults, including both semantic information and visual information, they performed just as well as young adults in identifying what was being said to them. Interestingly though, the accuracy boost found in the audio-visual context with the addition of meaningful semantic cues for older adults was not replicated in the audio-only context with semantic cues. Within the framework of FUEL, we can consider a few reasons why the added semantic information helped more in the audio-visual condition than the audio-only condition. For instance, prior research has suggested that

one benefit of the visual information is allowing the listener to lock onto the onset of the sentence. For instance, there are milliseconds of lip movement that can prime the listener to know that something is about to be spoken so to allocate the appropriate attentional resources to the speaker- a cue unavailable in the audio-only condition regardless of semantic information. In addition, the visual information may facilitate lip-reading throughout the rest of the speech signal beyond the onset and consequently reduce the temporal processing demands of the speech signal – a perceptual process. Additionally, the meaningful semantic context allows for the listener to create more accurate predictions about what is going to be said relative to a semantically anomalous sentence and thus requires less executive functioning – a cognitive process. Therefore, the addition of both semantic and visual contextual cues may reduce the input-related demands on the listener through message factors, since the executive functioning demand on the listener is reduced in meaningful contexts, and context factors because temporal processing perceptual demands are reduced in audio-visual contexts. Contrastingly, in the audio-only semantically meaningful context only the message factor demand (but not context factor demand) is reduced, which alleviates less of the input-related demand on the listener. It is possible that the additive benefit of semantic and visual information suggest different supporting mechanisms (one cognitive, one perceptual) in the listener's use of either cue, and would be an interesting test of the neural mechanisms of FUEL's input-related demands structure.

In Paper 2 (Smayda et al., 2015), we test the cognitive and perceptual benefits of intense long-term music training in a young adult population. Native English-speaking musicians and non-musicians (ages 18-35) learned to categorize the four Mandarin tone in the context of phonemes spoken by a male and female speaker. Mandarin tones convey semantic information and research suggests pitch height and pitch direction (slope of the

tone) are integral features for learning to distinguish between the four tones. On each trial, the participant listened to a tone and made a categorical judgment as to which of the four Mandarin tones they heard using the computer keyboard. Feedback was given after each trial. Results showed a robust musician advantage in learning to correctly classify the four Mandarin tones across multiple talkers, replicating prior work in the field (Alexander et al., 2005; Lee & Hung, 2008; Wong & Perrachione, 2007). We also showed the musician advantage generalizes to new contexts (i.e. a novel speaker). Computational modeling of participant responses suggest that musicians' improved accuracy may be related to their use of both pitch direction and pitch height information earlier and more often during learning relative to non-musicians who tend to rely on only the pitch height information. The differing uses of relevant information to learn novel speech sounds between the two groups represents a difference in decisional processing since the placement of decision boundaries in a perceptual stimulus space is a complex cognitive process (Maddox & Ashby, 1993). The computational modeling of the perceptual variance between groups also suggests that musicians are, on average, more accurate in their perceptual representation of the speech stimuli as well. In addition to its value in replicating prior work, Smayda et al. (2015) also provides a theoretical foundation for implementing music training as a means of improving cognitive and perceptual abilities and the more complex behavior those processes support.

One interesting finding in Paper 2 is that non-musicians who use a multidimensional strategy by block five of training perform just as well as musicians (all of whom used a multidimensional strategy by block five). Furthermore, we found that block five multidimensional users (regardless of music background) had significantly higher working memory scores relative to block five unidimensional users. Although these results suggest that working memory to be a confound (such that only an enhanced

working memory would be necessary to successfully learn to distinguish between Mandarin tones), our perceptual noise results hint that there is an added perceptual advantage supporting the musician advantage. Musicians' perceptual representation of the tones decreases along both pitch height and pitch direction with learning, whereas the non-musicians' perceptual representation of the tones decreases along only the pitch height dimension. Therefore, it seems likely that the musicians' use of multidimensional strategies earlier and more often (conferring an accuracy advantage early in training) relative to non-musicians relates to both cognitive *and* perceptual enhancements. An interesting future test of this may be to recruit musicians and non-musicians matched on working memory ability to learn the four Mandarin tones, and model their use of cognitive decision strategies and perceptual noise in their representation of the tones.

Paper 3 (Smayda et al., 2017) lends additional support to the argument that music training confers cognitive enhancements observed in Paper 2. Smayda et al. (2017) tests the extent to music training confers long-lasting enhancements in decision-making. Prior research suggests that beginning music training during the time when the motor and sensory regions of the brain are developing provides long-lasting enhancements to sensorimotor skills (Bailey et al., 2014; Bailey & Penhune, 2012; Steele, Bailey, Zatorre, & Penhune, 2013) because of the increased engagement of those regions during music training. The neural regions supporting complex and important cognitive abilities like decision-making have a protracted development beginning late in childhood and extending into adulthood. Because learning to play an instrument requires intense engagement of cognitive abilities, we hypothesized that beginning music training during a time of rapid growth in regions (i.e. late childhood) would confer long-lasting enhancements in processes supported by cognition, such as decision-making. Therefore, in Paper 3, young adults (ages 18-35) who had begun music training at or before age 8,

after age 8, and those with no music training performed a gold-standard reward-based decision-making task, the Iowa Gambling Task (IGT). On every trial, participants are chose a card from one of four decks of cards and either received points or lost points with the goal of maximizing the number points achieved. Participants were unaware of the underlying reward structure at the onset of the experiment although two of the decks produce a net gain in points and the other two decks confer a net loss of points. Within each type of deck, one will produce frequent and small changes in points, whereas the other will produce infrequent and large changes in points.

Behavioral results suggest no overall effect of music training on decision-making. Rather, the musician advantage is mediated by the age at which the musician began studying music. Late-trained musicians produced expressed enhanced decision-making abilities as represented by a significantly larger number of points accumulated throughout the decision-making task relative to both early-trained musicians and non-musicians. Computational modeling of each group's deck choices on a trial-by-trial basis revealed that musicians used longer spans of deck-choice outcomes in order to inform the current trial's decision, suggestive of superior implementation of working memory resources in order to maximize points gained relative to early-trained musicians and non-musicians. The results of Smayda et al., 2017 suggest that music training begun later in childhood may confer long-lasting decision-making enhancements.

All three papers included in this thesis provide support for implementing music training as a tool for attenuating the age-related deficit in the aging population. Critically, in order to test a causal role of music training on older adult speech perception in a randomized control trial, logistical questions of such an intervention must first be addressed. In a proof-of-concept study presented next, we attempt to answer logistical questions such as: how much training is enough to confer a speech-in-noise benefit; what

constitutes an appropriate active control group; how frequently should the music lessons should be implemented; and lastly, what features of music training confer which cognitive and perceptual enhancements. A complete report of results from this proof-of-concept study will be published at a later date.

Exploratory Training Study: Music Training for the Enhancement of Speech-In-Noise Perception Ability in Older Adults

Although advancing age typically confers declines in speech-in-noise perception, there is evidence supporting the use of music training to offset such decline across the lifespan from both perceptual and cognitive frameworks (Anderson et al., 2014; Kraus et al., 2014; Parbery-Clark et al., 2011; Slater et al., 2015). In addition, there is a large literature suggesting that productively engaging activities in which the individual learns a new skill that requires cognitive resources is more likely to confer long-lasting cognitive enhancements in the brain (Park et al., 2014; Park et al., 2007). Music training is a perfect example of productive engagement – a critical component for inducing long-term change in the aging brain ("A Consensus on the Brain Training Industry From the Scientific Community," 2014).

Thirteen cognitively normal older adults received thirty hours of group piano lessons over the course of ten weeks. Two doctoral students taught music classes from the Butler School of Music at The University of Texas at Austin. Classes included playing technique, music theory, music history, and critical listening. Speech-in-noise perception ability was assessed immediately preceding the ten weeks of training, and immediately proceeding the ten weeks of training. Changes in speech-in-noise perception in the music training condition were compared to those of an active control group, which completed ten weeks of computerized hearing training (LACE, Listening and

Communication Enhancement), and a no-training control group. Although not reported on in this thesis, a large battery of cognitive, perceptual, and musical skills was administered pre- and post-training and will be written for publication at a later date.

We posit that music training provides a productively engaging experience that will broadly enhance both the cognitive and perceptual processes associated with successful SPIN processing, thereby enhancing SPIN processing across both informational and energetic noise types. No study to date has experimentally investigated music training's effect on SPIN processing in older adults, and we address this shortcoming in the current proof-of-concept study (however see Zendel et al., preprint) . Importantly, results presented below should be interpreted with caution and within the framework of a proof-of-concept study since our preliminary findings do not include a no-training control group and participants were not randomly assigned to their respective experimental conditions.

METHODS

Participants

Older adults, ages 60 - 80 years old were recruited from the Austin, TX community through a variety of methods including community centers and university-affiliated learning centers. Inclusion criteria for this project include 1) having three years or less of formal music lessons at any point in their lifetime, 2) are monolingual and 3) have no known neurological disorders and normal hearing. Due to logistical restraints, participants were first recruited for the music training condition, subsequently for the LACE training condition, and finally for the no-training control group. All participants adhered to the same inclusion criteria, and were screened for cognitive impairment using a battery of neuropsychological measures. Participants also completed a hearing

threshold assessment, which will be used as a variable of interest in our analyses. See group comparisons of key neuropsychological measures and hearing threshold assessments in Table 12. Lastly, participants completed a speech-in-noise perception task in which informational masking is parametrically modulated in the speech signal.

Fifteen participants were recruited for each training condition, however, due to attrition in both the LACE and music training conditions only thirteen participants completed the full training protocol per training condition. Ten people participated in the no-training control group but due to timing, their results will not be included in the current write-up of this study.

Table 12. Means and standard errors for demographic information across groups.

	n	Age (years)	Years of Training	Age Began (years)	Digit Span	Trails B	Vocabulary	Years of Education	PTA
Music Training	13	67.30 (1.12)	0.62 (0.27)	14.40 (3.71)	8.54 (0.40)	85.62 (10.72)	44.69 (1.69)	16.85 (0.65)	23.17 (2.79)
LACE	13	70.46 (1.39)	0.25 (0.23)	10.20 (1.50)	8.90 (0.42)	83.61 (8.08)	44.69 (2.39)	16.92 (0.47)	31.11 (2.53)
No-Training	10	70.40 (1.66)	0.47 (0.24)	11.00 (3.34)	10.19 (0.38)	67.98 (4.85)	49.80 (1.31)	17.44 (0.44)	30.43 (4.36)

Materials

Surveys

Participants completed a set of pre-training surveys characterizing demographics including age, pure tone threshold averages across both ears, and music and language experience (e.g. age of onset of music training, years of formal music lessons), a sample of which is provided in Table 12. At post-training, participants in the music training and LACE conditions also completed a survey assessing motivation to complete their respective training (see Figure 15 for group comparisons). The motivation survey asked participants to respond on a scale from one ("not at all true") to seven ("very true") the degree to which they agree with statements such as "I am motivated to continue learning the skills I developed," "I participated actively throughout the whole activity," "I enjoyed doing this activity very much," "I plan on spending time to be able to continue this activity," and "I plan on spending money to be able to continue this activity." A total motivation score is calculated for each participant in the LACE or music training condition by simply adding up a participant's responses for the 25 survey questions. Group averages of total motivation scores are compared below. Results of the complete set of assessments (both pre- and post-training) including surveys and tasks assessing musical ability (PROMS), physical activity (Yale Physical Activity Scale; Dipietro, Caspersen, Ostfeld, & Ethan, 1993), music and language experience, spatial hearing (Speech, Spatial and Qualities of Hearing Questionnaire), practice logs, and attendance will be published at a later date.

Neuropsychological and Hearing Testing

Participants completed a battery of cognitive testing to ensure that our samples were cognitively healthy. Group comparisons of a subset of tests spanning verbal fluency

(WAIS IV Vocabulary), working memory manipulation (WAIS IV digit span), and switching (Trails A/B) are presented in Table 12. The h test included pure tone threshold averages (PTA) collected at 500 Hz, 1000 Hz, 2000 Hz, 4000 Hz, and 8000 Hz. The full battery of neuropsychological and hearing tests (including bone conduction and QuickSIN) will be published at a later date.

Speech-In-Noise Task

The speech-in-noise task parametrically modulates the amount of informational masking comprised of 64 target sentences mixed with a range of maskers ranging from mostly informational masking to mostly energetic masking: 1-talker babble, 2-talker babble, 8-talker babble, and speech shaped noise (SSN). This task is the same as the one used in Chandrasekaran et al., 2015. Target sentences were recorded at Northwestern University with a sound-attenuated stage. Babble maskers were created by recording eight female students speaking 30 simple meaningful English sentences (Bradlow & Alexander, 2007). The thirty sentences for each speaker were concatenated without any space between the sentences and equalized for RMS amplitude. To produce the 1-talker noise, one talker's string of sentences was used. To create the 2-talker noise, another talker's string of sentences were mixed with the first string of sentences. To create the 8-talker noise, six more talkers were added to the 2-talker babble in Audacity (Audacity Developer Team, 2008). The speech-shaped noise was generated using white noise that was filtered so its spectrum matched the average spectrum of the full set of recorded sentences. Each target sentence was mixed with a random sample of noise, and included 500 ms of noise prior to the onset of the target sentence.

Procedure

Speech-in-noise perception ability was assessed using a task created in-laboratory using E-prime 2.0 on a PC in a quiet testing room. Participants viewed and listened to stimuli at a comfortable listening level through Sennheiser HD 280 Pro headphones. In addition to visual instructions presented at the beginning of the task, participants were also instructed by a researcher that on every trial, they would hear a target sentence preceded by 500 ms of noise, and that they were to type the target sentence using the computer's keyboard. During audio-only trials of both speech-in-noise tasks, a black fixation cross was presented on the screen. During audio-visual trials, a video clip of the speaker covered the entire computer screen. Each trial was participant-initiated, and responses were scored as the number of correct keywords correctly identified on each trial. Spelling mistakes will be permitted except when it creates a different English word than the target keyword.

Training Procedure

Participants in the music training condition received ten weeks of free group piano lessons taught by a Doctor of Musical Arts and Doctor of Musical Arts candidate from the Butler School of Music. Classes consisted of 14 adults ages 60-80, two teachers, an occasional guest, and myself. Two participants dropped out of the music condition (one before classes started), and their data are not included in the following analyses. Classes were structured according to a very successful framework and syllabus developed by the two music teachers in collaboration with mentors from the music school. A report including course materials and pedagogical background are published in detail as a conference proceedings in a separate publication (Chou, Straub, & Smayda, *in press*). The authors hope the proceedings will be used as a model for both future music training

interventions and general piano classes for the mature student. Participants received piano books and personal piano keyboards to take home for the duration of the project.

A total of 13 participants in the active control group received a subscription for an online hearing training program called LACE (Listening and Communication Enhancement; Sweetow & Sabes, 2006). Fourteen participants were recruited but two withdrew from the program. Participants were instructed to complete 30 hours of LACE training and received a pair of headphones, and a protocol for the training program. The primary objectives of LACE, as stated by Sweetow and Sabes, are to:

1) Enhance listening and communication skills.

2) Get the patient involved.

3) Improve the confidence levels.

4) Provide communication strategies.

5) Reduce unnecessary visits (to the doctors).

6) Help the patient recognize that hearing aids address hearing but do not correct listening and communication skills.

The LACE training is split into sections spanning various conditions under which the participant must report how many keywords they heard in a variety of difficult listening contexts such as in the presence of a competing talker, and rapid speech. LACE also incorporates applications of working memory training. For instance for the "Working Memory Training" module, participants heard a sentence before or after which they were told to attend to the word immediately preceding or proceeding a target word. For instance, on an example trial participants might have heard the sentence, "Walking every day can increase activity and promote a happier lifestyle," and were asked to respond with the word that directly precedes the target word "activity." This requires the participant to recall hearing not only the target word, but also the neighboring words in

order to be successful. Importantly, during most portions of the LACE training, no feedback of their accuracy is given to the participant. Optional homework was provided in the form of a curated list of educational podcasts and related comprehension questions. Podcasts offered for each week averaged one hour in duration, and served to provide additional opportunities to practice listening, mirroring the optional homework offered in the music training condition. Podcast topics loosely followed the topics used as listening exercises in the LACE training program.

Participants in the no-training control group were tested on all the same tasks as the two training conditions, and after ten week, participants returned to the laboratory for post-test measurements. In the interim, participants did not complete any training and were instructed to go about their typical lifestyle.

Analyses

To test the effect of music training on SPIN perception performance, keyword identification accuracy for each participant's responses to target sentences will be computed. Because accuracy for each keyword will be coded as 0 ("incorrect") or 1 ("correct"), we will use the *glmer* function in R, which fits a linear mixed effects logistic model to binomial data. Our model will also include Time (pre-training or post-training), Noise (1-Talker, 2-Talker, 8-Talker, SSN) which parametrically modulates the amount of information masking present in the signal, Condition (Music-Training, LACE, No-Training), and by-sentence and by-subject random intercepts. PTA will be used as a covariate, reference levels for each variable are set to "energetic" for noise type, "control" for training condition, and "pre-training" for time. Therefore, our model will be:

$$Response \sim PTA + Time \times Noise \times Condition + (1|Subject) + (1|Sentence)$$

Predictions

Due to the multimodal and rich experience that music training provides, we hypothesize that participants who receive music training will show enhancements in informational masking SPIN conditions (1-Talker, 2-Talker) to a greater extent than energetic masking conditions (8-Talker, Speech-Shaped Noise) because it relies on both perceptual *and* cognitive abilities, which we expect will both be trained during the music-learning process. We expect any improvements in SPIN perception in the LACE condition to be significantly less than the observed improvements in the music-training group. The age-related decline in the peripheral auditory system is less modifiable. Additionally, based on pilot data presented below, and because music is inherently rewarding (Zald & Zatorre, 2011), we predict that older adults who receive music training will be more motivated in their training and likely to continue their training upon completion of the study, making music training a sustainable and viable method for improving cognition, perception, and SPIN processing.

RESULTS

Presently, we only have results from 13 in the music training condition, 9 in the LACE training condition, and 10 in the no-training control group. Results suggest a significant three-way interaction between time, training condition, and noise type for both LACE training and music training conditions, suggesting these conditions conferred significantly greater benefit in the informational masking conditions relative to the no-training control group [LACE: $\beta = 0.71$, $p < 0.008$; music training: $\beta = 0.78$, $p < 0.002$]. In addition, a t-test between total motivation scores (max score = 175) of participants in the music training condition and LACE condition suggest that music participants were significantly more motivated to complete their training than LACE participants [$t(19) = 3.79$, $p < 0.005$]. Please note that these results include data from all

participants included in the logistic regression above except for one LACE participant because data collection is not complete as of now.

Preliminary results of the current proof-of-concept study suggest that group piano lessons may improve older adults' ability to understand conversations in difficult listening environments. Compared to a no-training control group, participants who completed ten weeks of group piano lessons and participants who completed ten weeks of computerized hearing training showed enhanced speech-in-noise perception, especially in informational masking conditions at the end of training (see Figure 14). Importantly, the participants in the music training condition also showed increased motivation to complete the intense training relative to participants who were engaged in the LACE active comparison training (See Figure 15). Complete group comparisons of motivation will be reported in the full report publication.

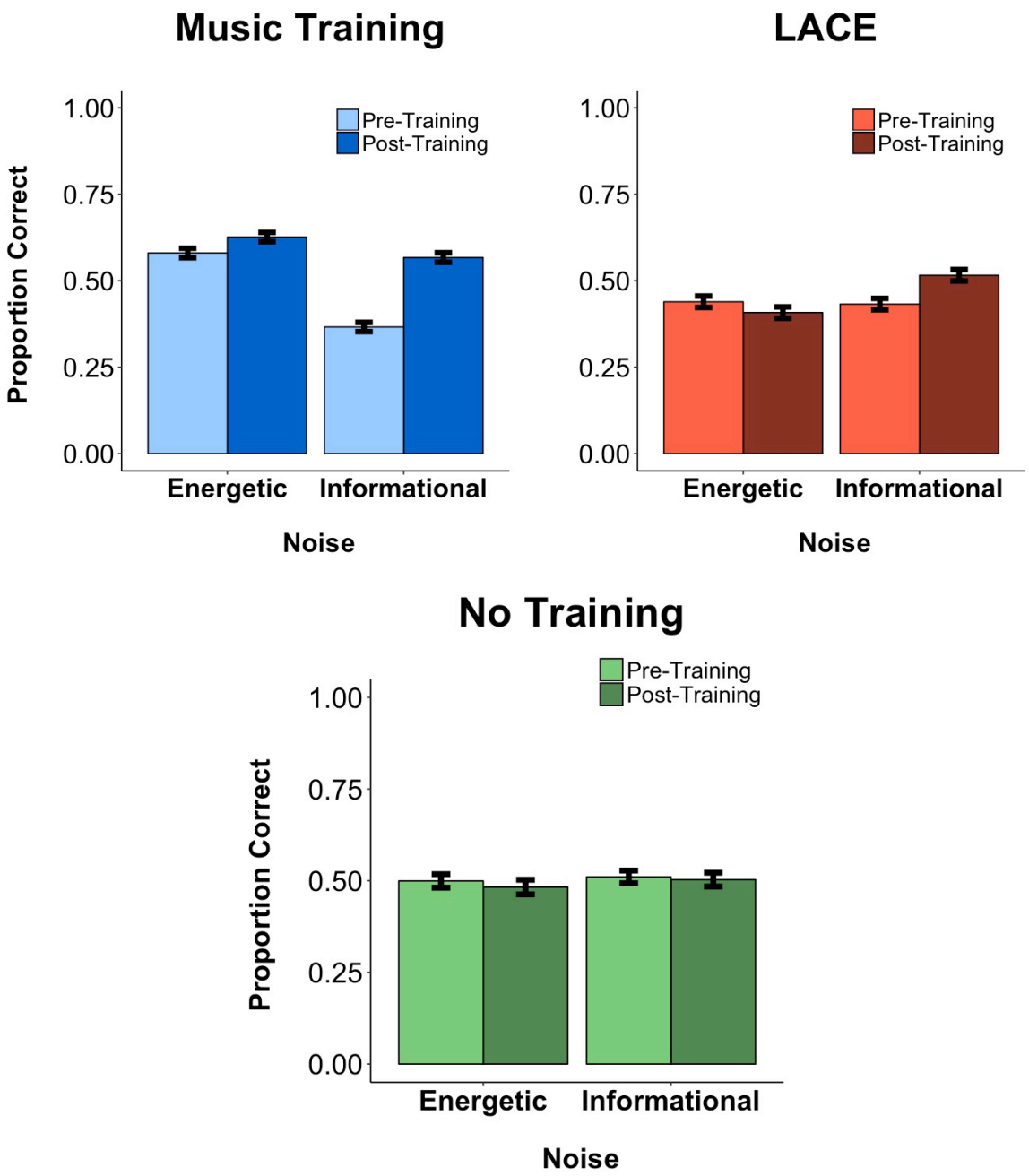


Figure 14. Average speech-in-noise accuracy by training condition (music training, LACE , and No Training), noise (Energetic, Informational) and testing time (Pre-training, Post-training). Bars represent standard error of the mean.

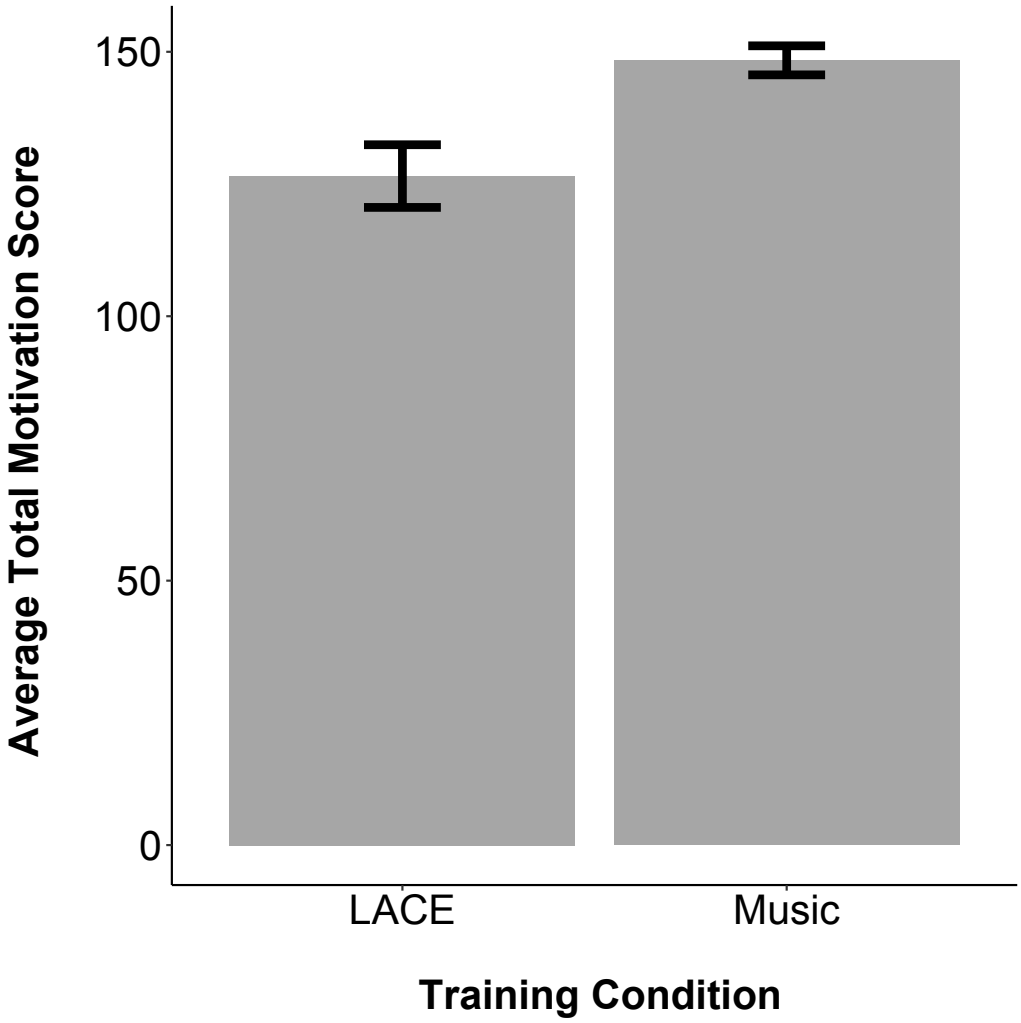


Figure 15. Average overall motivation scores for respective group's training activities.

DISCUSSION

Although there has been much discussion regarding the therapeutic uses of music in combatting age-related changes in cognition and perception, there have been few experimental tests of such an effect. Therefore, the goal of the current exploratory study was to fill the critical gap in the literature and provide a proof of concept for using music training to improve older adult speech-in-noise perception. To that end, thirteen non-musical, cognitively health adults ages 60 - 80 learned to play piano for three hours a week for ten weeks. Music-related improvements were compared with an active training comparison group and a no-training control group (the latter of which is not included in the current report). In the first comparison group, thirteen age-matched non-musical adults completed thirty hours of a computerized hearing training targeting the specific outcome we were interested in improving with music training. In the second comparison group, ten non-musical participants were tested at baseline, and were again tested ten weeks later. These participants were instructed to not change anything about their normal daily routine.

Results of the proof-of-concept music training intervention suggest that music training may be a viable method by which to improve older adults' speech-in-noise ability, especially in informational masking conditions. Not only did participants in the music training condition show a robust improvement in overall accuracy in a speech-in-noise task, but they were also more motivated to complete the training. Motivation is an important factor to consider when testing an intervention with the aging population, because in order to sustain long lasting benefit from the activity, they must regularly engage in the activity.

The results of this proof-of-concept study provide critical information for designing a randomized control trial to directly test the effect of music training on older

adult speech perception. First, our preliminary results suggest that thirty hours of music lessons is an acceptable amount of time to improve older adult speech-in-noise perception ability. However, to better understand the rate of improvement, a mid-test of speech perception ability should be implemented in future studies. We initially scheduled the piano classes to meet three times a week for an hour each session, but based on participant feedback, we realized that two sessions a week for 1.5 hours each session is much more amenable to the schedules and driving concerns of the participants.

Second, our results suggest LACE training is an appropriate training for the active training control group. Our preliminary results suggest LACE training moderately improves speech-in-noise perception. There were several reasons we chose to use LACE. First, as our world becomes more and more driven by technology and computer-based activities, it was important to include a test of a time-relevant methodology, although the inherently minimalized social component of such a methodology should be critically inspected. Research suggests that older adults thrive on social engagement and is fundamentally important to improving communication in an older population. Second, it is important to consider the participant's experience. We wanted to offer both training groups the opportunity to improve their hearing, so as to recruit groups of participants equally motivated for the intended outcome of the training. Therefore, LACE was chosen because of it is a time-relevant method of training in this technology-driven era, and because of the research supporting its effectiveness in improving speech-in-noise perception (Sweetow & Sabes, 2006).

There are several important limitations of the current study worth discussing. First, participant recruitment occurred serially for each condition, which may bias the people who volunteered for each condition. Second, the participants who would be motivated and engaged to learn to play piano on a university campus may have different

mobility abilities than those who would sign up for an intervention performed at home. Participation in the LACE condition would require fewer travel obligations to a congested, high-traffic part of Austin. This concern may contribute to an unanticipated difference in sample demographics between training groups. A third limitation of the current study is that condition assignment was neither single-nor double-blind. Each participant knew what condition they would be volunteering for before baseline testing occurred, and the research team also knew the participant's training condition assignments. Fourth, it is important to note that the amount of social components of each condition varied widely. The music training condition met as a group multiple times a week for ten weeks at a top-tier music institution. During their time in the study, participants in the music condition made friends, shared frustrations, learned new things, and built a bond with one another that has extended beyond the limited duration of their involvement in our study. The LACE training group did include multiple friend/partner pairs, however, participants worked on their training independently. The opportunity for social engagement in the LACE condition was limited to only the interactions the participants had with the excited and motivated researcher team coordinating their involvement (which is not a negligible amount).

Lastly, although the present study seeks to test the role of music learning as a general methodology of improving speech-in-noise perception, there may be more optimal control conditions. For instance, it would be relevant to understand the specific and unique role that social engagement plays in the relationship between music learning and speech perception in the aging population above and beyond any other active training group including a social component (such as a drama or painting course). In addition, it is imperative we as a field understand the unique contribution of the motoric and non-motoric components of learning to play an instrument on speech perception. To do this,

one might consider designing a double-blind intervention study in which mature adults take a course in learning to play an instrument using a teacher/classroom model, and a course wthat trains critical listening, interpretation, theory, and historical relevance of music (all of which would be inherent in the first course option, as well as in the current study's music training condition) but void of learning how to actually play an instrument, or a no-training waitlist group.

The results of the current proof-of-concept study corroborate a recent experimental test of music training's effect on speech-in-noise perception in the older adult population (Zendel, West, Belleville, and Peretz, biorxiv). Published online before peer-review, the study design implements a single-blind method to assign participants to receive either six months of computerized music training, visuo-spatially demanding video game training, or participate in the no-training condition in which they did not complete any training. Participants completed both behavioral tests and EEG recordings at pre-training, mid-training, and post-training. Participants were instructed to complete at least 30 minutes of training for at least five days a week for six months in contrast to our ten weeks of 3 hours a week (plus additional personal practice time). Behavioral results suggest that six months of music training improved speech-in-noise perception to a greater extent than both an active control group that received training on a visuo-spatially demanding video game, and a no-contact control group. Interestingly, there was no "social" component to the music training condition, which lends support to our finding that music training improves speech perception above and beyond the benefit the older adults in the current study may have received from the social aspects of our music training.

Experimental tests of the impact of music training on speech perception have also been applied in the developing brain. Slater et al. (2015) found that after two years of

group music lessons, speech-in-noise perception significantly improved in a cohort of elementary-aged children relative to children placed on a waitlist. Across multiple papers, results included neural and behavioral support for improved speech-in-noise perception ability in children who learned to play music relative to a waitlist, no-training control group.

In conclusion, the preliminary results of our proof-of-concept exploratory study suggest that ten weeks of group piano lessons can improve the ability of older adults to hear in noisy environments. The complete report of these findings, including cognitive and perceptual subtests will be published at a later date.

Implications and Conclusions

Our nation's growing population of adults over the age of 50 as a result of the baby boomer generation will not revert back to the level it was before the baby-boomer generation. Recent estimates suggest that the number of Americans aged 65 and older will more than double by 2060 (Mather, Jacobsen, & Pollard, 2015), meaning that hearing ability in the aging population will only become a greater national concern from here forward. Therefore, it is imperative that academic and industry enterprises collaborate, test, and commercialize engaging, cost-effective methods for reducing the effect of such a widespread health issue encumbered by the nation's healthcare system. The three papers and proof-of-concept exploratory study suggest that group music classes may be a way to address this growing health concern.

A recent framework developed by Pichora-Fuller and colleagues presents an approach for understanding the ways in which music training might improve speech perception in the aging brain. FUEL, or the Framework for Understanding Effortful Listening, expands upon Kahneman's 1973 Capacity Model for Attention that originally

included (and is maintained in FUEL) an input of cognitive capacity, and other unspecified "miscellaneous determinants", and an evaluation of demands on capacity. FUEL updates the 1973 model with components that measure high and low arousal on attention (and therefore performance). A specified description of what Kahneman termed "miscellaneous determinants" is also included in FUEL which includes source factors (such as accented speech), transmission factors (noise, reverberation), listener factors (such as sensory/cognitive abilities), message factors (i.e. semantic context), and context factors (such as visual information and familiarity of situation). More specifically, FUEL proposes that speech comprehension relies on an interaction between task demands, listener motivation, and listener effort. As motivation increases (engagement in a fascinating conversation, even in the presence of a some background noise), effort will also increase which can help direct attentional and cognitive resources towards the conversation. However, as demand increases (perhaps the noise remains, but the conversation topic ends and the listener is therefore left without a semantic context), motivation might remain high but effort sharply decreases. In such a framework, greater cognitive resources or more efficient perceptual processing might be expected to modify the demand of a given listening task.

The FUEL model is directly applicable to the current thesis. Paper 1 tests three of these input-related demands, specifically the interactions of transmission factors (signal-to-noise ratio), message factors (meaningful/anomalous semantic contexts), and context factors (audio-only/audio-visual modalities). Papers 2 and 3 of this thesis test listener factors such as music training experience, cognitive abilities, and perceptual abilities. See Figure 16 for the FUEL model.

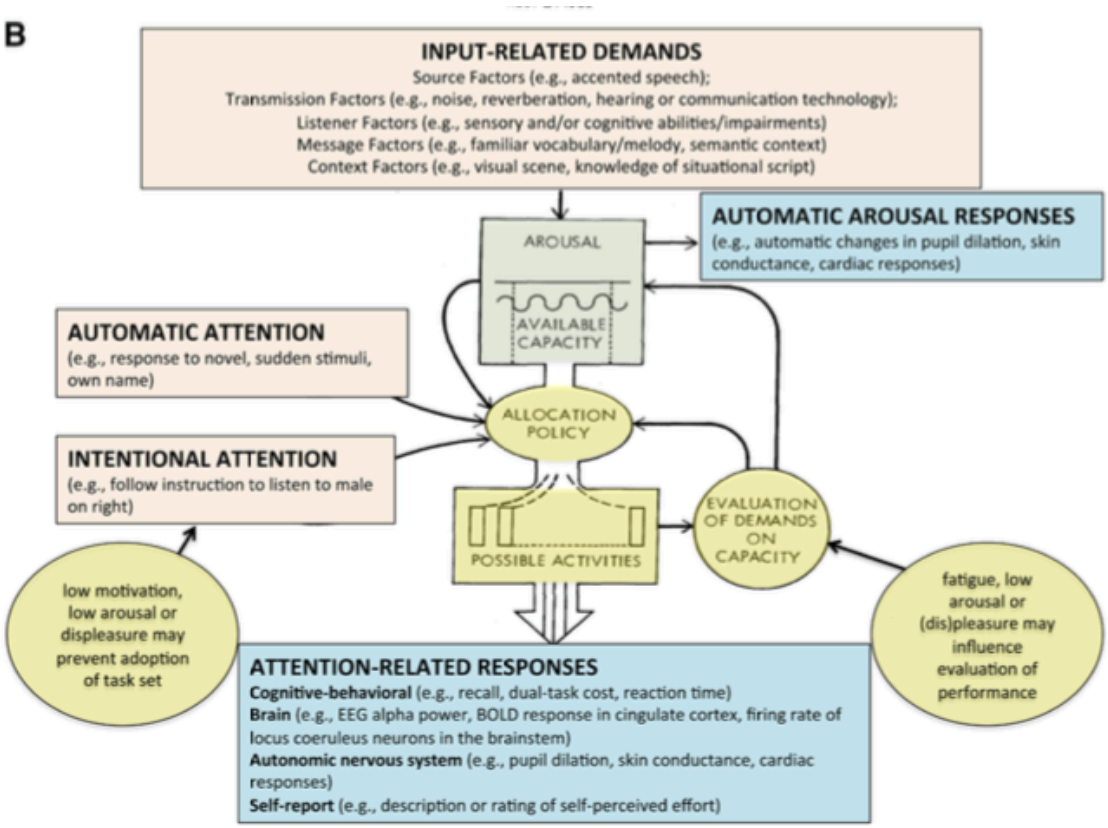


Figure 16. Model proposed within the Framework for Understanding Effortful Listening (FUEL).

In conclusion, the present thesis sought to provide a theoretical and practical framework for testing the effectiveness of music training as a viable intervention for improving older adult hearing in difficult listening environments. Paper 1 tested the integrative effect of contextual cues such as semantic context and visual information to show that older adults use these cues to the same extent as young adults, however, still show a performance accuracy deficit in any context condition other than the most supportive (auditory-visual information plus meaningful semantic information). Paper 2 provides a basis for using music training to improve cognitive and perceptual ability in the brain as computational modeling suggests musicians employ cognitive strategies and more perceptual representations of sound as related to learning non-native speech sounds. Paper 3 extends the musician-related cognitive enhancements introduced in Paper 2 by testing the age at which music training began and its long-term effect on decision-making. Lastly, the exploratory training study provides a proof-of-concept for utilizing music learning as an ecologically valid training tool for improving speech perception in the aging population. The four papers presented in this thesis should be used in service of developing our understanding of the cognitive and perceptual components of speech perception in the older adult brain, as well as a model for future music training intervention studies.

APPENDIX

Paper 3 Supplementary Material

1. Computational Modeling

1.1 Model Descriptions

In addition to our behavioral analysis we also fit a series of computational models to isolate and identify specific psychological components underpinning choice behavior that may have differed between groups. Computational models of decision-making processes provide insights into the mechanisms underlying the effects of WM load. In the current work we fit a series of models that have recently been shown to provide the best fit to IGT data. As stated in the main text we fit a total of four reinforcement learning models (RL), a win-stay-lose-shift model (WSLS; Worthy, Hawthorne, et al., 2013), and a baseline or null model.

A common assumption to all of the RL models is that people make decisions by comparing the expected values, or payoffs for each option. All of the RL models also account for the tendency to perseverate, which has been shown to be a critical component of decision-making beyond expected value (Worthy, Pang, et al., 2013). Aside from these commonalities the models differ in the ways in which they assume that expected values for each option are tracked and updated over time. The RL models we fit were the Prospect Valence 2 Delta and Decay (PVL2-Delta and PVL2 Decay) models (Dai, Kerestes, Upton, Busemeyer, & Stout, 2015), the Valence-Plus-Perseveration (VPP) model (Worthy, Pang, et al., 2013), and a newly developed model called the Expectancy-Frequency-Perseveration (EFP) model that explicitly accounts for the frequency of net

gains versus losses outcomes.

1.1.1 PVL2 Delta and Decay Models

The PVL2 Delta and Decay models differ in their value updating rules, which determine how the utility $u(t)$ is used to update expected values or expectancies $E_j(t)$ for the selected option, *i,* on trial *t*. The Delta rule used in the PVL2-Delta model assumes that expected values are recency-weighted averages of the rewards received for each option:

$$E_i(t) = E_i(t-1) + \phi \cdot \delta_j(t) \cdot [u(t) - E_i(t-1)] \quad (1)$$

Here $\delta_j(t)$ is a dummy variable equal to 1 for the chosen option and 0 for all other options. The recency parameter ϕ ($0 \leq \phi \leq 1$) defines the weight given to recent outcomes in updating expected values. Higher values of ϕ denote a greater weight to recent outcomes, while lower values of ϕ indicate that a greater weight to past outcomes.

The utility function in the PVL2-Delta model is derived from Prospect Theory (Ahn, Busemeyer, Wagenmakers, & Stout, 2008; Kahneman & Tversky, 1979) in which the evaluation of the outcome on each trial has diminishing sensitivity to increases in magnitude and different sensitivity to gains than to losses. The utility, $u(t)$, on trial t, for both PVL2 models is given by:

$$u(t) = [win(t)]^{\alpha} - \gamma|loss(t)|^{\alpha} \quad (2)$$

Here *win(t)* and *loss(t)* represent the number of points gained and lost on each trial. The shape parameter α ($0 < \alpha < 1$) determines the shape of the utility function, and λ represents the loss aversion parameter ($0 < \lambda < 5$) that governs loss sensitivity compared to gain sensitivity. A value of λ greater than 1 indicates that an individual is more

sensitive to losses than gains. Similarly, a λ value less than 1 signifies enhanced sensitivity to gains as compared to losses.

The Decay rule used in the PVL2-Decay model assumes that expected values decay across trials, and that the updated expected value for each option is the sum of the decayed expectancy and the utility on the current trial. Specifically it is:

$$E_i(t) = \phi \cdot E_i(t-1) + \delta_j(t) \cdot u(t) \quad (3)$$

As in Equation 1 $\delta_j(t)$ is a dummy variable and ϕ is a learning rate parameter. The key difference between the PVL2-Delta and PVL2-Decay models is that in the Decay model expectancies are assumed to decrease. This allows the model to indirectly account for tendencies to perseverate, by picking the same option on successive trials because the selected option's expected value will become increasingly larger than non-selected alternatives (Worthy, Pang, et al., 2013).

The PVL2 models also include an action-selection rule that controls the predicted probability that deck *j* will be chosen on trial *t,* Pr[$G_j(t)$] and is calculated using a Softmax rule (Sutton & Barto, 1998). The PVL2-Delta model also includes a parameter *p* in Equation 4 which controls for individual differences in tendencies toward perseveration ($p > 0$) or switching ($p < 0$) on each trial (Daw, Gershman, Seymour, Dayan, & Dolan, 2011; Lau & Glimcher, 2005):

$$Pr(G_i(t)) = \frac{e^{(\theta(t)[E_i(t)+p\cdot rep(i)])}}{\sum_{j=1}^{4} e^{(\theta(t)[E_j(t)+p\cdot rep(j)])}} \quad (4)$$

Here *rep(i)* is equal to 1 if the same action was chosen on the previous trial and zero otherwise. We allowed *p* to vary from -100 to 100, effectively adding or subtracting 100 points from the net value of the action that was selected on the previous trial. The

addition of the *p* parameter follows from previous results where we showed that perseveration is a critical component to account for in models of the IGT (Worthy et al., 2013). The PVL2-Decay model's action selection rule was identical to that given in Equation 4 except that *p* was set to zero as the model's updating rule (Equation 3) already accounts for perseverative tendencies. In both models the trial-independent action-selection rule governs expected values and is represented as:

$$\theta(t) = 3^c - 1 \tag{5}$$

where c ($0 \leq c \leq 5$) is the choice consistency or exploitation parameter. Larger values of c indicate that an individual has a greater tendency to choose options with higher expected values. Similarly, smaller c values indicate a greater tendency explore options with lower expected values.

1.1.2 VPP Model

The VPP model utilizes Equations 3 and 4 to update the expected values for each option. This model included perseverative tendencies as a second term, the perseveration ($P_j(t)$) strengths for each *j* option, which were determined by a more general form of the Decay rule that has been used to model perseveration or autocorrelation among choices in recent work (Kovach et al., 2012; Schönberg, Daw, Joel, & O'Doherty, 2007). The perseveration term for chosen option *i*, on trial *t,* differs based on whether the net outcome, *x(t)*, is positive or negative:

$$P_i(t) = \begin{cases} k \cdot P_i(t-1) + \varepsilon_{pos} & \text{if } x(t) \geq 0 \\ k \cdot P_i(t-1) + \varepsilon_{neg} & \text{if } x(t) < 0 \end{cases} \tag{6}$$

Here k ($0 \leq k \leq 1$) is a decay parameter similar to ϕ in Equation 4 above. The tendency to perseverate or switch is incremented each time an option is chosen by ε_{pos} and ε_{neg} which we allowed to vary between -1 and 1. Positive values indicate a tendency to persevere by picking the same option on succeeding trials, while negative values indicate a tendency to switch.

The overall value of each option is determined by taking a weighted average of the two terms in the model, the expected value and the perseveration strength of each *j* option:

$$V_j(t) = w_{E_j} \cdot E_j(t) + (1 - w_{E_j}) \cdot P_j(t) \tag{7}$$

Where w_{E_j}($0 \leq w_{E_j} \leq 1$) is the weight given to the expected value for each option. Values greater than .5 indicate greater weight based on the expected value of each option, and values less than .5 indicate greater weight based on the perseverative strength of each option.

These values $V_j(t)$ were entered into a Softmax rule to determine the probability of selecting each option, *j,* on each trial, *t*:

$$Pr\big(G_j(t)\big) = \frac{e^{[\theta(t) \cdot V_j(t)]}}{\sum_{j=1}^{4} e^{[\theta(t) \cdot V_j(t)]}} \tag{8}$$

where $\theta(t)$ was determined based on Equation 5 above.

1.1.3 EFP model

The EFP model includes three terms to account for three critical components of choice behavior: expected value, gain-loss frequency, and perseveration. Increasing the number of terms may ostensibly improve the fit of the model or lead to overfitting simply because the model has too many parameters. Considering this, we sought to design a

model to capture these three important psychological components while keeping it as parsimonious as possible.

The first model assumption is that after a choice is made and feedback ($win(t)$ and $loss(t)$) is presented, the utility $u(t)$ for the choice made on trial *t* is given by:

$$u(t) = win(t) - \gamma \cdot |loss(t)| \tag{9}$$

here γ represents a loss aversion parameter ($0 \leq \gamma \leq 5$) that governs the sensitivity of losses compared to gains. A value of γ greater than 1 indicates that an individual is more sensitive to losses than gains, and a value less than 1 indicates greater sensitivity to gains than to losses. Note that the EFP model assumes that the subjective utility is linearly proportional to the actual payoff amount, in contrast to the PVL models that use a nonlinear function. One major reason for the nonlinear function in the PVL models is to implicitly account for the gain-loss frequency (Ahn, Busemeyer, Wagenmakers, & Stout, 2008). The EFP model, however, explicitly captures the gain-loss frequency and thus a shape parameter is not necessary. The EFP model then assumes that the utility $u(t)$ is used to update expected values or expectancies $E_j(t)$ for the chosen option, *i,* on trial *t*. It utilizes the Delta rule from Equation 1.

The perseveration term in the VPP model presented above was designed to model the tendency to perseverate following gains and to switch following losses. Thus, it also implicitly captures the frequency of gains and losses. In the EFP model we decompose the tendency to select the option with infrequent losses and frequent gains and the tendency to perseverate. The frequency term for chosen option *i*, on trial *t*, differed based on whether the net outcome, *x(t)*, was positive or negative:

$$F_i(t) = \begin{cases} (1-\phi)\cdot F_i(t-1)+1 & \text{if } x(t)\geq 0 \\ (1-\phi)\cdot F_i(t-1)-1 & \text{if } x(t)<0 \end{cases} \quad (10)$$

The frequency value increases by 1 following a net gain or decreases by 1 following a net loss. Instead of using a separate parameter to capture the weight to previous information as in the VPP model, the EFP model utilizes the term: $1-\phi$. Here ϕ is the same as in Equation 1, accounting for weight given to recent information. Thus utilizing the same recency parameter for both the value updating function and the gain-loss frequency function increases the parsimony of the EFP model and restricts the model to assume that attention to recent outcomes is the same for both value and frequency information.

The perseveration term for chosen option *i*, on trial *t*, is determined by:

$$P_i(t) = \gamma \quad (11)$$

The tendency to perseverate or switch is denoted by γ which varies between -100 and 100. Thus the perseveration term here is the same as the one utilized in the PVL2-Delta model. Note that in the VPP model, tendencies to stay or switch were conflated with attention to the frequency of net gains versus losses, while here our goal is to account for frequency and perseveration processes separately.

The overall value of each option was determined by taking a weighted average of the expected value and the frequency value plus the perseveration strength of each *j* option:

$$V_j(t) = \omega \cdot E_j(t) + (1-\omega)\cdot F_j(t) + P_i(t) \quad (12)$$

where ω ($0 \leq \omega \leq 1$) quantifies the weight given to the expected value for each option versus the weight given to the frequency of losses versus gains provided by each option.

Finally, these overall values $V_j(t)$ were entered into a Softmax rule identical to Equation 8 above.

1.1.4 WSLS Model

The WSLS has two free parameters and is identical to the model used in prior work from our lab (Worthy, Hawthorne, et al., 2013). The first parameter represents the probability of staying with the same option on the next trial if the net gain received on the current trial is equal to or greater than zero:

$$P\left(G_j(t)\middle|choice_{t-1} = G_j \,\&\, r(t-1) \geq 0\right) = P(stay|win) \tag{13}$$

In Equation, 13 *r* represents the net payoff received on a given trial where any loss is subtracted from the gain received. The probability of switching to another option following a win trial is 1-$P(stay|win)$. To determine a probability of selecting each of the other three options we divide this probability by three, so that the probabilities for selecting each of the four options sum to one.

The second parameter represents the probability of shifting to the other option on the next trial if the reward received on the current trial is less than zero:

$$P\left(G_j,(t)\middle|choice_{t-1} = G_j \,\&\, r(t-1) < 0\right) = P(shift|loss) \tag{14}$$

This probability is divided by three and assigned to each of the other three options. The probability of staying with an option following a "loss" is 1-$P(shift|loss)$. This is a critical difference between the WSLS and RL models in that the WSLS model makes no assumption about which of the other three options should be selected on switch trials,

while the RL models assume that switch decisions are governed by the learned expected values for each option.

1.1.5 Baseline Model

Finally, the Baseline model assumes equal fixed choice probabilities for each option (Gureckis & Love, 2009; Worthy & Maddox, 2012; Yechiam, Busemeyer, Stout, & Bechara, 2005). The Baseline model has three free parameters that signify the probability of selecting Deck A, B, or C. The probability of choosing Deck D is the sum of the probability of Deck A, B, and C subtracted from 1.

1.2 Model Comparisons

For the four RL models, the PVL2-Delta model had five free parameters, the PVL2-Decay model had four free parameters, the VPP model had eight free parameters, and the EFP model had five free parameters. The WSLS model had two free parameters, and the Baseline model had three. We fit each participant's data by maximizing the log-likelihood for each model's prediction on each trial. We used Bayesian Information Criterion (Schwarz, 1978) to examine the relative fit of the model. BIC penalizes models with more free parameters. For each model, i, BIC_i is defined as:

$$\mathrm{BIC}_i = -2\log \mathrm{L}_i + \mathrm{V}_i \log(\mathrm{n}) \tag{15}$$

where L_i is the maximum likelihood for model i, V_i is the number of free parameters in the model, and n is the number of trials. Smaller BIC values indicate a better fit to the data.

We used BIC or Schwarz weights to compare the relative fit of each model (Wagenmakers & Farrell, 2004). BIC weights are calculated from BIC to obtain a

continuous measure of goodness-of-fit. A difference score is computed by subtracting the BIC of the best fitting model for each data set from the BIC of each model for the same data set:

$$\Delta_i(BIC) = BIC_i - minBIC \tag{16}$$

From the differences in BIC we then computed the relative likelihood, L, of each model, i, with the transform:

$$L(M_i|data) \propto \exp\left\{-\frac{1}{2}\Delta_{\mathrm{i}}(\mathrm{BIC})\right\} \tag{17}$$

Finally, the relative model likelihoods are normalized by dividing the likelihood for each model by the sum of the likelihoods for all models. This yields Akaike weights:

$$w_i(BIC) = \frac{\exp\left\{-\frac{1}{2}\Delta_{\mathrm{i}}(\mathrm{BIC})\right\}}{\exp\left\{-\frac{1}{2}\Delta_{\mathrm{k}}(\mathrm{BIC})\right\}} \tag{18}$$

These weights can be interpreted as the probability that the model is the best model given the data set and the set of candidate models (Wagenmakers & Farrell, 2004). We computed the BIC weights for each model for each participant. The average BIC values for each model and participant group can be found in Table 13**.**

Table 13. Average BIC values for each model

Model	**PVL2-Delta**	**PVL2-Decay**	**VPP**	**EFP**	**WSLS**	**Baseline**
Non-Musician	232.9 (69.7)	225.0 (67.2)	238.6 (67.1)	232.5 (68.9)	226.8 (71.9)	264.1 (56.6)
Early-Trained Musician	261.7 (44.4)	260.1 (46.1)	263.7 (43.8)	261.5 (45.1)	255.8 (46.1)	280.74 (10.3)
Late-Trained Musician	204.5 (78.5)	207.9 (78.2)	208.2 (74.0)	206.7 (77.7)	209.9 (77.4)	256.4 (44.1)

Standard deviations are listed in parentheses.

Music Training Condition Resources

Video recordings of group piano classes: https://www.youtube.com/channel/UCIvB-zje6n5bA73Ls7cXvtA

Chou, R., Straub, J., & Smayda, K.E. (2017). Behind the Scenes: How group piano pedagogy contributes to older adults' speech-in-noise processing ability. Proceedings from the Australasian Piano Pedagogy Conference. Adelaide, Australia. PDF

- Includes course description, theoretical framework, course materials

Participant Feedback Video (presented at Australasian Piano Pedagogy Conference).

REFERENCES

"A Consensus on the Brain Training Industry from the Scientific Community," Max Planck Institute for Human Development and Stanford Center on Longevity, accessed (10/29/17), http://longevity3.stanford.edu/blog/2014/10/15/the-consensus-on-the-brain-training-industry-from-the-scientific-community/

Ahn, W.-Y., Busemeyer, J. R., Wagenmakers, E.-J., & Stout, J. C. (2008). Comparison of decision learning models using the generalization criterion method. *Cognitive Science*, *32*(8), 1376–402. https://doi.org/10.1080/03640210802352992

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–723. https://doi.org/10.1109/TAC.1974.1100705

Akeroyd, M. A. (2008). Are individual differences in speech reception related to individual differences in cognitive ability? A survey of twenty experimental studies with normal and hearing-impaired adults. *International Journal of Audiology*, *47*(s2), S53–S71. https://doi.org/10.1080/14992020802301142

Alexander, J. A., Wong, P. C., & Bradlow, A. R. (2005). Lexical tone perception in musicians and non-musicians. In *Proceedings of Interspeech* (pp. 397–400). Retrieved from http://groups.linguistics.northwestern.edu/speech_comm_group/publications/2005/Alexander-Wong-Bradlow-2005.pdf

Anderson, S., White-Schwoch, T., Choi, H. J., & Kraus, N. (2014). Partial maintenance of auditory-based cognitive training benefits in older adults. *Neuropsychologia*, *62*, 286–296. https://doi.org/10.1016/j.neuropsychologia.2014.07.034

Anderson, S., White-Schwoch, T., Parbery-Clark, A., & Kraus, N. (2013). A dynamic auditory-cognitive system supports speech-in-noise perception in older adults. *Hearing Research*, *300*, 18–32. https://doi.org/10.1016/j.heares.2013.03.006

Ashby, F. G. (1988). Estimating the parameters of multidimensional signal detection theory from simultaneous ratings on separate stimulus components. *Perception & Psychophysics*, *44*(3), 195–204.

Ashby, F. G., & Maddox, W. (1993). Relations between prototype, exemplar, and decision 5 bound models of categorization. *Journal of Mathematical Psychology*, *37*, 372–400.

Ashby, F. G., & Maddox, W. T. (2005). Human Category Learning. *Annual Review of Psychology*, *56*(1), 149–178. https://doi.org/10.1146/annurev.psych.56.091103.070217

Ashby, F. G., & Maddox, W. T. (2011). Human category learning 2.0: Human category learning 2.0. *Annals of the New York Academy of Sciences*, *1224*(1), 147–161. https://doi.org/10.1111/j.1749-6632.2010.05874.x

Ashby, F. G., & Townsend, J. T. (1986). Varieties of perceptual independence. *Psychological Review*, *93*(2), 154.

Ashby, G. F., Maddox, W. T., & Lee, W. W. (1994). On the dangers of averaging across subjects when using multidimensional scaling or the similarity-choice model. *Psychological Science*, *5*(3), 144–151.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Bailey, J. A., Zatorre, R. J., & Penhune, V. B. (2014a). Early Musical Training Is Linked to Gray Matter Structure in the Ventral Premotor Cortex and Auditory–Motor Rhythm Synchronization Performance. *Journal of Cognitive Neuroscience*, *26*(4), 755–767. https://doi.org/10.1162/jocn_a_00527

Bailey, J. A., Zatorre, R. J., & Penhune, V. B. (2014b). Early Musical Training Is Linked to Gray Matter Structure in the Ventral Premotor Cortex and Auditory–Motor Rhythm Synchronization Performance. *Journal of Cognitive Neuroscience*, *26*(4), 755–767. https://doi.org/10.1162/jocn_a_00527

Bailey, J., & Penhune, V. B. (2012). A sensitive period for musical training: contributions of age of onset and cognitive abilities: Bailey & Penhune. *Annals of the New York Academy of Sciences*, *1252*(1), 163–170. https://doi.org/10.1111/j.1749-6632.2011.06434.x

Ball, K., Berch, D., Helmers, K., Jobe, J., Leveck, M., Marsiske, M., … Willis, S. (2002). Effects of cognitive training interventions with older adults: A randomized controlled trial. *Journal of American Medical Association*, *288*(18).

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Barraclough, D. J., Conroy, M. L., & Lee, D. (2004). Prefrontal cortex and decision making in a mixed-strategy game. *Nature Neuroscience*, *7*(4), 404–410. https://doi.org/10.1038/nn1209

Basak, C., Boot, W. R., Voss, M. W., & Kramer, A. F. (2008). Can training in a real-time strategy video game attenuate cognitive decline in older adults? *Psychology and Aging*, *23*(4), 765–777. https://doi.org/10.1037/a0013494

Bates, D., Maechler, M., & Bolker, B. (2012). *lme4: Linear mixed-effects models using S4 classes*.

Bechara, A., Damasio, A., Damasio, H., & Anderson, S. (1994). Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition*, (50), 7–15.

Bechara, A., Damasio, H., Tranel, D., & Anderson, S. W. (1998). Dissociation of working memory from decision making within the human prefrontal cortex. *Journal of Neuroscience*, *18*(1), 428–437.

Benz, S., Sellaro, R., Hommel, B., & Colzato, L. S. (2016). Music Makes the World Go Round: The Impact of Musical Training on Non-musical Cognitive Functions—A Review. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.02023

Bergman Nutley, S., Darki, F., & Klingberg, T. (2014). Music practice is associated with development of working memory during childhood and adolescence. *Frontiers in Human Neuroscience*, *7*. https://doi.org/10.3389/fnhum.2013.00926

Bermudez, P., Lerch, J. P., Evans, A. C., & Zatorre, R. J. (2009). Neuroanatomical Correlates of Musicianship as Revealed by Cortical Thickness and Voxel-Based Morphometry. *Cerebral Cortex*, *19*(7), 1583–1596. https://doi.org/10.1093/cercor/bhn196

Besson, M., Chobert, J., & Marie, C. (2011). Transfer of Training between Music and Speech: Common Processing, Attention, and Memory. *Frontiers in Psychology*, *2*. https://doi.org/10.3389/fpsyg.2011.00094

Bialystok, E., & DePape, A.-M. (2009). Musical expertise, bilingualism, and executive functioning. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(2), 565–574. https://doi.org/10.1037/a0012735

Bidelman, G. M., Krishnan, A., & Gandour, J. T. (2011). Enhanced brainstem encoding predicts musicians' perceptual advantages with pitch. *The European Journal of Neuroscience*, *33*(3), 530–538. https://doi.org/10.1111/j.1460-9568.2010.07527.x

Bitterman, Y., Mukamel, R., Malach, R., Fried, I., & Nelken, I. (2008). Ultra-fine frequency tuning revealed in single neurons of human auditory cortex. *Nature*, *451*(7175), 197–201. https://doi.org/10.1038/nature06476

Boersma, P., & Weenink, D. (2009). Praat: Doing phonetics by computer (Version 5.1). Retrieved from www.praat.org

Bonneville-Roussy, A., & Bouffard, T. (2015). When quantity is not enough: Disentangling the roles of practice time, self-regulation and deliberate practice in musical achievement. *Psychology of Music*, *43*(5), 686–704.

Borella, E., Carretti, B., Riboldi, F., & De Beni, R. (2010). Working memory training in older adults: Evidence of transfer and maintenance effects. *Psychology and Aging*, *25*(4), 767–778. https://doi.org/10.1037/a0020683

Bradlow, A. R. (2008). Training non-native language sound patterns Lessons from training Iapanese adults on the English. *Phonology and Second Language Acquisition*, *36*, 287.

Bradlow, A. R., & Alexander, J. A. (2007). Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners. *The Journal of the Acoustical Society of America*, *121*(4), 2339. https://doi.org/10.1121/1.2642103

Broche-Pérez, Y., Herrera Jiménez, L. F., & Omar-Martínez, E. (2016). Neural substrates of decision-making. *Neurologia*, *31*(5), 319–325.

Brosch, M., Selezneva, E., Bucks, C., & Scheich, H. (2004). Macaque monkeys discriminate pitch relationships. *Cognition*, *91*(3), 259–272. https://doi.org/10.1016/j.cognition.2003.09.005

Brosch, M., Selezneva, E., & Scheich, H. (2005). Nonauditory Events of a Behavioral Procedure Activate Auditory Cortex of Highly Trained Monkeys. *Journal of Neuroscience*, *25*(29), 6797–6806. https://doi.org/10.1523/JNEUROSCI.1571-05.2005

Bugos. (2010). The benefits of music instruction on processing speed, verbal fluency, and cognitive control in aging. *Music Educ. Res. Int*, *4*, 1–9.

Bugos, J., & Mostafa, W. (2011). Musical Training Enhances Information Processing Speed. *Bulletin of the Council for Research in Music Education*, *187*, 7–18.

Cabeza, R., Anderson, N. D., Locantore, J. K., & McIntosh, A. R. (2002). Aging Gracefully: Compensatory Brain Activity in High-Performing Older Adults. *NeuroImage*, *17*(3), 1394–1402. https://doi.org/10.1006/nimg.2002.1280

Calandruccio, L., & Smiljanic, R. (2012). New sentence recognition materials developed using a basic non-native English lexicon. *Journal of Speech, Language, and Hearing Research*, *55*, 1342–1355.

Caspary, D. M. (2005). Age-Related Changes in the Inhibitory Response Properties of Dorsal Cochlear Nucleus Output Neurons: Role of Inhibitory Inputs. *Journal of Neuroscience*, *25*(47), 10952–10959. https://doi.org/10.1523/JNEUROSCI.2451-05.2005

Chandrasekaran, B., Gandour, J. T., & Krishnan, A. (2007). Neuroplasticity in the processing of pitch dimensions: A multidimensional scaling analysis of the mismatch negativity. *Restorative Neurology & Neuroscience*, *25*(3/4), 195–210.

Chandrasekaran, B., Krishnan, A., & Gandour, J. (2009). Relative influence of musical and linguistic experience on early cortical processing of pitch contours. *Brain and Language*, *108*(1), 1–9. https://doi.org/10.1016/j.bandl.2008.02.001

Chandrasekaran, B., Sampath, P. D., & Wong, P. C. M. (2010). Individual variability in cue-weighting and lexical tone learning. *The Journal of the Acoustical Society of America*, *128*(1), 456–465. https://doi.org/10.1121/1.3445785

Chandrasekaran, B., Yi, H.-G., & Maddox, W. T. (2013). Dual-learning systems during speech category learning. *Psychonomic Bulletin & Review*. https://doi.org/10.3758/s13423-013-0501-5

Chang, Y., Lee, S.-H., Lee, Y.-J., Hwang, M.-J., Bae, S.-J., Kim, M.-N., … Kang, D.-S. (2004). Auditory neural pathway evaluation on sensorineural hearing loss using diffusion tensor imaging: *NeuroReport*, *15*(11), 1699–1703. https://doi.org/10.1097/01.wnr.0000134584.10207.1a

Chen, J. L., Penhune, V. B., & Zatorre, R. J. (2008). Moving on time: Brain network for auditory-motor synchronization is modulated by rhythm complexity and musical training. *Journal of Cognitive Neuroscience*, *20*(2), 226–239.

Chobert, J., Francois, C., Velay, J.-L., & Besson, M. (2014). Twelve Months of Active Musical Training in 8- to 10-Year-Old Children Enhances the Preattentive Processing of Syllabic Duration and Voice Onset Time. *Cerebral Cortex*, *24*(4), 956–967. https://doi.org/10.1093/cercor/bhs377

Chou, R., Straub, J., & Smayda, K. E. (in press). Behind the scenes: Pedagogical insights from an adult group piano course in a speech perception study. *Proceedings from the Australasian Piano Pedagogy Conference*.

Ciorba, A., Bianchini, C., Pelucchi, S., & Pastore, A. (2012). The impact of hearing loss on the quality of life of elderly adults. *Clinical Interventions in Aging*, 159. https://doi.org/10.2147/CIA.S26059

Corrigall, K. A., Schellenberg, E. G., & Misura, N. M. (2013). Music Training, Cognition, and Personality. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00222

Costa-Giomi, E. (1999). The Effects of Three Years of Piano Instruction on Children's Cognitive Development. *Journal of Research in Music Education*, *47*(3), 198. https://doi.org/10.2307/3345779

Crone, E., & van der Molen, M. (2004). Developmental changes in real life decision making: Performance on a gambling task previously shown to depend on the ventromedial prefrontal cortex. *Developmental Neuropsychology*, *25*(3), 251–279.

Crummer, G. C., Walton, J. P., Wayman, J. W., Hantz, E. C., & Frisina, R. D. (1994). Neural processing of musical timbre by musicians, nonmusicians, and musicians possessing absolute pitch. *The Journal of the Acoustical Society of America*, *95*(5), 2720–2727. https://doi.org/10.1121/1.409840

Dai, J., Kerestes, R., Upton, D. J., Busemeyer, J. R., & Stout, J. C. (2015). An improved cognitive model of the Iowa and Soochow Gambling Tasks with regard to model fitting performance and tests of parameter consistency. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00229

Dalton, D. S., Cruickshanks, K. J., Klein, B. E., Klein, R., Wiley, T. L., & Nondahl, D. M. (2003). The impact of hearing loss on quality of life in older adults. *The Gerontologist*, *43*(5), 661–668.

Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P., & Dolan, R. J. (2011). Model-Based Influences on Humans' Choices and Striatal Prediction Errors. *Neuron*, *69*(6), 1204–1215. https://doi.org/10.1016/j.neuron.2011.02.027

Delis, D. C., Kramer, J. W., Kaplan, E., & Ober, B. A. (1987). *The California verbal learning test: Adult version manual*. San Antonio, TX: The Psychological Corporation.

Dipietro, L., Caspersen, C., Ostfeld, A., & Ethan, N. (1993). A survey for assessing physical activity among older adults. *Medicine & Science in Sports & Exercise*, *25*(5), 628–642.

Dubno, J. R., Ahlstrom, J. B., & Horwitz, A. R. (2000). Use of context by young and aged adults with normal hearing. *The Journal of the Acoustical Society of America*, *107*(1), 538. https://doi.org/10.1121/1.428322

Estes, W. K. (1956). The problem of inference from curves based on group data. *Psychological Bulletin*, *53*(2), 134–140.

Evans, S., Meekings, S., Nuttall, H. E., Jasmin, K. M., Boebinger, D., Adank, P., & Scott, S. K. (2014). Does musical enrichment enhance the neural coding of syllables? Neuroscientific interventions and the importance of behavioral data. *Frontiers in Human Neuroscience*, *8*. https://doi.org/10.3389/fnhum.2014.00964

Filoteo, J. V., & Maddox, W. T. (1999). Quantitative modeling of visual attention processes in patients with Parkinson's disease: Effects of stimulus integrality on selective attention and dimensional integration. *Neuropsychology*, *13*(2), 206.

Francis, A. L., Ciocca, V., Ma, L., & Fenn, K. (2008). Perceptual learning of Cantonese lexical tones by tone and non-tone language speakers. *Journal of Phonetics*, *36*(2), 268–294. https://doi.org/10.1016/j.wocn.2007.06.005

Francis, A. L., & Nusbaum, H. C. (2002). Selective attention and the acquisition of new phonetic categories. *Journal of Experimental Psychology: Human Perception and Performance*, *28*(2), 349–366. https://doi.org/10.1037//0096-1523.28.2.349

Frisina, D. R., & Frisina, R. D. (1997). Speech recognition in noise and presbycusis: relations to possible neural mechanisms. *Hearing Research*, *106*(1–2), 95–104.

Gandour, J. (1983). Tone perception in Far Eastern languages. *Journal of Phonetics*, *11*, 149–175.

Gandour, J. T., & Harshman, R. A. (1978). Crosslanguage differences in tone perception: A multidimensional scaling investigation. *Language and Speech*, *21*(1), 1–33.

George, E. M., & Coch, D. (2011). Music training and working memory: An ERP study. *Neuropsychologia*, *49*(5), 1083–1094. https://doi.org/10.1016/j.neuropsychologia.2011.02.001

Gilbert, R. C., Chandrasekaran, B., & Smiljanic, R. (2014). Recognition memory in noise for speech of varying intelligibilitya). *The Journal of the Acoustical Society of America*, *135*(1), 389–399.

Gogtay, N., Giedd, J. N., Lusk, L., Hayashi, K. M., Greenstein, D., Vaituzis, A. C., … others. (2004). Dynamic mapping of human cortical development during childhood through early adulthood. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(21), 8174–8179.

Goldstone, R. (1994). Influences of categorization on perceptual discrimination. *Journal of Experimental Psychology: General*, *123*(2), 178–200.

Gottfried, T. L., & Riester, D. (2000). Relation of pitch glide perception and Mandarin tone identification. *The Journal of the Acoustical Society of America*, *115*(5), 2545.

Gottfried, T. L., Staby, A. M., & Ziemer, C. J. (2004). Musical experience and Mandarin tone discrimination and imitation. *The Journal of the Acoustical Society of America*, *115*(5), 2545–2545.

Grahn, J. A., & Rowe, J. B. (2012). Finding and Feeling the Musical Beat: Striatal Dissociations between Detection and Prediction of Regularity. *Cerebral Cortex*. https://doi.org/10.1093/cercor/bhs083

Green, D. M., & Swets, J. A. (1967). *Signal detection theory and psychophysics*. New York: Wiley.

Gureckis, T. M., & Love, B. C. (2009). Learning in noise: Dynamic decision-making in a variable environment. *Journal of Mathematical Psychology*, *53*(3), 180–193. https://doi.org/10.1016/j.jmp.2009.02.004

Hanna-Pladdy, B., & MacKay, A. (2011). The relation between instrumental musical activity and cognitive aging. *Neuropsychology*, *25*(3), 378–386. https://doi.org/10.1037/a0021895

Hare, T. A., O'Doherty, J., Camerer, C. F., Schultz, W., & Rangel, A. (2008). Dissociating the Role of the Orbitofrontal Cortex and the Striatum in the Computation of Goal Values and Prediction Errors. *Journal of Neuroscience*, *28*(22), 5623–5630. https://doi.org/10.1523/JNEUROSCI.1309-08.2008

He, N., Dubno, J. R., & Mills, J. H. (1998). Frequency and intensity discrimination measured in a maximum-likelihood procedure from young and aged normal-hearing subjects. *The Journal of the Acoustical Society of America*, *103*(1), 553–565. https://doi.org/10.1121/1.421127

Heaton, R. K., Chelune, G. J., Talley, J. L., Kay, G. G., & Curtiss, G. (1993). *Wisconsin Card Sorting Test manual*. Odessa, FL: Psychological Assessment Resources, Inc.

Heine, C., & Browning, C. J. (2002). Communication and psychosocial consequences of sensory loss in older adults: overview and rehabilitation directions. *Disability & Rehabilitation*, *24*(15), 763–773.

Helfer, K. S. (1997). Auditory and auditory-visual perception of clear and conversational speech. *Journal of Speech, Language, and Hearing Research*, *40*(2), 432–443.

Helfer, K. S. (1998). Auditory and auditory-visual recognition of clear and conversational speech by older adults. *JOURNAL-AMERICAN ACADEMY OF AUDIOLOGY*, *9*, 234–242.

Helfer, K. S., & Freyman, R. L. (2007). Aging and Speech-on-Speech Masking: *Ear and Hearing*, *PAP*. https://doi.org/10.1097/AUD.0b013e31815d638b

Hensch, T. K. (2005). Critical period plasticity in local cortical circuits. *Nature Reviews Neuroscience*, *6*(11), 877–888. https://doi.org/10.1038/nrn1787

Herholz, S. C., & Zatorre, R. J. (2012). Musical Training as a Framework for Brain Plasticity: Behavior, Function, and Structure. *Neuron*, *76*(3), 486–502. https://doi.org/10.1016/j.neuron.2012.10.011

Hooper, C. J., Luciana, M., Conklin, H. M., & Yarger, R. S. (2004). Adolescents' Performance on the Iowa Gambling Task: Implications for the Development of Decision Making and Ventromedial Prefrontal Cortex. *Developmental Psychology*, *40*(6), 1148–1158. https://doi.org/10.1037/0012-1649.40.6.1148

Hornickel, J., & Kraus, N. (2013). Unstable Representation of Sound: A Biological Marker of Dyslexia. *Journal of Neuroscience*, *33*(8), 3500–3504. https://doi.org/10.1523/JNEUROSCI.4205-12.2013

Horstmann, A. (2012). Iowa Gambling Task: there is more to consider than long-term outcome. Using a linear equation model to disentangle the impact of outcome and frequency of gains and losses. *Frontiers in Neuroscience*, *6*. https://doi.org/10.3389/fnins.2012.00061

Howie, J. (1976). *Acoustical studies of Mandarin vowels and tones*. Cambridge: Cambridge University Press.

Humes, L. E., Busey, T. A., Craig, J. C., & Kewley-Port, D. (2009). The effects of age on sensory thresholds and temporal gap detection in hearing, vision, and touch.

Attention, Perception, & Psychophysics, *71*(4), 860–871. https://doi.org/10.3758/APP.71.4.860

Humes, L. E., Dubno, J. R., Gordon-Salant, S., Lister, J. J., Cacace, A. T., Cruickshanks, K. J., … Wingfield, A. (2012). Central Presbycusis: A Review and Evaluation of the Evidence. *Journal of the American Academy of Audiology*, *23*(8), 635–666. https://doi.org/10.3766/jaaa.23.8.5

Husain, F. T., Medina, R. E., Davis, C. W., Szymko-Bennett, Y., Simonyan, K., Pajor, N. M., & Horwitz, B. (2011). Neuroanatomical changes due to hearing loss and chronic tinnitus: A combined VBM and DTI study. *Brain Research*, *1369*, 74–88. https://doi.org/10.1016/j.brainres.2010.10.095

Jin, S.-H., Liu, C., & Sladen, D. P. (2014). The Effects of Aging on Speech Perception in Noise: Comparison between Normal-Hearing and Cochlear-Implant Listeners. *Journal of the American Academy of Audiology*, *25*(7), 656–665. https://doi.org/10.3766/jaaa.25.7.4

Karni, A., Meyer, G., Rey-Hipolito, C., Jezzard, P., Adams, M. M., Turner, R., & Ungerleider, L. G. (1998). The acquisition of skilled motor performance: fast and slow experience-driven changes in primary motor cortex. *Proceedings of the National Academy of Sciences*, *95*(3), 861–868.

Kempe, V., Bublitz, D., & Brooks, P. J. (2013). Musicality and non-native speech sound processing are linked through temporal, pitch and spectral acuity. In *CogSci*. Retrieved from

https://pdfs.semanticscholar.org/636e/dda7442b3391a782a14123fb67673b120441.pdf

Kidd Jr, G., Mason, C. R., Richards, V. M., Gallun, F. J., & Durlach, N. I. (2008). Informational masking. In *Auditory perception of sound sources* (pp. 143–189). Springer. Retrieved from http://link.springer.com/chapter/10.1007/978-0-387-71305-2_6

Kim, B. J., & Oh, S.-H. (2013). Age-Related Changes in Cognition and Speech Perception. *Korean Journal of Audiology*, *17*(2), 54. https://doi.org/10.7874/kja.2013.17.2.54

Kjome, K. L., Lane, S. D., Schmitz, J. M., Green, C., Ma, L., Prasla, I., … Moeller, F. G. (2010). Relationship between impulsivity and decision making in cocaine dependence. *Psychiatry Research*, *178*(2), 299–304. https://doi.org/10.1016/j.psychres.2009.11.024

Koelewijn, T., Zekveld, A. A., Festen, J. M., Rönnberg, J., & Kramer, S. E. (2012). Processing Load Induced by Informational Masking Is Related to Linguistic Abilities. *International Journal of Otolaryngology*, *2012*, 1–11. https://doi.org/10.1155/2012/865731

Kovach, C. K., Daw, N. D., Rudrauf, D., Tranel, D., O'Doherty, J. P., & Adolphs, R. (2012). Anterior Prefrontal Cortex Contributes to Action Selection through Tracking of Recent Reward Trends. *Journal of Neuroscience*, *32*(25), 8434–8442. https://doi.org/10.1523/JNEUROSCI.5468-11.2012

Kraus, N., & Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nature Reviews Neuroscience*, *11*(8), 599–605.

Kraus, N., Slater, J., Thompson, E. C., Hornickel, J., Strait, D. L., Nicol, T., & White-Schwoch, T. (2014). Music Enrichment Programs Improve the Neural Encoding of Speech in At-Risk Children. *Journal of Neuroscience*, *34*(36), 11913–11918. https://doi.org/10.1523/JNEUROSCI.1881-14.2014

Kraus, N., Strait, D. L., & Parbery-Clark, A. (2012). Cognitive factors shape brain networks for auditory skills: spotlight on auditory working memory: Kraus et al. *Annals of the New York Academy of Sciences*, *1252*(1), 100–107. https://doi.org/10.1111/j.1749-6632.2012.06463.x

Lau, B., & Glimcher, P. W. (2005). Dynamic Response-by-Response Models of Matching Behavior in Rhesus Monkeys. *Journal of the Experimental Analysis of Behavior*, *84*(3), 555–579. https://doi.org/10.1901/jeab.2005.110-04

Lee, C.-Y., & Hung, T.-H. (2008). Identification of Mandarin tones by English-speaking musicians and nonmusicians. *The Journal of the Acoustical Society of America*, *124*(5), 3235. https://doi.org/10.1121/1.2990713

Levitin, D. J. (2012). What Does It Mean to Be Musical? *Neuron*, *73*(4), 633–637. https://doi.org/10.1016/j.neuron.2012.01.017

Li, X., Lu, Z.-L., D'Argembeau, A., Ng, M., & Bechara, A. (2009). The Iowa Gambling Task in fMRI images. *Human Brain Mapping*, NA-NA. https://doi.org/10.1002/hbm.20875

Lin, F. R., Thorpe, R., Gordon-Salant, S., & Ferrucci, L. (2011). Hearing loss prevalence and risk factors among older adults in the United States. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, *66*(5), 582–590.

Lin, Y., Wang, J., Wu, C., Wai, Y., Yu, J., & Ng, S. (2008). Diffusion tensor imaging of the auditory pathway in sensorineural hearing loss: Changes in radial diffusivity and diffusion anisotropy. *Journal of Magnetic Resonance Imaging*, *28*(3), 598–603. https://doi.org/10.1002/jmri.21464

Lutz, J., Hemminger, F., Stahl, R., Dietrich, O., Hempel, M., Reiser, M., & Jäger, L. (2007). Evidence of Subcortical and Cortical Aging of the Acoustic Pathway: A Diffusion Tensor Imaging (DTI) Study. *Academic Radiology*, *14*(6), 692–700. https://doi.org/10.1016/j.acra.2007.02.014

Maddox, W. T. (1999). On the dangers of averaging across observers when comparing decision bound models and generalized context models of categorization. *Attention, Perception, & Psychophysics*, *61*(2), 354–374.

Maddox, W. T. (2001). Separating perceptual processes from decisional processes in identification and categorization. *Perception & Psychophysics*, *63*(7), 1183–1200.

Maddox, W. T., & Ashby, F. G. (1993). Comparing decision bound and exemplar models of categorization. *Perception & Psychophysics*, *53*(1), 49–70.

Maddox, W. T., & Bogdanov, S. V. (2000). On the relation between decision rules and perceptual representation in multidimensional perceptual categorization. *Attention, Perception, & Psychophysics*, *62*(5), 984–997.

Maddox, W. T., & Chandrasekaran, B. (2014). Tests of a dual-system model of speech category learning. *Bilingualism: Language and Cognition*, 1–20. https://doi.org/10.1017/S1366728913000783

Maddox, W. T., Chandrasekaran, B., Smayda, K., & Yi, H.-G. (2013). Dual systems of speech category learning across the lifespan. *Psychology and Aging*, *28*(4), 1042–1056. https://doi.org/10.1037/a0034969

Maddox, W. T., & Dodd, J. L. (2003). Separating perceptual and decisional attention processes in the identification and categorization of integral-dimension stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(3), 467–480. https://doi.org/10.1037/0278-7393.29.3.467

Magne, C., Schön, D., & Besson, M. (2006). Musician children detect pitch violations in both music and language better than nonmusician children: behavioral and electrophysiological approaches. *Journal of Cognitive Neuroscience*, *18*(2), 199–211.

Maia, T. V., & McClelland, J. L. (2004). A reexamination of the evidence for the somatic marker hypothesis: what participants really know in the Iowa gambling task. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(45), 16075–16080.

Manes, F., Sahakian, B., Clark, L., Rogers, R., Antoun, N., Aitken, M., & Robbins, T. (2002). Decision-making processes following damage to the prefrontal cortex. *Brain*, *125*, 624–639.

Marie, C., Delogu, F., Lampis, G., Belardinelli, M. O., & Besson, M. (2011). Influence of musical expertise on segmental and tonal processing in Mandarin Chinese. *Journal of Cognitive Neuroscience*, *23*(10), 2701–2715.

Mather, M., Jacobsen, L., A., & Pollard, K., M. (2015). *Aging in the United States* (No. 70, no. 2). Population Reference Bureau.

Lezak, M.D. (1995). *Neuropsychological Assessment* (3rd ed.). New York, NY: Oxford University Press.

Miyachi, S., Hikosaka, O., & Lu, X. (2002). Differential activation of monkey striatal neurons in the early and late stages of procedural learning. *Experimental Brain Research*, *146*(1), 122–126. https://doi.org/10.1007/s00221-002-1213-7

Moreno, S., Bialystok, E., Barac, R., Schellenberg, E. G., Cepeda, N. J., & Chau, T. (2011). Short-term music training enhances verbal intelligence and executive function. *Psychological Science*, *22*(11), 1425–1433. https://doi.org/10.1177/0956797611416999

Moreno, S., Marques, C., Santos, A., Santos, M., Castro, S. L., & Besson, M. (2009). Musical Training Influences Linguistic Abilities in 8-Year-Old Children: More Evidence for Brain Plasticity. *Cerebral Cortex*, *19*(3), 712–723. https://doi.org/10.1093/cercor/bhn120

Nye, P. W., & Gaitenby, J. H. (1974). *The intelligibility of synthetic monosyllabic words in short, syntactically normal sentences. (Status Report on Speech Research SR-37/38)*. New Haven, CT: Haskins Laboratory.

Ohl, F. W., & Scheich, H. (2005). Learning-induced plasticity in animal and human auditory cortex. *Current Opinion in Neurobiology*, *15*(4), 470–477. https://doi.org/10.1016/j.conb.2005.07.002

Ohl, F. W., Scheich, H., & Freeman, W. J. (2001). Change in pattern of ongoing cortical activity with auditory category learning. *Nature*, *412*, 734–736.

Ono, Y., Kikuchi, M., Hirosawa, T., Hino, S., Nagasawa, T., Hashimoto, T., … Minabe, Y. (2015). Reduced prefrontal activation during performance of the Iowa Gambling Task in patients with bipolar disorder. *Psychiatry Research: Neuroimaging*, *233*(1), 1–8. https://doi.org/10.1016/j.pscychresns.2015.04.003

Otto, A. R., Taylor, E. G., & Markman, A. B. (2011). There are at least two kinds of probability matching: Evidence from a secondary task. *Cognition*, *118*(2), 274–279. https://doi.org/10.1016/j.cognition.2010.11.009

Ouda, L., Profant, O., & Syka, J. (2015). Age-related changes in the central auditory system. *Cell and Tissue Research*, *361*(1), 337–358. https://doi.org/10.1007/s00441-014-2107-2

Ouimet, T., Foster, N. E. V., & Hyde, K. L. (2012). Auditory global-local processing: Effects of attention and musical experience. *The Journal of the Acoustical Society of America*, *132*(4), 2536–2544. https://doi.org/10.1121/1.4747009

Pallesen, K. J., Brattico, E., Bailey, C. J., Korvenoja, A., Koivisto, J., Gjedde, A., & Carlson, S. (2010). Cognitive Control in Auditory Working Memory Is Enhanced in Musicians. *PLoS ONE*, *5*(6). https://doi.org/10.1371/journal.pone.0011120

Palmer, C. M. (2011). Challenges of access to post-secondary music education programs for people of color. *Visions of Research in Music Education*, *18*. Retrieved from http://www-usr.rider.edu/~vrme/v18n1/visions/Palmer-Access%20to%20Music%20Ed%20for%20People%20of%20Color.pdf

Parbery-Clark, A., Anderson, S., Hittner, E., & Kraus, N. (2012). Musical experience strengthens the neural representation of sounds important for communication in middle-aged adults. *Frontiers in Aging Neuroscience*, *4*, 30. https://doi.org/10.3389/fnagi.2012.00030

Parbery-Clark, A., Skoe, E., Lam, C., & Kraus, N. (2009). Musician enhancement for speech-in-noise. *Ear and Hearing*, *30*(6), 653–661.

Parbery-Clark, A., Strait, D. L., Anderson, S., Hittner, E., & Kraus, N. (2011). Musical Experience and the Aging Auditory System: Implications for Cognitive Abilities and Hearing Speech in Noise. *PLoS ONE*, *6*(5), e18082. https://doi.org/10.1371/journal.pone.0018082

Parbery-Clark, A., Strait, D. L., & Kraus, N. (2011). Context-dependent encoding in the auditory brainstem subserves enhanced speech-in-noise perception in musicians. *Neuropsychologia*, *49*(12), 3338–3345. https://doi.org/10.1016/j.neuropsychologia.2011.08.007

Park, D. C., Gutchess, A. H., Meade, M. L., & Stine-Morrow, E. A. (2007). Improving cognitive function in older adults: nontraditional approaches. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *62*(Special Issue 1), 45–52.

Park, D. C., Lodi-Smith, J., Drew, L., Haber, S., Hebrank, A., Bischof, G. N., & Aamodt, W. (2014). The Impact of Sustained Engagement on Cognitive Function in Older Adults: The Synapse Project. *Psychological Science*, *25*(1), 103–112. https://doi.org/10.1177/0956797613499592

Park, D. C., & Reuter-Lorenz, P. (2009). The Adaptive Brain: Aging and Neurocognitive Scaffolding. *Annual Review of Psychology*, *60*(1), 173–196. https://doi.org/10.1146/annurev.psych.59.103006.093656

Patel, A. D. (2011). Why would Musical Training Benefit the Neural Encoding of Speech? The OPERA Hypothesis. *Frontiers in Psychology*, *2*. https://doi.org/10.3389/fpsyg.2011.00142

Patel, A. D. (2012). The OPERA hypothesis: assumptions and clarifications: Patel. *Annals of the New York Academy of Sciences*, *1252*(1), 124–128. https://doi.org/10.1111/j.1749-6632.2011.06426.x

Patel, A. D. (2014). Can nonlinguistic musical training change the way the brain processes speech? The expanded OPERA hypothesis. *Hearing Research*, *308*, 98–108. https://doi.org/10.1016/j.heares.2013.08.011

Perrachione, T. K., Fedorenko, E. G., Vinke, L., Gibson, E., & Dilley, L. C. (2013). Evidence for Shared Cognitive Processing of Pitch in Music and Language. *PLoS ONE*, *8*(8), e73372. https://doi.org/10.1371/journal.pone.0073372

Perrachione, T. K., Lee, J., Ha, L. Y. Y., & Wong, P. C. M. (2011). Learning a novel phonological contrast depends on interactions between individual differences and

training paradigm design. *The Journal of the Acoustical Society of America*, *130*(1), 461–472. https://doi.org/10.1121/1.3593366

Pichora-Fuller, M. K. (2003). Processing speed and timing in aging adults: psychoacoustics, speech perception, and comprehension. *International Journal of Audiology*, *42*, S59–S67.

Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W., Humes, L. E., … others. (2016). Hearing Impairment and Cognitive Energy: The Framework for Understanding Effortful Listening (FUEL). *Ear and Hearing*, *37*, 5S–27S.

Pichora-Fuller, Schneider, & Daneman. (1995). How young and old adults listen to and remember speech. *The Journal of the Acoustical Society of America*, *97*(1), 593–608.

Premkumar, P., Fannon, D., Kuipers, E., Simmons, A., Frangou, S., & Kumari, V. (2008). Emotional decision-making and its dissociable components in schizophrenia and schizoaffective disorder: A behavioural and MRI investigation. *Neuropsychologia*, *46*(7), 2002–2012. https://doi.org/10.1016/j.neuropsychologia.2008.01.022

Profant, O., Škoch, A., Balogová, Z., Tintěra, J., Hlinka, J., & Syka, J. (2014). Diffusion tensor imaging and MR morphometry of the central auditory pathway and auditory cortex in aging. *Neuroscience*, *260*, 87–97. https://doi.org/10.1016/j.neuroscience.2013.12.010

Raina, P., Wong, M., & Massfeller, H. (2004). The relationship between sensory impairment and functional independence among elderly. *BMC Geriatrics*, *4*(1), 3.

Rodrigues, A. C., Loureiro, M. A., & Caramelli, P. (2013). Long-term musical training may improve different forms of visual attention ability. *Brain and Cognition*, *82*(3), 229–235. https://doi.org/10.1016/j.bandc.2013.04.009

Rogers, C. S., Jacoby, L. L., & Sommers, M. S. (2012). Frequent false hearing by older adults: The role of age differences in metacognition. *Psychology and Aging*, *27*(1), 33–45. https://doi.org/10.1037/a0026231

Ross, L. A., Saint-Amour, D., Leavitt, V. M., Javitt, D. C., & Foxe, J. J. (2006). Do You See What I Am Saying? Exploring Visual Enhancement of Speech Comprehension in Noisy Environments. *Cerebral Cortex*, *17*(5), 1147–1153. https://doi.org/10.1093/cercor/bhl024

Samanez-Larkin, G. R., Worthy, D. A., Mata, R., McClure, S. M., & Knutson, B. (2014). Adult age differences in frontostriatal representation of prediction error but not reward outcome. *Cognitive, Affective, & Behavioral Neuroscience*, *14*(2), 672–682. https://doi.org/10.3758/s13415-014-0297-4

Savion-Lemieux, T., Bailey, J. A., & Penhune, V. B. (2009). Developmental contributions to motor sequence learning. *Experimental Brain Research*, *195*(2), 293–306. https://doi.org/10.1007/s00221-009-1786-5

Schellenberg, E. G. (2005). Music and Cognitive Abilities. *Current Directions in Psychological Science*, *14*(6), 317–320. https://doi.org/10.1111/j.0963-7214.2005.00389.x

Schneider, B. A., & Pichora-Fuller, M. K. (2001). Age-related changes in temporal processing: implications for speech perception. In *Seminars in hearing* (Vol. 22, pp. 227–240). Copyright\copyright 2001 by Thieme Medical Publishers, Inc., 333 Seventh Avenue, New York, NY 10001, USA. Tel.:+ 1 (212) 584-4662.

Schön, D., Magne, C., & Besson, M. (2004). The music of speech: Music training facilitates pitch processing in both music and language: Music and prosody: An ERP study. *Psychophysiology*, *41*(3), 341–349. https://doi.org/10.1111/1469-8986.00172.x

Schönberg, T., Daw, N. D., Joel, D., & O'Doherty, J. P. (2007). Reinforcement Learning Signals in the Human Striatum Distinguish Learners from Nonlearners during Reward-Based Decision Making. *Journal of Neuroscience*, *27*(47), 12860–12867. https://doi.org/10.1523/JNEUROSCI.2496-07.2007

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*(2), 461–464.

Shurman, B., Horan, W. P., & Nuechterlein, K. H. (2005). Schizophrenia patients demonstrate a distinctive pattern of decision-making impairment on the Iowa Gambling Task. *Schizophrenia Research*, *72*(2–3), 215–224. https://doi.org/10.1016/j.schres.2004.03.020

Slater, J., Skoe, E., Strait, D. L., O'Connell, S., Thompson, E., & Kraus, N. (2015). Music training improves speech-in-noise perception: Longitudinal evidence from a community-based music program. *Behavioural Brain Research*, *291*, 244–252. https://doi.org/10.1016/j.bbr.2015.05.026

Smayda, K., Chandrasekaran, B., & Maddox, W. T. (2015). Enhanced cognitive and perceptual processing: A computational basis for the musician advantage in speech learning. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00682

Smayda, K. E., Van Engen, K. J., Maddox, W. T., & Chandrasekaran, B. (2016). Audio-Visual and Meaningful Semantic Context Enhancements in Older and Younger Adults. *PLOS ONE*, *11*(3), e0152773. https://doi.org/10.1371/journal.pone.0152773

Smiljanic, R., & Sladen, D. (2013). Acoustic and Semantic Enhancements for Children With Cochlear Implants. *Journal of Speech Language and Hearing Research*, *56*(4), 1085. https://doi.org/10.1044/1092-4388(2012/12-0097)

Sommers, M. S., Tye-Murray, N., & Spehar, B. (2005). Auditory-visual speech perception and auditory-visual enhancement in normal-hearing younger and older adults. *Ear and Hearing*, *26*(3), 263–275.

Steele, C. J., Bailey, J. A., Zatorre, R. J., & Penhune, V. B. (2013). Early Musical Training and White-Matter Plasticity in the Corpus Callosum: Evidence for a Sensitive Period. *Journal of Neuroscience*, *33*(3), 1282–1290. https://doi.org/10.1523/JNEUROSCI.3578-12.2013

Steingroever, H., Fridberg, D. J., Horstmann, A., Kjome, K. L., Kumari, V., Lane, S. D., … Wagenmakers, E.-J. (2015). Data from 617 Healthy Participants Performing the Iowa Gambling Task: A "Many Labs" Collaboration. *Journal of Open Psychology Data*, *3*. https://doi.org/10.5334/jopd.ak

Stepankova, H., Lukavsky, J., Buschkuehl, M., Kopecek, M., Ripova, D., & Jaeggi, S. M. (2014). The malleability of working memory and visuospatial skills: A randomized controlled study in older adults. *Developmental Psychology*, *50*(4), 1049–1059. https://doi.org/10.1037/a0034913

Stevenson, R. A., Nelms, C. E., Baum, S. H., Zurkovsky, L., Barense, M. D., Newhouse, P. A., & Wallace, M. T. (2015). Deficits in audiovisual speech perception in normal aging emerge at the level of whole-word recognition. *Neurobiology of Aging*, *36*(1), 283–291. https://doi.org/10.1016/j.neurobiolaging.2014.08.003

Strait, D. L., & Kraus, N. (2011). Can You Hear Me Now? Musical Training Shapes Functional Brain Networks for Selective Auditory Attention and Hearing Speech in Noise. *Frontiers in Psychology*, *2*. https://doi.org/10.3389/fpsyg.2011.00113

Strait, D. L., Kraus, N., Parbery-Clark, A., & Ashley, R. (2010). Musical experience shapes top-down auditory mechanisms: Evidence from masking and auditory attention performance. *Hearing Research*, *261*(1–2), 22–29. https://doi.org/10.1016/j.heares.2009.12.021

Strait, D. L., O'Connell, S., Parbery-Clark, A., & Kraus, N. (2013). Musicians' Enhanced Neural Differentiation of Speech Sounds Arises Early in Life: Developmental Evidence from Ages 3 to 30. *Cerebral Cortex*. https://doi.org/10.1093/cercor/bht103

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*(6), 643.

Sul, J. H., Kim, H., Huh, N., Lee, D., & Jung, M. W. (2010). Distinct Roles of Rodent Orbitofrontal and Medial Prefrontal Cortex in Decision Making. *Neuron*, *66*(3), 449–460. https://doi.org/10.1016/j.neuron.2010.03.033

Sumby, W. H., & Pollack, I. (1954). Visual Contribution to Speech Intelligibility in Noise. JASA Express.

Sutton, R., & Barto, A. (1998). *Reinforcement Learning: An Introduction*. Cambridge: MIT Press.

Sweetow, R. W., & Sabes, J. H. (2006). The need for and development of an adaptive listening and communication enhancement (LACE™) program. *Journal of the American Academy of Audiology*, *17*(8), 538–558.

Talamini, F., Carretti, B., & Grassi, M. (2016). The Working Memory of Musicians and Nonmusicians. *Music Perception: An Interdisciplinary Journal*, *34*(2), 183–191. https://doi.org/10.1525/mp.2016.34.2.183

Tervaniemi, M., Just, V., Koelsch, S., Widmann, A., & Schröger, E. (2004). Pitch discrimination accuracy in musicians vs nonmusicians: an event-related potential and behavioral study. *Experimental Brain Research*, *161*(1), 1–10. https://doi.org/10.1007/s00221-004-2044-5

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, *37*(3), 498–505.

Vaden, K. I., Kuchinsky, S. E., Ahlstrom, J. B., Dubno, J. R., & Eckert, M. A. (2015). Cortical Activity Predicts Which Older Adults Recognize Speech in Noise and

When. *Journal of Neuroscience*, *35*(9), 3929–3937. https://doi.org/10.1523/JNEUROSCI.2908-14.2015

Vaden, K. I., Kuchinsky, S. E., Ahlstrom, J. B., Teubner-Rhodes, S. E., Dubno, J. R., & Eckert, M. A. (2016). Cingulo-Opercular Function During Word Recognition in Noise for Older Adults with Hearing Loss. *Experimental Aging Research*, *42*(1), 67–82. https://doi.org/10.1080/0361073X.2016.1108784

Van Engen, K. J., Phelps, J. E. B., Smiljanic, R., & Chandrasekaran, B. (2014). Enhancing Speech Intelligibility: Interactions Among Context, Modality, Speech Style, and Masker. *Journal of Speech Language and Hearing Research*, *57*(5), 1908. https://doi.org/10.1044/JSLHR-H-13-0076

Vandervert, L. (2015). How music training enhances working memory: a cerebrocerebellar blending mechanism that can lead equally to scientific discovery and therapeutic efficacy in neurological disorders. *Cerebellum & Ataxias*, *2*(1). https://doi.org/10.1186/s40673-015-0030-2

Venables, W. N., & Ripley, B. D. (2002). Modern Applied Statistics with S. (Version 4th Edition). New York: Springer.

Wagenmakers, E.-J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin & Review*, *11*(1), 192–196. https://doi.org/10.3758/BF03206482

Walden, B. E., Busacco, D. A., & Montgomery, A. A. (1993). Benefit from visual cues in auditory-visual speech recognition by middle-aged and elderly persons. *Journal of Speech, Language, and Hearing Research*, *36*(2), 431–436.

Wang, Y., Spence, M. M., Jongman, A., & Sereno, J. A. (1999). Training American listeners to perceive Mandarin tones. *The Journal of the Acoustical Society of America*, *106*(6), 3649–3658.

Wayne, R. V., Hamilton, C., Jones Huyck, J., & Johnsrude, I. S. (2016). Working Memory Training and Speech in Noise Comprehension in Older Adults. *Frontiers in Aging Neuroscience*, *8*. https://doi.org/10.3389/fnagi.2016.00049

Wechsler, D. (1997). *Wechsler Adult Intelligence Scale - Third edition*. San Antonio, TX: Harcourt Brace & Company.

White-Schwoch, T., Carr, K. W., Anderson, S., Strait, D. L., & Kraus, N. (2013). Older Adults Benefit from Music Training Early in Life: Biological Evidence for Long-Term Training-Driven Plasticity. *Journal of Neuroscience*, *33*(45), 17667–17674. https://doi.org/10.1523/JNEUROSCI.2560-13.2013

Willis, S. L., Tennstedt, S. L., Marsiske, M., Ball, K., Elias, J., Koepke, K. M., … others. (2006). Long-term effects of cognitive training on everyday functional outcomes in older adults. *Jama*, *296*(23), 2805–2814.

Wong, P. C., Ettlinger, M., Sheppard, J. P., Gunasekera, G. M., & Dhar, S. (2010). Neuroanatomical characteristics and speech perception in noise in older adults. *Ear and Hearing*, *31*(4), 471.

Wong, P. C. M., Jin, J. X., Gunasekera, G. M., Abel, R., Lee, E. R., & Dhar, S. (2009). Aging and cortical mechanisms of speech perception in noise. *Neuropsychologia*, *47*(3), 693–703. https://doi.org/10.1016/j.neuropsychologia.2008.11.032

Wong, P. C. M., & Perrachione, T. K. (2007). Learning pitch patterns in lexical identification by native English-speaking adults. *Applied Psycholinguistics*, *28*(4), 565–585. https://doi.org/10.1017/S0142716407070312

Wong, P. C. M., Skoe, E., Russo, N. M., Dees, T., & Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nature Neuroscience*. https://doi.org/10.1038/nn1872

Wood, S., Busemeyer, J., Koling, A., Cox, C. R., & Davis, H. (2005). Older Adults as Adaptive Decision Makers: Evidence From the Iowa Gambling Task. *Psychology and Aging*, *20*(2), 220–225. https://doi.org/10.1037/0882-7974.20.2.220

Worthy, D. A., Hawthorne, M. J., & Otto, A. R. (2013). Heterogeneity of strategy use in the Iowa gambling task: A comparison of win-stay/lose-shift and reinforcement learning models. *Psychonomic Bulletin & Review*, *20*(2), 364–371. https://doi.org/10.3758/s13423-012-0324-9

Worthy, D. A., & Maddox, W. T. (2012). Age-Based Differences in Strategy Use in Choice Tasks. *Frontiers in Neuroscience*, *5*. https://doi.org/10.3389/fnins.2011.00145

Worthy, D. A., Otto, A. R., & Maddox, W. T. (2012). Working-memory load and temporal myopia in dynamic decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(6), 1640–1658. https://doi.org/10.1037/a0028146

Worthy, D. A., Pang, B., & Byrne, K. A. (2013). Decomposing the roles of perseveration and expected value representation in models of the Iowa gambling task. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00640

Xie, Z., Maddox, W. T., Knopik, V. S., McGeary, J. E., & Chandrasekaran, B. (2015). Dopamine receptor D4 (DRD4) gene modulates the influence of informational masking on speech recognition. *Neuropsychologia*, *67*, 121–131. https://doi.org/10.1016/j.neuropsychologia.2014.12.013

Yechiam, E., Busemeyer, J. R., Stout, J. C., & Bechara, A. (2005). Using cognitive models to map relations between neuropsychological disorders and human decision-making deficits. *Psychological Science*, *16*(12), 973–978.

Yi, H.-G., Maddox, W. T., Mumford, J. A., & Chandrasekaran, B. (2014). The Role of Corticostriatal Systems in Speech Category Learning. *Cerebral Cortex*. https://doi.org/10.1093/cercor/bhu236

Yin, H. H., Mulcare, S. P., Hilário, M. R. F., Clouse, E., Holloway, T., Davis, M. I., … Costa, R. M. (2009). Dynamic reorganization of striatal circuits during the acquisition and consolidation of a skill. *Nature Neuroscience*, *12*(3), 333–341. https://doi.org/10.1038/nn.2261

Zald, D. H., & Zatorre, R. J. (2011). *Neurobiology of Sensation and Reward*. (J. A. Gottfried, Ed.). Boca Raton: CRC Press/Taylor & Francis.

Zarate, J. M., Ritson, C. R., & Poeppel, D. (2012). Pitch-interval discrimination and musical expertise: Is the semitone a perceptual boundary? *The Journal of the*

Acoustical Society of America, *132*(2), 984–993. https://doi.org/10.1121/1.4733535

Zekveld, A. A., Rudner, M., Johnsrude, I. S., Heslenfeld, D. J., & Rönnberg, J. (2012). Behavioral and fMRI evidence that cognitive ability modulates the effect of semantic context on speech intelligibility. *Brain and Language*, *122*(2), 103–113. https://doi.org/10.1016/j.bandl.2012.05.006

Zendel, B. R., & Alain, C. (2012). Musicians experience less age-related decline in central auditory processing. *Psychology and Aging*, *27*(2), 410–417. https://doi.org/10.1037/a0024816

Zendel, B. R., West, G. L., Belleville, S., & Peretz, I. (preprint). Music training improves the ability to understand speech-in-noise in older adults. *bioRxiv*, 196030.