EFFECTS OF MUSICAL TRAINING ON SPEECH UNDERSTANDING IN NOISE

By

Jeremy Federman

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Hearing & Speech Sciences

December, 2011

Nashville, Tennessee

Approved:

Dan Ashmead, PhD

Linda Hood, PhD

Benjamin Hornsby, PhD

Todd Ricketts, PhD

John Rieser, PhD

To my son Jacob who reminds me what is most important

and

To Jessica for her love, patience, and loyalty

and

To my parents for always being there.

ACKNOWLEDGMENTS

This work was supported by financial support from the United States Department of Education. Of course, this work would not be possible without the voluntary participation of some of the most amazing and talented musicians in the world. Very special thanks to all of them for their participation in this and previous studies.

I am especially appreciative to Drs. Steven Camarata and John Riser and for their friendship and support from the very beginning of my graduate studies and for their help in guiding me to the end of this project. I am thankful to Drs. Todd Ricketts and Benjamin Hornsby for their time and energy with this project and over the years in the Dan Maddox Hearing Aid Research Lab. Special thanks to Dr. Dan Ashmead for teaching me how to balance my work with life while dissertating, Dr. Wes Grantham for his skill and patience, and Dr. Linda Hood for her support and direction. The members of my Dissertation Committee have provided extensive guidance and, for that, I am forever indebted to them. Also, thanks to Nelson Cowan at University of Missouri for providing sound files used during dichotic listening testing.

I am grateful to all of my past and present peers, both PhD and AuD students, for their collaboration, inspiration, and friendship. Special thanks to Doug Sladen, Hollea Ryan, Heather Porter, Katarina Ntourou, Kathryn Guillot, Jason Galster, Steven Marcrum, and Alex Parbery-Clark. You helped me complete this project every day.

My family members have been stalwart supporters from the beginning of this and every other project. Were it not for my parents, I would never have made it this far. I love you more than I can express. Were it not for your love and gentle guidance, I would have lost my way long ago. Thanks to my sister Erica, her husband Scott, and their daughters Rubyrose and Coco for putting up with my extended absences. Most importantly, thanks to Jessica, Jacob, and Jane for being my family—now and forever.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Abbreviation	Definition
AMMA	Advanced Measures of Music Audiation
ANOVA	Analysis of variance
BBN	Broadband noise
CST	Connected Speech Test
CST 2T Avg	CST results using 2-talker babble averaged across eccentricity
CST BBN	CST results from 0°/0° with broadband noise
dB	decibel
EMAT	Colwell Elementary Music Achievement Test
HINT	Hearing in Noise Test
HINT 1H Avg	HINT results with 1-hemisphere noise averaged across eccentricity
HINT 2H	HINT results from 0°/0° with 2-hemisphere noise
IMT	Interleaved Melody Task
IQ	Intelligence quotient
KBIT-2	Kaufman Brief Intelligence Test – 2nd Edition
MBEA	Montreal Battery of Evaluation of Amusia
SBR	Signal-to-babble ratio
SIN	Speech-in-noise
SNR	Signal-to-noise ratio
SPL	Sound pressure level

CHAPTER I

INTRODUCTION

Effects of musical training

Professional musicians with normal hearing have demonstrated superior performance on a wide variety of psychoacoustic and electrophysiological tasks when compared to untrained listeners. For example, professional musicians have performed better, demonstrated shorter reaction times, and/or exhibited larger cortical amplitude responses than non-musicians on tasks of timbre perception (Chartrand & Belin, 2006; Pitt, 1994; Shahin, Roberts, Pantev, Trainor, & Ross, 2005; Zendel & Alain, 2008), pitch perception and frequency discrimination (Akin & Belgin, 2009; Besson, Schon, Moreno, Santos, & Magne, 2007; Nikjeh, Lister, & Frisch, 2008, 2009; Tervaniemi, Just, Koelsch, Widmann, & Schroger, 2005), contour and interval processing (Fujioka, Trainor, Ross, Kakigi, & Pantev, 2004; Hantz, Crummer, Wayman, Walton, & Frisina, 1992; Pantev et al., 2003; Tervaniemi, Castaneda, Knoll, & Uther, 2006), spatial ability (Douglas & Bilkey, 2007; Schellenberg, 2005; Sluming, Brooks, Howard, Downes, & Roberts, 2007), and vocabulary and verbal sequencing (Piro & Ortiz, 2009). Musical training also has been shown to result in both anatomical (Gaser & Schlaug, 2003a, 2003b) and functional auditory system changes (Gaab & Schlaug, 2003a, 2003b; Tervaniemi et al., 2009). However, what is currently unknown is whether musicians' enhanced abilities generalize to other important, non-musical scenarios such as understanding speech in noise.

Recent studies (Bidebnan & Krishnan, 2010; Parbery-Clark, Skoe, Lam, & Kraus, 2009; Song, Skoe, Banai, & Kraus, 2011) suggest that musicians may indeed exhibit better speech understanding in noise than non-musicians. In a study by Parbery-Clark, Skoe, Lam, and Krauss, normal hearing musicians and non-musicians completed tasks of speech-in-noise in which speech and noise were presented from either same or different spatial locations. Results showed that the musician group performed better on the speech-in-noise (SIN) tasks when the speech and noise were presented at 0 degrees azimuth, the condition in which there was no spatial separation advantage. This finding suggests that musicians are better at separating the speech signal from the noise when other potential useful cues such as spatial separation are unavailable. It is possible this advantage is not solely from the effects of musical training but may be due to preexisting enhanced working memory capacity, attention ability, or some other factor.

Moreover, it may be that those possessing such ability(ies) simply gravitate toward musical careers. Further, it is acknowledged that at least some individuals without musical training may demonstrate similar enhanced underlying abilities. This may result in speech recognition in noise performance that is better than for average listeners. Regardless, the speech-in-noise scenario utilized by Parbery-Clark et al. seems an ecologically valid example of auditory scene analysis not unlike the cocktail party effect described by Cherry (1953), which is known to be confounded by energetic and informational masking (Kidd, Mason, & Gallun, 2005; Schneider, Li, & Daneman, 2007). Specifically, the ability to segregate a single target, such as a talker, from a group of distracting signals (commonly referred to as auditory stream segregation or auditory object formation) requires cognitively organizing the combined signal reaching the ears accurately into discrete auditory objects.

Accurate segregation can be impaired by a competing masker in two ways. First, segregation can be impaired when the target talker signal is simply overwhelmed by the competing masker spectro-temporally due to both signal and masker falling within the same auditory filter, commonly known as energetic masking (Arbogast, Mason, & Kidd, 2002). Second, segregation can be affected when the competing masker places a cognitive load on attention and/or memory (Kidd, Arbogast, Mason, & Walsh, 2002), otherwise known as informational masking. In either case, accurate segregation of the combined signal into auditory objects occurs over time, and is known to occur differently in musicians than non-musicians (Beauvois & Meddis, 1997; Snyder & Alain, 2007). However, in the presence of masking sounds, the degree of benefit that might remain for individuals with musical expertise or the impact of differences in attention, working memory, and auditory stream segregation ability for such individuals remains relatively unknown.

Preliminary work by Federman and Ricketts (in preparation) investigated the effects of musical training, hearing loss, and audibility on performance of tests of music perception and cognition. In one experiment, 32 participants were tested behaviorally on the Montreal Battery of Evaluation of Amusia (MBEA) and the Advanced Measures of Music Audiation (AMMA) tests in low and high audibility conditions. Results showed that musical training may have some benefits related to stream segregation. Specifically, musical training appeared to mediate the effects of hearing loss such that participants with musical training and hearing loss performed similarly to participants with normal hearing and no musical training. (See Figure 1.)

The musicians from the Parbery-Clark et al. study (2009) demonstrated significantly better frequency discrimination ability and greater working memory capacity than non-musicians, which is consistent with prior studies (Akin & Belgin, 2009; Gaab & Schlaug, 2003a, 2003b; Micheyl, Delhommeau, Perrot, & Oxenham, 2006; Trainor, Shahin, & Roberts, 2003). Gaab and Schlaug (2003a) measured brain activation patterns in performance-matched musicians and non-musicians during a pitch memory task, and showed that non-musicians rely on cortical pitch discrimination areas, whereas musicians' brains prefer to recruit working memory and recall areas suggesting that musical training influences the neural networks used for such tasks.

These findings suggest that musicians' superior ability to discriminate one frequency from another coupled with a different cortical processing strategy may aid better and faster auditory object formation leaving potentially greater cognitive resource reserve left to be used for other processes even during non-musical tasks. It is possible that when experiencing a demanding listening situation such as understanding speech in noise, the differences in cortical processing strategies between musicians and non-musicians' brains generalize and lead to the observed superior speech understanding results. However, possibilities to explain musicians' enhanced speech recognition in noise performance also include greater attention, more efficient working memory processing (as opposed to capacity), advanced auditory stream segregation ability (i.e., automatic and/or attention-dependent buildup), or some combination of them all. The role of each will be discussed further below.

Attention

Attention is known to impact the selection of information we process and is also known to interact with working memory (Cowan et al., 2005; Engle, 2002; Soto, Hodsoll, Rotshtein, & Humphreys, 2008). There seem to be three kinds of attention (Klingberg, 2009). First is arousal, which can affect performance based on level, and is considered non-selective. Performance is best with moderate levels of arousal; whereas too little or too much arousal results in poorer performance. The second type of attention is selective and exogenous (i.e., involuntary), and called stimulus-driven attention because it is instantly and involuntarily drawn to unexpected, novel and salient environmentally occurring events. The third type of attention is endogenous (i.e., under conscious control) and called controlled attention. It is also selective, and is the type of most interest for the current study.

Stimulus-driven and controlled attention systems have been shown to be somewhat independent (Corbetta & Shulman, 2002). It is possible that musicians are better at certain tasks because they have schema (i.e. prototypes) in long-term memory due to musical training that allow for more efficient comparisons to incoming stimuli than their non-musical counterparts. It is known that detection of visual objects is made easier and quicker if features about the object are known in advance (Corbetta, Miezin, Dobmeyer, Shulman, & Petersen, 1990) or if spatial location is cued (Smith et al., 2009).

This phenomenon has been dubbed biased competition, and has been shown to be a factor in other sensory modalities. Biased competition is thought to direct appropriate selective attention onto specific stimuli among the sea of ongoing exogenous and endogenous stimuli, which may be relevant for auditory stream segregation and speech in noise tasks. Attention, therefore, would be expected to moderate working memory processing and be based on one's ability to control attention for the purpose of maintaining information over time in working memory (i.e., to attend or suppress information).

Musicians may also have a superior ability to maintain attention or ignore distracters than non-musicians. One way to ascertain the presence of this ability is to test an individual's ability to attend to information while ignoring other information (Engle, 2002). For example, Conway, Cowan, and Bunting (2001) used a dichotic listening task during which they simultaneously presented attended words in one ear and ignored words in the other ear. In the ignored ear, the participant's name was included in the word list. Participants in their study with better working memory capacities were better able to ignore distracting stimuli and only 20% of those reported hearing their names in the ignored ear. Conversely, participants with poorer working memory capacities were less able to ignore distracting information and 65% of them reported hearing their name. It is speculated that musicians will possess greater working memory and therefore be less susceptible to distracting information. It is also of interest to determine whether such performance generalizes to non-musical tasks.

Working Memory

Regarding working memory, over the last few decades behavioral models have evolved to a multiple component system of memory whose purpose is to maintain the information necessary to perform a complex cognitive task (Baddeley, 2010; Baddeley & Hitch, 1974; Repovs & Baddeley, 2006). Some portion of the incoming sensory information is attended to and input into a short-term storage called the episodic buffer (Baddeley, 2007), where it is known to decay quickly without rehearsal (Miyake & Shah, 1999). Unless encoded into a long-term storage mechanism through rehearsal, the information is lost from memory (Baddeley, 2010). It is worth noting the difference observed in the literature between definitions of short-term memory and working memory. According to Jarrold and Towse (2006), short-term memory is typically considered a simple storage mechanism, whereas working memory is considered functionally (i.e., a processing and storage workspace), and is considered to have subsystems by some researchers. For the purposes of this study, working memory will be defined simply as the part of the memory system that allows for actively keeping information available for a short period of time. Tasks of short-term memory typically have investigated storage capacity (i.e., recall), whereas working memory tasks investigate processing. Research has also shown a strong correlation between working memory capacity and both high level cognition and predicted intelligence (Daneman & Carpenter, 1980; Daneman & Merikle, 1996).

The explanation as to why some individuals perform better on working memory tasks (i.e., those who have superior intelligence or cognitive abilities) is fairly straightforward. Daneman and Carpenter (1980) suggested that working memory has both processing and storage functions, that there is a tradeoff between the capacities of these two functions, and that performance differences across individuals can be explained as differences in processing efficiency. That is, the more efficiently one can process information, the greater the portion of working memory left available for storing information. Subsequent experiments have provided additional support for this hypothesis (Conway et al., 2001; Daneman & Merikle, 1996; Kane & Engle, 2003; Kyllonen & Christal, 1990), and have also suggested that working memory is a general mechanism (i.e., modality independent) used in many types of cognitive tasks (e.g., language, math).

Daneman and Carpenter (1980) developed a reading span test that had research participants read aloud a series of sentences and try to keep a running list of the last word in each sentence. They hypothesized that the test would be sensitive to two aspects of working memory (capacity and processing efficiency) such that reading skill level would correlate with recall of the list of final words. Findings showed that less skilled readers use more working memory capacity or processing resources than more skilled readers, who could use more of their capacity to maintain the list since their processing was more efficient. Tirre and Pena (1992) used a test similar to Daneman and Capenter's reading span test method and also showed that recall performance got worse as memory load increased. It was predicted that the greater the number of sentence lists presented, recall accuracy for the final word of each sentence would decrease. Although recall accuracy worsened as predicted, the number of items recalled increased with the number of sentences presented suggesting that working memory is a dual mechanism that has a tradeoff between capacity and processing mediated by attentional demands. According to the authors, dividing attention between two tasks is expected to lead to reduced recall as the limit of working memory is approached.

Conway, Cowan, and Bunting (2001) demonstrated that subjects with high working memory capacity did better on a dichotic listening task than did subjects with low working memory. Performance was shown to be attention and working memory capacity dependent. Specifically, performance was dependent on the ability (or inability) to ignore distracting information in the non-test ear, focus on the information in the test ear, and maintain the test ear information in working memory. The high working memory capacity subjects, who were better able to ignore the information in the non-test ear, performed better on the task.

Similar results comparing high working memory and low working memory capacity were observed for a Stroop task (Kane & Engle, 2003) during which low working memory subjects erred twice as often as high-working memory subjects when 75% of the trials were congruent (i.e., the condition requiring the greatest ability to ignore the differences in word meaning and word ink color). In light of the tradeoff between processing and storage, it is hypothesized that musical training leads to a superior ability to ignore non-target stimuli, and attend to stimuli of interest. Such ability would lead, therefore, to better performance on tasks that stress working memory capacity and controlled attention.

Auditory stream segregation

Potential cues related to successful auditory stream segregation include fundamental frequency discrimination (Moore & Gockel, 2002), frequency resolution (Darwin, 1997), common onset time (Darwin, 1997), harmonics (Darwin, 1997), signal duration (Assmann & Paschall, 1998), temporal envelope (Moore & Gockel, 2002), and spatial separation of a target signal from a non-target signal. Although there has been work to isolate these cues and their effects on performance, the contribution and the strength of each relative to auditory stream segregation is difficult to determine in a system that relies on both bottom-up and top-down processes (Alain & Tremblay, 2007).

Auditory stream can be defined as a perceptual representation (as compared to the physical cause of the percept, [Bregman, 1990, p10]), that is, a "perceptual grouping of the neural parts" of a sound that go together (Bregman, 1990). More generally, it can be thought of as the ability to form and distinguish auditory objects from one another (e.g., a target voice from noise, a singer from the band, etc.). There are two categories of potential cues: bottom-up (primitive) and top-down (schema-based) cues. Bottom-up contributions to auditory stream segregation (also called primitive stream segregation) are considered automatic such as detection of a signal and discriminating it from other sounds present (Alain & Tremblay, 2007; Bregman, 1990). This presumably occurs by utilizing cues such as F0 discrimination, frequency resolution, and onset/offset time. Once detected and discriminated, the sounds arriving at the ear must be organized into auditory objects (horn honk, footstep) or streams (speech, music).

Once streaming occurs, top down processes such as attention and working memory appear to play a significant role. That is, the ability to segregate one sound from another depends on the sound to which we want to listen and our ability to stream that sound over time. Although there is evidence that working memory is recruited for complex sound processing in musicians to a greater degree than non-musicians (Bermudez, Lerch, Evans, & Zatorre, 2009; Chen, Penhune, & Zatorre, 2008; Gaab & Schlaug, 2003a, 2003b; Hantz et al., 1992; Zatorre, Perry, Beckett, Westbury, & Evans, 1998), it is less understood how working memory operates in relation to, or conjunction with, auditory stream segregation ability.

It is also less known how musical training impacts auditory stream segregation ability. However, a study by Beauvois and Meddis (1997) concluded that musicians were more sensitive to stream segregation because they were more likely to report segregation following trials with silences less than 4 seconds between induction and test sequences, and they were able to maintain segregation despite silences greater than 4 seconds between a biasing sequence and a test sequence. Non-musician participants in their study reported fewer segregations and, for trials with silences greater than 4 seconds, non-musicians performed near chance level (i.e., responses were evenly divided between segregated and coherent). These findings suggest potential for musical training to positively impact difficult musical and non-musical cocktail party-like listening scenarios that require auditory stream segregation such as watching movies or television which commonly contain dialogue and simultaneous musical soundtracks.

They also suggest potential for determining which cues may contribute to improving auditory stream segregation, and the relative contribution of each cue.

Although sound segregation has been shown to be affected by attention (Alain & Izenberg, 2003; Carlyon, 2004; Carlyon, Plack, Fantini, & Cusack, 2003), and our understanding of the role of attention is burgeoning, it is possible that musicians perform better due to greater cognitive resource availability. Specifically, it may be that, at least for musical signals, musicians have better long-term representations with which to make comparisons in working memory of incoming stimuli. The function of top-down cognitive processes has been named schema-based auditory stream segregation (Alain & Bernstein, 2008; Bey & McAdams, 2002, 2003; Bregman, 1990), and is thought to be dependent on a listener's attention and comparisons of incoming stimuli with previously experienced sounds.

For clarification, non-speech schema-based auditory stream segregation will be differentiated from speech schema-based auditory stream segregation regarding the current study. Generally, it is speculated that better developed schema due to training (a person's name, musical constructs) may lead to faster, more efficient processing. What is currently unknown is whether the potential advantages obtained from musical training generalize to speech, which arguably all normal hearing individuals have similar experience and expertise.

While few studies have directly examined the impact of musical training on auditory stream segregation ability, it is possible that musicians experience both fusion and fission more easily and more quickly when advantageous than non-musicians.

Fusion, also known as coherence can be defined as the maintenance of a single auditory object or perceptual stream. Alternatively, fission can be defined as the point at which a single stream segregates into two perceptual streams. Support for this notion can be found from studies that used cortical potentials and structural and functional imaging (Alain & Tremblay, 2007; Snyder & Alain, 2007; Snyder, Alain, & Picton, 2006; Zendel & Alain, 2008). Results from these studies showed, for example, that musicians have more grey matter dedicated to processing complex sounds resulting in superior frequency discrimination and resolution. If musicians have superior complex sound processing systems, their ability to segregate one sound from another may also be superior. In addition, musician reaction times on similar tasks have been shown to be faster (Tervaniemi et al., 2005), and the auditory stream segregation buildup time has been shown to be shorter than non-musicians (Beauvois & Meddis, 1997).

It has also been shown that, at least in the general population, a relationship exists between working memory capacity and auditory stream segregation ability (Conway et al., 2001; Dalton, Santangelo, & Spence, 2009; Engle, 2002). Specifically, auditory stream segregation requires cognitive resources. Therefore, it is reasonable to expect that individuals with larger working memory capacities, more efficient processing, or better attention/ignore capabilities will perform better on auditory stream segregation tasks. Since musicians as a group have been shown to have larger working memory capacity, it is reasonable to predict they will perform better. Second, it may be that the combination of working memory and auditory stream segregation underlies musicians' superior performance since the two processes seem to be closely related. Therefore, examining performance on tasks that require segregating sounds from one another (e.g., speech-in-noise [SIN], a target melody from distracter tones) as a means to investigate whether there are crossover benefits from musical training to non-musical tasks is important because such inquiry may provide information that informs us about how to investigate in future both the effects of hearing loss on cognition and the potential for musical training to mediate such effects. However, it is currently unclear what factors would explain the superior performance predicted for musicians during such tasks.

Spatial release from masking

When a speech target and masking noise(s) are spatially separated, a known improvement in speech understanding performance results as compared to when the target and masker are collocated (Arbogast et al., 2002; Arbogast, Mason, & Kidd, 2005; Freyman, Balakrishnan, & Helfer, 2001; Kidd, Arbogast, Mason, & Gallun, 2005; Kidd et al., 2002; Marrone, Mason, & Kidd, 2008). This resulting improvement from spatial separation of the target and masker has been labeled “spatial release from masking”, and its magnitude has been shown to depend on the type of masking.

For energetic maskers, the release has been shown to be less than for informational maskers (Arbogast et al., 2002, 2005). Arbogast, Mason, and Kidd (2002) investigated the effect of spatial separation on both energetic and informational maskers using a speech-in-noise task to determine if the type of masker affected performance on a spatial separation task. They presented a target signal from 0° and a masker from either 0° or 90° azimuth and found that the release from masking, when the target and masker were pre-filtered to minimize spectral overlap, was 7 dB for an energetic masker and 18 dB for an informational masker. These results therefore showed a larger release from informational versus energetic masking suggesting that informational masking increases the difficulty of the listening task to a greater degree than energetic masking.

Investigations examining the effects of these two masker types have suggested that energetic masking is primarily occurring in the peripheral auditory pathway and is due to effects on the neuronal response (Freyman et al., 2001; Freyman, Helfer, McCall, & Clifton, 1999), whereas informational masking is primarily occurring at higher levels of the system and is due to target-masker similarity effects on segregation of auditory objects (both target and masker are audible but are similar sounding) and attention (Arbogast et al., 2002; Ihiefeld & Shinn-Cunningham, 2008; Kidd, Arbogast, et al., 2005). Despite the previous investigations into the effects of spatial separation of target and masker(s) on speech understanding, however, it is currently unknown whether musical training would result in an advantage at eccentricities other than 0°/0° and less than 0°/90° spatial separation. In addition, it is unknown whether performance as spatial separation increases changes differently than for those without training.

If the musical training advantage observed at 0°/0° by Parbery-Clark et al. (2009) is real, the inclusion of experimental conditions that incrementally separate the target from the masker(s) may illuminate more clearly the effect of musical training on the relationship between auditory stream segregation and spatial separation cues (i.e., head shadow and binaural interaction). In addition, it is important to determine if any existing advantage is limited by the type of masker. Therefore, the inclusion of separate conditions employing maskers that are either primarily energetic or primarily informational would be particularly informative.

As observed by Arbogast, Mason, and Kidd (2002), the predicted result for an informational masker would be a larger release from masking than for an energetic masker as spatial separation of target and masker increases. However, due to target and masker similarity, informational masking conditions would be expected to increase the difficulty of the task and stress the auditory stream segregation abilities of listeners differently than for energetic masking. If musical training results in a superior ability to segregate a target from a masker, participants with musical training would be expected to perform better in these conditions. Therefore, it is hypothesized that any advantage due to musical training will increase as the masker and target similarity is increased. By including multiple eccentricities, it will also be possible to quantify the effect of the spatial release from masking across the two groups for both masker types.

As previously stated, it was predicted that musical training or general preexisting abilities would result in better performance particularly in the more difficult listening conditions (i.e., 0°/0°), and that certain advantages from musical training or preexisting abilities may also be revealed in tests of working memory and attention. If so, the resulting behavioral data should not only reveal under which conditions any advantage exists, if present, but will better quantify the magnitude of any advantage. Using a statistical multiple regression approach, predictor variables can then be identified for the conditions where the largest advantage(s) was/were observed, and subsequently be used to predict performance of individuals based on those available predictor variables. The variables that may predict performance under investigation in this study include working memory, attention, and non-speech schema-based auditory stream segregation.

Summary

For the current study, primary aims included investigating the effects of musical training on attention, working memory, and auditory stream segregation as they relate to speech understanding in noise. Specifically, it was of interest to determine if there were performance differences on speech-in-noise and/or auditory stream segregation tasks, attention, and working memory abilities between those with and without musical training. In addition, it was of interest to determine if factors such as musical training, attention, non-speech based stream segregation and/or working memory could be used to predict speech recognition in noise differences across individuals. In order to examine these questions, an evaluation of performance was conducted for musicians and non-musician using tests of attention (dichotic listening task), working memory (automated operation span task), and auditory stream segregation (i.e., melodic schema-based task, music achievement test, speech-in-noise). By assessing participant performance on specific aspects of attention, working memory, and auditory stream segregation, critical information about the impact of musical training on these factors was gathered. In addition, a greater understanding about the differences between those with and without musical training and the role of attention, working memory, and auditory stream segregation was sought.

Summary of hypotheses

- Based on the work of Bregman (1990), Alain and Bernstein (2008), and Bey and McAdams (2002, 2003), it is predicted that musicians will perform better than non-musicians on a task of schema-based auditory stream segregation (i.e., melodic schema-based task). This predicted performance advantage is due to the fact that musicians are expected to have greater working memory capacity as indicated by higher scores on a test of working memory (consistent with Parbery-Clark et al., (2009)) as well as greater selective attention as demonstrated by fewer errors on a dichotic listening task (consistent with Conway, Cowan, and Bunting (2001) and Kane and Engle (2003). These effects employ cognitive rather than peripheral auditory processes (i.e., top down not bottom up).
- Based on the findings of Parbery-Clark et al. (2009) regarding speech and noise presented from the same spatial location, and findings by Beauvois and Meddis (1997) showing increased sensitivity to auditory stream segregation, it is hypothesized that musicians will outperform non-musicians as evidenced by lower required SNRs for 50% recognition on the HINT speech-in-noise test, and higher percent correct scores on the Connected Speech Test.

CHAPTER II

SPEECH UNDERSTANDING IN NOISE

Methods

Two groups differentiated by musical training status were assessed on tasks of non-speech schema-based auditory stream segregation (EMAT, interleaved melody task [IMT]), working memory capacity (automated operation span [AO Span]), attention (dichotic listening task [DAT]), and speech-based schema-based auditory stream segregation (HINT, CST). In order to minimize potential confounds of age, IQ, education level, and hearing ability, the first session included a hearing screening, an IQ screening (KBIT-2), and a demographic questionnaire that included age, education level, and hearing health history information. This information was used to verify candidacy and match these factors as closely as possible between the musician and non-musician groups. This first session lasted approximately 60 minutes.

Once qualified, each participant was asked to participate in up to three additional test sessions lasting a total of no more than 6 hours. Subsequent sessions included behavioral testing on specific measures of perception and cognition (i.e., as listed above and described further below). Except for the dichotic attention task (DAT) stimuli, which were presented over insert earphones (Etymotic ER-3), all stimuli were presented to participants seated in the center of a double-walled sound treated room (4 x 4 x 2.7m) from a distance of 1 meter via loudspeaker(s) (Tannoy Precision 6P). Except for IMT stimuli, which were generated and presented using MATLAB software, all stimuli were digitally stored (16 bit, 44.1 kHz sampling rate) on a desktop computer (Pentium PC), and subsequently presented using Adobe Audition software (v. 1.5). Except for the operation span stimuli, which were visually presented on a computer monitor, all stimuli were output from the computer via a soundcard (Echo Layla 3G) that converted the digital signals to analog and output them either to an audiometer (Madsen Orbiter 922) or to a crown power amp to be sent to the loudspeakers. All procedures were reviewed and approved by the Vanderbilt University Institutional Review Board.

Data were preliminarily analyzed using analysis of variance (ANOVA). An individual ANOVA was conducted on each screening factor (age, IQ, and education level) in order to examine how well they had been controlled for between the groups. A correlation analysis of the interleaved melody task (IMT) "adaptive" and "block" data results was planned in order to examine the accuracy of the adaptive results and their potential utility for subsequent between-groups analyses. A repeated measures ANOVA of the results from each speech-in-noise test was planned to examine the between groups effect of musical training status and the within groups effect of noise masker azimuth in degrees (i.e., eccentricity). For each speech-in-noise test set of conditions, if performance was found to be stable across eccentricity, results were averaged for any subsequent analyses. If not, using the results from the most difficult listening condition (0°/0°) was planned. This condition was chosen to allow comparisons with results from a previous study that employed a similar condition (Parbery-Clark et al., 2009).

Following preliminary data analyses and preparation, an overall ANOVA was completed. For this ANOVA, there was one between-subjects factor (musical training status) and eight dependent variables (schema-based auditory stream segregation [EMAT, IMT], working memory capacity (AO Span), attention (DAT), and auditory stream segregation speech recognition [HINT 1H, HINT 2H, CST 2T, CST BBN]). The main goal of this analysis was to determine if there were any between group differences among the test measures. In addition, a correlation analysis was conducted to examine the relationships among the eight dependent variables to specifically examine the relationships among the speech-in-noise tests and between the speech-in-noise tests and the other dependent variables.

Multiple linear regression analysis was also conducted to assess how well the factors (e.g., working memory, auditory stream segregation, attention) predicted speech recognition in noise performance.

Participants

A power analysis using data from Parbery-Clark et al. (2009) suggested that sample size of 8-10 total participants were needed to reach a statistical power level of 0.8 on a test of speech understanding in noise. However, for this study, a total of 32 participants aged 18 to 65 with and without musical training was proposed to offset the loss of statistical power due to the number of tests being conducted. In addition, the greater number of subjects is typically necessary to draw any meaningful conclusions using multiple linear regression techniques. For example, assuming three predictor variables, a desired power level of 0.8, an alpha of 0.05 and a large anticipated effect size, a sample population of 36 would be theoretically required. Participants were divided into two groups based on their degree of musical training. Specifically, the groups were participants without musical training and normal hearing (n = 15), and participants with musical training and normal hearing (n = 17). Participants with musical training had at least 10 years of formal training or equivalent experience. Participants were excluded if hearing loss (pure tone thresholds >20 dB HL from 250 to 8000 Hz) was present.

Materials & Procedures

To examine the role of intelligence, the Kaufman Brief Intelligence Test – 2nd Edition (KBIT-2) (Kaufman & Kaufman, 2004) was administered to each participant. The KBIT-2 consists of three subtests; two measure verbal (i.e., crystallized) intelligence, one measures non-verbal (i.e., fluid) intelligence. Verbal questions require only one-word responses and non-verbal (visual) items require pointing to select a choice. The first of two verbal subtests measures receptive vocabulary and general information about the world by either asking the participant to point to a picture that represents the meaning of a word or is the answer to a question. The second subtest measures verbal intelligence (e.g., verbal comprehension, reasoning, and verbal knowledge) without requiring reading by having the participant answer a riddle with a single word. The non-verbal portion of the test measures non-verbal reasoning, cognitive flexibility, and problem solving ability using pictures and abstract designs that follow a pattern. The participant is asked to select a picture from several options that would complete the pattern contained in each test picture/design. The verbal and non-verbal sections of the test each result in a score and are combined to calculate a composite intelligence quotient (IQ) score. Although not a comprehensive intelligence test, advantages of the KBIT-2 include the ability to quickly assess verbal intelligence without requiring reading or spelling, the inclusion of a measure of non-verbal (i.e., fluid) intelligence, and good reliability (i.e., test-retest) for adults aged 19 to 90 for the verbal (M = 0.91), non-verbal (M = 0.83) subtests and composite IQ score (M = 0.90), as well as strong validity when correlated with other intelligence tests (e.g., Wechsler Intelligence Scale for Children = 3rd and 4th Ed., and the Wechsler Adult Intelligence Scale - 3rd Ed.). In addition, the KBIT-2 scores are standardized for age and can be normalized to be an IQ score with a mean of 100 and a standard deviation of 15.

An interleaved melody task (IMT) (Bey & McAdams, 2002, 2003) was administered as a primary measure of schema-based auditory stream segregation. The purpose of this test was specifically to assess if participants could determine whether two melodies were the same or different when one melody was interleaved with distracter tones. For this study, only one condition from the original Bey and McAdams experiments (named "Before" since the target melody was played before the interleaved melody) was included. For each trial in this condition, a six-note pure tone melody was presented followed by a second six-note melody that was interleaved with a series of distracter tones. There are five distracter sequences that were presented in the test condition, and were chosen randomly by trial. The distracter sequences were either in the same frequency range as the melody, or transposed to lower mean frequencies (mean difference of 1, 2, 3, 4, 6, 8, 12 or 24 semitones). For each of these nine frequency separation conditions, 24 trials were completed three times for each condition after the conditions were randomized. Participants decided if the two melodies were the same or different. In addition, an adaptive version of the task was completed. Since previous work had only completed the IMT using time-intensive block trials, we created an adaptive version of the task to determine if an adaptive task would provide comparable results in a more time efficient manner. As such, both test methods were included. Discrimination thresholds were measured using a three-down, one-up adaptive paradigm which tracked the 71% correct point on the psychometric function (Levitt, 1971). That is, three correct responses were required before the task became more difficult, while the task became easier after one incorrect response. All signals were presented through a loudspeaker at an azimuth of 0 degrees, at a distance of 1 m, and at a level of 75 dB SPL in a sound-treated room. For half of the trials, the melodies were identical (same trials); for the other half, two notes were altered (different trials). Prior to the test conditions, a series of familiarization trials occurred with feedback providing the correct response to ascertain the ability to complete the experimental task.

The Colwell Elementary Music Achievement Test 3, Part 2 (EMAT) was administered as a second measure of schema-based auditory stream segregation. This test was included in addition to the interleaved melody task in order to assess a musical, non-speech schema-based auditory stream segregation ability. The purpose of the EMAT Test 3, Part 2 is to measure the ability to recognize in which part a melody is played within a three-part harmony (Colwell, 1967). Successful completion of the task requires participants to accurately segregate the target melody from the two other harmonies. Participants listened through a loudspeaker at a distance of 1m to short recorded melodies played on piano followed by the same melody with two added harmony parts presented at 75 dB SPL since music is often listened to at levels higher than average speech. Following presentation of each melody-harmony trial, participants indicated in which part they perceived the original melody was played (high, medium, or low). Reliability for college aged students on this test has been reported to be 0.94 (Marchand, 1975).

Attention was assessed using a dichotic listening task (i.e., selective attention task) that required the simultaneous ability to attend and ignore, and to recall (Conway et al., 2001). During this task, 300 one and two syllable target (attended) words spoken by a female were presented to the right ear via ER-3 insert earphones at a presentation level of 65 dB SPL.[1] Thirty seconds post onset of the target word list, a second, 300 distracter word list (ignored) spoken by a male was presented to the left ear. Word onsets for both lists were synchronized. The task of the participant was to listen to and repeat each target word following its presentation, to ignore information presented to the other ear, and to make as few errors as possible. Errors were recorded by the experimenter during testing. Following completion of the test, participants were asked about any attention shifts during the test and the cause of those shifts. Specifically, they

[1] The presentation level used during the Conway et al. study (2001) was not reported except to say “presented at a constant volume for all subjects”.

were asked whether they recalled any information from the distracter ear, and if so, they were asked to report what they recalled (e.g., male voice, specific words).

In order to assess their working memory capacity, participants completed an operation span task (Turner & Engle, 1989). During the operation span task (i.e., divided attention task), participants were presented with an arithmetic equation containing two operations followed by an unrelated letter on a computer screen (e.g., (6+4)/2 = 5 ? H). Approximately half of the equations were correct and half were incorrect. The participant's task was to read the equation aloud, to decide if the equation was mathematically correct or incorrect (correct in this example), and to remember the unrelated letter. Sets of equation-letter pairs that increase from 2 to 6 with three trials for each set size were presented. At the end of each set, the participant was asked to recall as many of the presented letters as possible. Each participant's score was calculated as the total number of letters recalled correctly in serial order from each series. No points are counted from a series from which the letters are recalled imperfectly.

Other complex span tasks such as the complex reading span task (Daneman & Carpenter, 1980) have been demonstrated to be an accurate measure of working memory and shown to be related to individual variability in speech understanding (Akeroyd, 2008), but were deemed redundant for the purposes of this study. Specifically, by varying the difficulty of the background task used during measurement of complex span, Turner and Engle (1989) have shown that complex span tasks that do not require reading (i.e., complex operation span) accurately measure verbal working memory (but not spatial working memory) because they prevent memory strategies such as rehearsal and grouping. Since this study was interested in the effects of musical training and hearing loss on working memory capacity, neither of which are considered spatial working memory, the reading span task was not expected to provide information that is directly salient to the specific question of interest.

In order to measure speech recognition-based auditory stream segregation, participants completed the Hearing In Noise Test (HINT) (Nilsson, Soli, & Sullivan, 1994) and modified versions of the Connected Speech Test (CST) (Cox, Alexander, & Gilmore, 1987; Cox, Alexander, Gilmore, & Pusakulich, 1988, 1989). Two separate tests (HINT and CST) were used to investigate the effects of both energetic and informational maskers since each is thought to operate differently on the auditory system's function (Arbogast et al., 2002, 2005; Brungart, 2001; Freyman et al., 2001; Kidd, Arbogast, et al., 2005).

Since steady state noise is thought to primarily affect the audibility/detection of a target signal at the level of the auditory periphery by overwhelming the neural response to the target signal (Freyman et al., 2001; Freyman et al., 1999; Kidd, Arbogast, et al., 2005), the HINT conditions were expected to provide insight into whether energetic masking impacts performance differently for those with and without musical training. For this study, participants completed the HINT in conditions that presented a target talker from 0° azimuth and an accompanying speech-shaped noise presented from 0°, +10°, +22.5°, and +90° azimuth as well as from both hemispheres at ±10°, ±22.5°, and ±90° azimuth. The 0°/0° condition is a difficult listening condition since no spatial separation or release from masking cues are available to aid segregation of the target from the noise, and this condition was predicted to reveal a significant difference between groups due to the advantage(s) musical training was hypothesized to provide.

The additional eccentricities for the noise were evaluated in order to determine if there was an observable effect on any musical training advantage (Parbery-Clark et al., 2009) as a spatial cue is introduced. In order to isolate the spatial separation cue from a SNR cue, test conditions as described above were included with maskers presented symmetrically (i.e., in both hemispheres) to limit differences in SNR at each ear. The HINT is comprised of 25 phonemically balanced lists of ten sentences each and is conducted with a simultaneous spectrally matched broadband noise (i.e., energetic masker), and uses an adaptive method to assess speech recognition in noise. The noise presentation level was fixed at 65 dB SPL. The speech presentation level was adjusted until the threshold signal-to-noise ratio (SNR) was obtained at which 50% of the sentences were repeated correctly. Stimuli for the HINT were taken from Bench and Bamford (1979), and modified by Nilsson et al. (1994) to equate sentence difficulty, eliminate British idioms, and to obtain uniform sentence lengths.

Additional conditions using modified versions of the connected speech test (CST) were also completed. The purpose of these conditions was to assess the impact of musical training on the spatial release of masking when the masker contains an informational masking component (i.e., 2-talker babble). That is, informational masking is thought to interfere with higher order cognitive function (Arbogast et al., 2002; Ihiefeld & Shinn-Cunningham, 2008; Kidd, Arbogast, et al., 2005) while energetic masking simply reduces the audibility and subsequent neural response of the target signal (Freyman et al., 2001; Freyman et al., 1999). Specifically, although spatial separation of the speech and masker has been shown to result in a larger release from an informational masker than from an energetic one, it is the target-masker similarity that is predicted to affect the difficulty of the task and stress the auditory stream segregation and/or attention abilities of listeners.

The CST includes 24 passage pairs of 10 sentences on a topic. Each topic contains 25 key words used for scoring which were selected based on difficulty regarding level of intelligibility of each word. Each passage contains 5 words in 5 categories of intelligibility. Scoring is based on the number of correct key words repeated by the listener. The CST noise is 6-talker speech babble. However, it has been shown that the original CST noise results in similar performance to a spectrally matched broadband noise such that the performance intensity functions of the two masker types were both shown to be within 2 rau per dB (R. L. Sherbecoe & Studebaker, 2002; Robert L. Sherbecoe & Studebaker, 2003).

Typically, the CST is completed with both the target and masking noise collocated. Since the objective was to assess the addition of an informational masking component (i.e., 2 distracter talkers) and the effect of spatial separation of target and masker, the CST was tested with a) broadband noise spectrally matched to the original CST 6-talker babble, and (b) 2-talker, same gender babble since this combination has been shown to result in the most effective informational masking and because signal-to-babble ratio (SBR) for an informational masker has been shown to have a small effect on intelligibility (Brungart, 2001). In order to maintain similarity with the HINT test conditions, a spectrally matched broadband noise was used in place of the original CST 6-talker babble, the noise level was fixed at 65 dB, and the maskers were symmetrical but uncorrelated.

Additional conditions using 2-talker babble were also included. The 2-talker babble was chosen to increase similarity of the target and masker, and was made up of two individual female talkers reciting CST sentences not used for testing. SBR settings were fixed at +2 dB for the CST testing. Pilot testing suggested that this particular SBR would likely result in performance within the 20 – 80% correct range for both groups. It was initially thought that a reduction in SBR would be required to avoid ceiling effects since performance was expected to increase dramatically from an energetic masker to the 2-talker babble. However, because silences had been removed from the babble maskers thereby reducing the dips within which to listen (Lorenzi, Gilbert, Carn, Garnier, & Moore, 2006; Peters, Moore, & Baer, 1998), it was found that a +2dB SBR was required to avoid both floor and ceiling effects for both experimental groups in all conditions.

The relative difficulty of the CST for listeners with normal hearing has been shown not to be affected by signal-to-babble ratio (SBR). That is, performance on the CST has been shown to change ~12% per 1 dB per change in SBR when scores are between 20% and 80% (Cox et al., 1987). As in the proposed HINT conditions, four eccentricity conditions were included for both masker types where the target was presented from 0° azimuth and accompanying noise presented from 0°, ±10°, ±22.5°, and ±90° azimuth. Maskers were presented symmetrically to limit differences in SNR at each ear for the same reasons outlined previously regarding the HINT.

The primary difference from the HINT conditions, aside from being a fixed SNR task, was that conditions using same-gender 2-talker babble (informational masking) were included with the CST. Participants listened to the combined speech and noise for both the HINT and the CST at a distance of 1m, and repeated each sentence after it was presented. All testing described above was completed in a sound treated room as previously described.

CHAPTER III

RESULTS

The objective of this study was to investigate the effect of musical training on auditory stream segregation, working memory, and attention as they relate to speech understanding in noise. In addition, it was of interest to determine whether any of these factors had predictive value for performance on the speech in noise tasks. While speech recognition was evaluated in several conditions, it was of interest to identify whether there was an interaction between speech recognition test condition and between group differences revealed on other tests. Specifically, it was the goal of this work to initially identify speech recognition conditions that led to the largest between group differences in order to maximize the likelihood that any significant predictive factors could be identified. Consequently, the following initial preliminary analyses were completed.

Data Preparation and Reduction

Planned preliminary statistical analyses including analyses of variance (ANOVA) and correlation analysis were conducted to examine the data within several of the tests to appropriately simplify subsequent analyses. ANOVAs were conducted on each screening factor (age, IQ, and education level) to determine how well these factors had been controlled between groups. A correlation analysis of the block trial and adaptive trial IMT was completed to determine whether results from the adaptive version of the task were comparable to the block trial version. A repeated measures ANOVA of the of the speech-in-noise tasks results was then completed to examine the effect of noise eccentricity on speech-in-noise performance.

Preliminary ANOVA results examining education level in years showed that non-musician study participants had significantly more education than the musician participants ($F_{1, 30}$ = 8.752, p = 0.006, η^2 = 0.226). However, the difference in years between the two groups was 2.11 (18.71 vs. 16.60 years, SD = 2.3 and 1.6, respectively for non-musicians and musicians), and all participants had at least two years of bachelors level education. ANOVA results of participants' IQs (M = 118.588 and 121.133 for non-musicians and musicians, respectively) and ages (M = 29.00 and 30.47 for non-musicians and musicians, respectively) did not reveal any significant differences between musicians and non-musicians. Taken together, these results suggest the groups were closely matched on these factors, and that it was not necessary to include them in later analyses.

Correlation coefficients were calculated for adaptive and block trial results for the interleaved melody task (IMT). Initial correlation analysis of the interleaved melody task (IMT) results revealed a strong correlation between performance on an adaptive version of the task and the 90% y-intercept value calculated from block trial data, $r(32)$ = 0.64, p < 0.001. In other words, it appears that these two test methods provided very similar information about participants' ability to correctly identify a target melody presented with an interleaved distracter. The 90% point was calculated for each participant using results from nine test conditions that employed distracters ranging from 0 to 24 semitones lower in average frequency than the target melody. Since the correlation coefficient was large, the adaptive threshold results were used to conduct subsequent analysis.

Once it was determined that age, education level, and IQ were similar across groups, and that the results from the adaptive version of the IMT were valid, a repeated measures ANOVA was completed for each speech-in-noise test. The purpose of these ANOVAs was to determine whether changes in the eccentricity of the noise had an impact on between group differences, and if so, how did performance change as eccentricity of the noise increased. For these initial ANOVAs, significance level was set at $p < 0.05$. Specifically, the CST noise types (BBN, 2-talker babble) and HINT 1-hemisphere noise and 2-hemisphere noise test results were analyzed using repeated measures analyses of variance (ANOVA) in order to examine the between groups effect of musical training status and the within groups effect of noise masker azimuth in degrees (i.e., eccentricity). Albeit slightly unusual, the within groups results are discussed here because they relate to data reduction and preparation for additional analyses. Between groups results are discussed in the overall analysis section below. There were no significant interactions for any of the speech-in-noise tests.

Figure 1 shows the results for the CST using broadband noise and 2-talker babble. Within-groups results for 2-talker babble showed a main effect of eccentricity ($F_{3, 90} = 310.08$, $p < 0.001$, $\eta^2 = 0.912$) for which follow-up testing revealed that all eccentricities were different from one another. There were no significant interactions. Results for the CST using broadband masking noise also showed a within-groups main effect of eccentricity ($F_{3, 90} = 23.206$, $p < 0.001$, $\eta^2 = 0.436$) for which follow-up testing revealed that conditions 0°/0° and 0°/90° were different from all other eccentricities and that 0°/10° and 0°/22.5° were not different from each other. There were no significant interactions.

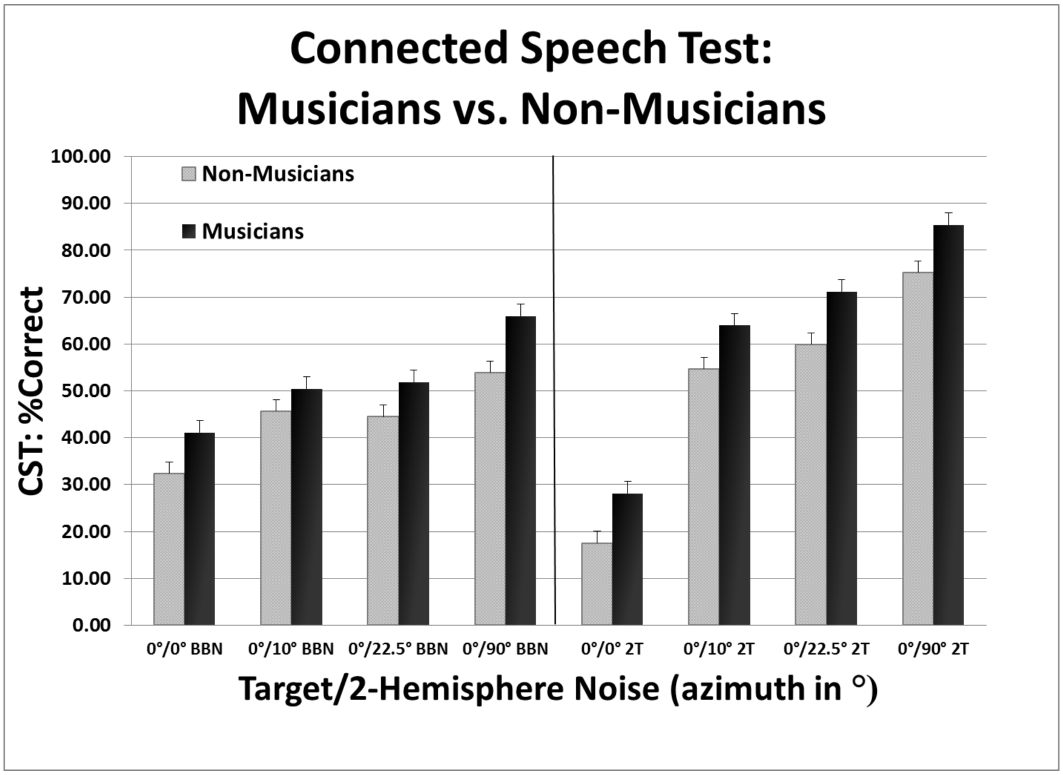


Figure 1. Musicians' and non-musicians' percent correct scores for connected speech test (CST) stimuli presented in a background of speech-shaped broadband noise (BBN) or two talker babble (2T) by azimuth in degrees. Error bars represent one standard error.

The two sets of CST noise conditions were highly correlated with a large correlation coefficient, $r(32) = 0.618$, $p < 0.001$. For the CST 2-talker babble conditions, results showed a stable performance advantage for musicians. The magnitude of this advantage was similar across eccentricities. For these reasons, results from the CST 2-talker babble were averaged across eccentricities and this average was used for the subsequent ANOVA and regression analyses ("CST 2T Avg"). For the CST broadband noise conditions, the most difficult listening eccentricity (target and masker at 0° azimuth) was used for subsequent analyses in order to eliminate any potential confound from the previously described lack of difference between 0°/10° and 0°/22.5° using this noise type ("CST BBN"). However, it should be noted that averaging results across eccentricities for the CST BBN conditions resulted in similar regression analysis results (see below).

Figure 2 shows the results from the Hearing in Noise Test. Initial HINT ANOVA results using 1-hemisphere masking noise showed a within-groups main effect of eccentricity ($F_{3, 90} = 221.85$, $p < 0.001$, $\eta^2 = 0.881$) for which follow-up testing revealed that all eccentricities were different from one another. For 2-hemisphere noise, a within-groups main effect of eccentricity was observed ($F_{3, 90} = 39.147$, $p < 0.001$, $\eta^2 = 0.566$) similar to the CST BBN conditions. Specifically, conditions 0°/0° and 0°/90° were different from all other eccentricities, and 0°/10° and 0°/22.5° were not different from each other. There were no significant interactions for HINT test results. Therefore, as with the CST 2-talker babble, the HINT 1-hemisphere noise results were averaged across eccentricity ("HINT 1H Avg"), and the results from the HINT 2-hemisphere 0°/0° condition were used for subsequent analyses ("HINT 2H").

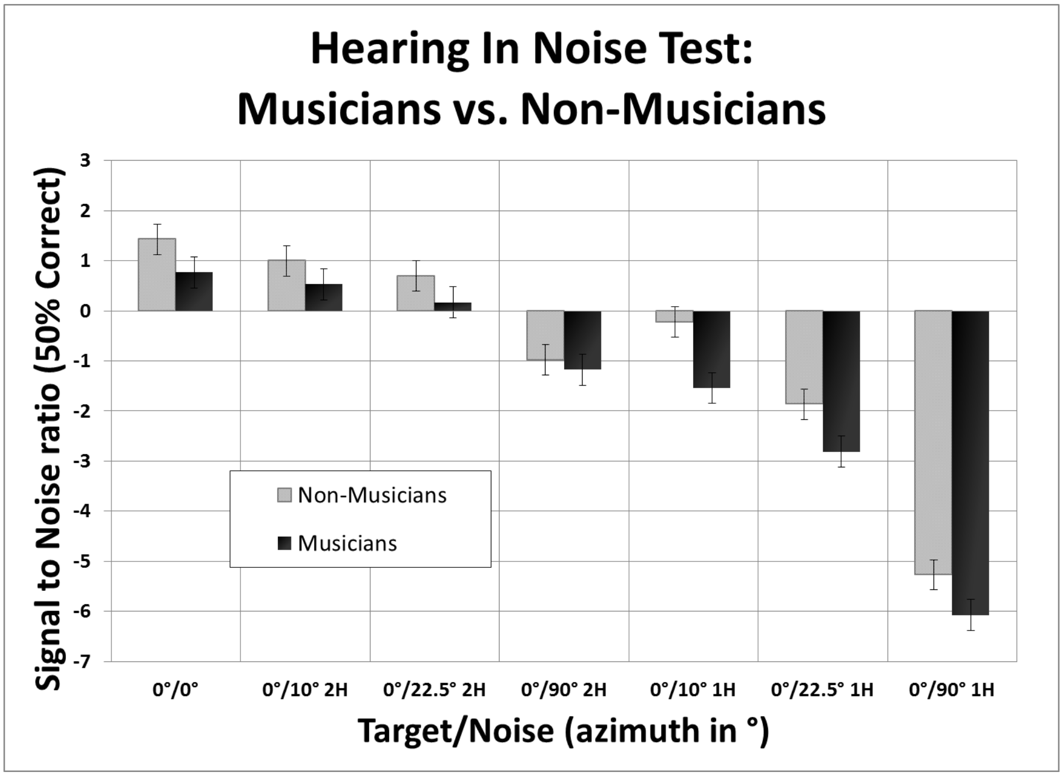


Figure 2. Hearing in Noise Test (HINT) performance. Musicians' and non-musicians' average sentence recognition 50% correct scores on the Hearing in Noise Test in speech-shaped broadband noise (BBN) by azimuth in degrees. Conditions with a '1H' designation represent performance with right hemisphere noise only. All other conditions included noise in both hemispheres (2H) except for 0°/0°. Error bars represent one standard error.

Overall Analysis

Results from the "pre-analyses" were used for additional statistical analyses discussed here. Specifically, an overall ANOVA was completed with one between-subjects factor (musical training status) and eight dependent variables (IMT, EMAT, DAT, AO Span, CST 2T Avg, CST BBN, HINT 1H Avg, HINT 2H), and 2-tailed correlations among the variables were calculated.

As detailed in the following, significant between-groups differences were found for the music-based auditory stream segregation tasks (IMT, EMAT), attention (DAT), and speech-based auditory stream segregation tasks (CST 2T Avg, CST BBN, HINT 1H

Avg, and HINT 2H). No significant differences between musician's and non-musicians were found for the AO Span. Results for the adaptive version of the IMT revealed a significant between-groups effect ($F_{1, 30} = 45.107$, $p < 0.001$, $\eta^2 = 0.601$) indicating that individuals with musical training were able to identify a target melody correctly in the presence of a distracter melody that was closer in average frequency (See Figure 3). That is, non-musicians, on average, needed distractor melodies to be 9.49 semitones lower in average frequency to accurately identify the target melodies as same or different. By comparison, to accurately identify the target melodies, on average, the musicians required the distracter melodies to be only 1.78 semitones lower in average frequency.

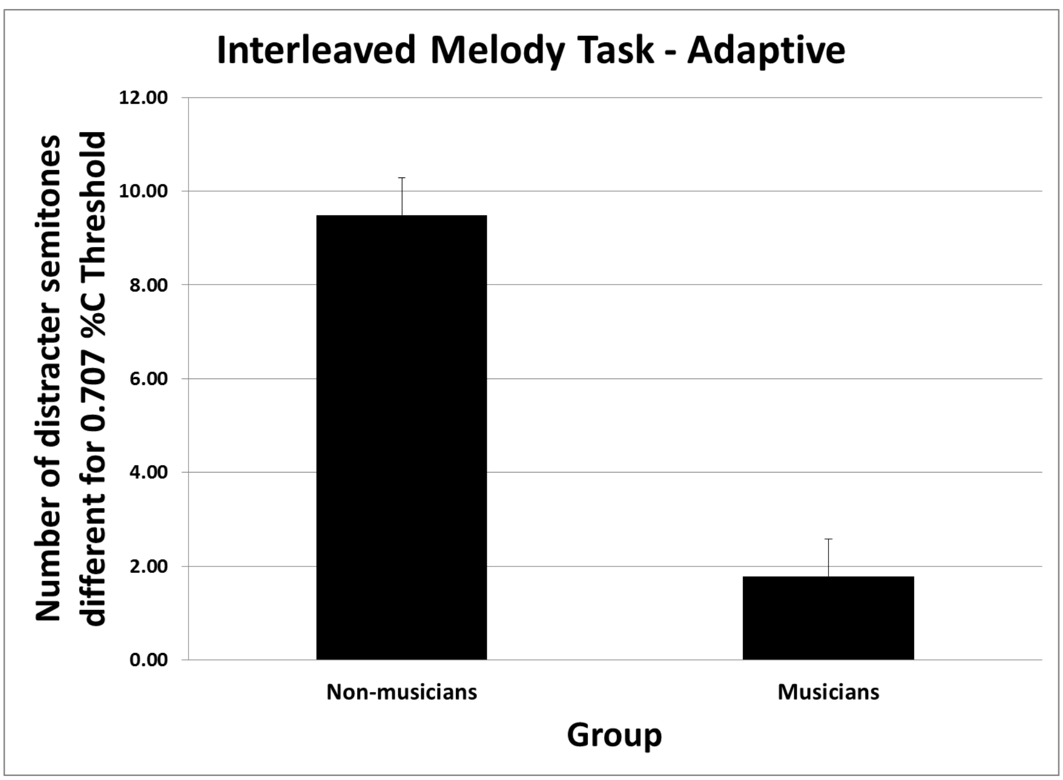


Figure 3. Performance on the adaptive version of the Interleaved Melody Task by group. Graph shows the number of distracter semitones different from the target average frequency necessary for 0.707% correct performance for musicians' and non-musicians'. Specifically, the smaller the values, the closer in average frequency were the target and distracter tones. Error bars represent one standard error.

Results from ANOVA testing for the EMAT showed a significant between-groups effect ($F_{1, 30} = 66.677$, $p < 0.001$, $\eta^2 = 0.690$) indicating individuals with musical training were better able to identify a target melody within a 3-part harmony (See Figure 4).

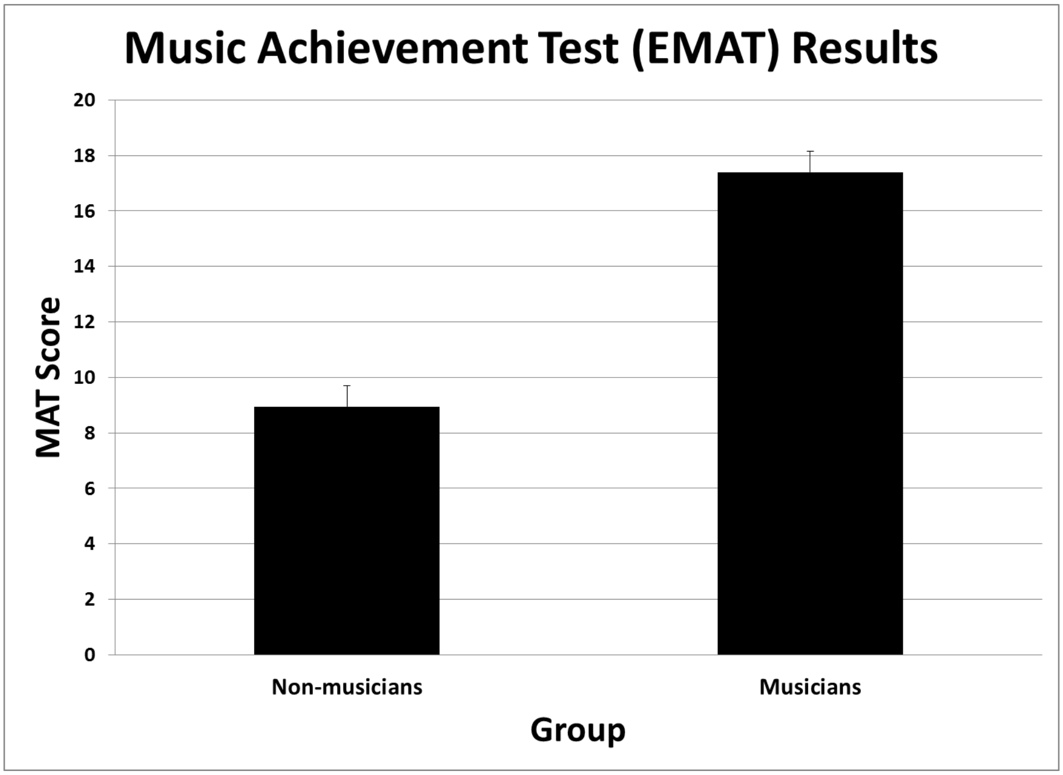


Figure 4. Results from the Music Achievement Test (EMAT). Participants were asked to indicate which melody of a 3 part harmony was the target played previously on piano. Error bars represent one standard error.

ANOVA results for the DAT showed a significant between-groups effect ($F_{1, 30} = 10.197$, $p = 0.003$, $\eta^2 = 0.254$) indicating musicians were better able to attend to words presented to the right ear while simultaneously ignoring other words presented to the left ear (See Figure 5).

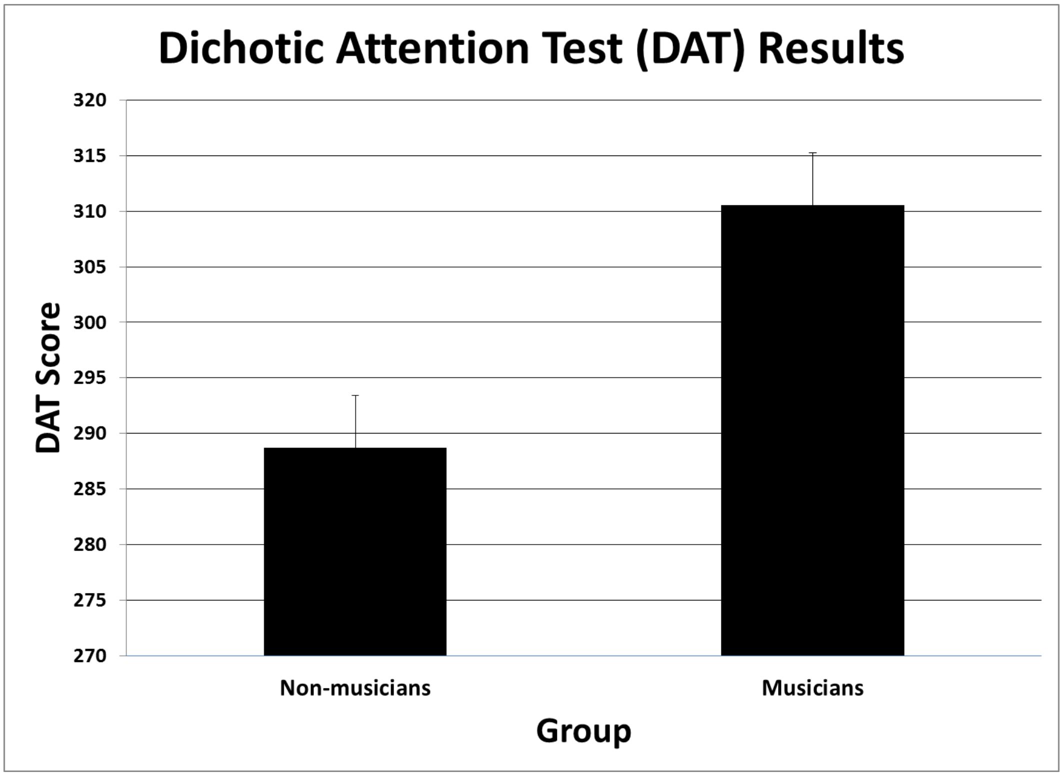


Figure 5. Musicians and non-musicians average performance on the Dichotic Attention Task (DAT). Three hundred target words were presented one per second to the right ear while distracter words were presented simultaneously to the left ear. Error bars represent one standard error.

For the CST 2-talker babble, results showed a between-groups effect of musical training ($F_{1, 30}$ = 7.274, p = 0.011, η^2 = 0.195) indicating musicians consistently performed better than non-musicians. Results for the CST using broadband masking noise also showed a between-groups main effect of musical training ($F_{1, 30}$ = 5.43, p = 0.027, η^2 = 0.153) indicating that musicians outperformed non-musicians. HINT ANOVA results using 1-hemisphere masking noise showed a between-groups effect of musical training ($F_{1, 30}$ = 4.885, p < 0.035, η^2 = 0.14) indicating that musicians performed better than non-musicians in these listening conditions. HINT results using masking noises in both hemispheres showed a between-groups trend similar to the 1-hemisphere results, but were not statistically significant.

Results from the working memory (AO Span) showed a trend similar to the other factors such that the musician participants scored higher on average than non-musician participants, but the difference was not found to be statistically significant ($F_{1,\,30} = 2.569$, $p = 0.117$, $\eta^2 = 0.079$). See Figure 6.

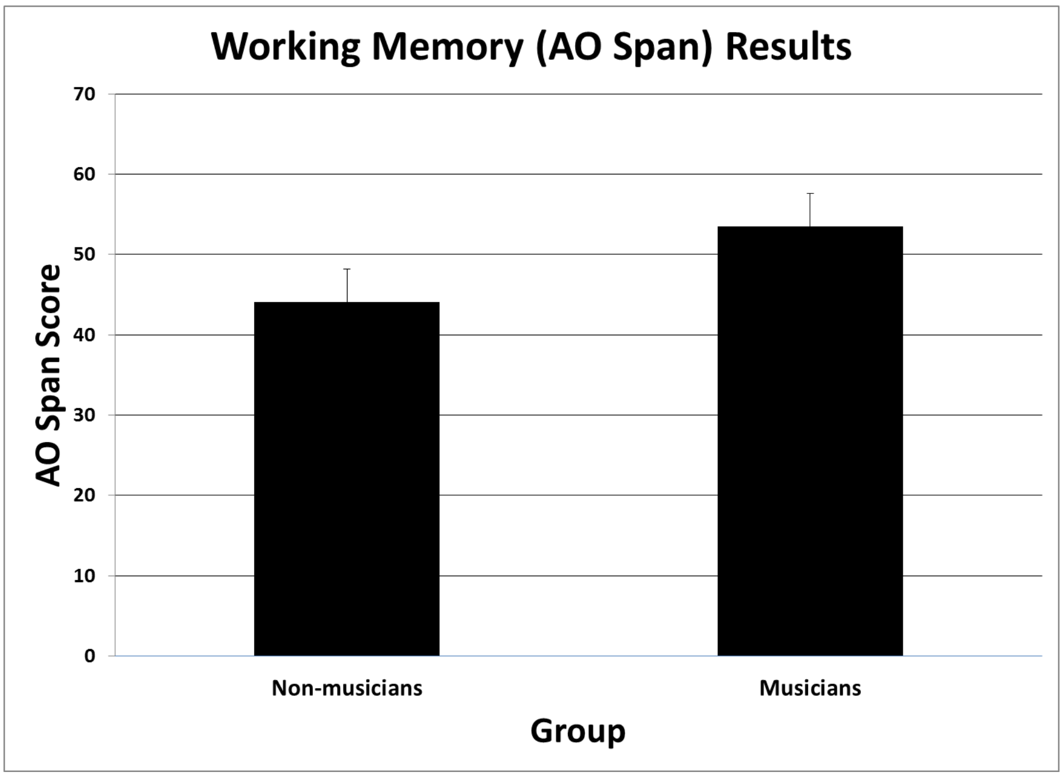


Figure 6. Group results from the visual working memory task (AO Span). Results were not statistically significant but showed a trend similar to the other variables such that musicians had higher scores on average. Error bars represent one standard error.

Significant results from the correlation analysis showed that the EMAT was correlated with all other tests except the AO Span, that the DAT was significantly correlated with the MAT and IMT, and that the speech-in-noise tests were all correlated with each other (See Table 1). Figure 7 and Figure 8 show scatterplots of the raw data for the EMAT and CST and HINT test, respectively. In other words, performance on speech recognition in noise and music based stream segregation were significantly related. However, speech recognition in noise performance was not significantly related to performance on the attention or working memory tasks.

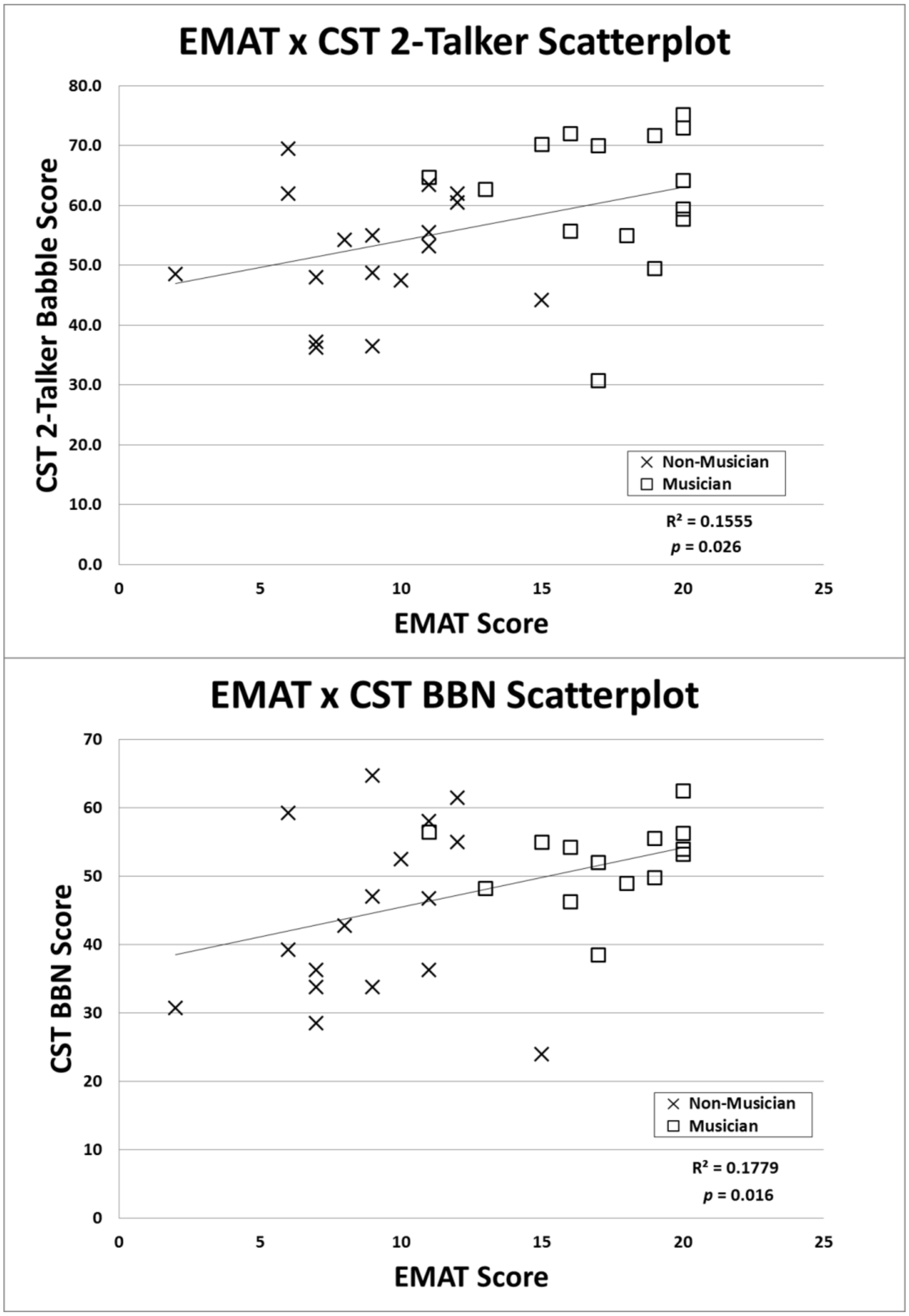
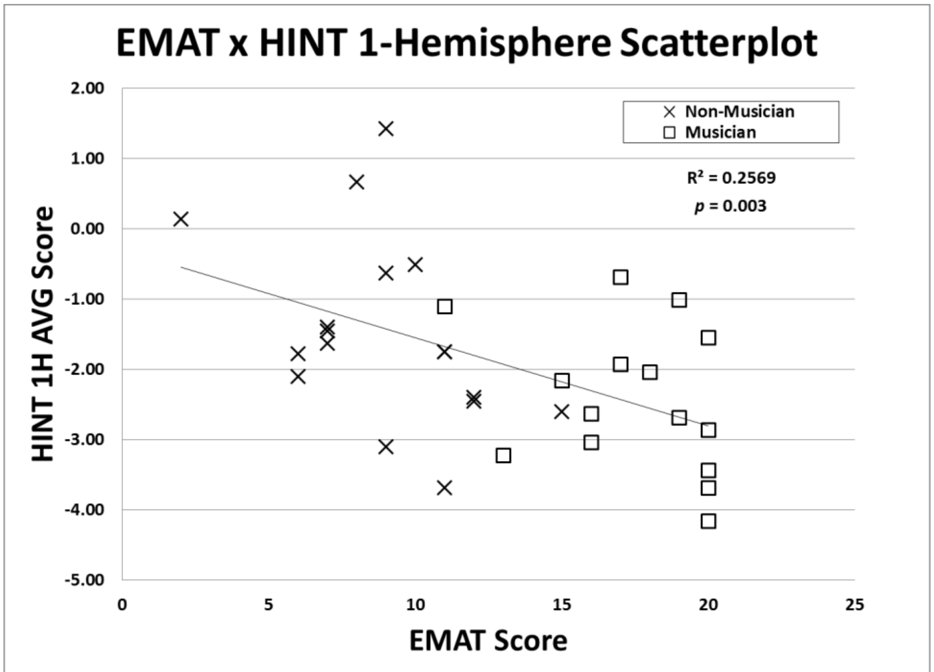


Figure 7. Scatterplots of raw data showing the correlations between the Elementary Music Achievement Test and each CST speech-in-noise test (2-talker babble, BBN).



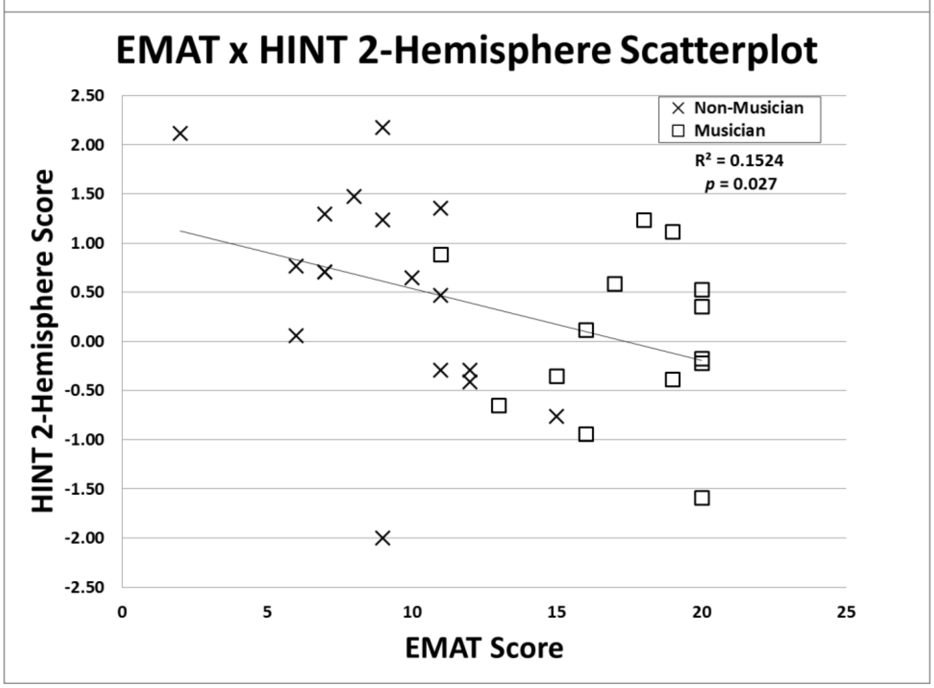


Figure 8. Scatterplots of raw data showing the correlations between the Elementary Music Achievement Test and each HINT speech-in-noise test (1- and 2-hemisphere BBN noise).

Correlations

		AOSPANSC	MAT	IMTADAPT	DAT	HINT1HAVG	HINT2H	CSTBBN	CST2TAVG
AOSPANSC	Pearson Correlatio	1	0.306	-.409*	0.283	-0.203	-0.271	.396*	0.249
	Sig. (2-tailed)	.	0.088	0.02	0.116	0.264	0.133	0.025	0.169
	N	32	32	32	32	32	32	32	32
MAT	Pearson Correlatio	0.306	1	-.760**	.354*	-0.507**	-.390*	.422*	.394*
	Sig. (2-tailed)	0.088	.	0	0.047	0.003	0.027	0.016	0.026
	N	32	32	32	32	32	32	32	32
IMTADAPT	Pearson Correlatio	-.409*	-.760**	1	-.604**	.253	.114	-.250	-.270
	Sig. (2-tailed)	.020	.000	.	.000	.163	.535	.168	.136
	N	32	32	32	32	32	32	32	32
DAT	Pearson Correlatio	.283	.354*	-.604**	1	-.143	.025	.058	.211
	Sig. (2-tailed)	.116	.047	.000	.	.436	.890	.751	.246
	N	32	32	32	32	32	32	32	32
HINT1HAVG	Pearson Correlatio	-.203	-.507**	.253	-.143	1	.801**	-.503**	-.576**
	Sig. (2-tailed)	.264	.003	.163	.436	.	.000	.003	.001
	N	32	32	32	32	32	32	32	32
HINT2H	Pearson Correlatio	-.271	-.390*	.114	.025	.801**	1	-.491**	-.447*
	Sig. (2-tailed)	.133	.027	.535	.890	.000	.	.004	.010
	N	32	32	32	32	32	32	32	32
CSTBBN	Pearson Correlatio	.396*	.422*	-.250	.058	-.503**	-.491**	1	.618**
	Sig. (2-tailed)	.025	.016	.168	.751	.003	.004	.	.000
	N	32	32	32	32	32	32	32	32
CST2TAVG	Pearson Correlatio	.249	.394*	-.270	.211	-.576**	-.447*	.618**	1
	Sig. (2-tailed)	.169	.026	.136	.246	.001	.010	.000	.
	N	32	32	32	32	32	32	32	32

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

Table 1. Correlation analysis results for the dependent variables.

Predictive Analyses

Multiple regression analysis was conducted on each of the SIN tests to determine whether performance on the EMAT, DAT, and/or AO Span could predict speech-in-noise performance. Although significant group differences were not observed for all test measures, the three predictive variables (EMAT, DAT, AO Span) were initially included because it is acknowledged that underlying factors, rather than, or in addition to, inclusion in one of the two groups, may be responsible for differences in speech recognition across the entire subject population. Despite showing a significant group difference, the IMT was not included in these analyses. Since the IMT theoretically measures non-speech schema-based auditory stream segregation, is highly correlated with the EMAT, and is less ecologically valid a measure than the EMAT, it was excluded from the regression analyses as redundant.

Using a stepwise method, regression analysis showed that only the EMAT significantly predicted speech-in-noise scores for the CST BBN (b = 0.382, $t(31)$ = 2.167, p = 0.039), HINT 1H (b = -0.507, $t(31)$ = -2.841, p = 0.008), and HINT 2H (b = -0.407, $t(31)$ = -2.221, p = .035). The EMAT also explained a significant proportion of variance in CST BBN scores, R^2 = 0.281, $F(3, 31)$ = 3.644, p < 0.025, and the HINT 1H scores, R^2 = 0.262, $F(3, 31)$ = 3.314, p < .034. No other factor was predicitive. There was no evidence for collinearity or important outliers for any of the regression analyses. These results are consistent with the correlation analyses which suggested that performance on speech recognition in noise and music based stream segregation were significantly related while speech recognition in noise was not significantly related to attention or working memory.

CHAPTER IV

DISCUSSION AND CONCLUSIONS

Discussion

The primary aims of this study included investigating the effects of musical training on auditory stream segregation, attention, and working memory as they relate to speech understanding in noise.

Speech in noise

Speech-in-noise test results revealed that musicians were better able to correctly identify a target talker in noise than non-musicians in all conditions. However, results reflected particular differences for the CST and HINT conditions when BBN was presented from both hemispheres. The intent of these conditions was to effectively isolate spatial separation effects from other binaural advantages present when a masking noise source is from one hemisphere only. In such listening situations, broadband noise would be expected to reduce access to temporal fine structure more so than 2-talker babble (Arbogast et al., 2002; Bernstein & Grant, 2009; Lorenzi et al., 2006). Our results support this hypothesis, which was also supported in other studies' results.

We had also predicted that musicians would do better than non-musicians as the similarity of target and masker increased. Results comparing performance on the CST BBN to 2-talker babble were used to address this prediction such that the 2-talker babble was one single talker from each hemisphere speaking CST sentences not used for testing. This 2-talker babble increased the similarity to the target as compared to broadband noise. Our results showing that the difference between musicians' and non-musicians' performance was larger at all eccentricities in the 2-talker babble than it was in BBN except for 0°/90° where the musician advantage was larger in BBN. Regardless of the 0°/90° result, we believe taken in total, these results provide support for the prediction that musicians would do better as target and masker similarity increased (See Figure 9). The lack of a significant between-groups result for the HINT 2-hemisphere conditions could be interpreted as conflicting with this interpretation. However, unlike the HINT, the CST provides contextual cues across each group of ten sentences. Perhaps the increase of contextual cues in the CST partially offsets the loss of audibility when noise is presented in both hemispheres.

In any case, the result at 0°/90° eccentricity is thought to be due to the spatial separation cue being large enough to dominate performance, or to make reliance on cues utilized at smaller eccentricities less necessary. In addition, our results suggest that spatial separation greater than 22.5° is required to offset the loss of these cue advantages and see a significant difference in performance for an energetic masker. Specifically, performance differences for the CST dependent on noise type (BBN, 2-talker babble) were observed as eccentricity for the masking noise was increased. All CST conditions had masking noise in both hemispheres.

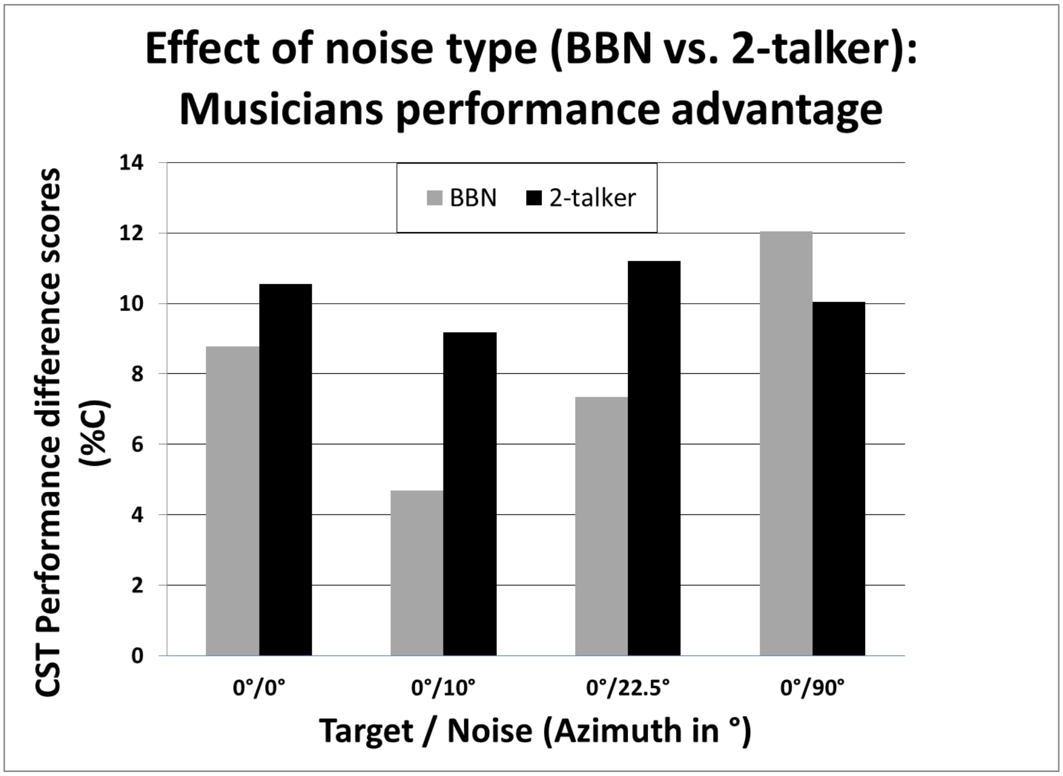


Figure 9. Connected Speech Test performance difference scores. Gray and black bars show results by eccentricity for speech-shaped broadband noise and 2-talker babble, respectively. Each bar shows the number of percentage points better in performance by musicians than non-musicians. Results suggest that musicians' are better able to auditory stream segregate target from masker when the target and masker are more similar (i.e., 2-talker babble).

As described above, when the masker was broadband noise, and access to temporal fine structure was reduced (compared to 2-talker babble), performance at 0°/10° and 0°/22.5° azimuth were not different from each other.

Regarding eccentricity as opposed to noise type, a similar results pattern was observed for the CST and HINT when masking noise was presented in both hemispheres. This suggests that a greater increase in spatial separation is needed to offset the loss of access to head shadow, binaural processing, and temporal fine structure cues. The lack of difference for the 0°/10° and 0°/22.5° eccentricities with broadband noise maskers is likely due to a tradeoff between the effect of increasing spatial separation and the decreasing signal-to-noise ratio. So, although the within-groups analysis did not show statistically significant differences between 0°/10° and 0°/22.5° azimuth, the trend was towards improved performance for both groups as spatial separation increased.

The results of our between-groups overall analysis showed that musicians are better able to do speech-based schema-based auditory stream segregation.

The calculated change in performance when the masker was moved from being collocated with the target at 0°/0° to 0°/10°resulted in a greater increase in performance for non-musicians than musicians. This change reflects the narrowing of the performance gap between musicians and non-musicians as non-musicians take advantage of the spatial separation cue. This effect was observed for both BBN and 2-talker noise. In other words, the benefit from spatially moving the noise 10 degrees away from 0° had greater benefit for non-musicians since non-musicians experienced more difficulty at 0°/0°. However, the difference in performance as a result of moving the noise from 0° to 90° shows a similar overall improvement for BBN and 2-talker babble for both groups. This pattern suggests the musician advantage for eccentricity is more detectable in the more difficult listening conditions that require better auditory stream segregation. See Figure 10.

We speculate that musical training strengthens the relationship between the general bottom-up process of auditory stream segregation and the top-down process of attention for non-speech-based schema-based (e.g., music) auditory stream segregation. This strengthening may explain the superior performance of musicians on both these measures. However, it may also explain the apparent generalization

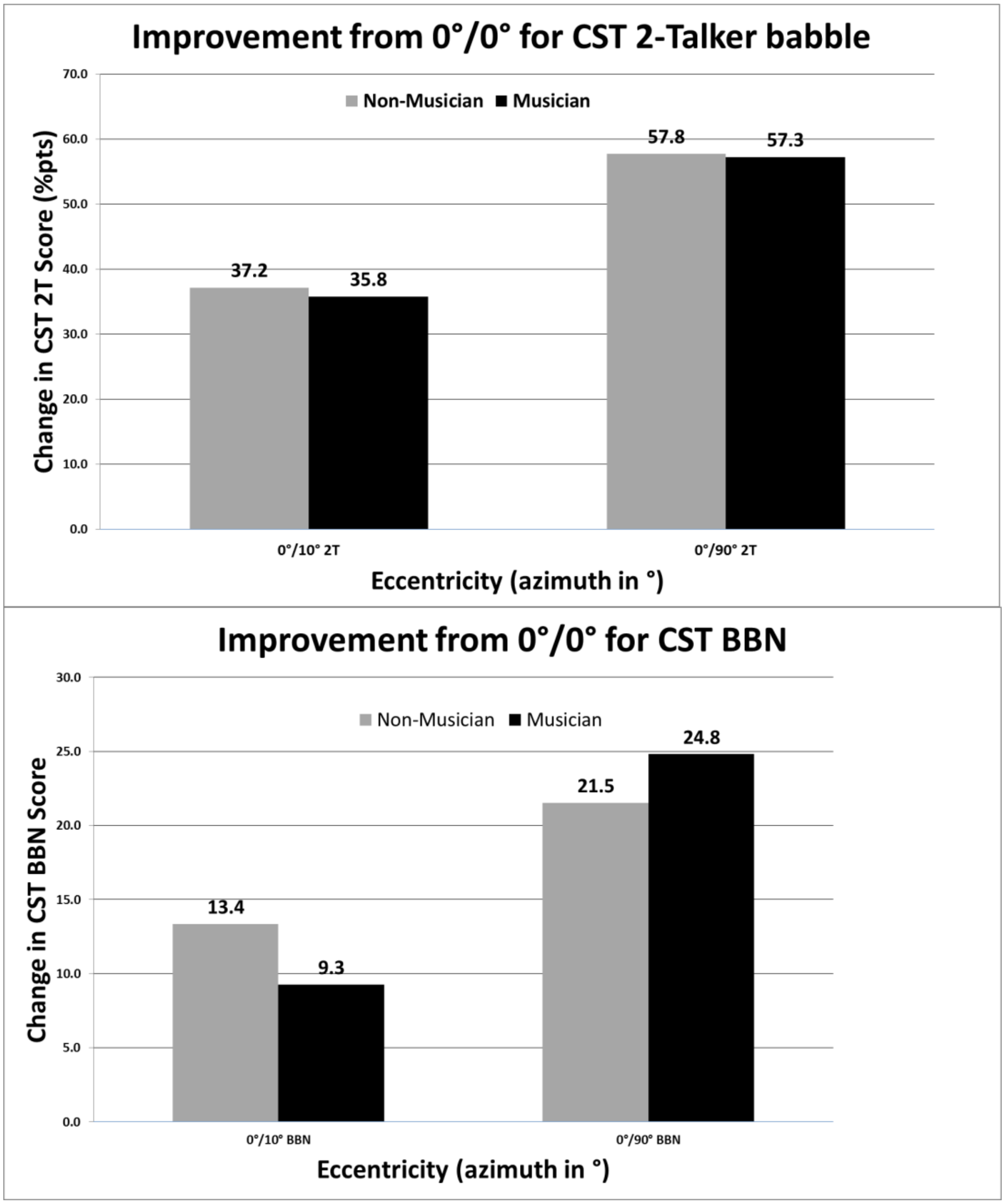


Figure 10. CST Change in performance by group as the noise masker is moved from 0° to 10° and 90° azimuth. The top panel shows the change in performance for 2-Talker babble. The bottom panel shows the change in performance for broadband noise.

to speech-based schema-based auditory stream segregation observed in this study since it has been shown previously that at least some of the speech and non-speech (i.e., music-related) cortical pathways are shared(Anderson & Kraus, 2011; Chan, Ho, & Cheung, 1998). We discuss auditory stream segregation below further in the "Auditory Stream Segregation" section.

HINT: One-hemisphere versus two-hemisphere noise

With the intent of replicating and extending the findings of previous research showing that musicians perform better when speech and noise are collocated (Parbery-Clark et al., 2009), we included similar 1-hemisphere HINT conditions at 0°/0° and 0°/90° azimuth, as well as additional eccentricities as described in the "Methods" and "Discussion" sections. We predicted that there would be a significant difference between musicians' and non-musicians' performance when speech and noise were collocated.

Our results for the condition with target and masker collocated at 0° azimuth showed that musicians were better able than non-musicians to stream segregate the target from the masker, which was a finding consistent with previous research. However, unlike previous research, we also showed significantly better performance by musicians at all other eccentricities including 0°/90°. Although our 0°/90° HINT condition was similar to previous work, we used the original HINT BBN, whereas previous work employed a newer version of the HINT that uses multi-talker babble. If not due to the minor differences in test materials and listening environment differences between studies, it should be noted that the previous work showed a trend toward our significant results indicating musicians require a smaller SNR to obtain 50% correct performance.

Moreover, when comparing the improvement in performance due to spatial release of masking for 1-hemisphere noise conditions at 0°/0° with 0°/90°, both groups showed almost identical magnitude in improvement (i.e., 6.7 dB and 6.8 dB for non-musicians and musicians, respectively). This indicates that performance differences for the HINT 1-hemisphere conditions used in the current study were relatively stable, and that musicians maintained their advantage across eccentricities.

Auditory stream segregation

In the current study, the observed group differences between musicians and non-musicians for speech-based and non-speech-based schema-based auditory stream segregation tasks, as well as the significant correlations between all speech-in-noise tests and the Elementary Music Achievement Test (EMAT), are interpreted as convincing evidence that listeners' stream segregation abilities for both musical and non-musical tasks may be related. In addition, participants who were better at stream segregation performed better whether the competing distracter signals were less similar (i.e., broadband noise), or more similar (i.e., speech maskers for speech targets, music distracters for music targets) to the targets.

Additional evidence for concluding that auditory stream segregation is a primary explanatory factor behind differences in speech recognition in noise across listeners in this study is found in the correlations between the speech-in-noise and the EMAT results. Both are auditory stream segregation tasks, although the former are speech-based, and the latter is non-speech-based. Significant correlation results showed that the better a participant performed on each speech-in-noise task, the more likely s/he would perform better on the EMAT regardless of musical training status. In addition, our regression analysis results showed that the only predictive factor for speech-in-noise performance was performance on the EMAT, which is a non-speech schema-based auditory stream segregation task. Simply stated, one's ability to accurately segregate target signal streams from other signals determines performance on any auditory stream segregation task regardless of signal type. Moreover, those with musical training as a group were better able to perform both speech based and non-speech-based auditory stream segregation.

Attention

The results from the attention (dichotic listening) testing revealed a between-groups difference but not predictability of performance on speech-in-noise testing. Musicians are required to engage their ability to attend and ignore as well as regularly shift their attention during performances. Monitoring intonation, listening for musical cues from other instruments, reading the music, following the conductor, etc. all while playing one's own instrument are some examples requiring musicians' attention and attention shifts. Strait et al. (2010) reported a significant musician versus non-musician difference for auditory attention. They compared reaction time of musicians and non-musicians to a target beeping tone as a measure of auditory attention. For some trials, a second auditory stimulus was presented to which participants were instructed not to respond. Their results showed a statistically significant result in the form of shorter reaction times by musicians to the beeping tones. However, the average group difference was ~45 msec. This finding was statistically significant, but generalizing this finding to real-world listening situations seems limited. For the current study, the dichotic listening task arguably represents a more ecologically valid demand on attention, which required participants to attend to words spoken in one ear while ignoring distracter words in the opposite ear. We posit that our task is more similar to the attention demands of real world listening situations (e.g., listening to French horns while ignoring the other brass players, listening to a specific talker in the presence of other talkers, etc.). Such listening situations require the listener to attend to a target auditory object of interest while ignoring other simultaneously occurring auditory objects. We acknowledge that, specifically for instrumental musical listening, no speech-based schema based (i.e., verbal) information would be present, and for the current study we used such stimuli to compare musicians and non-musicians. Perhaps if such stimuli were used, even larger group differences would have been observed. Regardless, results suggest musicians are better at using attention to positively impact the more peripheral process of auditory stream segregation.

Working memory

Musicians outperformed non-musicians on all tasks requiring auditory stream segregation. However, the results from the working memory testing (AO Span) did not reveal either a group difference or predictability of performance on speech-in-noise testing. Although not statistically significant, musicians scored high on average than non-musicians, and the interleaved melody results were correlated with working memory suggesting a relationship. Lee, Lu, and Ko (2007) showed an effect of musical training in children on an operation span task; however, they did not show such an effect with adults. Therefore, it is possible that the visual working memory test used in this study was not sensitive enough to effects of musical training on auditory working memory in adults.

Some research has suggested that working memory is modality independent (Crottaz-Herbette, Anagnoson, & Menon, 2004; Schumacher et al., 1996). However, recent work by Schulz, Mueller, and Koelsch (2011) showed that musicians had different cortical activation patterns than non-musicians for atonal and tonal sequences. Their functional magnetic resonance imaging (fMRI) measures showed that a separate neural network is involved in non-verbal auditory working memory, and that such a network is more strongly activated in musicians. Schulze, Zysset, Mueller, Frederici, and Koelsch (2011) used fMRI to examine verbal and tonal working memory in musicians and non-musicians. They presented simultaneous pure tones and syllables followed by a rehearsal period of up to 4.2 - 6.2 sec. At the end of each trial, and depending on whether it was a verbal or tonal condition, a syllable or a probe tone was presented, and the participant would indicate if the stimulus had been presented in the initial sequence. Their findings revealed neural structures for both verbal and tonal working memory with different weightings depending on the type of signal. In addition, they showed that only musician participants activated specific cortical subcomponents for each type suggesting the existence of two working memory systems in musicians. In the current study, it may be that no group difference was observed on the AO Span working memory task because the activation of tonal working memory as posited by Schulze et al (2011) possibly used to complete the non-speech schema-based tasks was not required to complete the AO Span task. Therefore, perhaps if instead of the AO Span, an alternative auditory working memory task had been used, a group difference would have been observed. If one accepts that the non-significant result for WM as an anomaly, then an alternative interpretation is that working memory may not be as contributory as the relationship between auditory stream segregation and attention following the acquisition of musical training.

Tradeoff

Our between-groups overall ANOVA results suggest that both auditory stream segregation and attention are contributory towards musicians' superior performance compared to non-musicians'. However, our ANOVA results do not tell us what underlies this advantage. Our regression analyses result suggest that auditory stream segregation is what underlies this advantage, and does not support attention as predictive. Perhaps musical training can result in different weightings for how much WM and attention affect auditory stream segregation. Perhaps there is a trading relationship among these factors such that musicians employ attention more than non-musicians, but those with stronger working memory are able to compensate when attention is not vastly improved from training. Therefore, in general (across groups), another alternative interpretation is that there is a tradeoff between the factors. That is, one person's attention and auditory stream segregation may be better than her working memory, while another's working memory is better. Since previous research has shown all of the factors to be active during complex listening scenarios, perhaps different strategies are employed based on the strengths and weaknesses present in the individual listener. Such strategies could lead to different performance patterns on individual measures of each factor, but similar overall outcomes during ecologically valid listening situations.

Although our results show better performance by musicians than non-musicians on tasks of auditory stream segregation for both speech- (HINT, CST) and non-speech schema-based auditory stream segregation (EMAT, IMT), as well as a task of attention (DAT), they are not completely consistent with the conclusions of some other previous studies that superior performance by musicians can be at least partially explained by working memory capacity and/or processing efficiency (George & Coch, 2011; Parbery-Clark et al., 2009). Our results may not necessarily conflict, but may have revealed in auditory stream segregation a complementary factor and its effects. See above for additional discussion.

Conclusion

An evaluation of performance was conducted for musicians and non-musicians using tests of attention (dichotic listening task), working memory (automated operation span task), and auditory stream segregation (i.e., melodic schema-based task, music achievement test, and speech-in-noise tests). Results indicated that musicians performed better than non-musicians on speech-based and non-speech-based schema-based auditory stream segregation measures and an attention task, but not on measures of working memory. In addition, auditory stream segregation ability was the only factor able to predict speech-in-noise performance. Perhaps some of the inconsistences in previous studies were due to poorer matching of subjects between the groups, and that auditory stream segregation governed by attention is primarily responsible for performance in any difficult listening situation. That is, perhaps individuals with better auditory stream segregation and attention ability are better able to identify and attend to signals of interest. We believe it is reasonable to conclude that musical training results in better auditory stream segregation (EMAT, IMT, CST, HINT) and attention (DAT) since musicians performed significantly better on these tasks. However, it cannot be unequivocally determined from these results whether improved auditory stream segregation and attention abilities are a result of musical training or individuals innately better at it pursue musical activities and careers.

REFERENCES

Akin, O., & Belgin, E. (2009). Hearing Characteristics and Frequency Discrimination Ability in Musicians and Nonmusicians. [Article]. *Journal of International Advanced Otology, 5*(2), 195-202.

Alain, C., & Bernstein, L. J. (2008). From sounds to meaning: the role of attention during auditory scene analysis. [Article]. *Current Opinion in Otolaryngology & Head and Neck Surgery, 16*(5), 485-489. doi: 10.1097/MOO.0b013e32830e2096

Alain, C., & Izenberg, A. (2003). Effects of attentional load on auditory scene analysis. *Journal of Cognitive Neuroscience, 15*, 1063-1073.

Alain, C., & Tremblay, K. (2007). The role of event-related brain potentials in assessing central auditory processing. [Article]. *Journal of the American Academy of Audiology, 18*(7), 573-589.

Arbogast, T. L., Mason, C. R., & Kidd, G. (2002). The effect of spatial separation on informational and energetic masking of speech. [Proceedings Paper]. *Journal of the Acoustical Society of America, 112*(5), 2086-2098. doi: 10.1121/1.1510141

Arbogast, T. L., Mason, C. R., & Kidd, G. (2005). The effect of spatial separation on informational masking of speech in normal-hearing and hearing-impaired listeners. [Proceedings Paper]. *Journal of the Acoustical Society of America, 117*(4), 2169-2180. doi: 10.1121/1.1861598

Assmann, P. F., & Paschall, D. D. (1998). Pitches of concurrent vowels. *Journal of the Acoustical Society of America, 103*(2), 1150-1160.

Baddeley, A. (2007). *Working memory, thought, and action*. Oxford ; New York: Oxford University Press.

Baddeley, A. (2010). Working memory. *Current Biology, 20*(4), R136-R140.

Baddeley, A., & Hitch, G. (1974). Working Memory. In G. Bower (Ed.), *Recent advances in learning and motivation* (Vol. 8, pp. 47-90). New York: Academic Press.

Beauvois, M. W., & Meddis, R. (1997). Time decay of auditory stream biasing. *Perception & Psychophysics, 59*(1), 81-86.

Bench, J., & Bamford, J. (Eds.). (1979). *Speech-hearing tests and the spoken language of hearing-impaired children*. London: Academic Press.

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. [Review]. *Cerebral Cortex, 19*(7), 1583-1596. doi: 10.1093/cercor/bhn196

Besson, M., Schon, D., Moreno, S., Santos, A., & Magne, C. (2007). Influence of musical expertise and musical training on pitch processing in music and language. *Restor Neurol Neurosci, 25*(3-4), 399-410.

Bey, C., & McAdams, S. (2002). Schema-based processing in auditory scene analysis. [Article]. *Perception & Psychophysics, 64*(5), 844-854.

Bey, C., & McAdams, S. (2003). Postrecognition of interleaved melodies as an indirect measure of auditory stream formation. [Article]. *Journal of Experimental Psychology-Human Perception and Performance, 29*(2), 267-279. doi: 10.1037/0096-1523.29.2.267

Bregman. (1990). *Auditory scene analysis : the perceptual organization of sound*. Cambridge, Mass.: MIT Press.

Brungart, D. S. (2001). Informational and energetic masking effects in the perception of two simultaneous talkers. *Journal of the Acoustical Society of America, 109*(3), 1101-1109.

Carlyon, R. P. (2004). How the brain separates sounds. [Review]. *Trends in Cognitive Sciences, 8*(10), 465-471. doi: 10.1016/j.tics.2004.08.008

Carlyon, R. P., Plack, C. J., Fantini, D. A., & Cusack, R. (2003). Cross-modal and non-sensory influences on auditory streaming. *Perception, 32*(11), 1393-1402. doi: 10.1068/p5035

Chartrand, J.-P., & Belin, P. (2006). Superior voice timbre processing in musicians. *Neuroscience Letters, 405*(3), 164-167.

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. [Article]. *Journal of Cognitive Neuroscience, 20*(2), 226-239.

Cherry, E. (1953). Some experiments on the recognition of speech, with one and two ears. *Journal of the Acoustical Society of America, 25(5)*, 975-979.

Colwell, R. (1967). *Elementary Music Achievement Tests: Administration and scoring manual; Guide to the tests*. Chicago: Follett.

Conway, A. R. A., Cowan, N., & Bunting, M. F. (2001). The cocktail party phenomenon revisited: The importance of working memory capacity. *Psychonomic Bulletin & Review, 8*(2), 331-335.

Corbetta, M., Miezin, F. M., Dobmeyer, S., Shulman, G. L., & Petersen, S. E. (1990). ATTENTIONAL MODULATION OF NEURAL PROCESSING OF SHAPE, COLOR, AND VELOCITY IN HUMANS. [Article]. *Science, 248*(4962), 1556-1559.

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. [Review]. *Nature Reviews Neuroscience, 3*(3), 201-215. doi: 10.1038/nrn755

Cowan, N., Elliott, E. M., Saults, J. S., Morey, C. C., Mattox, S., Hismjatullina, A., & Conway, A. R. A. (2005). On the capacity of attention: Its estimation and its role in working memory and cognitive aptitudes. [Review]. *Cognitive Psychology, 51*(1), 42-100. doi: 10.1016/j.cogpsych.2004.12.001

Cox, R. M., Alexander, G. C., & Gilmore, C. (1987). DEVELOPMENT OF THE CONNECTED SPEECH TEST CST. *Ear and Hearing, 8*(5 SUPPL), 118S-126S.

Cox, R. M., Alexander, G. C., Gilmore, C., & Pusakulich, K. M. (1988). USE OF THE CONNECTED SPEECH TEST (CST) WITH HEARING-IMPAIRED LISTENERS. *Ear and Hearing, 9*(4), 198-207.

Cox, R. M., Alexander, G. C., Gilmore, C., & Pusakulich, K. M. (1989). THE CONNECTED SPEECH TEST VERSION-3 - AUDIOVISUAL ADMINISTRATION. *Ear and Hearing, 10*(1), 29-32.

Dalton, P., Santangelo, V., & Spence, C. (2009). The role of working memory in auditory selective attention. [Article]. *Quarterly Journal of Experimental Psychology, 62*(11), 2126-2132. doi: 10.1080/17470210903023646

Daneman, M., & Carpenter, P. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning & Verbal Behavior. Vol, 19*(4), 450-466.

Daneman, M., & Merikle, P. (1996). Working memory and language comprehension: A meta-analysis. [Article]. *Psychonomic Bulletin & Review, 3*(4), 422-433.

Darwin, C. J. (1997). Auditory grouping. [Review]. *Trends in Cognitive Sciences, 1*(9), 327-333.

Douglas, K. M., & Bilkey, D. K. (2007). Amusia is associated with deficits in spatial processing. [Article]. *Nature Neuroscience, 10*(7), 915-921. doi: 10.1038/nn1925

Engle, R. W. (2002). Working memory capacity as executive attention. [Article]. *Current Directions in Psychological Science, 11*(1), 19-23.

Freyman, R. L., Balakrishnan, U., & Helfer, K. S. (2001). Spatial release from informational masking in speech recognition. *Journal of the Acoustical Society of America, 109*(5), 2112-2122.

Freyman, R. L., Helfer, K. S., McCall, D. D., & Clifton, R. K. (1999). The role of perceived spatial separation in the unmasking of speech. *Journal of the Acoustical Society of America, 106*(6), 3578-3588.

Fujioka, T., Trainor, L. J., Ross, B., Kakigi, R., & Pantev, C. (2004). Musical training enhances automatic encoding of melodic contour and interval structure. *J Cogn Neurosci, 16*(6), 1010-1021.

Gaab, N., & Schlaug, G. (2003a). The effect of musicianship on pitch memory in performance matched groups. [Article]. *Neuroreport, 14*(18), 2291-2295. doi: 10.1097/01.wnr.0000093587.33576.f7

Gaab, N., & Schlaug, G. (2003b). Musicians differ from nonmusicians in brain activation despite performance matching. *Neurosciences and Music, 999*, 385-388. doi: 10.1196/annals.1284.048

Gaser, C., & Schlaug, G. (2003a). Brain structures differ between musicians and non-musicians. *J Neurosci, 23*(27), 9240-9245.

Gaser, C., & Schlaug, G. (2003b). Gray matter differences between musicians and nonmusicians. *Ann N Y Acad Sci, 999*, 514-517.

Hantz, E. C., Crummer, G. C., Wayman, J. W., Walton, J. P., & Frisina, R. D. (1992). EFFECTS OF MUSICAL TRAINING AND ABSOLUTE PITCH ON THE NEURAL PROCESSING OF MELODIC INTERVALS - A P3 EVENT-RELATED POTENTIAL STUDY. *Music Perception, 10*(1), 25-42.

Ihiefeld, A., & Shinn-Cunningham, B. (2008). Spatial release from energetic and informational masking in a selective speech identification task. *Journal of the Acoustical Society of America, 123*(6), 4369-4379. doi: 10.1121/1.2904826

Jarrold, C., & Towse, J. N. (2006). Individual differences in working memory. *Neuroscience, 139*(1), 39-50.

Kane, M. J., & Engle, R. W. (2003). Working-memory capacity and the control of attention: The contributions of goal neglect, response competition, and task set to Stroop interference.

Journal of Experimental Psychology-General, 132(1), 47-70. doi: 10.1037/0096-3445.132.1.47

Kaufman, A., & Kaufman, N. (2004). Kaufman Brief Intelligence Test (2nd ed.). Circle Pines, MN: AGS Publishing.

Kidd, G., Arbogast, T. L., Mason, C. R., & Gallun, F. J. (2005). The advantage of knowing where to listen. *Journal of the Acoustical Society of America, 118*(6), 3804-3815. doi: 10.1121/1.2109187

Kidd, G., Arbogast, T. L., Mason, C. R., & Walsh, M. (2002). Informational masking in listeners with sensorineural hearing loss. *Jaro, 3*(2), 107-119. doi: 10.1007/s1016200210095

Kidd, G., Mason, C. R., & Gallun, F. J. (2005). Combining energetic and informational masking for speech identification. *Journal of the Acoustical Society of America, 118*(2), 982-992. doi: 10.1121/1.1953167

Klingberg, T. (2008). *The overflowing brain: information overload and the limits of working memory*: Oxford University Press, USA.

Kyllonen, P. C., & Christal, R. E. (1990). REASONING ABILITY IS (LITTLE MORE THAN) WORKING-MEMORY CAPACITY. [Article]. *Intelligence, 14*(4), 389-433.

Marchand, D. J. (1975). A Study of Two Approaches to Developing Expressive Performance. *Journal of Research in Music Education, 23*(1), 14-22. doi: 10.2307/3345199

Marrone, N., Mason, C. R., & Kidd, G. (2008). The effects of hearing loss and age on the benefit of spatial separation between multiple talkers in reverberant rooms. *Journal of the Acoustical Society of America, 124*(5), 3064-3075. doi: 10.1121/1.2980441

Micheyl, C., Delhommeau, K., Perrot, X., & Oxenham, A. J. (2006). Influence of musical and psychoacoustical training on pitch discrimination. *Hear Res, 219*(1-2), 36-47.

Miyake, A., & Shah, P. (1999). *Models of working memory*: Citeseer.

Moore, B. C. J., & Gockel, H. (2002). Factors influencing sequential stream segregation. [Proceedings Paper]. *Acta Acustica United with Acustica, 88*(3), 320-333.

Nikjeh, D. A., Lister, J. J., & Frisch, S. A. (2008). Hearing of note: An electrophysiologic and psychoacoustic comparison of pitch discrimination between vocal and instrumental musicians. *Psychophysiology, 45*(6), 994-1007. doi: 10.1111/j.1469-8986.2008.00689.x

Nikjeh, D. A., Lister, J. J., & Frisch, S. A. (2009). The relationship between pitch discrimination and vocal production: Comparison of vocal and instrumental musicians. [Article]. *Journal of the Acoustical Society of America, 125*(1), 328-338. doi: 10.1121/1.3021309

Nilsson, M., Soli, S. D., & Sullivan, J. A. (1994). DEVELOPMENT OF THE HEARING IN NOISE TEST FOR THE MEASUREMENT OF SPEECH RECEPTION THRESHOLDS IN QUIET AND IN NOISE. *Journal of the Acoustical Society of America, 95*(2), 1085-1099.

Pantev, C., Ross, B., Fujioka, T., Trainor, L. J., Schulte, M., & Schulz, M. (2003). Music and learning-induced cortical plasticity. In G. Avanzini, C. Faienza, D. Minciacchi, L. Lopez & M. Majno (Eds.), *Neurosciences and Music* (Vol. 999, pp. 438-450).

Parbery-Clark, A., Skoe, E., Lam, C., & Kraus, N. (2009). Musician Enhancement for Speech-In-Noise. *Ear and Hearing, 30*(6), 653-661.

Piro, J. M., & Ortiz, C. (2009). The effect of piano lessons on the vocabulary and verbal sequencing skills of primary grade students. [Review]. *Psychology of Music, 37*(3), 325-347. doi: 10.1177/0305735608097248

Pitt, M. A. (1994). Perception of pitch and timbre by musically trained and untrained listeners. *J Exp Psychol Hum Percept Perform, 20*(5), 976-986.

Plomp, R., & Mimpen, A. M. (1981). Effect of the orientation of the speakers head and the azimuth of a noise source on the speech-reception threshold for sentences. [Article]. *Acustica, 48*(5), 325-328.

Repovs, G., & Baddeley, A. (2006). The multi-component model of working memory: Explorations in experimental cognitive psychology. *Neuroscience, 139*(1), 5-21.

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

Schneider, B. A., Li, L., & Daneman, M. (2007). How competing speech interferes with speech comprehension in everyday listening situations. *Journal of the American Academy of Audiology, 18*(7), 559(514).

Shahin, A., Roberts, L. E., Pantev, C., Trainor, L. J., & Ross, B. (2005). Modulation of P2 auditory-evoked responses by the spectral complexity of musical sounds. *Neuroreport, 16*(16), 1781-1785.

Sherbecoe, R. L., & Studebaker, G. A. (2002). Audibility-index functions for the connected speech test. [Article]. *Ear and Hearing, 23*(5), 385-398. doi: 10.1097/01.aud.0000034412.18975.eb

Sherbecoe, R. L., & Studebaker, G. A. (2003). Audibility-index predictions of normal-hearing and hearing-impaired listeners' performance on the connected speech test. *Ear and Hearing, 24*(1), 71-88.

Sluming, V., Brooks, J., Howard, M., Downes, J. J., & Roberts, N. (2007). Broca's Area Supports Enhanced Visuospatial Cognition in Orchestral Musicians. *J. Neurosci., 27*(14), 3799-3806. doi: 10.1523/jneurosci.0147-07.2007

Smith, D. V., Davis, B., Niu, K., Healy, E. W., Bonilha, L., Fridriksson, J., . . . Rorden, C. (2009). Spatial Attention Evokes Similar Activation Patterns for Visual and Auditory Stimuli. [Article]. *Journal of Cognitive Neuroscience, 22*(2), 347-361.

Snyder, J. S., & Alain, C. (2007). Toward a neurophysiological theory of auditory stream segregation. *Psychological Bulletin, 133*, 780-799. doi: 10.1037/0033-2909.133.5.780

Snyder, J. S., Alain, C., & Picton, T. W. (2006). Effects of attention on neuroelectric correlates of auditory stream segregation. *J Cogn Neurosci, 18*(1), 1-13.

Soto, D., Hodsoll, J., Rotshtein, P., & Humphreys, G. W. (2008). Automatic guidance of attention from working memory. *Trends in Cognitive Sciences, 12*(9), 342-348.

Tervaniemi, M., Castaneda, A., Knoll, M., & Uther, M. (2006). Sound processing in amateur musicians and nonmusicians: event-related potential and behavioral indices. *Neuroreport, 17*(11), 1225-1228.

Tervaniemi, M., Just, V., Koelsch, S., Widmann, A., & Schroger, E. (2005). Pitch discrimination accuracy in musicians vs nonmusicians: an event-related potential and behavioral study. *Exp Brain Res, 161*(1), 1-10.

Tervaniemi, M., Kruck, S., De Baene, W., Schroger, E., Alter, K., & Friederici, A. D. (2009). Top-down modulation of auditory processing: effects of sound context, musical expertise and attentional focus. [Article]. *European Journal of Neuroscience, 30*(8), 1636-1642. doi: 10.1111/j.1460-9568.2009.06955.x

Tirre, W. C., & Pena, C. M. (1992). INVESTIGATION OF FUNCTIONAL WORKING MEMORY IN THE READING SPAN TEST. [Article]. *Journal of Educational Psychology, 84*(4), 462-472.

Trainor, L. J., Shahin, A., & Roberts, L. E. (2003). Effects of musical training on the auditory cortex in children. *Neurosciences and Music, 999*, 506-513. doi: 10.1196/annals.1284.061

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language, 28*(2), 127-154.

Zatorre, R. J., Perry, D. W., Beckett, C. A., Westbury, C. F., & Evans, A. C. (1998). Functional anatomy of musical processing in listeners with absolute pitch and relative pitch. *Proc Natl Acad Sci U S A, 95*(6), 3172-3177.

Zendel, B. R., & Alain, C. (2008). Concurrent Sound Segregation Is Enhanced in Musicians. *J Cogn Neurosci*. doi: 10.1162/jocn.2009.21140