

EFFECTS OF CHOIR PARTICIPATION AND MUSICAL TRAINING
ON AUDITORY PROCESSING IN OLDER ADULTS

by

Ella Dubinsky

Honours Bachelor of Science, Dalhousie University, 2014

A thesis presented to

Ryerson University

in partial fulfillment of the
requirements for the degree of

Master of Arts

in the program of

Psychology

Toronto, Ontario, Canada, 2017

©Ella Dubinsky 2017

AUTHOR'S DECLARATION FOR ELECTRONIC SUBMISSION OF A THESIS

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I authorize Ryerson University to lend this thesis to other institutions or individuals for the purpose of scholarly research.

I further authorize Ryerson University to reproduce this thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

I understand that my thesis may be made electronically available to the public.

Effects of Choir Participation and Musical Training
on Auditory Processing in Older Adults

Ella Dubinsky

Master of Arts, 2017

Psychology

Ryerson University

Abstract

Prior studies have demonstrated musicianship enhancements of various aspects of auditory and cognitive processing in older adults, but musical training has not been examined as an intervention for mitigating age-related declines in these abilities. The current study investigates whether ten weeks of choir participation can improve aspects of auditory processing in older adults, particularly speech-in-noise (SIN) perception. Three groups of participants underwent pre- and post-testing: choir singing (n=50), music appreciation (n=13), and an age- and audiometry-matched control group (n=32). Linear mixed effects modelling in a multilevel regression analysis showed that choir participants demonstrated the most improvements across auditory measures. Choir participants' gains in SIN perception were partially mediated by pitch discrimination, suggesting a possible mechanism for this perceptual improvement. These findings support the hypothesis that short-term choir participation is an effective intervention for mitigating age-related hearing losses.

Acknowledgments

First and foremost, I would like to thank my supervisor, Dr. Frank Russo, for his tremendous support, encouragement, and patience throughout this project, and for giving me the opportunity to do such motivating and stimulating research. The past two years have been an incredible journey, and I have felt immensely fortunate to have Dr. Russo as a guide and teacher. I would also like to extend my gratitude to Dr. Lixia Yang and Dr. Julia Spaniol, who provided insightful and meaningful feedback throughout this process.

In the Science of Music, Auditory Research and Technology (SMART) Laboratory, I would like to thank my incredible research assistants, Emily Wood, Joseph Rovetti, Rebecca Nurgitz, and Lauren Smith, without whom I would have been buried under 600+ hours of data collection. In particular, Emily's enthusiasm, positivity, and commitment to this project over the past year has been impressive and essential, and I am immensely grateful. The SMART Lab has felt like my academic home for the past two years, and I want to thank the many members who contributed their knowledge and experience as I developed and ran this research protocol, and who supported me throughout the project. In particular, I'd like to thank Gabriel Nespoli for his time and tutelage, brain-storming and troubleshooting, and the encouragement he offered throughout. To my whole SMART Lab family – thank you for the support, engagement, and friendship.

This research would not have been possible without the efforts of Sandra Kerr and Mena Caravetta at the 50+ Program at Ryerson University, who supported the development of the choir and the role of this research project within it. I'd like to thank Sina Fallah, the remarkable and talented choir director, for creating such a positive environment, and shaping the choir into the community it has become. To all the choir members and research participants – getting to know

each of you has been an inspiration and a gift. Thank you for your time and interest in this work, and for embracing me as a part of the community.

Last (but certainly not least), I want to thank my big, musical family, who have inspired and supported me all my life, and have given me endless joy and love over the past two years. I'd like to thank my partner Kyle, for helping me overcome all obstacles with patience and dedication. Finally, I want to thank my parents, who give me music and light, and are the best people I know. Your support and love are the greatest gifts, and I am so grateful to be your daughter.

I'd like to dedicate this work to the memory of my aunt Evelyn Carnat, a choir director and musical soul who filled the world with warmth, laughter, and music.

Table of Contents

Title Page	i
Author's Declaration.....	ii
Abstract.....	iii
Acknowledgments	iv
List of Tables	vii
List of Figures.....	ix
Introduction.....	1
Age-related hearing loss	2
Age-related declines in central processing	3
Speech perception in noise	4
Neural underpinnings of auditory processing	5
Frequency following response (FFR)	6
Musicians vs. non-musicians.....	8
Musical training as an intervention	10
The benefits of group singing.....	11
The current study.....	13
Materials and Methods.....	15
Participants	15
Study design	16
Experimental procedure	18
Questionnaires.....	18
Auditory measures	19
EEG measure: the frequency following response (FFR)	22
Cognitive measures.....	23
Statistical analyses.....	26
Results.....	28
Participants	28
Auditory measures.....	28
Pure-tone audiometry.....	28
Speech-in-noise (SIN) perception.....	29
Pitch discrimination	30
Frequency following response (FFR).....	32
Cognitive measures	33
Working memory	33
Inhibitory control of attention.....	34
Overview of results	35
Moderation and mediation analyses	37
Discussion.....	40
Conclusion	44
Appendix.....	45
References.....	64

List of Tables

Table 1: Experimental conditions and counterbalancing..... 45

Table 2: Pre- and post-testing thresholds (mean \pm standard deviation) for perceiving speech in noise (signal-to-noise ratio; SNR) for choir singing participants (n = 50), music appreciation class participants (n = 13), and do-nothing control participants (n = 32). 45

Table 3: Baseline-matched pre- and post-testing thresholds (mean \pm standard deviation) for perceiving speech in noise (signal-to-noise ratio; SNR) for choir singing participants (n = 32), music appreciation class participants (n = 13), and do-nothing control participants (n = 32). 45

Table 4: Pre- and post-testing thresholds (mean \pm standard deviation) for pitch discrimination (frequency difference limens; FDL in Hz) for choir singing participants (n = 39), music appreciation class participants (n = 11), and do-nothing control participants (n = 18). 46

Table 5: Baseline-matched pre- and post-testing thresholds (mean \pm standard deviation) for pitch discrimination (frequency difference limens; FDL in Hz) for choir singing participants (n = 11), music appreciation class participants (n = 11), and do-nothing control participants (n = 11). 46

Table 6 Pre- and post-testing scores (mean \pm standard deviation) in listening working memory span (Listening Span Task; LSpan) for choir singing participants (n = 36), music appreciation class participants (n = 12), and do-nothing control participants (n = 24). 46

Table 7 Pre- and post-testing scores (mean \pm standard deviation) in total correct letters recalled (TCL; LSpan) for choir singing participants (n = 36), music appreciation class participants (n = 12), and do-nothing control participants (n = 24). 47

Table 8 Pre- and post-testing mean reaction times (mean \pm standard deviation) for Flanker task (ms) for choir singing participants (n = 36), music appreciation class participants (n = 12), and do-nothing control participants (n = 24). 47

List of Figures

Figure 1: Group mean audiograms (top) and average dB HL (bottom) by group. 48

Figure 2: Binaural hearing loss (average dB HL; measured using pure-tone audiometry) as a function of age. 49

Figure 3: Pre- and post-testing thresholds for perceiving speech in noise (signal-to-noise ratio; SNR) for choir singing participants (n = 50), music appreciation class participants (n = 13), and do-nothing control participants (n = 32). 50

Figure 4: Baseline matched pre- and post-testing thresholds for perceiving speech in noise (signal-to-noise ratio; SNR) for choir singing participants (n = 32), music appreciation class participants (n = 13), and do-nothing control participants (n = 32). 51

Figure 5: Pre- and post-testing thresholds for pitch discrimination (frequency difference limens; FDL in Hz) for choir singing participants (n = 39), music appreciation class participants (n = 11), and do-nothing control participants (n = 18). 52

Figure 6: Baseline matched pre- and post-testing thresholds for pitch discrimination (frequency difference limens; FDL in Hz) for choir singing participants (n = 11), music appreciation class participants (n = 11), and do-nothing control participants (n = 11). 53

Figure 7: Pre- and post-testing LSpan scores (auditory working memory) for choir singing participants (n = 36), music appreciation class participants (n = 12), and do-nothing control participants (n = 24). 54

Figure 8: Pre- and post-testing total correct letter scores from the LSpan task (auditory working memory) for choir singing participants (n = 36), music appreciation class participants (n = 13), and do-nothing control participants (n = 23). 55

Figure 9: Pre- and post-testing Flanker effect (inhibitory control of attention, calculated as reaction time (incongruent) – reaction time (congruent); msec) for choir singing participants (n = 35), music appreciation class participants (n = 13), and do-nothing control participants (n = 24)..... 56

Figure 10: Pre- and post-testing mean reaction times (msec) of Flanker task responses of choir singing participants (n = 35), music appreciation class participants (n = 13), and do-nothing control participants (n = 24)..... 57

Figure 11: Pre- and post-testing reaction times (msec) of Flanker task responses of choir singing participants (n = 36), music appreciation class participants (n = 13), and do-nothing control participants (n = 24) within congruent (left) and incongruent (right) Flanker conditions.... 57

Figure 12: SNR ~ Session x Group interaction plot: baseline matched pre- and post-testing thresholds for perceiving speech in noise (signal-to-noise ratio; SNR) for choir singing participants (n = 32) and do-nothing control participants (n = 32)..... 58

Figure 13: 2D (top) and 3D (bottom) representations of the SNR ~ Audiometry (dB HL) x FDL (Hz) interaction plot: effect of increasing hearing loss on SNR, as influenced by FDL (two visualizations)..... 59

Figure 14: SNR ~ Session x FDL (Hz) interaction plot: marginal effect of FDL x Session on SNR (two visualizations). 59

Figure 15: FDL ~ Session x Group interaction plot: pre- and post-testing thresholds for perceiving speech in noise (signal-to-noise ratio; SNR) for choir singing participants (n = 38) and do-nothing control participants (n = 16). 60

Figure 16: Effect of musicianship on FDL (Hz): interaction plot of FDL ~ age of onset of music training x years of continuous musical training. 61

Figure 17: Effects of hearing loss on efficacy of intervention; trends within choir participants.. 62

Figure 18: Exploring FDL (Hz) as a potential mediator for the effects of choir participation on SNR (mediation model). 62

Figure 19: Distribution of indirect mediation effect of FDL (Hz) on the relationship between choir training and speech perception in noise; Monte Carlo approach, resampling = 20000, 95% CI [-0.5400542, -0.02167482]. 63

Introduction

As the population ages, and the expectation of longevity increases, a growing concern in healthcare is the promotion of *healthy aging* – the maintenance of mental, social, and physical wellbeing as one ages, in order to retain independence and lead a high-quality life (Fillit et al., 2002; Kramer, Bherer, Colcombe, Dong, & Greenough, 2004; Peel, McClure, & Bartlett, 2005). Aging is associated with declines in cognitive functioning (e.g., decreased working memory and attentional control; for review, see Fabiani, 2012), and deteriorating sensory-perceptual processes (e.g., worsening vision and hearing; for review, see Fozard, 1990). These declines – in particular, hearing problems – can make it difficult for aging individuals to maintain personal relationships and engage socially, which have in turn been linked to feelings of isolation and depression (Arlinger, 2003; Djernes, 2006). Although assistive technologies (e.g., hearing aids) can target aspects of peripheral hearing loss, persistent perceptual deficits are widely reported, and associated impairments remain (e.g., Killion, 1997). One prevalent example is the loss of the ability to perceive speech in a noisy environment, which severely impacts quality of life and is only minimally remediated by hearing aids (Betlejewski, 2006; Chmiel & Jerger, 1996; Gomez & Madey, 2001; Ricketts & Hornsby, 2005; Salomon, 1986). As such, there is a great demand for effective and accessible interventions that target and remediate auditory losses which are resistant to peripheral assistance – likely occurring further up the auditory processing pathway – in order to promote auditory health and mitigate associated age-related declines (Alain, Zendel, Hutka, & Bidelman, 2014; Anderson, Parbery-Clark, Yi, & Kraus, 2011; Bidelman, Villafuerte, Moreno, & Alain, 2014; Tremblay, Piskosz, & Souza, 2003).

Age-related hearing loss

Hearing impairment is the most prevalent sensory deficit in humans (Mathers, Smith, & Concha, 2000), and is a loss which most adults will experience to some degree as they age (Gates & Mills, 2005). Hearing loss can occur at different stages in the *auditory system* – the mechanisms and pathways responsible for turning acoustic information in the environment into a comprehensible percept of sound in the brain (e.g., Chisolm et al., 2003). *Peripheral hearing loss* refers to problems with the ear structures that transmit and transduce sounds, including the loss of efficient sound transmission through the eardrum and bones in the middle ear (*conductive hearing loss*; Wingfield, Tun, & McCoy, 2005; Yueh, Shapiro, MacLean, & Shekelle, 2003), and the deterioration of the mechanisms responsible for transducing mechanical impulses into neural signals, in the inner ear and cochlea (*sensorineural hearing loss*; Arlinger, 2003; Wingfield et al., 2005; Yueh et al., 2003). *Central hearing loss* refers to the degradation of neural mechanisms that relay sound information from the cochlea to the brain, resulting from long-term attenuation or removal of neural input from the cochlea, or age-related changes in neuronal responses to sound (Frisina & Walton, 2006; Syka, 2002; Yamasoba et al., 2013). Although peripheral losses can be remediated to some degree through the use of assistive technologies, such as hearing aids (or, in extreme cases, cochlear implants), central processing deficits seem to persist in spite of such interventions (Chmiel & Jerger, 1996; Killion, 1997). These central deficits are thought to contribute to losses in key perceptual abilities – such as the perception of speech in a noisy environment – which can severely impact quality of life in aging individuals (Pichora-Fuller, Schneider, MacDonald, Pass, & Brown, 2007; Ricketts & Hornsby, 2005; Schneider, Daneman, & Murphy, 2005; Strouse, Ashmead, Ohde, & Grantham, 1998). Declines in central processing appear to be mediated by age-related disruptions in neural timing mechanisms, which can occur

at multiple stages in the central auditory system (Chisolm et al., 2003; F. A. Russo, Ives, Goy, Pichora-Fuller, & Patterson, 2012).

Age-related declines in central processing

After sound is transduced into neural responses in the cochlea, acoustic information is sent to the auditory cortex by way of the *central auditory system* – tracts and nuclei involved in the afferent neural relay of acoustic information (e.g., Chisolm et al., 2003). This pathway includes the cochlear nerve and nuclei, trapezoid body, superior olivary nuclei, lateral lemniscus, inferior colliculi, medial geniculate nucleus of the thalamus, and finally the brain's primary auditory cortex, where acoustic information is resolved into a meaningful auditory percept (e.g., Phillips, 2002).

At each stage in this central auditory processing pathway, age-related disruptions in neural synchrony can introduce noise into the signal and decrease the fidelity of the neural representation of sound (F. A. Russo et al., 2012; Syka, 2002). The auditory system relies on extremely sensitive neuronal timing mechanisms to encode and transmit acoustic information (Carr, 1993); in the auditory nerve, for instance, the dominant time interval of the phase-locked neuronal firing pattern corresponds with the period of the stimulus, allowing groups of neurons to effectively fire at the frequency of acoustic information. This temporal processing, in conjunction with the spectral contribution of the place code originating in the basilar membrane, is thought to be responsible for encoding *pitch information* – a key component of many auditory perceptual processes (Carney, 1994; Ives & Patterson, 2008; Patterson, Allerhand, & Giguere, 1995; F. A. Russo et al., 2012).

Because neural timing relays key information about the auditory signal, age-related disruptions in temporal processing can have serious impacts on the acoustic representation in the

brain (e.g., F. A. Russo et al., 2012). Aging adults' central auditory systems demonstrate increased temporal variance, affecting neural firing and synchrony (Frisina & Frisina, 1997; Pichora-Fuller & Schneider, 1992; Pichora-Fuller et al., 2007); longer neural recovery time (Walton, Frisina, & O'Neill, 1998); and decreased numbers of neurons in auditory nuclei (Frisina & Walton, 2006). These age-related declines lead to impaired processing of acoustic information, whereby temporal and spectral information are not preserved in the neural representation of sounds (Parbery-Clark, Anderson, Hittner, & Kraus, 2012). Such persistent central processing deficits have been associated with age-related losses in key auditory perceptual abilities, including sound localization (Abel, Giguere, Consoli, & Papsin, 2000); pitch discrimination (Raz, Millman, & Moberg, 1989); melodic discrimination (F. A. Russo et al., 2012); duration (Fitzgibbons & Gordon-Salant, 1994), gap (Schneider, Pichora-Fuller, Kowalchuk, & Lamb, 1994), and mistuned harmonic (Alain, McDonald, Ostroff, & Schneider, 2001) detection; and speech perception in noise (Pichora-Fuller, Schneider, & Daneman, 1995; F. A. Russo & Pichora-Fuller, 2008; B. Schneider, Pichora-Fuller, & Daneman, 2010). Of the perceptual deficits, the loss of speech-in-noise (SIN) perception seems to have the most severe impact on the aging adult's quality of life (e.g., Alain et al., 2014; Anderson et al., 2011; Pichora-Fuller et al., 1995, 2007).

Speech perception in noise

Speech perception in noise refers to the ability to track a voice in a complex acoustic environment, such as a crowded room with many people talking. Vital in social settings and everyday interactions, the loss of this skill can immensely impact an individual's ability to maintain independence, emotional wellbeing, and quality of life as they age (Betlejewski, 2006; Gomez & Madey, 2001; Salomon, 1986). This age-related decline also appears to persist in spite

of peripheral remediation, and can occur in adults with normal peripheral hearing (Alain et al., 2014; Cruickshanks et al., 1998; Gordon-Salant, 2005; Souza, Boike, Witherell, & Tremblay, 2007; Tremblay et al., 2003); in research studies involving older individuals, audiometry was not found to predict speech-in-noise perception (Dubno, Dirks, & Morgan, 1984; Hargus & Gordon-Salant, 1995; Kim, Frisina, Mapes, Hickman, & Frisina, 2006; Souza et al., 2007). As such, there has been a push in the research community to better understand predictors of this and other age-related losses, to inform interventions that may prevent and mitigate declines in auditory health and quality of life in aging adults (Alain et al., 2014; Bidelman & Krishnan, 2010; Frisina & Frisina, 1997; Johnson, Nicol, & Kraus, 2005; Kraus & Chandrasekaran, 2010; Syka, 2002; Zendel, 2011).

Neural underpinnings of auditory processing

One method by which researchers can elucidate the underlying central processing mechanisms involved in auditory perception is by recording an electroencephalogram (EEG) during an auditory perceptual task (e.g., Anderson et al., 2011; Tremblay et al., 2003).

Electroencephalography is a neuroimaging technique in which electrodes placed on the scalp detect changes in polarization that occur with neuronal activity (Bratzlavsky, Van Zandijcke, & Vander Eecken, 1981). When the EEG signal is amplified, digitized, and plotted, it appears as a complex oscillatory pattern of changes in voltage over time, corresponding to neural activity (e.g., Angelakis et al., 2007). The signal can then be filtered, processed and analyzed to extract meaningful information about neural processes (Winkler, Debener, Muller, & Tangermann, 2015). EEG recordings have an extremely high degree of temporal resolution, meaning that rapid changes in neuronal activation can be tracked over time (e.g., Luck, 2012; Skoe & Kraus, 2010; Woodman, 2010). Because auditory processing involves the ultra-rapid integration of acoustic

features, this temporal sensitivity makes electroencephalography uniquely well-suited to the study of neural mechanisms in central auditory processes.

Frequency following response (FFR). EEGs have been used to study auditory processes at different levels in the central auditory pathway, through the examination of cortical and subcortical responses to acoustic stimuli (Musacchia, Strait, & Kraus, 2008; Zendel, 2011). In particular, the *auditory brainstem* – a collection of nuclei involved in afferent and efferent auditory processing – has been shown to encode spectral and temporal acoustic information with a high degree of precision (e.g., Skoe & Kraus, 2010). The electrical potentials generated by synchronous neural activity in the brainstem can be measured using EEGs (Hood 1998 Moushegian, Rupert, & Stillman, 1973). EEG recordings of subcortical responses to auditory stimuli can provide insight into the precise neural mechanisms responsible for sound encoding at this level of the auditory system (Clinard, Tremblay, & Krishnan, 2010; Skoe & Kraus, 2010).

One component of the auditory brainstem response (ABR; Skoe & Kraus, 2010) that has been implicated in perceptual deficits – in particular, speech perception in noise – is the *frequency following response* (Johnson et al., 2005; Skoe & Kraus, 2010). The FFR is thought to be generated primarily in the *inferior colliculus*, and consists of phase-locked neural activation, wherein the inter-spike intervals correspond to the fundamental frequency (F0) of the sound input (Hoormann, Falkenstein, Hohnsbein, & Blanke, 1992). The FFR represents the capacity of neurons in the auditory brainstem to track and encode changes in frequency during the solid state of a complex auditory signal – such as a sinusoidal tone, or the vowel component of a speech stimulus – through sustained synchronous neural phase-locking (for summary, see Skoe & Kraus, 2010). An FFR can be obtained by recording an EEG during presentation of a repeated auditory stimulus, and analyzing the strength and degree of neural synchrony in the evoked

response (Skoe & Kraus, 2010). Spectral and temporal features of the FFR, obtained through signal analysis, are associated with different aspects of neural pitch encoding. A *fast Fourier transform* (FFT) of the signal yields a spectral analysis that can be used to assess response *fidelity* – the degree to which AB neurons faithfully encode frequency information present in the steady state component of the auditory signal, as indexed by the peak magnitude of the stimulus’ fundamental frequency within the neural response (Skoe & Kraus, 2010). Another feature, the *inter-trial phase coherence* (ITPC), is a frequency-domain measure that can be used to assess *consistency* of neural responses to a repeated auditory stimulus, by assessing the synchronization of the response activity and latency with time-locked features of the auditory stimulus (e.g., Delorme & Makeig, 2004). The FFT and ITPC features of the FFR have been proposed as indices of the fidelity and consistency with which neurons in the auditory brainstem respond to sound (for ABR & FFR summary, see Skoe & Kraus, 2010).

Research findings suggest that age-related disruptions in neural timing mechanisms lead to less consistent and lower fidelity frequency following responses to auditory stimuli (Clinard et al., 2010; Werff & Burns, 2011). Signal perturbations introduced by neurons in the auditory brainstem are retained in the signal as it travels from the brainstem to the cortex, leading to a cortical representation of sound that has more *noise* than the original acoustic signal (e.g., Tremblay et al., 2003).

In the perception of speech cues, the ability to discern and track changes in pitch over time gets significantly more difficult when the signal-to-noise ratio (SNR) decreases (e.g., Killion, Niquette, Gudmundsen, Revit, & Banerjee, 2004). By the time an acoustic signal reaches the brain of an aging adult, it is likely to have undergone both peripheral and subcortical distortion (due to age-related declines in sensorineural hearing, and neural noise introduced in

the brainstem, respectively), leading to diminished preservation of key temporal and spectral characteristics (e.g., Clinard et al., 2010; Yueh et al., 2003). This suggests a possible mechanism for age-related declines speech perception in noise (and other auditory perceptual abilities which rely on pitch discrimination), whereby age-related central processing deficits (e.g., reduced FFR features) result in downstream perceptual impairments which persist in spite of peripheral remediation.

As such, researchers working towards developing interventions that mitigate age-related declines in auditory perception have increasingly been looking to predictors of efficient central auditory processing mechanisms, as indexed by measures like the FFR (Alain et al., 2014; Besson, Schön, Moreno, Santos, & Magne, 2007; Herdener et al., 2010; Musacchia, Sams, Skoe, & Kraus, 2007). One predictor that seems to improve both perceptual and neural experiences of audition is *musical experience* (Kraus & Chandrasekaran, 2010; Parbery-Clark, Skoe, & Kraus, 2009; Zendel, 2011).

Musicians vs. non-musicians

Musicianship has been shown to confer many benefits outside the musical domain, but has the most profound positive effects on the auditory system (for review, see Herholz & Zatorre, 2012). Over the course of training, musicians are taught to attend to extremely fine-grained acoustic features – including pitch, timing, and timbre – that contribute to human perception of sound (Kraus & Chandrasekaran, 2010; Kraus, Skoe, Parbery-Clark, & Ashley, 2009). This trained sensitivity to minute acoustic changes is thought to promote the enhancement of auditory perceptual abilities, including those which decline throughout the aging process (Alain et al., 2014; Musacchia et al., 2007; Parbery-Clark, Skoe, & Kraus, 2009). In studies comparing auditory perception in musicians and non-musicians, musical experience has been

shown to enhance pitch discrimination (Micheyl, Delhommeau, Perrot, & Oxenham, 2006; Schellenberg & Moreno, 2009); mistuned harmonic (Koelsch, Schröger, & Tervaniemi, 1999; Zendel & Alain, 2009) and gap (Rammsayer & Altenmüller, 2006) detection; and perception of speech-in-noise (e.g., Parbery-Clark, Skoe, Lam, & Kraus, 2009; Zendel, 2011). Importantly, these benefits appear to be largely resistant to age-related declines (Alain et al., 2014; Parbery-Clark et al., 2012; Zendel, 2011; Zendel & Alain, 2012).

Musicians also demonstrate structural and functional differences in the neural substrates of auditory, sensory-motor, and visuospatial processing (Hyde et al., 2009; Kraus & Chandrasekaran, 2010; Musacchia et al., 2007; Schlaug, 2015). Among musicians, musical aptitude is correlated with an increase in grey matter volume in the primary auditory cortex (Gaser & Schlaug, 2003; P. Schneider et al., 2002); musicianship is also associated with greater volumes of grey matter around the central sulcus, including somatosensory and motor areas, the inferior temporal gyrus, hippocampus, and corpus callosum regions (Gaser & Schlaug, 2003; Herdener et al., 2010; Schlaug, Jäncke, Huang, & Steinmetz, 1995). In addition to structural changes in associated brain regions, musicians also demonstrate functional improvements in neural responses to sound, at cortical and subcortical levels in the auditory processing pathway. Compared with non-musicians, musicians demonstrate enhanced neural responses and activation in the auditory cortex (Besson et al., 2007; Hyde et al., 2009; Koelsch et al., 1999; Kuriki, 2006; Pantev et al., 2003; P. Schneider et al., 2002; Shahin, Bosnyak, Trainor, & Roberts, 2003) and the auditory brainstem (Bidelman & Krishnan, 2010; Lee, Skoe, Kraus, & Ashley, 2009; Musacchia et al., 2007; Parbery-Clark, Skoe, Lam, et al., 2009; Strait, Kraus, Skoe, & Ashley, 2009; Wong, Skoe, Russo, Dees, & Kraus, 2007; Zendel, 2011). Notably, musicians demonstrate improvements in both the fidelity (Bidelman, Krishnan, & Gandour, 2011; Musacchia et al.,

2007, 2008) and consistency (Bidelman, Gandour, & Krishnan, 2011; Bidelman, Krishnan, et al., 2011; Skoe & Kraus, 2013; Strait et al., 2009) of the frequency following response, benefits which appear largely resistant to normal age-related declines (Alain et al., 2014; Parbery-Clark et al., 2012; Parbery-Clark, Skoe, & Kraus, 2009; Zendel, 2011; Zendel & Alain, 2012). Because of the importance of pitch processing across auditory perceptual domains, FFR improvements have been suggested as the mechanism through which musicianship enhances auditory perceptual abilities (Bidelman, Krishnan, et al., 2011; Kraus & Chandrasekaran, 2010).

Similar results have been produced in longitudinal studies of music training in early childhood, suggesting that differences between musicians and non-musicians are mediated by experience-dependent neuroplasticity, rather self-selection due to pre-existing genetic differences (Fujioka, Ross, Kakigi, Pantev, & Trainor, 2006; Hyde et al., 2009). These findings lend support to the idea that musical training alters brain structure and function, and suggest that musical training has the capacity to promote enhancements in the same auditory abilities that decline as we age (Alain et al., 2014; Hanna-Pladdy & MacKay, 2011; Solé Resano, Mercadal-Brotons, Gallego Matas, & Riera, 2010; Zendel, 2011).

Musical training as an intervention

Several studies have examined the effect of musical training in a controlled experimental context, despite the difficulties involved in this type of paradigm (e.g., tracking and controlling practice over the time it takes to ‘become a musician’; Zendel, 2011). These studies provided musical training to non-musicians, and examined structural and functional differences pre- and post-training. Musical training was found to enhance both structure (Hyde et al., 2009) and function (Fujioka et al., 2006; Lappe, Herholz, Trainor, & Pantev, 2008) of the auditory cortex in young children who received musical training, compared to those who did not (control

participants). In addition, neural coding in the auditory brainstem – including fidelity and consistency of the FFR – were found to be enhanced in children following musical training, as well as in adult musicians (Musacchia et al., 2007; Skoe & Kraus, 2013; Wong et al., 2007; Zendel, 2011).

In terms of mitigating age-related declines, the effect of musical training on perceptual, cognitive, and neuro-physiological measures appears to be a function of the interaction between experience and age, whereby the rate of age-related decline is slower in experts (*differential preservation*; e.g., Zendel & Alain, 2012). Studies finding differential preservation of auditory abilities lend support to the protective effect of musical training and expertise, demonstrating an additive benefit of experience over time (Alain et al., 2014; Salthouse, 2006). This is supported by computational models simulating normal aging (Mireles & Charness, 2002), and studies directly examining the effect of age and musicianship on auditory processes (e.g., Zendel, 2011). Previous research shows that older adults with musical experience demonstrate enhanced subcortical functioning (as indexed by features of the frequency following response), and improved auditory perceptual abilities, such as pitch discrimination and speech perception in noise (Alain et al., 2014; Herholz & Zatorre, 2012; Hyde et al., 2009; Kraus & Chandrasekaran, 2010; Micheyl et al., 2006; Pantev et al., 2003; Strait et al., 2009; Zendel, 2011). Although these findings strongly support the capacity of musical experience to mitigate age-related declines in auditory processing, studies have not yet examined the effects of musical training on auditory functioning in an intervention context with older adults.

The benefits of group singing

Singing – compared to instrumental music – is a highly intuitive form of music making, which emerges spontaneously during the first months of life (Papousek, 1996). Singing is a

universal form of expression across many cultures (Mithen, Morley, Wray, Tallerman, & Gamble, 2006), and is as natural as speaking for many adults (Dalla Bella & Berkowska, 2009; Dalla Bella, Giguère, & Peretz, 2007; Pfordresher & Brown, 2007). Group singing, a common form of musical collaboration, has been shown to lead to improvements in cooperation (Good & Russo, 2016) and social and emotional wellbeing (Bailey, 2005; Clift & Morrison, 2011), as well as positive physical and creative outcomes (Beck, Cesario, Yousefi, & Enamoto, 2000; Clift & Hancox, 2001). Some of these benefits may be mediated by changes in hormonal levels that occur during choral singing: after choir practice, choristers demonstrate decreased stress hormones (e.g., cortisol; Beck et al., 2000) and enhanced immune system functioning (e.g., increased immunoglobulin A; Kreutz, Bongard, Rohrman, Hodapp, & Grebe, 2004). Older adults involved in choral groups demonstrate improved emotional and cognitive functioning, and report positive interpersonal relationships, sustained social engagement, and decreased incidence of depression as they age (Coffman, 2002; Darrough, 1992; Hillman, 2002; Solé Resano et al., 2010).

In addition to the social, cognitive, and emotional benefits, singing – in particular, group singing – provides the opportunity for more immediate gains than does instrumental musical training, due to the mechanisms and feedback in place for speech production and monitoring (e.g., Pruitt & Pfordresher, 2015). Vocal production and perception involves the rapid integration of sensory and vocal-motor systems (Hickok, 2001; Pfordresher & Dalla Bella, 2011; Pruitt & Pfordresher, 2015); these feedback loops are necessary in order to be able to monitor speech online (vocal perception), and make real-time adjustments to what is being said (vocal production; Brainard & Doupe, 2000; Houde & Jordan, 1998; Zheng, Munhall, & Johnsrude, 2010). In group singing contexts, the singer not only listens to his or her own voice, but also to

surrounding voices, for changes in pitch and cues to modulate production (e.g., keeping in tune with neighbouring singers; following the correct melodic or harmonic line; Pruitt & Pfordresher, 2015). The use of group singing as a method for musical training could capitalize on those pre-existing auditory feedback mechanisms, and quickly hone auditory perceptual abilities while promoting a positive social environment that is enjoyable and motivating. Intervention studies with older adults often find participant compliance to be a major detractor and shortcoming (Coon & Thompson, 2003); an enjoyable, sustainable, and motivating intervention is more likely to be continued, and greater gains achieved. Finally, running a choir is an immensely scalable intervention, requiring minimal cost and equipment (Clift & Morrison, 2011); this paradigm could easily be implemented almost anywhere. For these reasons, choir participation as a form of musical training appears to be uniquely suited to yield auditory, cognitive, and emotional benefits in a short period of time.

The current study

The current study investigates the effect of short-term musical training and choir participation on older adults using various measures of auditory perception, including speech-in-noise perception (SIN) and pitch discrimination (frequency difference limens; FDL); exploratory cognitive measures, including working memory (LSpan) and inhibitory control of attention (Flanker task); and neural responses to sound, as indexed by features of the frequency following response (FFR) to a repeated speech stimulus. We hypothesized that ten weeks of group choral practice (two hours weekly) and individual online musical training (up to one hour weekly) would improve all measures of auditory processing, including perceptual and neural indices of audition. We predicted that compared with their pre-training assessments, choir participants would achieve significantly lower post-training thresholds on SIN and FDL tasks, indicating

improved perceptual abilities as a result of the intervention. We also predicted that post-training, choir participants would demonstrate higher fidelity and more consistent neural responses to sound, as indexed (respectively) by improvements in F0 amplitude in fast Fourier transforms (FFTs) and inter-trial phase coherence (ITPC) of the FFR component of the auditory brainstem response. Furthermore, we predicted that training-related changes in speech-in-noise perception and pitch discrimination may have been mediated by improved neural processing; i.e., gains in FFTs and ITPCs may have directly or indirectly mediated the intervention effects on auditory perceptual measures. Exploratory cognitive measures of working memory and attention were assessed in relation to training outcomes, as potential dependent variables, and also as potential predictors or mediators of training effect on auditory and neural measures. Background measures of age, peripheral hearing loss (dB HL), and musicianship were included as potential predictors of choir effect. We also hypothesized that choir participants would experience greater post-training gains than individuals participating in a music appreciation class (music perception group), and that both musically active groups would experience greater post-training effects than an age-matched do-nothing control group.

Materials and Methods

Participants

Sixty-one choir participants (aged 50+) were recruited from a ten-week choral class run through the 50+ program at Ryerson University (CSMU100). In addition to the choir singing group, fifty-two control participants were recruited from two populations: seventeen *music perception* participants (aged 50+) were recruited from 8-10 week music appreciation classes offered through the Royal Conservatory of Music (RCM) and the 50+ program at Ryerson University, and thirty-five *do-nothing control* participants (aged 50+) were recruited through advertisements placed in the Toronto Star and hearing clinics in the greater Toronto area. The choir and control groups were matched on age, sex, peripheral hearing loss, and number of classes attended; dependent variables included auditory perceptual measures (speech-in-noise perception, pitch discrimination, and the frequency following response) and cognitive measures (working memory and inhibitory control of attention), so baseline scores in these measures were allowed to vary between groups. Participants had normal or corrected-to-normal vision; participants with medical conditions and neurological issues were accepted into the study, but details were noted for consideration in future analysis. Participants who used assistive technology (e.g., hearing aids) in daily life were accepted into the study; hearing aids were allowed during the behavioural assessments, to replicate real-life experiences, but were not worn during the audiometry or EEG components of the experiment. Informed consent was obtained from each volunteer prior to their participation in the study, in accordance with the Ryerson Research Ethics Board (REB) guidelines. Each participant was compensated \$60 upon completion of the post-training assessment, or \$30 per 3-hour assessment. Participants who

wished to undergo follow-up testing after additional choir sessions and musical training were compensated \$30 per subsequent data collection session.

The current study and participant breakdown included pilot data collected by Saul Moshé-Steinberg (2015), involving a subset of data collection (speech perception in noise and pitch discrimination threshold scores) for 14 choir participants, and 10 control participants (Moshé-Steinberg, 2015).

Study design

Each choir participant (n = 61) came into the Science of Music, Auditory Research and Technology (SMART) Laboratory for a pre-training assessment that took approximately 3 hours, during which time they completed several questionnaires and auditory and cognitive assessments, and underwent an EEG during presentation of repeated auditory stimuli. If possible, this pre-training assessment took place before the beginning of their choir course; if not, it occurred within two weeks of the first choir session.

Choir singing participants took part in weekly two-hour group choral sessions over the course of ten weeks, during which time they received pitch training and vocal direction in an open and encouraging environment. In addition to the weekly group choir sessions, participants were assigned weekly individual online musical and vocal training exercises (up to one hour weekly). This training consisted of pitch discrimination and vocal production exercises designed to target and improve the participants' abilities to perceive and produce small changes in pitch (Theta Music Trainer; <https://trainer.thetamusic.com/en>).

After ten weeks of choir participation and online musical training, each choir participant returned to the SMART Lab for a post-training assessment that lasted approximately 2.5 hours. During this session, participants completed different versions of the original assessments, and

underwent a post-training EEG of auditory brainstem activity. To account for possible differences in version difficulty within the matched behavioural tasks, participants were assigned one of four possible counterbalanced configurations of assessments or *experimental conditions* (see Tables 1 and 2) to be included in the regression analysis.

Control participants (n = 52) underwent the same battery of pre- and post-testing over the same time frame as the choir participants, but received different (group-specific) interventions over the ten-week training period.

In the *music perception* group, participants (n = 17) took part in either an 8-week or a 10-week musical appreciation class which emphasized analytic listening to musical excerpts. The inclusion of this control group in the analysis intended to elucidate the contribution of musical engagement to intervention outcomes, by contrasting experimental outcomes of a group being trained to sing (*music production*) with a group being trained to listen (*music perception*). This was an attempt to account for the unique contribution of *active musical production* to intervention outcomes, in a context where benefits may also be conferred by musical exposure and perceptual engagement.

The *do-nothing control group* (n = 35) underwent the same battery of pre- and post-testing, with 8-10 weeks between data collection sessions, but did not receive any active training during this time. The inclusion of this control group in the analysis intended to account for any practice effects within the repeated measures, enabling a controlled examination of the unique effects of the musical intervention on experimental outcomes.

In terms of the use of the term ‘intervention’ in this context, this study examined the efficacy of musical engagement in mitigating age-related declines in auditory processing in older adults, in a controlled experimental setting without random group assignment. The study design

and participant recruitment relied on older adults enrolling and completing various courses – in particular, the choir singing course offered through Ryerson University’s 50+ Program. This design was chosen due to the time commitment and intensive nature of the intervention, as a viable way to recruit a large sample of participants; as such, this design lacked randomized control group assignment (a common practice in clinical intervention studies). The use of ‘intervention study’ in this case refers to a multi-arm pre-post study examining possible changes in outcome measures – primarily, auditory processes – that occurred as a result of group involvement in either a choir singing group, a music appreciation group, or a do-nothing control group (e.g., Thiese, 2014).

Experimental procedure

Apart from the questionnaires, all assessments were completed in an Industrial Acoustics Company (IAC) double-walled sound-attenuating booth. Computerized assessments were presented using a Mac mini (Apple, 2010), with visual components of the experiment presented on a 24” Acer LCD display (Acer X243w, 1920 x 1200) placed at eye level approximately 0.5m in front of the participant. Audiometric testing and FFR auditory stimuli were administered through binaural foam insert headphones (Electro-Medical Instruments, 3A) connected to a GSI 61 Clinical Audiometer (VIASYS Healthcare). All other auditory assessments were administered binaurally through Koss SB40 headphones.

Before the experiment began, participants were familiarized with task requirements and response methods for each assessment. Participants were monitored throughout the data collection session.

Questionnaires. After signing the consent form and going over experimental expectations and volunteer rights, participants were given background, tinnitus and music history

questionnaires. The background questionnaire elicited demographics and medical history; the tinnitus questionnaire assessed the incidence and degree of tinnitus symptoms. The music history questionnaires assessed the extent of musical training and experience, including age of onset of musical training, years of continuous formal training, lifelong experience (approximate hours practiced per week during each five-year period), and current musical involvement (approximate hours practiced per week during the past year and past decade). These aspects of musical experience have been shown to influence auditory perceptual measures (Zendel, 2011) and neural responses to sound (Kraus & Chandrasekaran, 2010); these measures were included in the analysis to account for the influence of baseline musical experience on training outcomes, and to enable future analyses of individual variability in baseline measures.

Auditory measures. *Pure-tone audiometry.* Peripheral hearing loss was assessed through the conduction of a pure-tone audiometry test, to determine sound sensation at the level of the ear. Participants seated in a sound-attenuated booth equipped with a GSI 61 Clinical Audiometer (VIASYS Healthcare) were presented with monaural pulsed pure tones at test frequencies of 250 Hz, 500 Hz, 1000 Hz, 2000 Hz, 3000 Hz, 4000 Hz, 6000 Hz, and 8000 Hz. Participants were asked to press a response button whenever they heard a tone. The assessment began with presentation of a 1000 Hz tone to the left ear at 40 dB HL; if the participant indicated that they heard a tone, the amplitude was reduced by 5 dB HL. This continued until the participant no longer responded to presented tones; at this point, the amplitude of the tone was increased by 5 dB HL. If the participant responded, the amplitude was again lowered by 5 dB HL; with no response, the amplitude was increased by 5 dB HL. A researcher alternated increasing and decreasing the amplitude of the pure tone presentation until they were confident they achieved consistent responses; the lowest amplitude presentation that the participant

consistently responded to was recorded as the pure-tone threshold for that frequency in that ear. If the participant did not respond to the initial tone presented at 40 dB HL, the amplitude was increased in 5 dB increments until the participant was able to detect the tone; at this point, it was increased an additional 10 dB, and threshold detection procedure was continued as stated above (alternating around the lowest detectable amplitude until a consistent threshold was established). Each tone was presented at irregular time intervals to prevent false positives in responses. This procedure was repeated for all test frequencies, and the 1000 Hz frequency was tested a second time to ensure that the initial threshold measurement was not compromised due to lack of familiarity with the test procedure. The entire procedure was then repeated for the right ear.

Speech-in-noise perception: Signal-to-Noise Ratio (SNR). Ability to track speech in a noisy environment was assessed using the QuickSIN test (Speech-In-Noise; Etymotic Research; Killion, Niquette, Gudmundsen, Revit, & Banerjee, 2004). Participants were presented with four sets of six pre-recorded sentences, with five key words per sentence embedded in four-talker babble noise. In this assessment, the sentences were presented binaurally with a decreasing signal-to-noise ratio (SNR): the first sentence was presented with an SNR of 25 dB (i.e., the target sentence was twenty-five times louder than the background noise; *very easy*), each subsequent sentence was presented with a -5 dB SNR reduction, to an SNR of 0 dB for the final sentence (i.e., no difference between signal and background babble volume; *very difficult*). Participants were asked to repeat back the target sentences as closely to what they heard as possible, and were awarded one point for each correctly repeated target word, for a possible total of thirty points per set. The sentences in the QuickSIN do not contain many semantic or contextual cues, despite being syntactically correct (Wilson, McArdle, & Smith, 2007). Out of the four sets of sentences presented, the first two lists were treated as practice sets, to familiarize

participants with the task requirements, and the second two lists were scored as experimental data. SNR loss for each list was calculated by subtracting the total number of correct words from 25.5; this number represents the signal-to-noise ratio required for the participant to correctly identify 50% of the key words in the target sentences (Killion, 1997; Killion et al., 2004). Final scores were calculated by averaging the scores of the two experimental lists; since this is a threshold assessment, a more negative SNR score indicates better performance. Participants' responses were scored online by a researcher, and were also recorded using Audacity software in case response ambiguity necessitated further review. The pre- and post-testing used different sentence sets in order to avoid practice effects, and participants' exposure to the sets were counterbalanced across experimental conditions, to be included in the regression analysis to account for any differences in overall set difficulty (see Table 1).

Pitch discrimination: Frequency Difference Limens (FDL). Participants' ability to distinguish different frequencies was measured using a computerized assessment of frequency difference limens (FDL). In this task, participants were presented with 3 pure tones, each lasting 200ms, with amplitude envelopes of 20ms rise and delay times. A three-alternative forced choice paradigm was used, in which each presented set contained two pure tones at the standard 500 Hz frequency, and one stimulus at a randomly selected higher frequency (Parbery-Clark, Skoe, Lam, et al., 2009; F. A. Russo et al., 2012; B. Schneider, 1997). The participant was instructed to identify which tone was higher than the other two by pressing the corresponding number on a computer keyboard (i.e., 1 = first tone is higher; 2 = second tone is higher; 3 = third tone is higher). An adaptive staircase paradigm was used to determine the pitch discrimination threshold, whereby the difference between standard and comparison frequencies was halved after three correct responses, or doubled after one incorrect response. After five reversals, the step was

changed, so that the frequency difference was divided by 1.414 after 3 correct responses or multiplied by 1.414 after one incorrect response. The least detectable difference was considered an index of pitch discrimination ability.

EEG measure: the frequency following response (FFR). *Stimulus.* Auditory presentation of a repeated /da/ syllable ($F_0 = 114$ Hz) was used to elicit the frequency following response (FFR), following methodological conventions described by Skoe & Kraus (2010). This stimulus was selected because it is a complex speech sound that has been extensively used in this area of research, and robustly elicits clear auditory brainstem responses (Parbery-Clark et al., 2012; Parbery-Clark, Skoe, & Kraus, 2009; N. M. Russo, Nicol, Zecker, Hayes, & Kraus, 2005; Skoe & Kraus, 2010). Each participant heard 6000 repetitions of this 170ms sound, presented at alternating polarities. Stimuli were presented binaurally through insert headphones; stimulus volume was set to 45 dB SPL for normal hearers. For individuals with hearing loss above 25 dB, presentation volume was set to $45 \text{ dB} + (\text{dB HL} - 25 \text{ dB})$, controlling stimulus levels for sensory loss across all participants.

EEG administration and data collection. EEG data were collected using a vertical one-channel montage configuration, using three electrodes; in which active and reference electrodes were placed on the mastoids, and a ground electrode was placed on forehead. A researcher applied 1" square cloth solid gel electrodes (EL504, BIOPAC Systems, Inc.) to the mastoids and forehead; electrodes were connected to a BIOPAC MP150 data acquisition system and ERS100C Evoked Response Amplifier (BIOPAC Systems, Inc.). Data were recorded at a sampling rate of 20 kHz, with a low-pass filter of 10 kHz and a high-pass filter of 1 Hz, and the signal was recorded using Acknowledge software (AcqKnowledge, version 4.1).

Stimuli were presented for 25 minutes in total, during which time participants shown a silent film (<http://www.openculture.com/free-silent-films>), to promote relaxation and stillness during the EEG.

EEG data processing. EEG data were processed in MATLAB, using the PHZLAB toolbox (Nespoli, 2016). A 100-3000 Hz band pass filter was applied, and data were segmented according to individual stimulus responses (i.e., 6000 segments), with epoch windows extending 40ms pre-stimulus onset and 10ms post-stimulus onset (Skoe & Kraus, 2010). The 40ms preceding stimulus onset was used as a baseline of ambient EEG activity, against which to compare the response activation. Peak amplitudes in the response waveform were compared to the baseline; response peaks with absolute amplitudes that did not exceed the baseline were not considered 'reliable' (Skoe & Kraus, 2010). Myogenic artifacts, which are many times larger than the brainstem response, were accounted for by rejecting all trials with amplitudes that exceeded a threshold of 35 μ V. Responses that remained after artifact rejection were averaged (accounting for polarity differences). Peak amplitude of the fundamental frequency (a measure of response fidelity) was calculated by applying a fast Fourier transform (FFT) to the averaged signal; inter-trial phase coherence (ITPC; a measure of response consistency) was calculated by finding the latency variations across each participant's un-averaged signal.

Cognitive measures. Cognitive assessments were administered electronically and automatically on stimulus computer (see experimental procedure). Assessment scripts, coded in HTML-5, were retrieved from the Millisecond online database (<http://www.millisecond.com/download/library/>, 2016) and run using Inquisit software (version 5.0.6). The original assessment scripts were adapted to conform to our experimental protocol, including altered task exposure and duration. Before running each automated assessment, a

researcher informed the participant as to the task instructions and protocol; spoken instructions were supplemented with visual presentation of task instructions. Each assessment had self-paced components, during which participants were able to take breaks while remaining in the booth. Due to the self-paced nature, duration of each assessment will be variable; participants were expected to spend between seven and fifteen minutes on each cognitive assessment.

Working memory: Listening Span (LSpan). Working memory was assessed using a computerized version of the listening span task (LSpan), an auditory adaptation of the reading span task developed by Daneman and Carpenter (1980). In this task, participants were asked to memorize sequences of letters spoken through their headphones, while also listening to sentences and determining sentence sensibility. The participant would hear a sentence, indicate whether it made sense, and then hear a letter to remember. This was followed by another sentence, sensibility response screen, and another spoken letter; this process was repeated to generate sets of letters and sentences ranging from length 2 (i.e., two sentences to judge and two letters to remember) to length 7 (i.e., seven sentences to judge and seven letters to remember). At the end of each set of letters and sentences, participants were shown a letter recall screen, on which they were instructed to indicate all the letters they heard in the correct order. Letters in the wrong position were counted as errors; the participant had the option of selecting BLANK as a placeholder for a letter they were unable to remember, to maintain overall letter order. After each response screen, participants received feedback on both the number of letters recalled correctly and the number of correct sentence judgments for that set. Throughout the task, the overall percentage of correct sentence judgments for the entire experiment was shown; participants were instructed to maintain 85% correct or higher. This was to ensure that participants paid attention to the distractor task (sentence judgments) in order to ascertain an accurate span measure. LSpan

was calculated as the sum of the lengths of all perfectly recalled sets, as per the absolute scoring method. However, partial-credit load scoring was also applied to calculate the total number of letters recalled in the correct position, to be considered in future analyses and capture another aspect of working memory, as suggested by Conway et al. (2005). Sentence accuracy was also reported, and data from participants whose accuracy fell below 85% were discarded.

Inhibitory control of attention: Flanker task. Aspects of attentional processing were assessed using a computerized version of an adapted Flanker task (Ridderinkhof, van der Molen, Band, & Bashore, 1997). In this experiment, participants were presented with five arrows inside a box on the computer screen, and were instructed to identify the direction of the center arrow by pressing one of two buttons on the computer keyboard. Arrows on either side ('flankers') of the central arrow were either *congruent* (facing the same direction as the central arrow) or *incongruent* (facing a different direction than the central arrow); difference in mean reaction time between congruent and incongruent conditions was calculated to determine the *Flanker effect*. In addition to experimental manipulation of congruency, the center arrow was sometimes *bigger* or *smaller* than the other arrows, and response keys were either *compatible* (e.g., the response key on the left was used to indicate left pointing arrows) or *incompatible* (e.g., the response key on the left was used to indicate right pointing arrows). Response key compatibility and size of central arrow was noted and considered for future analyses, as suggested by Ridderinkhof et al., (1997). The program reported the proportion of correct responses overall, the mean reaction time for the entire experiment, and the mean reaction times for both congruent and incongruent conditions, to be examined as indices of attentional processing.

Statistical analyses

Linear mixed effects (LME) modelling in a multilevel regression format was used to assess whether ten weeks of choir practice and musical training resulted in improvements to auditory measures, including speech-in-noise (SIN) perception and pitch discrimination (frequency difference limens; FDL); neural measures, including fidelity (fast Fourier transform; FFT) and consistency (inter-trial phase coherence; ITPC) of the frequency following response (FFR); and exploratory cognitive measures, including working memory (LSpan) and attentional control (Flanker task). LME were used to examine the extent to which different predictors contribute to gains in key auditory and neural abilities while accounting for individual variability across participants. Interaction effects and mediation analyses investigated the relationships between variables; in particular, pitch discrimination was examined as a potential mediator for training-induced changes in speech in noise perception. Post-hoc analyses, in the form of t-tests on pre- and post-testing scores of each group, were conducted to determine directionality and magnitude of effects.

Unique effects of choir participation and musical training were determined by examining the differences between intervention conditions; choir singing group and music perception group were contrasted against the do-nothing control group. This was intended to determine whether the effects of either music production training or music perception training significantly differed from any practice effects of the repeated measures. Statistical analyses were conducted in R; the *nlme* package was used to conduct linear mixed effects modelling in a multilevel regression format (Pinheiro et al., 2017).

Using LME with multilevel modelling in a regression format relies upon the comparison of the goodness of fit of various models, finding significance when the inclusion of an additional

predictor or interaction term significantly improves the fit of the model (accounting for significantly more unique variance), as reported in a chi-squared test. As such, all significant regression findings will be reported as chi-square statistics with associated p-values, as is customary in LME / MLM contexts.

Results

Participants

Of the 113 older adults originally recruited for the study, 95 participants completed both pre- and post-testing. These individuals were distinguished on the basis of their participation: choir participants ($n = 50$, 43 female), music listening participants ($n = 13$, 9 female), and do-nothing control participants ($n = 32$, 26 female). The choir singing group (aged 54 – 87; $M = 68.28$, $SD = 6.28$), music listening group (aged 53 – 82, $M = 69.08$, $SD = 6.97$), and do-nothing control group (aged 60 – 76, $M = 68.06$, $SD = 4.81$) were matched in terms of age; males (aged 54 – 75, $M = 68.59$, $SD = 5.42$) and females (aged 53-87, $M = 68.26$, $SD = 6.00$) were also age-matched both between and within groups. The do-nothing control group was recruited to match the choir singing group in terms of age, sex, and peripheral hearing loss; as the music listening group was recruited from existing classes, this between-group matching was coincidental.

Auditory measures

Pure-tone audiometry. Average binaural pure-tone audiometric thresholds did not differ significantly between groups (see Figure 1). A regression analysis showed that age accounted for significant variance in pure-tone audiometry thresholds, $\chi^2(1) = 36.110$, $p < 0.001$ (see Figure 2). This relationship between age and audiometry (overall model: $b = 1.1628$, $t(82) = 6.6363$, $p < 0.0001$) did not vary significantly between the choir singing group ($b = 1.2210$, $t(41) = 5.1260$, $p < 0.0001$), music listening group ($b = 0.8434$, $t(11) = 2.9679$, $p = 0.0128$), and do-nothing control group ($b = 1.2658$, $t(26) = 3.1248$, $p = 0.0043$); see Figure 2. Because of the collinearity

between age and audiometry, only audiometry measures will be used as a predictor in regressions considering dependent variables.

Speech-in-noise (SIN) perception. A linear mixed effects model was used to assess the effects of session and group on signal-to-noise ratio (SNR), the participants' thresholds for perceiving speech in noisy settings. Fixed effects were Session, Group, and the Session x Group interaction term. The relationship between Session and SNR showed significant variance in intercepts across participants, $SD = 2.47$ (95% CI: 2.10, 2.91), $\chi^2(1) = 79.15$, $p < .0001$, so random effects of Participant were included in the model to account for individual variability.

There were no main effects of Session ($\chi^2(1) = 2.141$, $p = 0.1435$) or Group ($\chi^2(2) = 1.431$, $p = 0.4889$) on SNR, but there was a significant effect of the Session x Group interaction on SNR, $\chi^2(2) = 6.948$, $p = 0.0310$. In order to break down this interaction, contrasts were assigned to Session and Group, comparing pre- and post-training scores of each intervention group with the pre-training SNR scores of the do-nothing control group. The contrasts revealed that a post-training effect for the choir singing group was the only significant relationship, $b = -1.0812$, $t(92) = -2.453$, $p = 0.0160$. Post-hoc pairwise t-tests confirmed that choir participants' post-training SNR scores ($M = 2.91$, $SD = 3.38$) were significantly better than their pre-training SNR scores ($M = 3.71$, $SD = 3.54$); $t(49) = 3.3831$, $p = 0.0014$, while no pre-post SNR differences were found in either the music perception class ($t(12) = -0.2855$, $p = 0.7801$) or do-nothing control group ($t(31) = -0.7460$, $p = 0.4612$; see Figure 3 and Table 2).

While there was no significant main effect of Group on baseline SNR scores ($\chi^2(2) = 3.667$, $p = 0.1599$), the pre-training SNR scores of the choir singing group ($M = 3.71$, $SD = 3.54$) were numerically higher than pre-training SNR scores of the music perception ($M = 2.31$, $SD = 2.48$) and do-nothing control groups ($M = 2.70$, $SD = 1.85$); $b = 1.006$, $t(92) = 1.515$, $p = 0.1333$

(see Figure 3). In order to determine whether choir-induced changes in SNR were due to a restoration of speech-in-noise perception to age-normal levels, or an improvement effect, a subset of 32 choir participants were matched with the 32 do-nothing control participants in terms of baseline SNR, and re-analyzed. This analysis yielded a significant session x group interaction ($\chi^2(2) = 10.295$, $p = 0.0058$), and contrasts revealed that a post-training effect for the choir singing group was the only significant difference from the do-nothing control group, $b = -1.4219$, $t(74) = -3.062$, $p = 0.0031$. A post-hoc pairwise t-test confirmed that choir participants' post-training SNR scores ($M = 1.58$, $SD = 1.51$) were significantly better than their pre-training SNR scores ($M = 2.72$, $SD = 1.67$); $t(31) = 5.4396$, $p < 0.0001$. This showed that when baseline SNR was controlled for, the choir singing group still experienced significant improvements, while the other two groups did not (see Figure 4 and Table 3). An additional baseline matching analysis was conducted, in which all choir participants ($n = 42$) were included if they had a baseline SNR score that was matched by that of a participant in the do-nothing control group ($n = 32$). This analysis was consistent with the previous baseline matching analysis, whereby choir participants' post-training SNR scores ($M = 1.96$, $SD = 1.71$) were significantly better than their pre-training SNR scores ($M = 2.52$, $SD = 1.54$; $t(41) = 2.3092$, $p = 0.0260$), while neither control group demonstrated post-testing improvements.

Pitch discrimination. A linear mixed effects model was used to assess the effects of session and group on frequency difference limens (FDL; Hz), the participants' thresholds for discriminating different pitches. Session, group, and the session x group interaction were considered as fixed effects. The relationship between Session and FDL showed significant variance in intercepts across participants, $SD = 7.59$ (95% CI: 6.17, 9.35), $\chi^2(2) = 44.99$, $p < .0001$. In addition, the slopes varied across participants, $SD = 4.87$, $\chi^2(4) = 13.71$, $p = 0.0083$,

and the slopes and intercepts were negatively and significantly correlated, $cor = -.70$; for these reasons, random effects of ~Session|Participant/Group (participants nested within each intervention group) were included in the model to account for individual variability.

There was a main effect of Session ($\chi^2(1) = 14.066$, $p = 0.0002$), but no main effect of Group ($\chi^2(2) = 2.862$, $p = 0.2391$) and no significant Session x Group interaction ($\chi^2(2) = 0.499$, $p = 0.7791$). A contrast analysis revealed an overall post-training improvement for all participants (main effect of post-training session; $b = -3.4223$, $t(65) = -2.225$, $p = 0.0296$), suggesting the possibility of practice effects for this task. Post-hoc pairwise t-tests confirmed that choir participants' post-training FDL scores ($M = 8.99$, $SD = 7.82$) were significantly better than their pre-training FDL scores ($M = 12.25$, $SD = 10.67$); $t(38) = 2.6343$, $p = 0.01213$; do-nothing control participants also demonstrated significantly better scores in their second session ($M = 12.11$, $SD = 8.58$) compared with their first session ($M = 15.54$, $SD = 11.41$; $t(17) = 3.0268$, $p = 0.0076$), and while the music perception class didn't achieve significance, they appeared to exhibit the same trend, with numerically better FDL scores post-training ($M = 7.78$, $SD = 4.10$) than pre-training ($M = 9.61$, $SD = 4.95$; see Figure 5 and Table 4).

While there was no significant main effect of Group on baseline FDL scores ($\chi^2(2) = 2.534$, $p = 0.2816$), the intercepts of the three groups varied, and the group with the lowest intercept (music perception group) was the only group to not achieve significant post-training gains in FDL (see Figure 5). In order to control for intercept differences, a subset of 11 choir participants and 11 do-nothing control participants were matched with the 11 music perception participants in terms of baseline FDL scores, and re-analyzed. This analysis yielded a main effect of Session ($\chi^2(1) = 15.918$, $p < 0.0001$), but no main effect of Group ($\chi^2(2) = 0.741$, $p = 0.6903$) and no significant Session x Group interaction ($\chi^2(2) = 4.436$, $p = 0.1088$). However, a contrast

analysis revealed that while all groups experienced marginally significant post-training gains as a result of the interventions ($b = -1.7220$, $t(30) = -1.758$, $p = 0.0889$), a post-training effect for the choir singing group was the only marginally significant difference from the do-nothing control group, $b = -2.5458$, $t(30) = -1.838$, $p = 0.0760$. This suggests that the choir singing group experienced greater improvements in FDL scores as a result of the intervention than the do-nothing control group, while the gains of the music perception group did not significantly differ from the do-nothing control group. Post-hoc pairwise t-tests confirmed that when baselines were matched, choir participants' post-training FDL scores ($M = 5.44$, $SD = 2.24$) were significantly better than their pre-training FDL scores ($M = 9.71$, $SD = 4.61$); $t(10) = 3.7454$, $p = 0.0038$, and do-nothing control participants' post-training FDL scores ($M = 7.78$, $SD = 2.98$) were significantly better than their pre-training FDL scores ($M = 9.50$, $SD = 4.11$); $t(10) = 2.3203$, $p = 0.04275$. While the music perception class didn't achieve significance, they appeared to exhibit the same trend as the do-nothing control group, with numerically better FDL scores post-training ($M = 7.78$, $SD = 4.10$) than pre-training ($M = 9.61$, $SD = 4.95$; see Figure 6 and Table 5). Post-training FDL scores for the choir singing group were numerically better than post-training scores for the other music perception and do-nothing control groups (see Figure 6 and Table 5).

Frequency following response (FFR)

Due to issues with the data processing toolbox used, only preliminary FFR data from the choir singing group was available at the time of analysis. The toolbox (PHZLAB; Nespoli, 2016) is in the early stages of development, and, after preliminary analyses were conducted, underwent an update that precluded ongoing data processing with the current FFR files. However, this issue will likely be remediated in a future update of the toolbox (the developer is aware of the problem and is working to correct it); for this reason, FFR data from the do-nothing control group, music

perception group, and additional choir participants will be processed and analyzed in a future study. Pairwise t-tests conducted on preliminary EEG data for choir participants revealed a marginally significant improvement in the magnitude of the neural representation of the fundamental frequency (FFT at 114Hz) between pre-training (M = 0.0135, SD = 0.0132) and post-training (M = 0.0140, SD = 0.00999); $t(12) = -2.0353$, $p = 0.06452$. Pairwise analyses also revealed a marginally significant improvement in neural response consistency, as measured by inter-trial phase coherence (ITPC), between the pre-training (M = 0.0928, SD = 0.0530) and post-training (M = 0.1492, SD = 0.0782) assessments for choir participants; $t(13) = -2.0399$, $p = 0.06222$.

Cognitive measures

Working memory. A linear mixed effects model was used to assess the effects of session and group on listening span (LSpan), a measure of auditory working memory. Session, group, and the session x group interaction were considered fixed effects. The relationship between Session and LSpan showed significant variance in intercepts across participants, SD = 12.59 (95% CI: 10.32, 15.36), $\chi^2(1) = 50.021$, $p < .0001$, so random effects of Participant were included in the model to account for individual variability. There was a significant main effect of Session ($\chi^2(1) = 8.369$, $p = 0.0038$), but no significant effect of Group ($\chi^2(2) = 2.550$, $p = 0.2794$) or the Session x Group interaction ($\chi^2(2) = 1.867$, $p = 0.3933$) on Listening Span score. Contrasts revealed that there were no significant differences in the way the different interventions affected the LSpan scores, with all groups improving. Post-hoc t-tests revealed that the music perception group showed significant improvements from pre-training (M = 28.00, SD = 11.28) to post-training (M = 35.75, SD = 11.86), $t(11) = -2.389$, $p = 0.03593$, while the choir singing group demonstrated marginally significant improvements from pre-training (M = 25.03,

SD = 15.91) to post-training (M = 28.58, SD = 15.55), $t(35) = -1.8867$, $p = 0.06752$, and the do-nothing control group did not significantly improve (see Figure 7 and Table 6).

Total correct letters (TCL) were also assessed, to see whether partial credit scoring yielded any more detailed or sensitive findings with respect to auditory working memory. The analysis was set up the same way as it was for LSpan scores, but with TCL as the dependent variable. The relationship between Session and LSpan showed significant variance in intercepts across participants, SD = 13.37 (95% CI: 11.21, 15.95), $\chi^2(1) = 96.349$, $p < .0001$, so random effects of Participant were included in the model to account for individual variability. The initial analysis yielded similar findings as the overall LSpan regression, with a significant main effect of Session ($\chi^2(1) = 14.413$, $p = 0.0001$), but no significant effects of Group or the Session x Group interaction term in the linear mixed effects analysis. Contrasts revealed that post-training TCL was marginally improved over pre-training ($b = 2.458$, $t(69) = 1.698$, $p = 0.0941$), but the relationship did not vary between groups. Post-hoc t-tests revealed that while all groups demonstrated improved TCL scores over the intervention period, in this case, the choir singing group showed significant improvements from pre-training (M = 43.11, SD = 15.90) to post-training (M = 46.81, SD = 14.40), $t(35) = -2.9807$, $p = 0.005203$; and the do-nothing control group also demonstrated marginally significant improvements from pre-training (M = 41.54, SD = 15.12) to post-training (M = 44.00, SD = 13.61), $t(23) = -1.872$, $p = 0.07398$, while the music perception group did not significantly improve (see Figure 8 and Table 7).

Inhibitory control of attention. A linear mixed effects model was used to assess the effects of session and group on reaction time measures from the Flanker task, a measure of inhibitory control of attention (calculated as mean RT incongruent - mean RT congruent). Session, group, and the session x group interaction were considered fixed effects. There were no

significant effects of Session, Group, or the Session x Group interaction term on the Flanker effect; however, the choir singing group appeared to demonstrate a different trend than the other two groups (see Figure 9). In order to elucidate this trend, this analysis was conducted on overall mean reaction times, as well as mean reaction times for congruent and incongruent conditions.

The relationship between Session and Flanker mean RT showed significant variance in intercepts across participants, $SD = 92.78$ (95% CI: 76.42, 112.63), $\chi^2(1) = 56.128$, $p < .0001$, so random effects of Participant were included in the model to account for individual variability. There was a significant main effect of session on mean reaction time ($\chi^2(1) = 4.059$, $p = 0.0439$); contrast analyses showed that while all groups became significantly faster over the course of the intervention period ($b = -36.760$, $t(69) = -2.345$, $p = 0.0219$), the baseline mean RT measure varied significantly in the choir singing group ($b = -58.991$, $t(70) = -2.089$, $p = 0.0404$) and marginally significantly in the music perception group ($b = -70.488$, $t(70) = -1.910$, $p = 0.0603$), as compared with the do-nothing control group. Post-hoc pairwise t-tests revealed that only the do-nothing control group showed marginally significant differences between pre-training ($M = 815.61$, $SD = 119.20$) and post-training ($M = 778.85$, $SD = 115.67$) scores, $t(23) = 2.0446$, $p = 0.0525$ (see Figure 10 and Table 8). These findings were consistent when the analysis was broken down based on congruency condition (see Figure 11), suggesting that participants with worse baseline reaction times were likely to experience gains in this measure, as a result of practice, but that no benefits were conferred by either music perception training or choir singing.

Overview of results

Analyses showed that choir singing participants experienced significant post-testing gains in speech-in-noise perception (SNR score improvements), while the music perception and do-nothing control participants did not (with and without between-group baseline matching).

Although all groups demonstrated post-testing gains in pitch discrimination ability (FDL score improvements), when baselines scores were matched, the choir participants had marginally significantly better post-training scores than the other two groups. The choir singing group also demonstrated marginally significant improvements in both the FFT and ITPC features of the frequency following response (FFR); the other groups were not tested in this domain. These findings support the use of choir participation to improve aspects of auditory processing in older adults.

All groups demonstrated some post-training improvements in working memory span (Lspan), with the music perception group showing the most (significant) improvement, followed by the choir singing group (marginally significant improvements), and finally the do-nothing control group (numeric improvement not reaching significance). These findings were not replicated in the partial scoring method of the same task (TCL), with the choir singing group demonstrating the most significant improvements, followed by the do-nothing control group (marginally significant improvements), and finally the music perception group (numeric improvement not reaching significance). In the Flanker task, all groups improved post-testing, but no additional gains were conferred by either training group (choir singing or music appreciation classes) relative to the do-nothing control group

These results support the use of choir singing to improve measures of auditory, but not cognitive, processing in older adults. Because speech-in-noise perception was the main outcome of interest in this study, and was shown to significantly improve as a result of choir participation, moderation and mediation analyses were conducted to further elucidate gains in SNR scores. These findings are presented below.

Moderation and mediation analyses

Moderation models of SNR were limited to the baseline matched subset of the choir singing group ($n = 32$) and do-nothing control group ($n = 32$), omitting the music perception group due to its small sample size. Linear mixed effects modelling in a multilevel regression format was used to examine possible predictors of SNR scores, including auditory measures (dB HL and FDL in Hz); cognitive measures (LSpan and Flanker effect), and background musicianship (age of onset and years of continuous training), along with effects of session and group (previously found to be significant; see above). The relationship between Session and SNR showed significant variance in intercepts across participants, $SD = 2.47$ (95% CI: 2.10, 2.91), $\chi^2(1) = 79.15$, $p < .0001$, so random effects of Participant were included in the model to account for individual variability.

The regression analysis found significant main effects of group ($\chi^2(1) = 4.604$, $p = 0.0319$) and dB HL ($\chi^2(1) = 6.082$, $p = 0.0137$) on SNR scores, as well as significant interaction effects of Session x Group ($\chi^2(1) = 9.594$, $p = 0.0020$; see Figure 12) and dB HL x FDL ($\chi^2(1) = 4.014$, $p = 0.0451$; see Figure 13), and a marginally significant interaction of Session x FDL ($\chi^2(1) = 2.884$, $p = 0.0894$; see Figure 14). When contrasts were applied, significant variance was accounted for by Group (choir singing group vs do-nothing control group; $b = -7.554$, $t(30) = -2.158$, $p = 0.0390$), and marginally significant variance was accounted for by dB HL, $b = -0.211$, $t(30) = -1.759$, $p = 0.0887$; FDL (Hz), $b = -0.313$, $t(18) = -2.097$, $p = 0.0504$. As well, the interaction between dB HL and FDL accounted for significant variance in SNR scores, $b = 0.0197$, $t(18) = 2.329$, $p = 0.0317$. Interaction effects were plotted for elucidation; see Figures 11-13. Simple slopes analyses revealed that in addition to the beneficial effect of choir training

on SNR score, individuals with better peripheral hearing (i.e., lower dB HL) and better pitch discrimination (i.e., lower FDL) were better at perceiving speech in noise (see Figures 11-13).

The mediation model investigated whether choir improvements in SNR scores were mediated by gains in pitch discrimination ability (FDL (Hz); see Figure 18); as such, this analysis was limited to the choir singing group ($n = 50$). Path a in this model tested whether the independent variable, Session, accounted for significant variance in the mediator, FDL (Hz). This was calculated by regressing FDL on Session. The results indicated a significant relationship between FDL and Session; as the choir participants progressed from pre-training to post-training, FDL scores decreased by 2.761 Hz ($b = -2.761$, $SE = 1.246$, $t(38) = -2.216$, $p = 0.0327$). This supports the previous finding that choir participants demonstrate significant improvements in pitch discrimination over the course of the intervention. Path b was calculated by regressing SNR on both FDL and Session, and tested whether the mediator, FDL (Hz), accounts for significant variance in the dependent variable, SNR, when Session is included in the model. The results suggest that while holding Session constant, there is a significant relationship between FDL and SNR ($b = 0.087$, $SE = 0.026$, $t(36) = 3.390$, $p = 0.0017$). This model also includes the c' path, denoted by the estimate, standard error, and p-value of the Session coefficient in the 'mediated' model (i.e., including FDL). In this case, when SNR was regressed on both FDL and Session, there was a marginally significant relationship between Session and SNR ($b = -0.458$, $SE = 0.237$, $t(36) = -1.932$, $p = 0.0613$). Path c was calculated by regressing SNR on Session; as expected, Session accounted for significant variance in the dependent variable, SNR, $b = -0.861$, $SE = 0.231$, $t(48) = -3.726$, $p = 0.0005$. Aroian's test was conducted to determine whether the indirect mediation effect ($a*b$) was significantly different from zero. This test yielded a marginally significant mediation effect ($z = -1.801$, $SE = 0.136$, $p = 0.0717$).

Given the sample size, and the fact that Aroian's test is known to be overly conservative (i.e., has an inflated Type II error rate), a Monte Carlo resampling technique was applied to elucidate the mediation effect of FDL on SNR~Session. Using a resampling of 20000 in the Monte Carlo approach, zero was not included in the 95% confidence interval; therefore, FDL (Hz) significantly mediated the relationship between Session and SNR, 95% CI [-0.54005, -0.02167] (see Figure 19).

Because FDL was found to mediate choir-induced changes in SNR, linear mixed effects modelling in a multilevel regression format was used to examine possible predictors and moderation of FDL scores, by audiometry (dB HL), cognitive measures (Lspan and Flanker effect), and background musicianship (age of onset and years of continuous training), along with effects of session and group (previously found to be significant; see above). The regression analysis found significant main effects of session ($\chi^2(1) = 8.712, p = 0.0032$), age of onset of musical training ($\chi^2(1) = 5.130, p = 0.0235$) and years of continuous musical training ($\chi^2(1) = 4.128, p = 0.0422$) on FDL scores, as well as significant interaction effects of Session x Group ($\chi^2(1) = 10.303, p = 0.0058$; see Figure 15) and age of onset x years of continuous musical training ($\chi^2(1) = 8.054, p = 0.0045$; see Figure 16). When contrasts were applied, only the variance accounted for by the musicianship interaction (age of onset x years of continuous musical training; $b = -0.106, t(18) = -2.722, p = 0.0140$) remained significant. Interaction effects were plotted for elucidation; see Figures 14-15. Simple slopes analyses revealed that in addition to the beneficial effect of choir training on FDL score, individuals who began musical training earlier (i.e., earlier age of onset of musical training) and had more years of continuous musical training were better at perceiving differences in pitch (see Figures 14-15).

Discussion

This study demonstrated for the first time that short-term choir participation can result in improved speech-in-noise perception in older adults. Because these improvements were not observed in the music perception group or the do-nothing control, it would appear that active engagement in music was critical. Choir participation also led to improved pitch discrimination and neural representation of frequency. Although there were some practice effects of the pitch discrimination task (i.e., comparable gains were observed in all groups), when baselines FDL scores were matched, the choir singing group demonstrated significantly greater gains than the other two groups.

The study also showed that improvements in speech-in-noise perception are partially mediated by gains in pitch discrimination; this suggests that the musicianship enhancement of speech-in-noise perception may be in part due to overarching improvements in pitch-tracking ability, which transfer to perceptual benefits in tracking and discerning speech in noisy environments. One mechanism by which such an enhancement might be occurring is through improved neural processing of auditory stimuli, as indexed by the frequency following response (FFR). This study found positive trends in the neural representation of sound, in marginal improvements of response magnitude (FFT at F0) and consistency (ITPC) as a result of choir training. An important future direction for this work will be the inclusion and analysis of additional participants' FFR data, including multi-group data, to elucidate the unique contribution of music production training to auditory processing gains, and the role that this might play in facilitating auditory perceptual abilities.

Although it does not appear that baseline musicianship directly influences the effect of choir participation on SNR outcomes, it may indirectly mediate the relationship through its

influence on FDL. The analysis showed that participants with earlier age of onset of musical training and more years of continuous musical training demonstrated improved pitch discrimination ability. In turn, the gains in pitch discrimination ability partially mediated choir-induced improvements in speech-in-noise perceptual ability. This suggests that there is a musicianship advantage in pitch discrimination ability (supporting previous research findings), which may even be conferred in a short-term musical training intervention context. As individuals gain more musical experience, they get better at tracking changes in pitch, which may lead to downstream perceptual advantages. In particular, the ability to track changes in pitch may confer gains to tasks which require pitch-tracking in complex acoustic environments, such as trying to discern speech in noisy settings. This finding can serve as a starting point for future investigations into the relationship between pitch discrimination and speech perception in noise, as well as the elucidation of a possible neural substrate of this relationship (e.g., musicianship advantages to the FFR). Importantly, training-induced improvements in these abilities are visible over a short period of time (10 weeks).

Peripheral hearing loss also seemed to play a role in auditory perceptual processes, and appeared to moderate the relationship between pitch discrimination and speech-in-noise perception. Individuals with lower levels of peripheral hearing loss had better thresholds for detecting speech in noise and changes in pitch, and trended towards more benefits as a result of the choir intervention. This finding opens the door to more extensive research investigating the relationship between peripheral hearing loss and downstream perceptual measures, and may inform future intervention applications in terms of target populations (e.g., if choir singing will only benefit individuals with lower levels of hearing impairments, audiologists may want to refrain from prescribing it as a means of auditory rehabilitation for older adults with severe

hearing impairments). One important future direction for this work is a follow-up study involving hearing-aided individuals, to see whether peripheral remediation helps people with more severe hearing impairments be receptive to the benefits of musical training as an intervention. The sample of hearing aided individuals included in this study was not large enough to conduct statistical analyses, but this is a critical future direction of this work, in support of the goal to improve auditory perception and promote healthy aging in older adults.

In terms of the exploratory cognitive measures, interestingly, it appeared that the music perception class experienced the most significant benefits to their listening memory span (followed by the choir participants), while the choir singing group demonstrated the most significant improvements in the total number of letters recalled (followed by the do-nothing control group). However, the fact that the do-nothing control group also experienced some benefits suggests that a component of the gains (particularly in TCL, a more sensitive measure) may be due to practice effects. That being said, one explanation for the span findings is that a music appreciation class puts greater emphasis on a listening and / or memorization aspect, while a choir singing class, although it includes a memory component, is more focused on vocal improvement and the act of singing as compared to memorization. One limitation of this study was that the music appreciation participants were recruited from existing classes, so the intervention was not subject to the same degree of control as the choir participants and do-nothing control groups, which were both facilitated by a researcher involved in this study. Another limitation was the small sample size of the music perception vs. the other two groups. These limitations could be addressed in a future study in which the music-perception classes are administered and / or supervised by a researcher, in order to control for task demands and other

aspects of the intervention. These findings do suggest that working memory may be improved by musical interventions, but this needs to be investigated further before strong claims can be made.

In terms of the attentional task, while reaction times tended to improve in the second session of data collection for all groups, there was no apparent benefit to either of the music classes. However, the baselines of the different groups varied, with the do-nothing control participants starting out significantly worse than the other two groups, and then achieving more significant gains than the other groups. This could be addressed by recruiting a group of control participants that have more matched RT baselines with the other two groups, or could be addressed in an analysis of a subset of the participants. However, there did not appear to be significant differences between the choir and music perception groups, suggesting that this ability is not influenced by musical training.

Overall, running a choir is an immensely scalable intervention, requiring minimal cost and equipment, and this study demonstrated that training-induced improvements in auditory perception can appear after a very short intervention period (10 weeks of choir singing). This paradigm could easily be implemented almost anywhere, providing both a rich opportunity for further investigation into auditory benefits of choir singing in different populations, and a framework for a scalable and effective intervention that can target and improve auditory abilities in aging adults.

Conclusion

Group singing is an intuitive, engaging, and motivating form of music making, that has in previous studies been shown to contribute to social, emotional, cognitive, and physical well-being. The current findings suggest that choir singing can be used as an effective intervention to mitigate age-related losses in auditory perceptual abilities, in as short a time as ten weeks. Importantly, these findings showed that this intervention improved older adults' abilities to perceive speech in noisy environments, a key concern in promoting healthy aging. This work provides an empirical basis for a highly scalable and effective intervention, that could significantly improve quality of life in older adults.

Appendix

Table 1

Experimental conditions and counterbalancing.

Experiment #	QuickSIN #		FDL #		Lspan #		Flanker #	
	Pre	Post	Pre	Post	Pre	Post	Pre	Post
1	1	3	1	2	1	2	1	2
2	2	4	2	1	1	2	2	1
3	3	1	2	1	2	1	1	2
4	4	2	1	2	2	1	2	1

Table 2

Pre- and post-testing thresholds (mean \pm standard deviation) for perceiving speech in noise (signal-to-noise ratio; SNR) for choir singing participants ($n = 50$), music appreciation class participants ($n = 13$), and do-nothing control participants ($n = 32$).

	Pre-training	Post-training
Choir singing class	3.710 \pm 3.539	2.910 \pm 3.385
Music appreciation class	2.308 \pm 2.479	2.500 \pm 1.1902
Do-nothing control group	2.703 \pm 1.849	2.984 \pm 2.104

Table 3

Baseline-matched pre- and post-testing thresholds (mean \pm standard deviation) for perceiving speech in noise (signal-to-noise ratio; SNR) for choir singing participants ($n = 32$), music appreciation class participants ($n = 13$), and do-nothing control participants ($n = 32$).

	Pre-training	Post-training
Choir singing class	2.719 \pm 1.670	1.578 \pm 1.509
Music appreciation class	2.308 \pm 2.479	2.500 \pm 1.1902
Do-nothing control group	2.703 \pm 1.849	2.984 \pm 2.104

Table 4

Pre- and post-testing thresholds (mean \pm standard deviation) for pitch discrimination (frequency difference limens; FDL in Hz) for choir singing participants ($n = 39$), music appreciation class participants ($n = 11$), and do-nothing control participants ($n = 18$).

	Pre-training	Post-training
Choir singing class	12.248 \pm 10.671	8.992 \pm 7.818
Music appreciation class	9.610 \pm 4.953	7.777 \pm 4.104
Do-nothing control group	15.535 \pm 11.407	12.113 \pm 8.584

Table 5

Baseline-matched pre- and post-testing thresholds (mean \pm standard deviation) for pitch discrimination (frequency difference limens; FDL in Hz) for choir singing participants ($n = 11$), music appreciation class participants ($n = 11$), and do-nothing control participants ($n = 11$).

	Pre-training	Post-training
Choir singing class	9.701 \pm 4.612	5.438 \pm 2.242
Music appreciation class	9.610 \pm 4.953	7.777 \pm 4.104
Do-nothing control group	9.502 \pm 4.114	7.780 \pm 2.975

Table 6

Pre- and post-testing scores (mean \pm standard deviation) in listening working memory span (Listening Span Task; LSpan) for choir singing participants ($n = 36$), music appreciation class participants ($n = 12$), and do-nothing control participants ($n = 24$).

	Pre-training	Post-training
Choir singing class	25.028 \pm 15.911	28.583 \pm 15.548
Music appreciation class	28.000 \pm 11.277	35.750 \pm 18.859
Do-nothing control group	22.875 \pm 15.295	25.250 \pm 11.241

Table 7

Pre- and post-testing scores (mean \pm standard deviation) in total correct letters recalled (TCL; LSpan) for choir singing participants (n = 36), music appreciation class participants (n = 12), and do-nothing control participants (n = 24).

	Pre-training	Post-training
Choir singing class	43.111 \pm 15.904	46.806 \pm 14.402
Music appreciation class	47.769 \pm 9.909	52.333 \pm 14.118
Do-nothing control group	41.542 \pm 15.126	44.000 \pm 13.606

Table 8

Pre- and post-testing mean reaction times (mean \pm standard deviation) for Flanker task (ms) for choir singing participants (n = 36), music appreciation class participants (n = 12), and do-nothing control participants (n = 24).

	Pre-training	Post-training
Choir singing class	756.620 \pm 95.853	744.073 \pm 103.727
Music appreciation class	745.123 \pm 116.581	736.976 \pm 94.563
Do-nothing control group	815.611 \pm 119.199	778.851 \pm 115.673

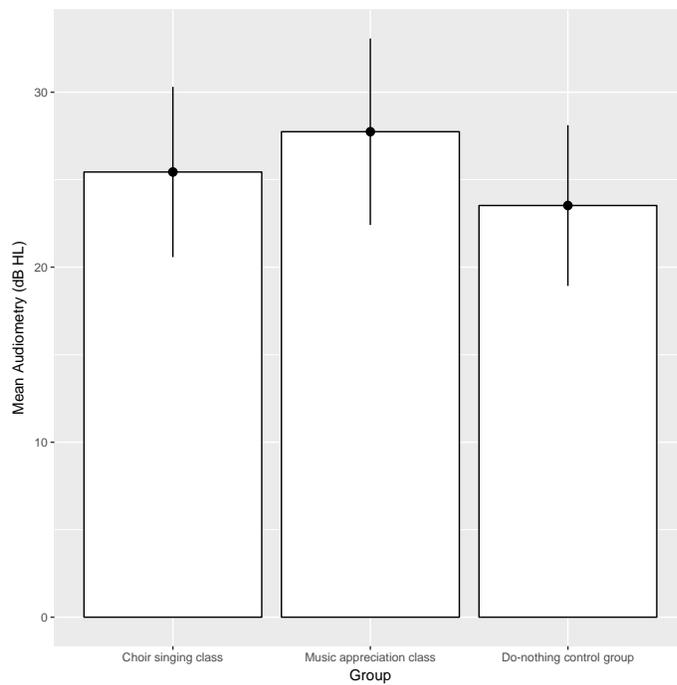
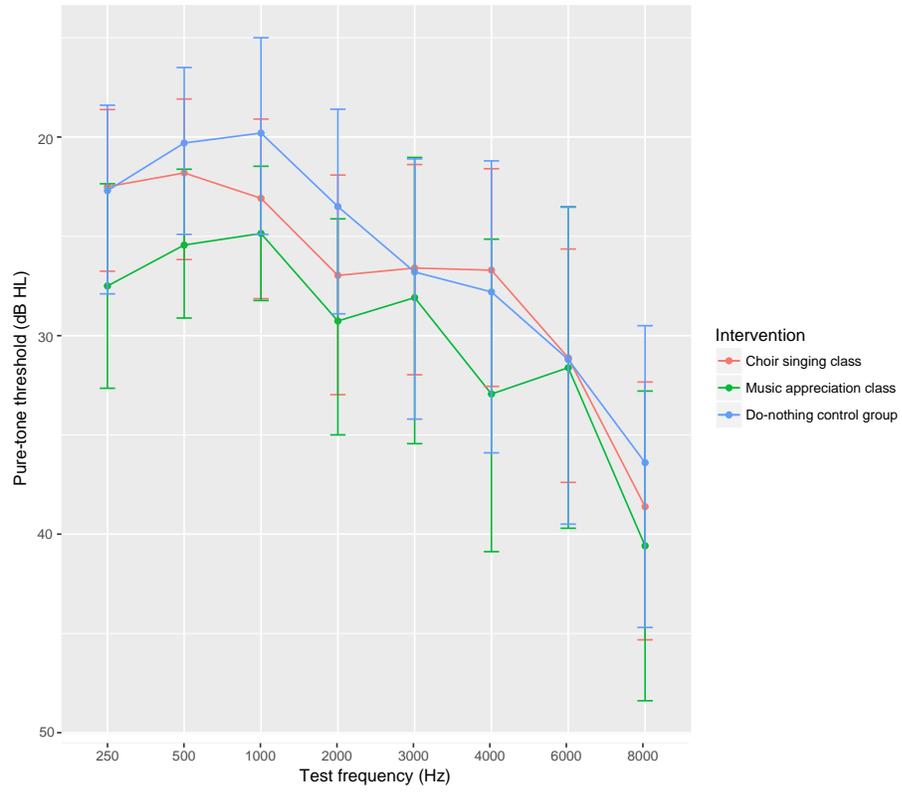


Figure 1: Group mean audiograms (top) and average dB HL (bottom) by group.

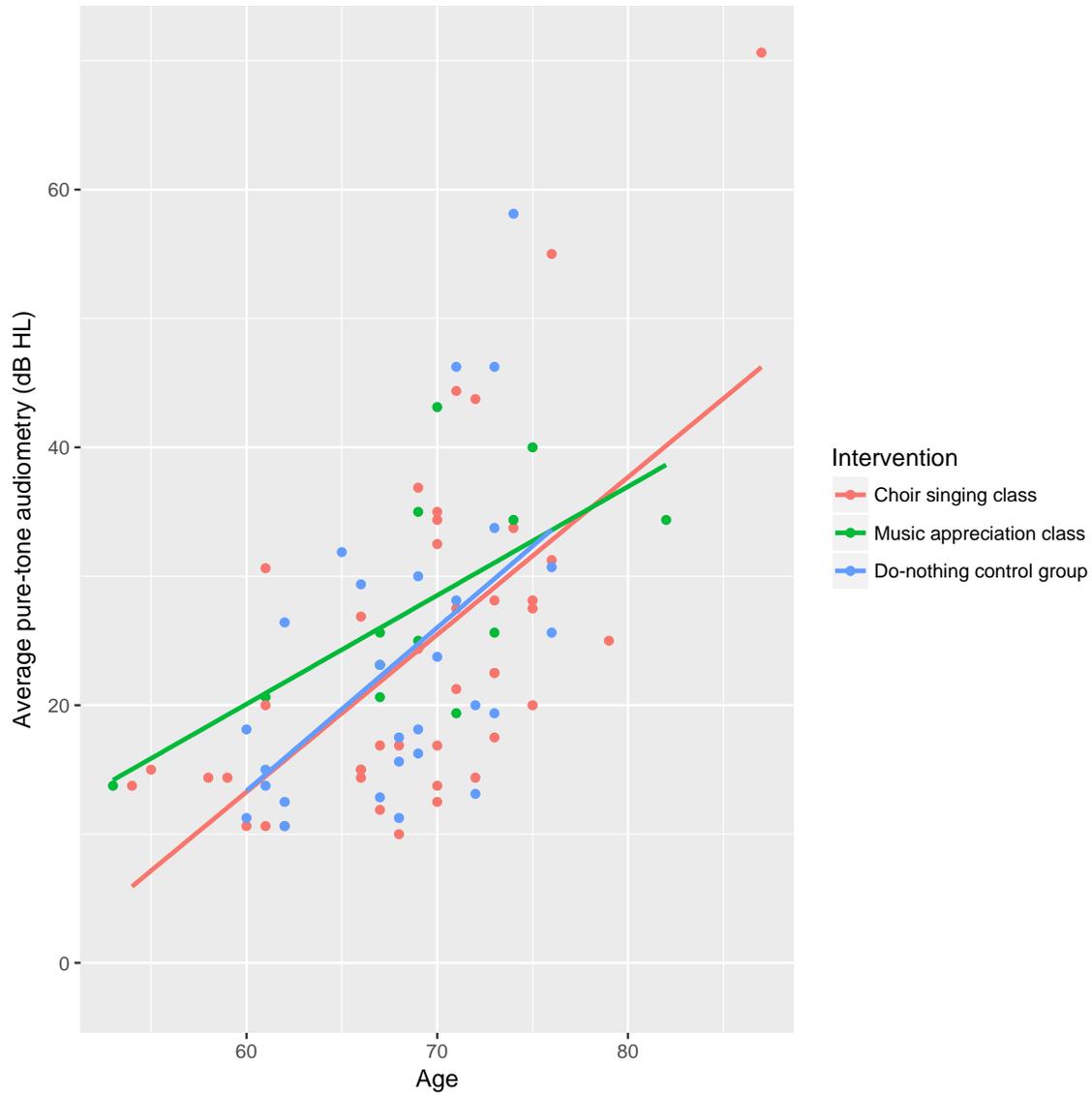


Figure 2: Binaural hearing loss (average dB HL; measured using pure-tone audiometry) as a function of age.

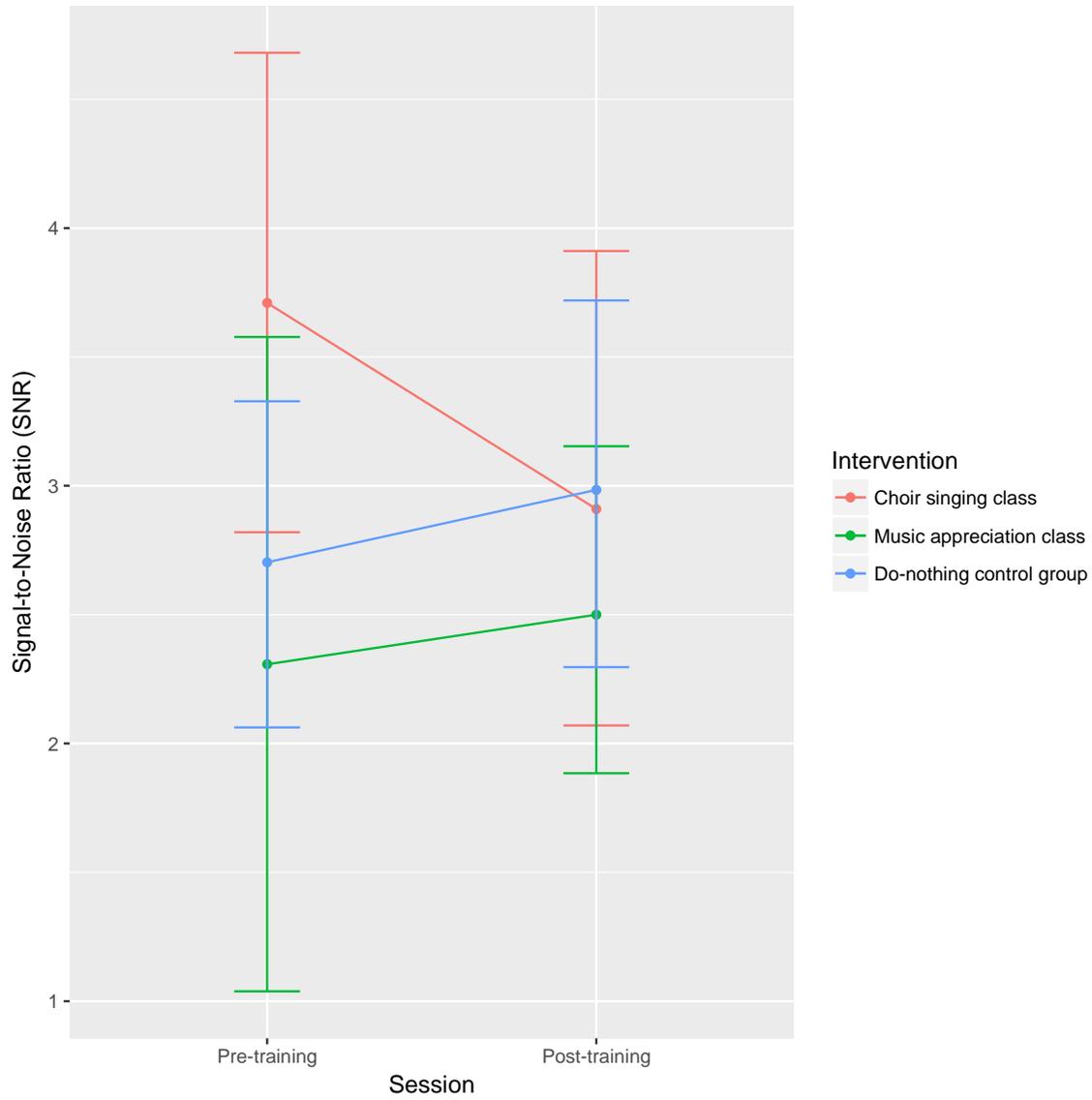


Figure 3: Pre- and post-testing thresholds for perceiving speech in noise (signal-to-noise ratio; SNR) for choir singing participants (n = 50), music appreciation class participants (n = 13), and do-nothing control participants (n = 32).

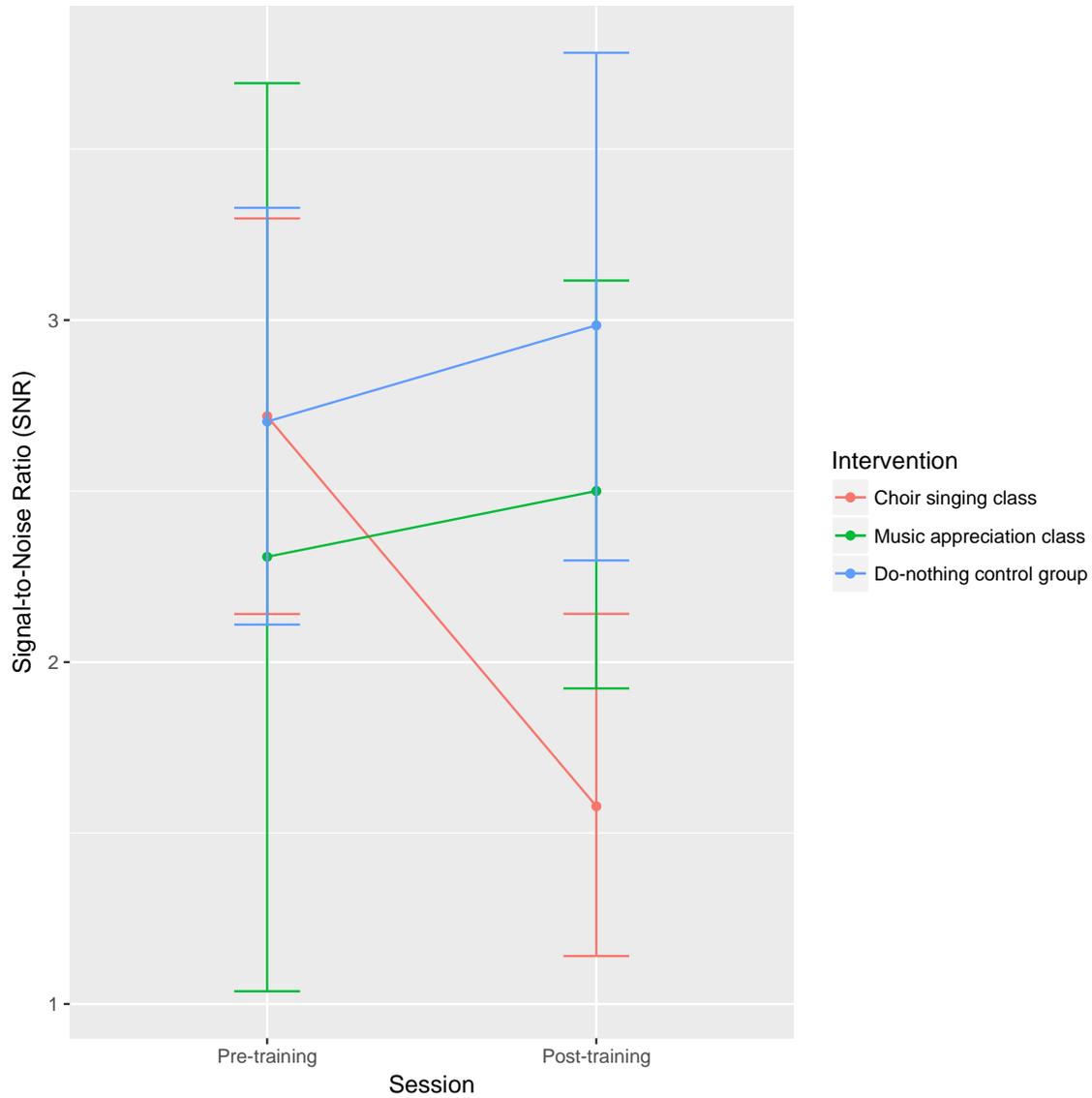


Figure 4: Baseline matched pre- and post-testing thresholds for perceiving speech in noise (signal-to-noise ratio; SNR) for choir singing participants ($n = 32$), music appreciation class participants ($n = 13$), and do-nothing control participants ($n = 32$).

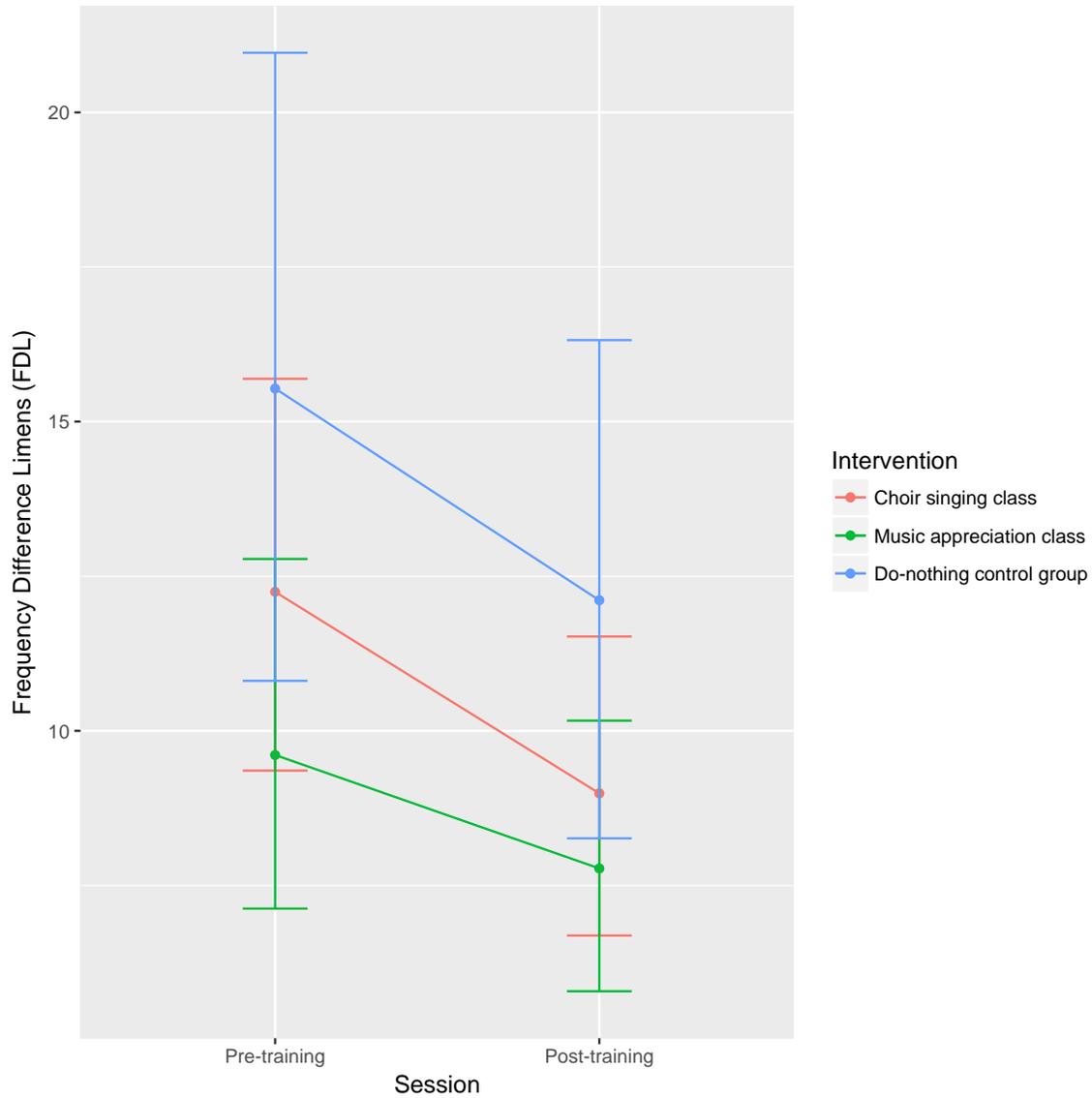


Figure 5: Pre- and post-testing thresholds for pitch discrimination (frequency difference limens; FDL in Hz) for choir singing participants ($n = 39$), music appreciation class participants ($n = 11$), and do-nothing control participants ($n = 18$).

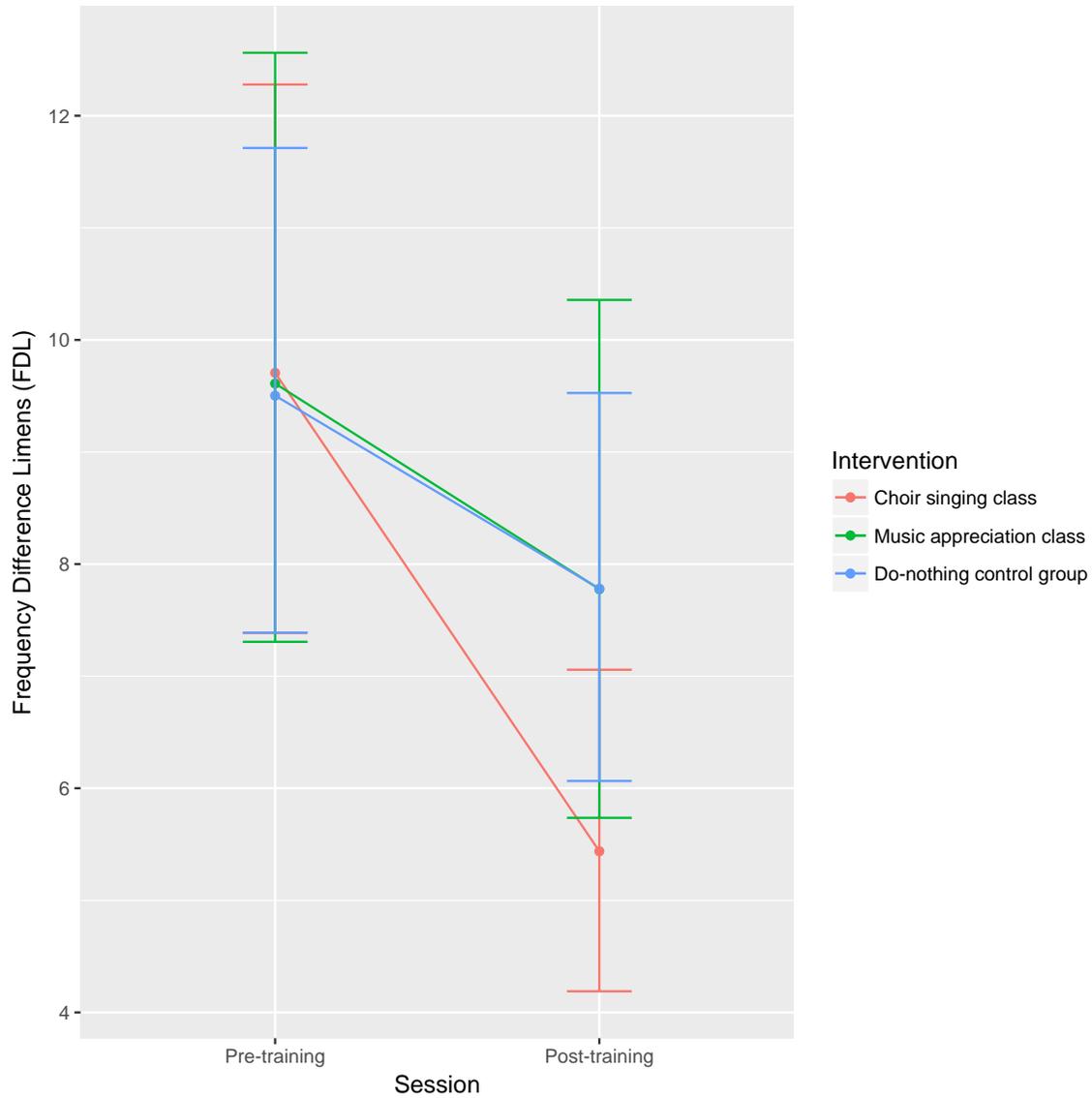


Figure 6: Baseline matched pre- and post-testing thresholds for pitch discrimination (frequency difference limens; FDL in Hz) for choir singing participants (n = 11), music appreciation class participants (n = 11), and do-nothing control participants (n = 11).

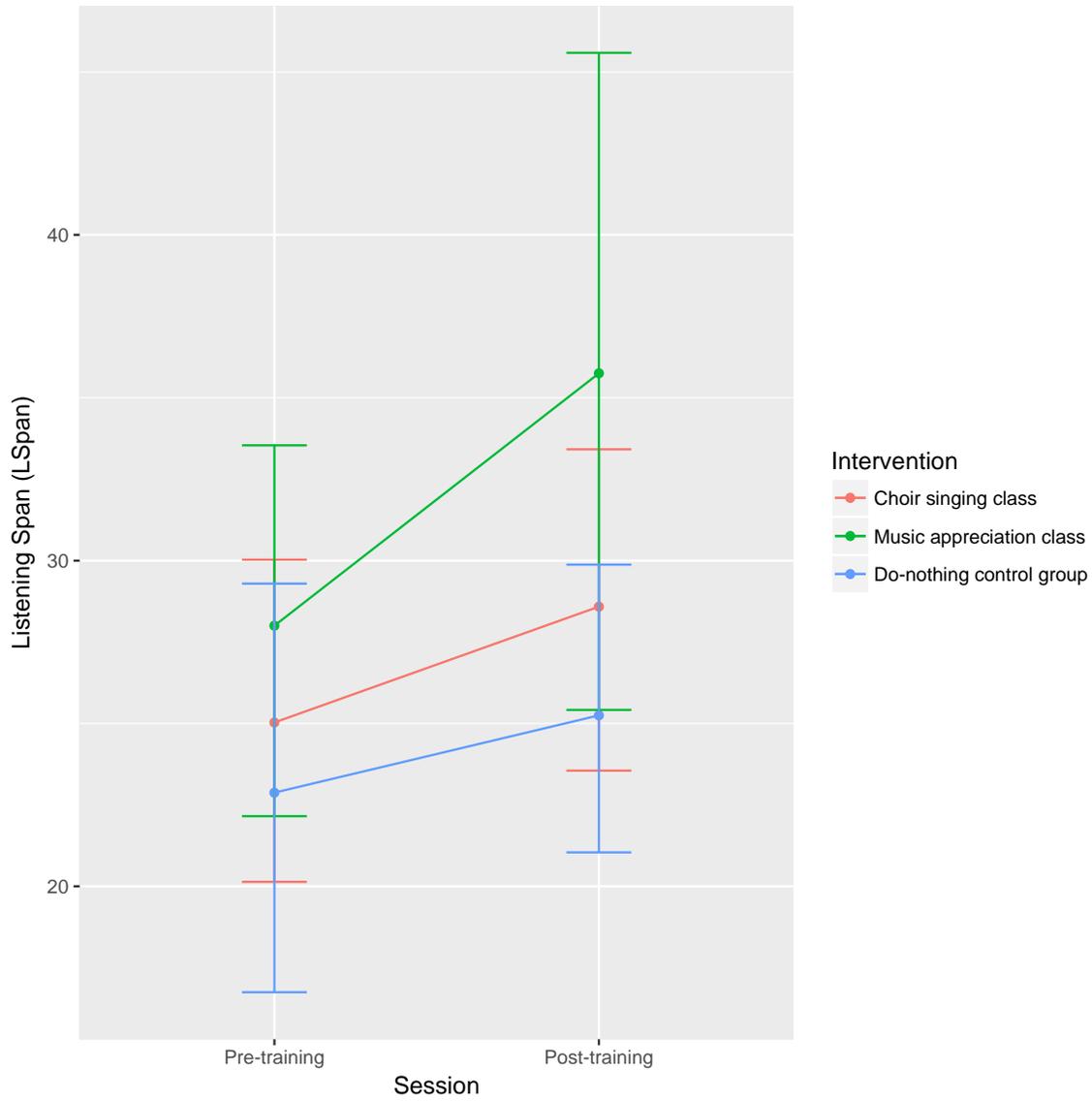


Figure 7: Pre- and post-testing LSpan scores (auditory working memory) for choir singing participants (n = 36), music appreciation class participants (n = 12), and do-nothing control participants (n = 24).

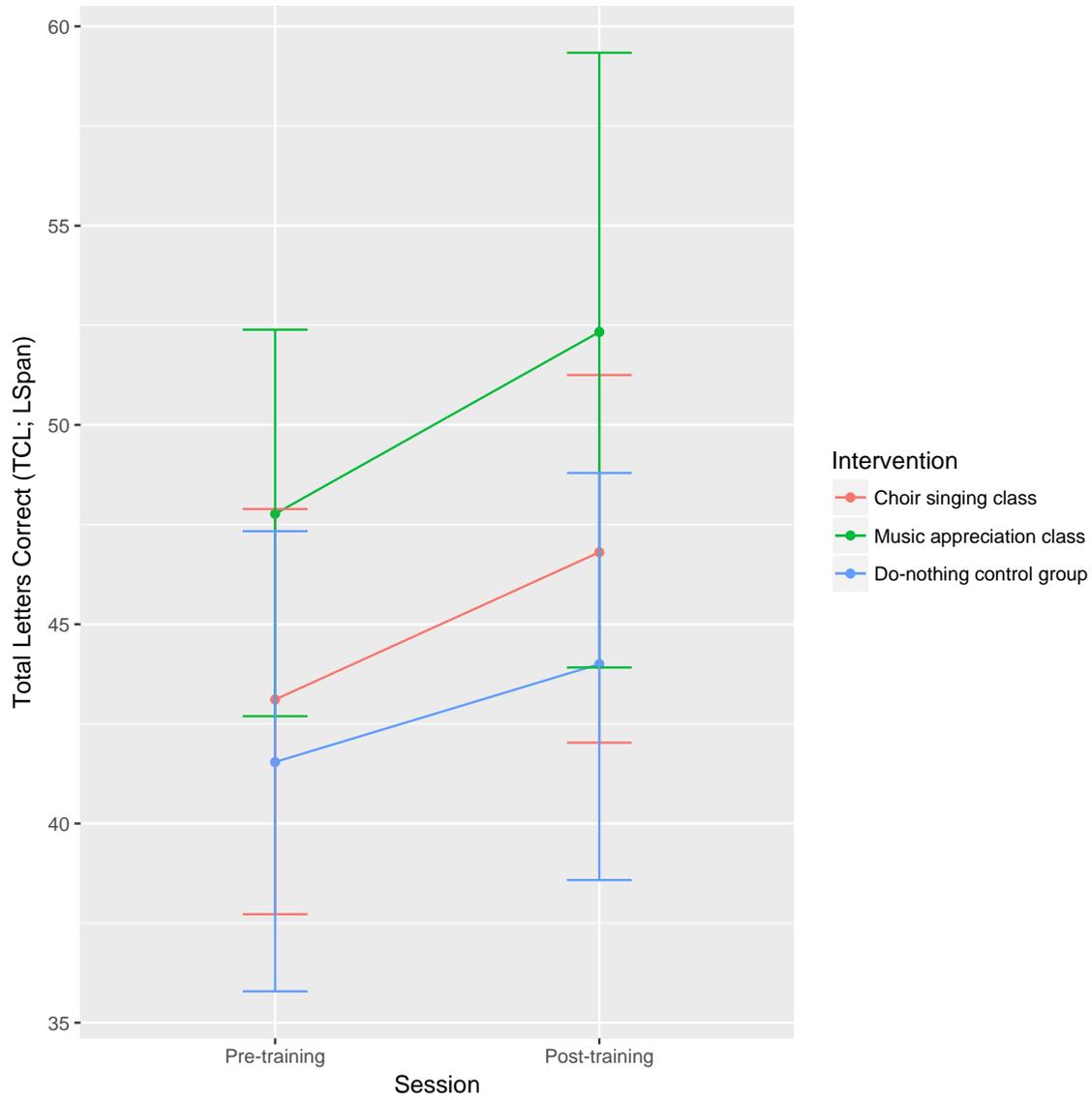


Figure 8: Pre- and post-testing total correct letter scores from the LSpan task (auditory working memory) for choir singing participants (n = 36), music appreciation class participants (n = 13), and do-nothing control participants (n = 23).

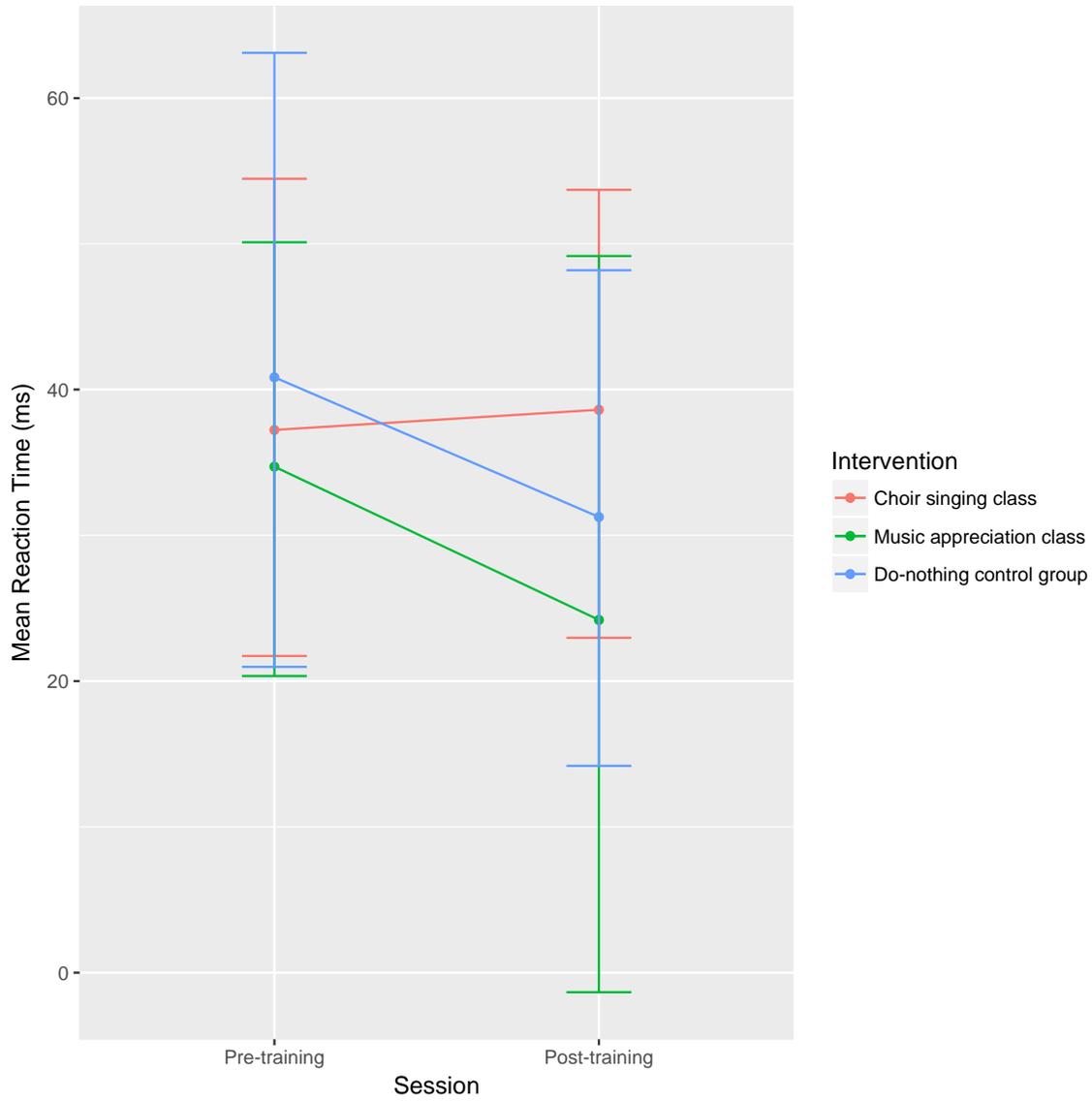


Figure 9: Pre- and post-testing Flanker effect (inhibitory control of attention, calculated as reaction time (incongruent) – reaction time (congruent); msec) for choir singing participants (n = 35), music appreciation class participants (n = 13), and do-nothing control participants (n = 24).

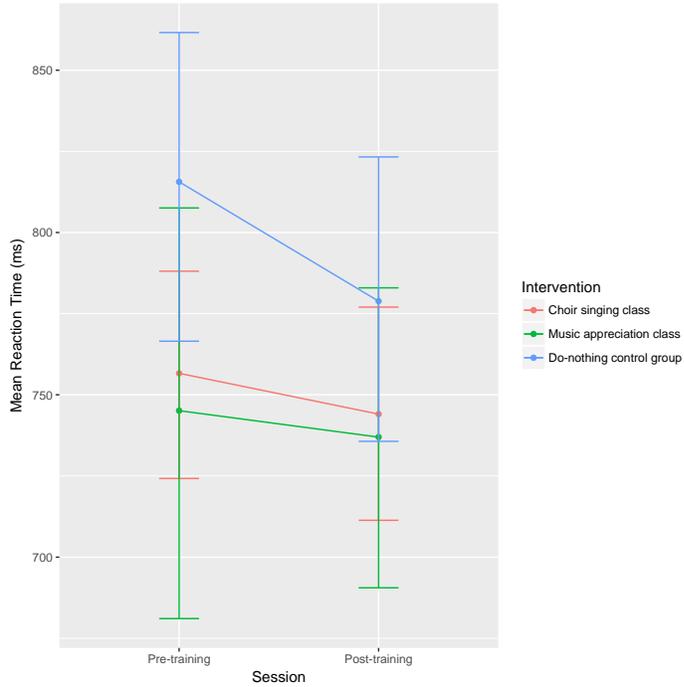


Figure 10: Pre- and post-testing mean reaction times (msec) of Flanker task responses of choir singing participants (n = 35), music appreciation class participants (n = 13), and do-nothing control participants (n = 24).

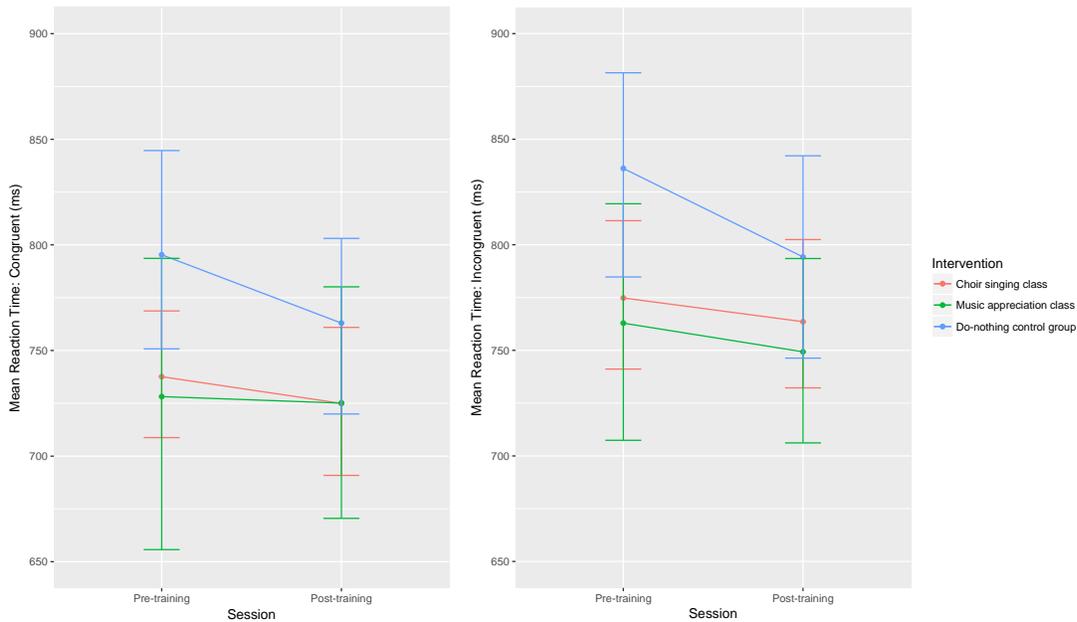


Figure 11: Pre- and post-testing reaction times (msec) of Flanker task responses of choir singing participants (n = 36), music appreciation class participants (n = 13), and do-nothing control participants (n = 24) within congruent (left) and incongruent (right) Flanker conditions.

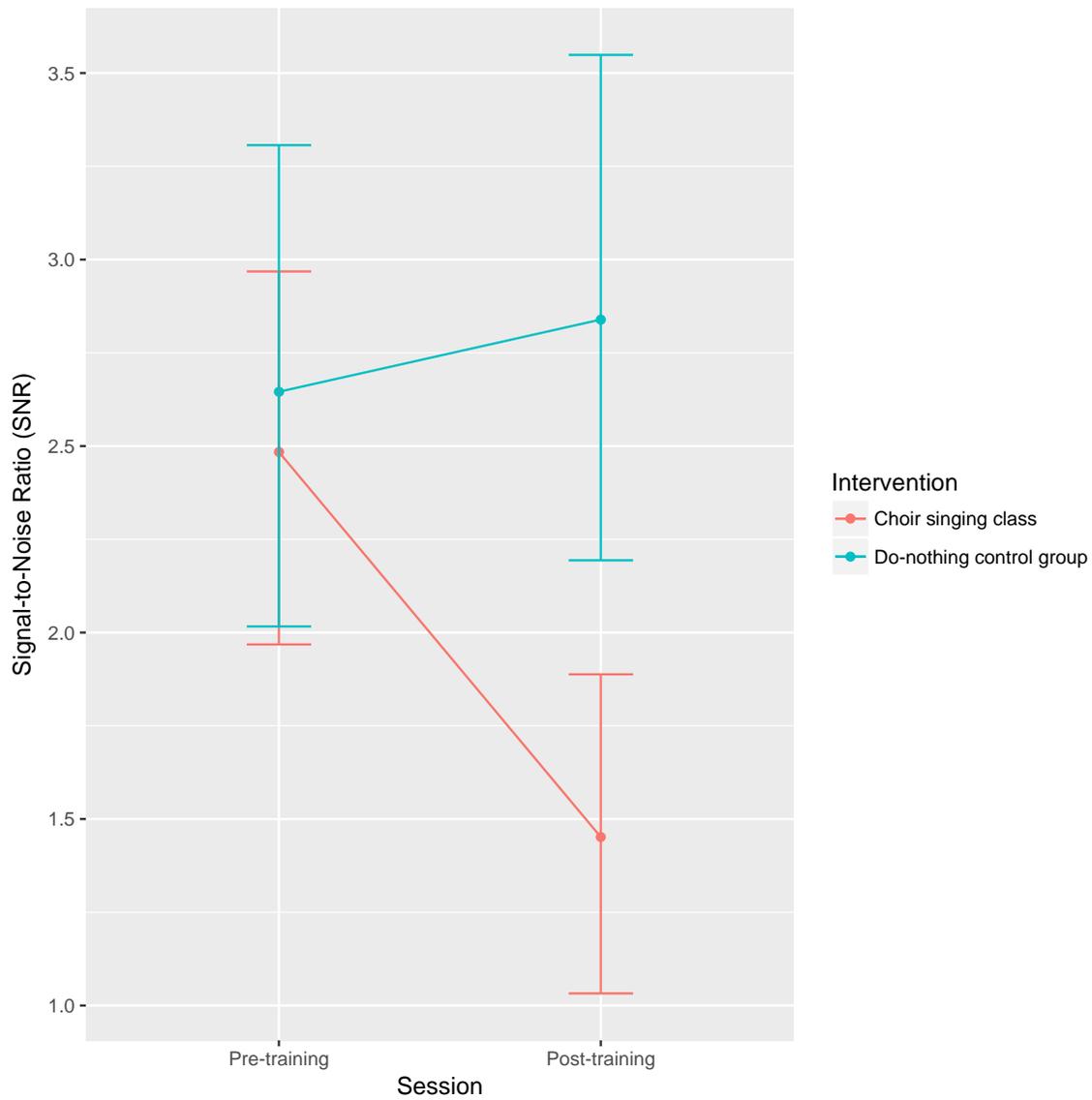


Figure 12: SNR ~ Session x Group interaction plot: baseline matched pre- and post-testing thresholds for perceiving speech in noise (signal-to-noise ratio; SNR) for choir singing participants (n = 32) and do-nothing control participants (n = 32).

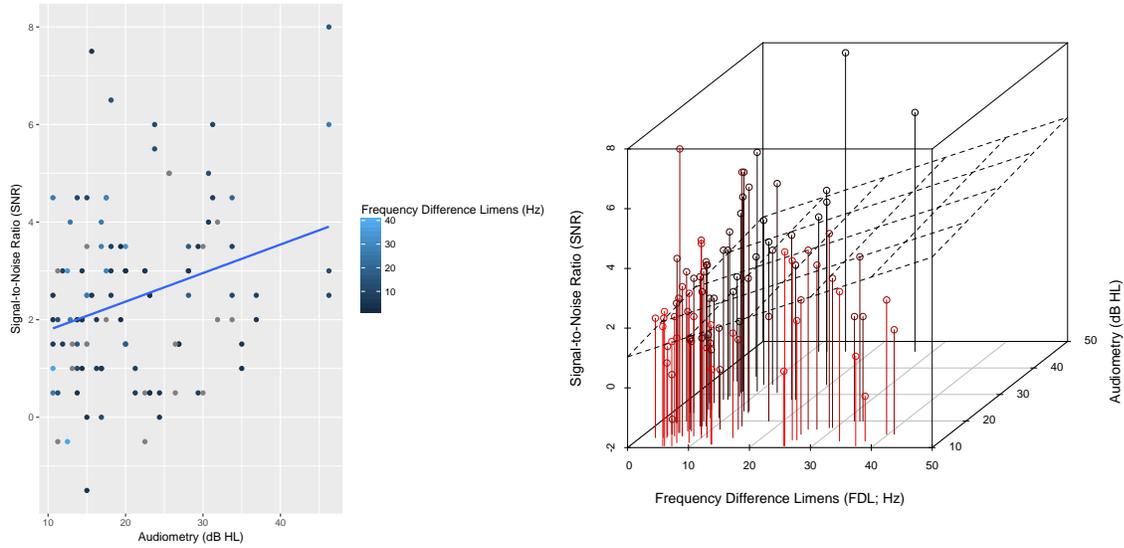


Figure 13: 2D (top) and 3D (bottom) representations of the SNR ~ Audiology (dB HL) x FDL (Hz) interaction plot: effect of increasing hearing loss on SNR, as influenced by FDL (two visualizations).

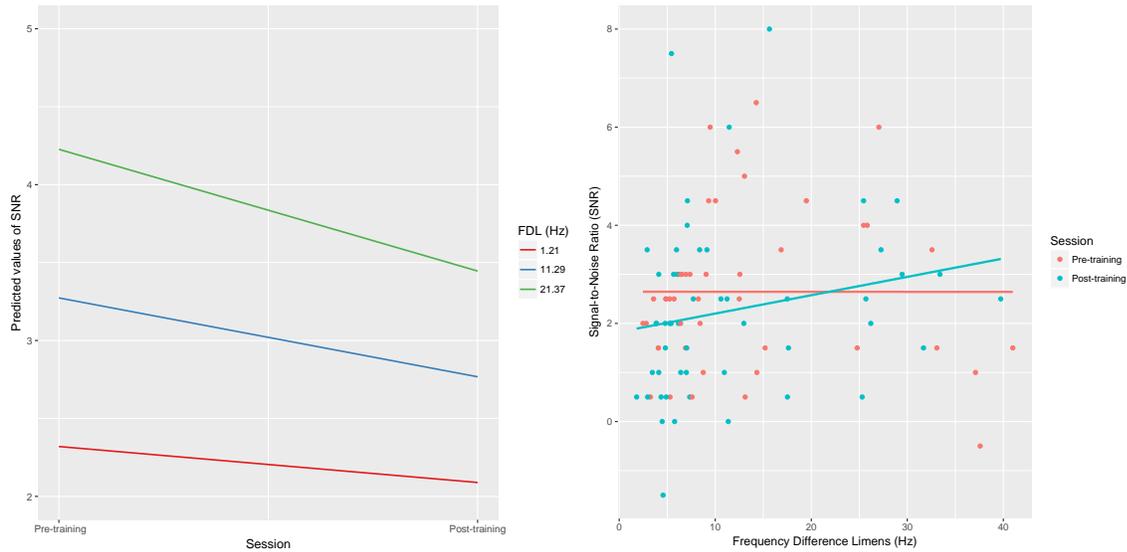


Figure 14: SNR ~ Session x FDL (Hz) interaction plot: marginal effect of FDL x Session on SNR (two visualizations).

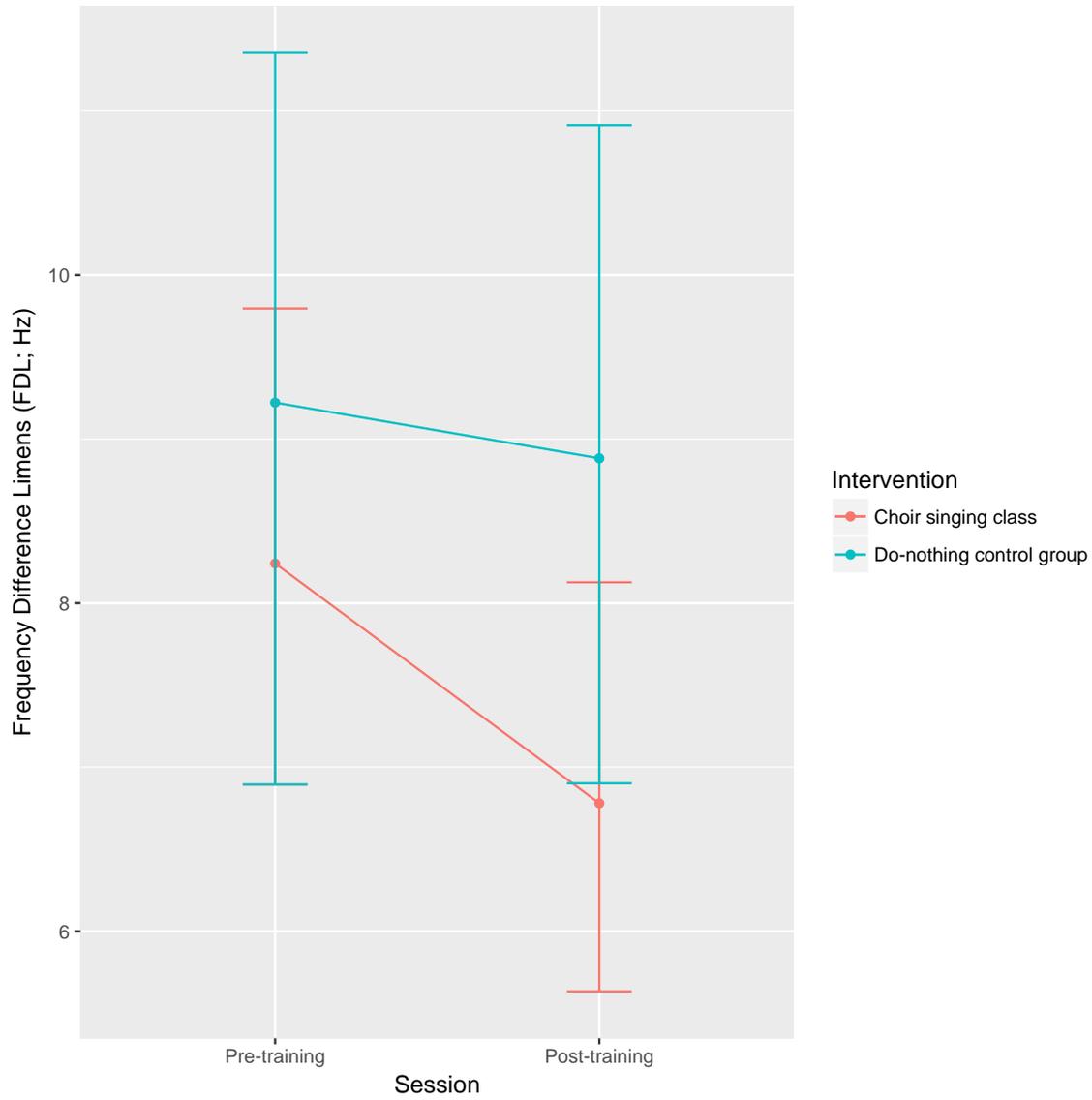


Figure 15: FDL ~ Session x Group interaction plot: pre- and post-testing thresholds for perceiving speech in noise (signal-to-noise ratio; SNR) for choir singing participants (n = 38) and do-nothing control participants (n = 16).

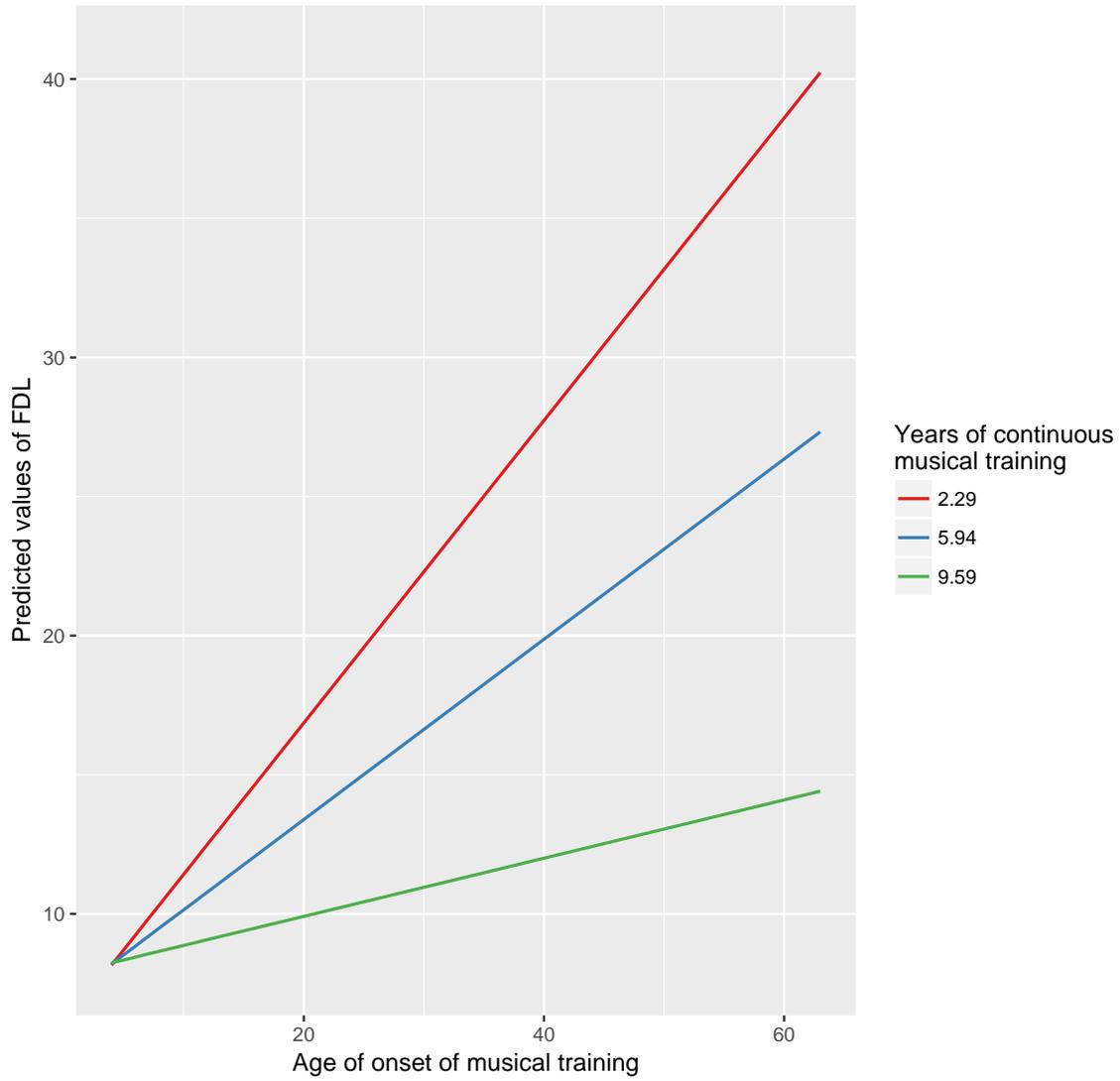


Figure 16: Effect of musicianship on FDL (Hz): interaction plot of FDL ~ age of onset of music training x years of continuous musical training.

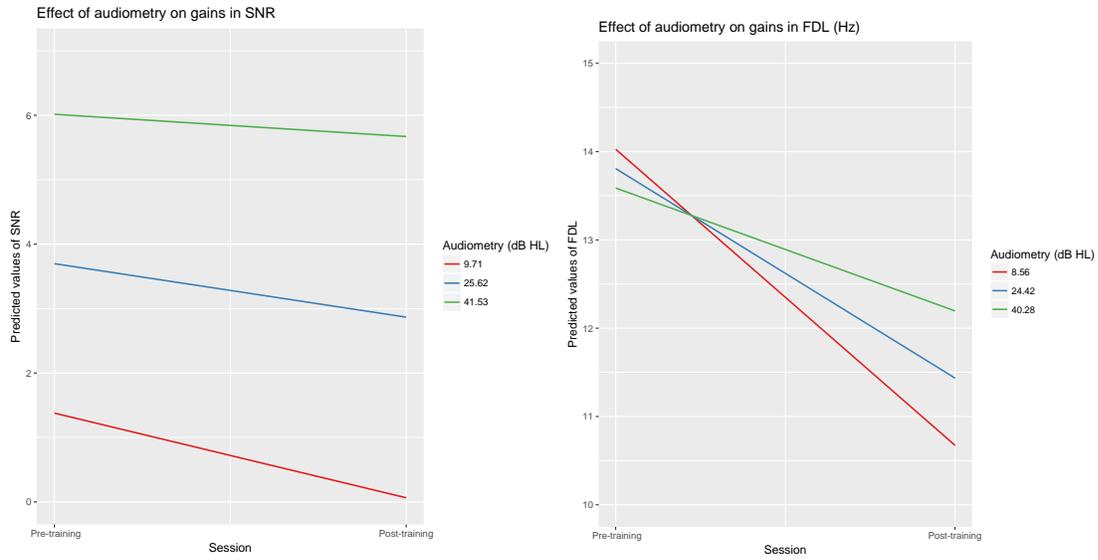


Figure 17: Effects of hearing loss on efficacy of intervention; trends within choir participants.

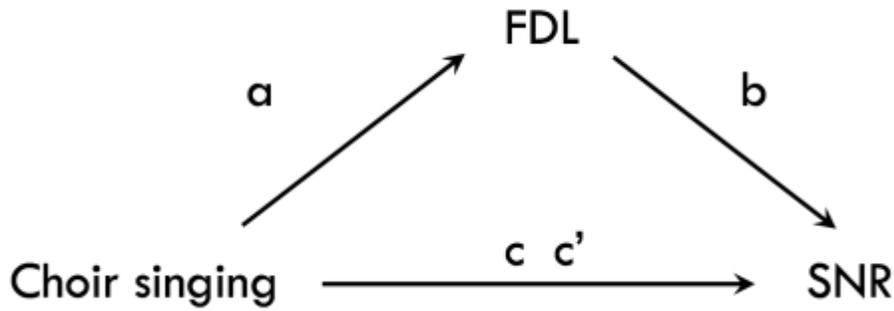


Figure 18: Exploring FDL (Hz) as a potential mediator for the effects of choir participation on SNR (mediation model).

Distribution of Indirect Effect of FDL (Hz) on SNR~Session

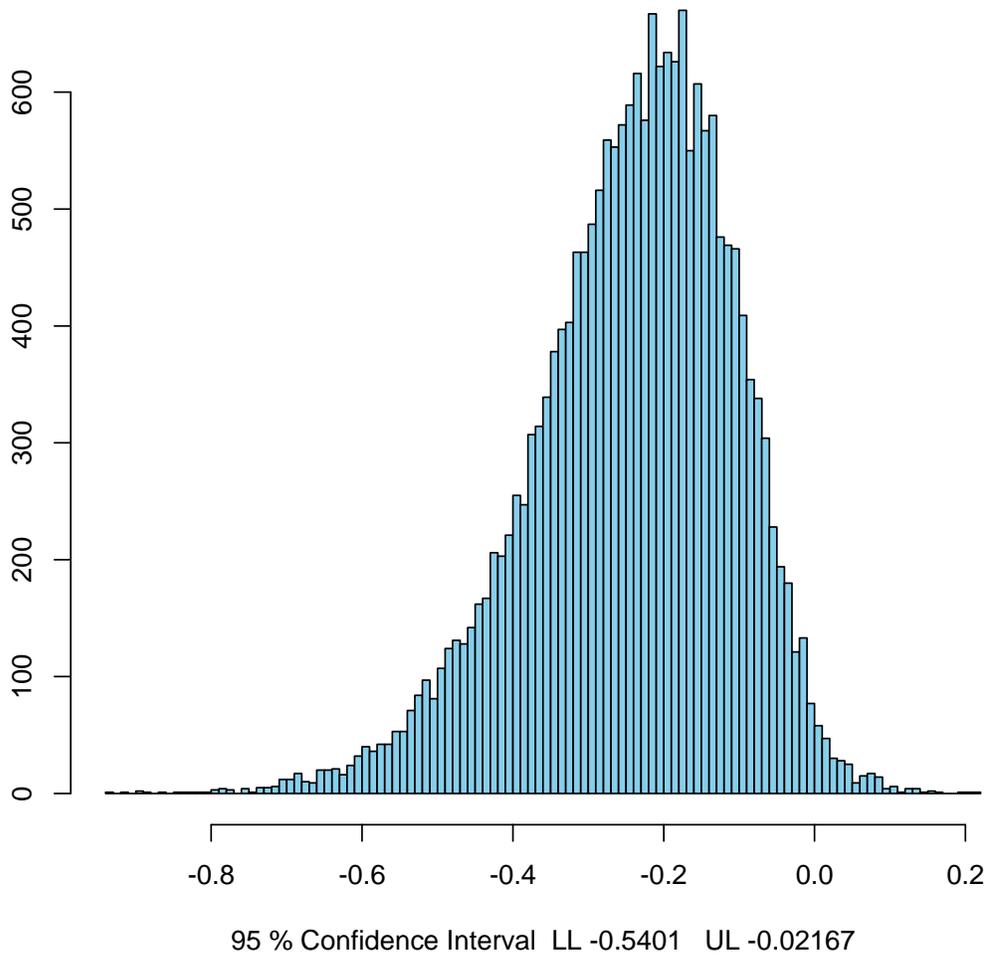


Figure 19: Distribution of indirect mediation effect of FDL (Hz) on the relationship between choir training and speech perception in noise; Monte Carlo approach, resampling = 20000, 95% CI [-0.5400542, -0.02167482].

References

- Abel, S. M., Giguere, C., Consoli, A., & Papsin, B. C. (2000). The effect of aging on horizontal plane sound localization. *J Acoust Soc Am*, *108*(2), 743–752.
<https://doi.org/10.1121/1.429607>
- Alain, C., McDonald, K. L., Ostroff, J. M., & Schneider, B. (2001). Age-related changes in detecting a mistuned harmonic. *The Journal of the Acoustical Society of America*, *109*(5 Pt 1), 2211–2216. <https://doi.org/10.1121/1.1367243>
- Alain, C., Zendel, B. R., Hutka, S., & Bidelman, G. M. (2014). Turning down the noise: The benefit of musical training on the aging auditory brain. *Hearing Research*. Elsevier B.V.
<https://doi.org/10.1016/j.heares.2013.06.008>
- Anderson, S., Parbery-Clark, A., Yi, H.-G., & Kraus, N. (2011). A neural basis of speech-in-noise perception in older adults. *Ear and Hearing*, *32*(6), 750–7.
<https://doi.org/10.1097/AUD.0b013e31822229d3>
- Angelakis, E., Stathopoulou, S., Frymiare, J. L., Green, D. L., Lubar, J. F., & Kounios, J. (2007). EEG neurofeedback: a brief overview and an example of peak alpha frequency training for cognitive enhancement in the elderly. *The Clinical Neuropsychologist*, *21*(1), 110–29.
<https://doi.org/10.1080/13854040600744839>

- Arlinger, S. (2003). Negative consequences of uncorrected hearing loss - a review. *International Journal of Audiology*, 42(2), S17–S20. <https://doi.org/10.3109/14992020309074639>
- Bailey, B. A. (2005). Effects of group singing and performance for marginalized and middle-class singers. *Psychology of Music*, 33(3), 269–303.
<https://doi.org/10.1177/0305735605053734>
- Beck, R. J., Cesario, T. C., Yousefi, a., & Enamoto, H. (2000). Choral singing, performance perception, and immune system changes in salivary immunoglobulin a and cortisol. *Music Perception*, 18(1), 87–106. <https://doi.org/10.2307/40285902>
- Besson, M., Schön, D., Moreno, S., Santos, A., & Magne, C. (2007). Influence of musical expertise and musical training on pitch processing in music and language. *Restorative Neurology and Neuroscience*, 25(3–4), 399–410. <https://doi.org/10.1162/jocn.2010.21585>
- Betlejewski, S. (2006). [Age connected hearing disorders (presbycusis) as a social problem]. *Otolaryngologia Polska = The Polish Otolaryngology*, 60(6), 883–886.
- Bidelman, G. M., Gandour, J. T., & Krishnan, A. (2011). Musicians and tone-language speakers share enhanced brainstem encoding but not perceptual benefits for musical pitch. *Brain and Cognition*, 77(1), 1–10. <https://doi.org/10.1016/j.bandc.2011.07.006>

- Bidelman, G. M., & Krishnan, A. (2010). Effects of reverberation on brainstem representation of speech in musicians and non-musicians. *Brain Research, 1355*, 112–125.
<https://doi.org/10.1016/j.brainres.2010.07.100>
- Bidelman, G. M., Krishnan, A., & Gandour, J. T. (2011). Enhanced brainstem encoding predicts musicians' perceptual advantages with pitch. *European Journal of Neuroscience, 33*(3), 530–538. <https://doi.org/10.1111/j.1460-9568.2010.07527.x>
- Bidelman, G. M., Villafuerte, J. W., Moreno, S., & Alain, C. (2014). Age-related changes in the subcortical-cortical encoding and categorical perception of speech. *Neurobiology of Aging, 35*(11), 2526–2540. <https://doi.org/10.1016/j.neurobiolaging.2014.05.006>
- Brainard, M. S., & Doupe, A. J. (2000). Auditory feedback in learning and maintenance of vocal behaviour. *Nature Reviews. Neuroscience, 1*(1), 31–40. <https://doi.org/10.1038/35036205>
- Bratzlavsky, M., Van Zandijcke, M., & Vander Eecken, H. (1981). Neurophysiologic basis of the EEG. *Tijdschrift Voor Geneeskunde, 37*(6), 365–369.
<https://doi.org/10.1097/01.wnp.0000220079.61973.6c>
- Carney, L. H. (1994). Spatiotemporal encoding of sound level: Models for normal encoding and recruitment of loudness. *Hearing Research, 76*(1–2), 31–44. [https://doi.org/10.1016/0378-5955\(94\)90084-1](https://doi.org/10.1016/0378-5955(94)90084-1)

- Carr, C. E. (1993). Processing of temporal information in the brain. *Annual Review of Neuroscience*, 16(Konishi 1991), 223–243.
<https://doi.org/10.1146/annurev.ne.16.030193.001255>
- Chisolm, T. H., Willott, J. F., & Lister, J. J. (2003). The aging auditory system: anatomic and physiologic changes and implications for rehabilitation. *International Journal of Audiology*, 42, 3–10. <https://doi.org/10.3109/14992020309074637>
- Chmiel, R., & Jerger, J. (1996). Hearing aid use, central auditory disorder, and hearing handicap in elderly persons. *Journal of the American Academy of Audiology*, 7, 190–202. Retrieved from http://www.audiology.org/sites/default/files/journal/JAAA_07_03_08.pdf
- Clift, S., & Hancox, G. (2001). The perceived benefits of singing: findings from preliminary surveys of a university college choral society. *The Journal of the Royal Society for the Promotion of Health*, 121(4), 248–256. <https://doi.org/10.1177/146642400112100409>
- Clift, S., & Morrison, I. (2011). Group singing fosters mental health and wellbeing: findings from the East Kent “singing for health” network project. *Mental Health & Social Inclusion*, 15(2), 88–97. <https://doi.org/10.1108/20428301111140930>
- Clinard, C. G., Tremblay, K. L., & Krishnan, A. R. (2010). Aging alters the perception and physiological representation of frequency: evidence from human frequency-following response recordings. *Hearing Research*, 264(1–2), 48–55.
<https://doi.org/10.1016/j.heares.2009.11.010>

- Coffman, D. D. (2002). Music and quality of life in older adults. *Psychomusicology*, *18*(1–2), 76–88. <https://doi.org/http://dx.doi.org/10.1037/h0094050>
- Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user’s guide. *Psychonomic Bulletin & Review*, *12*(5), 769–786. <https://doi.org/10.3758/BF03196772>
- Coon, D. W., & Thompson, L. W. (2003). The Relationship Between Homework Compliance and Treatment Outcomes Among Older Adult Outpatients With Mild-to-Moderate Depression. *American Journal of Geriatric Psychiatry*, *11*(1), 53–61. <https://doi.org/10.1176/appi.ajgp.11.1.53>
- Cruickshanks, K. J., Wiley, T. L., Tweed, T. S., Klein, B. E. K., Klein, R., Mares-Perlman, J. A., & Nondahl, D. M. (1998). Prevalence of Hearing Loss in Older Adults in Beaver Dam , Wisconsin. *American Journal of Audiology*, *148*(9), 879–886. <https://doi.org/10.1093/oxfordjournals.aje.a009713>
- Dalla Bella, S., & Berkowska, M. (2009). Singing proficiency in the majority: Normality and “phenotypes” of poor singing. *Annals of the New York Academy of Sciences*, *1169*, 99–107. <https://doi.org/10.1111/j.1749-6632.2009.04558.x>
- Dalla Bella, S., Giguère, J.-F., & Peretz, I. (2007). Singing proficiency in the general population. *The Journal of the Acoustical Society of America*, *121*(2), 1182–1189. <https://doi.org/10.1121/1.2427111>

- Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory during reading. *Journal Of Verbal Learning And Verbal Behavior*, *19*(4), 450–466.
[https://doi.org/10.1016/S0022-5371\(80\)90312-6](https://doi.org/10.1016/S0022-5371(80)90312-6)
- Darrough, G. P. (1992). Making Choral Music with Older Adults. *Music Educators Journal*, *79*(4), 27. <https://doi.org/10.2307/3398526>
- Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21. <https://doi.org/10.1016/j.jneumeth.2003.10.009>
- Djernes, J. K. (2006). Prevalence and predictors of depression in populations of elderly: A review. *Acta Psychiatrica Scandinavica*, *113*(5), 372–387. <https://doi.org/10.1111/j.1600-0447.2006.00770.x>
- Dubno, J. R., Dirks, D. D., & Morgan, D. E. (1984). Effects of age and mild hearing loss on speech recognition in noise. *Journal of the Acoustical Society of America*, *76*(1), 87–96.
<https://doi.org/10.1121/1.391011>
- Fabiani, M. (2012). It was the best of times, it was the worst of times: a psychophysiology's view of cognitive aging. *Psychophysiology*, *49*(3), 283–304. <https://doi.org/10.1111/j.1469-8986.2011.01331.x>

- Fillit, H. M., Butler, R. N., O'Connell, A. W., Albert, M. S., Birren, J. E., Cotman, C. W., ... Tully, T. (2002). Achieving and maintaining cognitive vitality with aging. *Mayo Clinic Proceedings. Mayo Clinic*, 77(7), 681–696. <https://doi.org/10.4065/77.7.681>
- Fitzgibbons, P. J., & Gordon-Salant, S. (1994). Age effects on measures of auditory duration discrimination. *Journal of Speech and Hearing Research*, 37(3), 662–670. <https://doi.org/10.1044/jshr.3703.662>
- Fozard, J. L. (1990). Vision and hearing in aging. *Handbook of the Psychology of Aging*. <https://doi.org/http://dx.doi.org/10.1016/B978-0-12-101280-9.50015-2>
- Frisina, R. D., & Frisina, R. D. (1997). Speech recognition in noise and presbycusis: Relations to possible neural mechanisms. *Hearing Research*, 106(1–2), 95–104. [https://doi.org/10.1016/S0378-5955\(97\)00006-3](https://doi.org/10.1016/S0378-5955(97)00006-3)
- Frisina, R. D., & Walton, J. P. (2006). Age-related structural and functional changes in the cochlear nucleus. *Hearing Research*, 216–217(1–2), 216–223. <https://doi.org/10.1016/j.heares.2006.02.003>
- Fujioka, T., Ross, B., Kakigi, R., Pantev, C., & Trainor, L. J. (2006). One year of musical training affects development of auditory cortical-evoked fields in young children. *Brain*, 129(10), 2593–2608. <https://doi.org/10.1093/brain/awl247>

- Gaser, C., & Schlaug, G. (2003). Brain structures differ between musicians and non-musicians. *The Journal of Neuroscience*, 23(27), 9240–9245. <https://doi.org/23/27/9240> [pii]
- Gates, G. A., & Mills, J. H. (2005). Presbycusis. *Lancet (London, England)*, 366(9491), 1111–20. [https://doi.org/10.1016/S0140-6736\(05\)67423-5](https://doi.org/10.1016/S0140-6736(05)67423-5)
- Gomez, R. G., & Madey, S. F. (2001). Coping-With-Hearing-Loss Model for Older Adults, 56(4), 223–225.
- Good, A., & Russo, F. A. (2016). Singing Promotes Cooperation in a Diverse Group of Children. *Social Psychology*, 1–5. <https://doi.org/10.1027/1864-9335/a000282>
- Gordon-Salant, S. (2005). Hearing loss and aging: new research findings and clinical implications. *Journal of Rehabilitation Research & Development*, 42(4 Suppl 2), 9–24. <https://doi.org/10.1682/JRRD.2005.01.0006>
- Hanna-Pladdy, B., & MacKay, A. (2011). The relation between instrumental musical activity and cognitive aging. *Neuropsychology*, 25(3), 378–86. <https://doi.org/10.1037/a0021895>
- Hargus, S. E., & Gordon-Salant, S. (1995). Accuracy of speech intelligibility index predictions for noise-masked young listeners with normal hearing and for elderly listeners with hearing impairment. *Journal of Speech and Hearing Research*, 38(1), 234–43. <https://doi.org/10.1044/jshr.3801.234>

Herdener, M., Esposito, F., di Salle, F., Boller, C., Hilti, C. C., Habermeyer, B., ... Cattapan-Ludewig, K. (2010). Musical training induces functional plasticity in human hippocampus. *J Neurosci*, *30*(4), 1377–1384. <https://doi.org/10.1523/JNEUROSCI.4513-09.2010>

Herholz, S. C., & Zatorre, R. J. (2012). Musical training as a framework for brain plasticity: behavior, function, and structure. *Neuron*, *76*(3), 486–502. <https://doi.org/10.1016/j.neuron.2012.10.011>

Hickok, G. (2001). Functional anatomy of speech perception and speech production: psycholinguistic implications. *Journal of Psycholinguistic Research*, *30*(3), 225–235. <https://doi.org/10.1023/A:1010486816667>

Hillman, S. (2002). Participatory singing for older people: a perception of benefit. *Health Education*, *102*(4), 163–171. <https://doi.org/10.1108/09654280210434237>

Hoormann, J., Falkenstein, M., Hohnsbein, J., & Blanke, L. (1992). The human frequency-following response (FFR): Normal variability and relation to the click-evoked brainstem response. *Hearing Research*, *59*(2), 179–188. [https://doi.org/10.1016/0378-5955\(92\)90114-3](https://doi.org/10.1016/0378-5955(92)90114-3)

Houde, J. F., & Jordan, M. I. (1998). Sensorimotor adaptation in speech production. *Science*, *279*(5354), 1213–1216. <https://doi.org/10.1126/science.279.5354.1213>

- Hyde, K. L., Lerch, J., Norton, A., Forgeard, M., Winner, E., Evans, A. C., & Schlaug, G. (2009). Musical training shapes structural brain development. *Journal of Neuroscience*, 29(10), 3019–3025. <https://doi.org/10.1523/JNEUROSCI.5118-08.2009>
- Ives, D. T., & Patterson, R. D. (2008). Pitch strength decreases as F0 and harmonic resolution increase in complex tones composed exclusively of high harmonics. *The Journal of the Acoustical Society of America*, 123(5), 2670–2679. <https://doi.org/10.1121/1.2890737>
- Johnson, K. L., Nicol, T. G., & Kraus, N. (2005). Brain stem response to speech: a biological marker of auditory processing. *Ear and Hearing*, 26(5), 424–434. <https://doi.org/10.1097/01.aud.0000179687.71662.6e>
- Killion, M. C. (1997). Hearing aids: Past, present, future: Moving toward normal conversation in noise. *British Journal of Audiology*, 31(June), 141–148. <https://doi.org/10.3109/03005364000000016>
- Killion, M. C., Niquette, P. A., Gudmundsen, G. I., Revit, L. J., & Banerjee, S. (2004). Development of a quick speech-in-noise test for measuring signal-to-noise ratio loss in normal-hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America*. <https://doi.org/10.1121/1.1784440>
- Kim, S. H., Frisina, R. D., Mapes, F. M., Hickman, E. D., & Frisina, D. R. (2006). Effect of age on binaural speech intelligibility in normal hearing adults. *Speech Communication*, 48(6), 591–597. <https://doi.org/10.1016/j.specom.2005.09.004>

- Koelsch, S., Schröger, E., & Tervaniemi, M. (1999). Superior pre-attentive auditory processing in musicians. *Neuroreport*, *10*(6), 1309–1313. <https://doi.org/10.1097/00001756-199904260-00029>
- Kramer, A. F., Bherer, L., Colcombe, S. J., Dong, W., & Greenough, W. T. (2004). Environmental influences on cognitive and brain plasticity during aging. *J.Gerontol.A Biol.Sci.Med.Sci.*, *59*(9), M940–M957. <https://doi.org/10.1093/gerona/59.9.M940>
- Kraus, N., & Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nat Rev Neurosci*, *11*(8), 599–605. <https://doi.org/nrn2882> [pii]\n10.1038/nrn2882
- Kraus, N., Skoe, E., Parbery-Clark, A., & Ashley, R. (2009). Experience-induced malleability in neural encoding of pitch, timbre, and timing: Implications for language and music. *Annals of the New York Academy of Sciences*, *1169*, 543–557. <https://doi.org/10.1111/j.1749-6632.2009.04549.x>
- Kreutz, G., Bongard, S., Rohrman, S., Hodapp, V., & Grebe, D. (2004). Effects of choir singing or listening on secretory immunoglobulin A, cortisol, and emotional state. *Journal of Behavioral Medicine*, *27*(6), 623–635. <https://doi.org/10.1007/s10865-004-0006-9>
- Kuriki, S. (2006). Effects of Musical Experience on Different Components of MEG Responses Elicited by Sequential Piano-Tones and Chords. *Journal of Neuroscience*, *26*(15), 4046–4053. <https://doi.org/10.1523/JNEUROSCI.3907-05.2006>

- Lappe, C., Herholz, S. C., Trainor, L. J., & Pantev, C. (2008). Cortical Plasticity Induced by Short-Term Unimodal and Multimodal Musical Training. *Journal of Neuroscience*, 28(39), 9632–9639. <https://doi.org/10.1523/JNEUROSCI.2254-08.2008>
- Lee, K. M., Skoe, E., Kraus, N., & Ashley, R. (2009). Selective subcortical enhancement of musical intervals in musicians. *The Journal of Neuroscience*, 29(18), 5832–40. <https://doi.org/10.1523/JNEUROSCI.6133-08.2009>
- Luck, S. J. (2012). Event-related potentials. *APA Handbook of Research Methods in Psychology*, 1, 1–18.
- Mathers, C., Smith, A., & Concha, M. (2000). Global burden of hearing loss in the year 2000. *Global Burden of Disease*, 18(4), 1–30. Retrieved from http://www.who.int/healthinfo/statistics/bod_hearingloss.pdf
http://www.who.int/entity/healthinfo/statistics/bod_hearingloss.pdf
- Micheyl, C., Delhommeau, K., Perrot, X., & Oxenham, A. J. (2006). Influence of musical and psychoacoustical training on pitch discrimination. *Hearing Research*, 219(1–2), 36–47. <https://doi.org/10.1016/j.heares.2006.05.004>
- Mireles, D. E., & Charness, N. (2002). Computational explorations of the influence of structured knowledge on age-related cognitive decline. *Psychology and Aging*, 17(2), 245–259. <https://doi.org/10.1037/0882-7974.17.2.245>

- Mithen, S., Morley, I., Wray, A., Tallerman, M., & Gamble, C. (2006). The Singing Neanderthals: The Origins of Music, Language, Mind and Body. *Cambridge Archaeological Journal*, *16*(1), 97–112. <https://doi.org/10.1017/S0959774306000060>
- Moshé-Steinberg, S. (2015). *Effect of Choir Participation on Speech-in-Noise Perception and Neural Timing in Hearing-Impaired Older Adults*. Ryerson University.
- Musacchia, G., Sams, M., Skoe, E., & Kraus, N. (2007). Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(40), 15894–15898. <https://doi.org/10.1073/pnas.0701498104>
- Musacchia, G., Strait, D., & Kraus, N. (2008). Relationships between behavior, brainstem and cortical encoding of seen and heard speech in musicians and non-musicians. *Hearing Research*, *241*(1–2), 34–42. <https://doi.org/10.1016/j.heares.2008.04.013>
- Pantev, C., Ross, B., Fujioka, T., Trainor, L. J., Schulte, M., & Schulz, M. (2003). Music and Learning-Induced Cortical Plasticity. *Annals of the New York Academy of Sciences*, *999*(December), 438–450. <https://doi.org/10.1196/annals.1284.054>
- Parbery-Clark, A., Anderson, S., Hittner, E., & Kraus, N. (2012). Musical experience offsets age-related delays in neural timing. *Neurobiology of Aging*, *33*(7), 1–4. <https://doi.org/10.1016/j.neurobiolaging.2011.12.015>

- Parbery-Clark, A., Skoe, E., & Kraus, N. (2009). Musical experience limits the degradative effects of background noise on the neural processing of sound. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, 29(45), 14100–7. <https://doi.org/10.1523/JNEUROSCI.3256-09.2009>
- Parbery-Clark, A., Skoe, E., Lam, C., & Kraus, N. (2009). Musician enhancement for speech-in-noise. *Ear and Hearing*, 30(6), 653–661. <https://doi.org/10.1097/AUD.0b013e3181b412e9>
- Patterson, R. D., Allerhand, M. H., & Giguere, C. (1995). Time-domain modeling of peripheral auditory processing: A modular architecture and a software platform. *J. Acoust. Soc. Am.* <https://doi.org/10.1121/1.414456>
- Peel, N. M., McClure, R. J., & Bartlett, H. P. (2005). Behavioral determinants of healthy aging. *American Journal of Preventive Medicine*, 28(3), 298–304. <https://doi.org/10.1016/j.amepre.2004.12.002>
- Pfordresher, P. Q., & Brown, S. (2007). Poor-Pitch Singing in the Absence of “Tone Deafness.” *Music Perception*, 25(2), 95–115. <https://doi.org/10.1525/mp.2007.25.2.95>
- Pfordresher, P. Q., & Dalla Bella, S. (2011). Delayed auditory feedback and movement. *Journal of Chemical Information and Modeling*, 53(2), 160. <https://doi.org/10.1017/CBO9781107415324.004>

- Phillips, D. P. (2002). Central Auditory System and Central Auditory Processing Disorders: Some Conceptual Issues. *Management of Auditory Processing Disorders Ph.D. Seminars in Hearing, 23*(4). Retrieved from [https://www.brainmaster.com/software/pubs/brain/Phillips-auditory-processing_4_23_2002\(1\).pdf](https://www.brainmaster.com/software/pubs/brain/Phillips-auditory-processing_4_23_2002(1).pdf)
- Pichora-Fuller, M. K., & Schneider, B. (1992). The effect of interaural delay of the masker on masking-level differences in young and old adults. *The Journal of the Acoustical Society of America, 91*(4 Pt 1), 2129–2135. <https://doi.org/10.1121/1.403673>
- Pichora-Fuller, M. K., Schneider, B., & Daneman, M. (1995). How young and old adults listen to and remember speech in noise. *The Journal of the Acoustical Society of America, 97*(1), 593–608. <https://doi.org/10.1007/s13398-014-0173-7.2>
- Pichora-Fuller, M. K., Schneider, B., MacDonald, E., Pass, H. E., & Brown, S. (2007). Temporal jitter disrupts speech intelligibility: A simulation of auditory aging. *Hearing Research, 223*(1–2), 114–121. <https://doi.org/10.1016/j.heares.2006.10.009>
- Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., Heisterkamp, S., & Van Willigen, B. (2017). nlme: Linear and Nonlinear Mixed Effects Models. *R Package 3rd Edn*. Retrieved from <https://cran.r-project.org/web/packages/nlme/nlme.pdf>
- Pruitt, T. A., & Pfordresher, P. Q. (2015). The role of auditory feedback in speech and song. *Journal of Experimental Psychology. Human Perception and Performance, 41*(1), 152–166. <https://doi.org/10.1037/a0038285>

- Rammsayer, T., & Altenmüller, E. (2006). Temporal information processing in musicians and nonmusicians. *Music Perception, 24*(1), 37–48. <https://doi.org/10.1525/mp.2006.24.1.37>
- Raz, N., Millman, D., & Moberg, P. J. (1989). Auditory memory and age-related differences in two-tone frequency discrimination: trace decay and interference. *Experimental Aging Research, 15*(1–2), 43–7. <https://doi.org/10.1080/03610738908259757>
- Ricketts, T. A., & Hornsby, B. W. Y. (2005). Sound quality measures for speech in noise through a commercial hearing aid implementing digital noise reduction. *Journal of the American Academy of Audiology, 16*(5), 270–277. <https://doi.org/10.3766/jaaa.16.5.2>
- Ridderinkhof, K. R., van der Molen, M. W., Band, G. P., & Bashore, T. R. (1997). Sources of interference from irrelevant information: a developmental study. *Journal of Experimental Child Psychology, 65*(3), 315–341. <https://doi.org/10.1006/jecp.1997.2367>
- Russo, F. A., Ives, D. T., Goy, H., Pichora-Fuller, M. K., & Patterson, R. D. (2012). Age-Related Difference in Melodic Pitch Perception Is Probably Mediated by Temporal Processing. *Ear and Hearing, 33*(2), 177–186. <https://doi.org/10.1097/AUD.0b013e318233acee>
- Russo, F. A., & Pichora-Fuller, M. K. (2008). Tune in or tune out: age-related differences in listening to speech in music. *Ear and Hearing, 29*(5), 746–760. <https://doi.org/10.1097/AUD.0b013e31817bdd1f>

- Russo, N. M., Nicol, T. G., Zecker, S. G., Hayes, E. A., & Kraus, N. (2005). Auditory training improves neural timing in the human brainstem. *Behavioural Brain Research, 156*(1), 95–103. <https://doi.org/10.1016/j.bbr.2004.05.012>
- Salomon, G. (1986). Hearing problems and the elderly. *Danish Medical Bulletin, 33 Suppl 3*, 1–22.
- Salthouse, T. A. (2006). Mental Exercise and Mental Aging: Evaluating the Validity of the "Use It or Lose It" Hypothesis. *Perspectives on Psychological Science, 1*(1), 68–87. <https://doi.org/10.1111/j.1745-6916.2006.00005.x>
- Schellenberg, E. G., & Moreno, S. (2009). Music lessons, pitch processing, and g. *Psychology of Music, 38*(2), 209–221. Retrieved from <http://www.erin.utoronto.ca/~w3psygs/SchellenbergMoreno2010.pdf>
- Schlaug, G. (2015). *Musicians and music making as a model for the study of brain plasticity. Progress in Brain Research* (1st ed., Vol. 217). Elsevier B.V. <https://doi.org/10.1016/bs.pbr.2014.11.020>
- Schlaug, G., Jäncke, L., Huang, Y. Y., & Steinmetz, H. (1995). In vivo evidence of structural brain asymmetry in musicians. *Science (New York, N.Y.), 267*, 699–701. <https://doi.org/10.1126/science.7839149>

- Schneider, B. (1997). Psychoacoustics and aging: Implications for everyday listening. *Journal of Speech-Language Pathology and Audiology*, 21(2), 111–124. Retrieved from <http://www.library.gatech.edu:2048/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=1997-05664-002&site=ehost-live>
- Schneider, B., Daneman, M., & Murphy, D. R. (2005). Speech Comprehension Difficulties in Older Adults: Cognitive Slowing or Age-Related Changes in Hearing? *Psychology and Aging*, 20(2), 261–271. <https://doi.org/10.1037/0882-7974.20.2.261>
- Schneider, B., Pichora-Fuller, K., & Daneman, M. (2010). Effects of Senescent Changes in Audition and Cognition on Spoken Language Comprehension. *Springer Handbook of Auditory Research*, 34, 167–210. <https://doi.org/10.1007/978-1-4419-0993-0>
- Schneider, B., Pichora-Fuller, M. K., Kowalchuk, D., & Lamb, M. (1994). Gap detection and the precedence effect in young and old adults. *The Journal of the Acoustical Society of America*, 95(2), 980–991. <https://doi.org/10.1121/1.408403>
- Schneider, P., Scherg, M., Dosch, H. G., Specht, H. J., Gutschalk, A., & Rupp, A. (2002). Morphology of Heschl's gyrus reflects enhanced activation in the auditory cortex of musicians. *Nature Neuroscience*, 5(7), 688–694. <https://doi.org/10.1038/nn871>
- Shahin, A., Bosnyak, D. J., Trainor, L. J., & Roberts, L. E. (2003). Enhancement of Neuroplastic P2 and N1c Auditory Evoked Potentials in Musicians. *J. Neurosci.*, 23(13), 5545–5552. <https://doi.org/23/13/5545> [pii]

- Skoe, E., & Kraus, N. (2010). Auditory brainstem response to complex sounds : a tutorial. *Ear Hear, 31*(3), 302–324. <https://doi.org/10.1097/AUD.0b013e3181cdb272>. Auditory
- Skoe, E., & Kraus, N. (2013). Musical training heightens auditory brainstem function during sensitive periods in development. *Frontiers in Psychology, 4*(SEP). <https://doi.org/10.3389/fpsyg.2013.00622>
- Solé Resano, C., Mercadal-Brotons, M., Gallego Matas, S., & Riera, M. (2010). Contribution of music to aging adults' quality of life. *Journal of Music Therapy, 47*(3), 264. Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&db=rih&AN=2010-06016&site=ehost-live>
- Souza, P. E., Boike, K. T., Witherell, K., & Tremblay, K. (2007). Prediction of speech recognition from audibility in older listeners with hearing loss: effects of age, amplification, and background noise. *Journal of the American Academy of Audiology, 18*(1), 54–65. <https://doi.org/10.3766/jaaa.18.1.5>
- Strait, D. L., Kraus, N., Skoe, E., & Ashley, R. (2009). Musical experience and neural efficiency - Effects of training on subcortical processing of vocal expressions of emotion. *European Journal of Neuroscience, 29*(3), 661–668. <https://doi.org/10.1111/j.1460-9568.2009.06617.x>

- Strouse, A., Ashmead, D. H., Ohde, R. N., & Grantham, D. W. (1998). Temporal processing in the aging auditory system. *Journal of the Acoustical Society of America*, *104*(4), 2385–2399. <https://doi.org/10.1121/1.423748>
- Syka, J. (2002). Plastic Changes in the Central Auditory System After Hearing Loss, Restoration of Function, and During Learning. *Physiological Reviews*, *82*(3), 601–636. <https://doi.org/10.1152/physrev.00002.2002>
- Thiese, M. S. (2014). Observational and interventional study design types; an overview. *Biochemia Medica*, *24*(2), 199–210. <https://doi.org/10.11613/BM.2014.022>
- Tremblay, K. L., Piskosz, M., & Souza, P. (2003). Effects of age and age-related hearing loss on the neural representation of speech cues. *Clinical Neurophysiology*, *114*(7), 1332–1343. [https://doi.org/10.1016/S1388-2457\(03\)00114-7](https://doi.org/10.1016/S1388-2457(03)00114-7)
- Walton, J. P., Frisina, R. D., & O'Neill, W. E. (1998). Age-related alteration in processing of temporal sound features in the auditory midbrain of the CBA mouse. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *18*(7), 2764–76. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/9502833> <http://www.jneurosci.org/content/18/7/2764.short>
- Werff, K. R. Vander, & Burns, K. S. (2011). Brain Stem Responses to Speech in Younger and Older Adults, (May), 168–180. <https://doi.org/10.1097/AUD.0b013e3181f534b5>

- Wilson, R. H., McArdle, R. a., & Smith, S. L. (2007). An Evaluation of the BKB-SIN, HINT, QuickSIN, and WIN Materials on Listeners With Normal Hearing and Listeners With Hearing Loss. *Journal of Speech, Language, and Hearing Research : JSLHR*, 50(4), 844–856. [https://doi.org/10.1044/1092-4388\(2007/059\)](https://doi.org/10.1044/1092-4388(2007/059))
- Wingfield, A., Tun, P. A., & McCoy, S. L. (2005). Hearing loss in older adulthood: What it is and how it interacts with cognitive performance. *Current Directions in Psychological Science*. <https://doi.org/10.1111/j.0963-7214.2005.00356.x>
- Winkler, I., Debener, S., Muller, K. R., & Tangermann, M. (2015). On the influence of high-pass filtering on ICA-based artifact reduction in EEG-ERP. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS, 2015–Novem*, 4101–4105. <https://doi.org/10.1109/EMBC.2015.7319296>
- Wong, P. C. M., Skoe, E., Russo, N. M., Dees, T., & Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nature Neuroscience*, 10(4), 420–422. <https://doi.org/10.1038/nn1872>
- Woodman, G. F. (2010). A brief introduction to the use of event-related potentials (ERPs) in studies of perception and attention. *Attention and Perceptual Psychophysiology*, 72(8), 1–29. <https://doi.org/10.3758/APP.72.8.2031.A>

- Yamasoba, T., Lin, F. R., Someya, S., Kashio, A., Sakamoto, T., & Kondo, K. (2013). Current concepts in age-related hearing loss: Epidemiology and mechanistic pathways. *Hearing Research, 303*, 30–38. <https://doi.org/10.1016/j.heares.2013.01.021>
- Yueh, B., Shapiro, N., MacLean, C. H., & Shekelle, P. G. (2003). Screening and management of adult hearing loss in primary care: scientific review. *JAMA : The Journal of the American Medical Association, 289*(15), 1976–1985. <https://doi.org/10.1001/jama.289.15.1976>
- Zendel, B. R. (2011). *The effect of lifelong musicianship on age-related changes in auditory processing by auditory processing*. (Doctoral dissertation, University of Toronto).
- Zendel, B. R., & Alain, C. (2009). Concurrent sound segregation is enhanced in musicians. *Journal of Cognitive Neuroscience, 21*(8), 1488–98. <https://doi.org/10.1162/jocn.2009.21140>
- Zendel, B. R., & Alain, C. (2012). Musicians experience less age-related decline in central auditory processing. *Psychology and Aging, 27*(2), 410–417. <https://doi.org/10.1037/a0024816>
- Zheng, Z. Z., Munhall, K. G., & Johnsrude, I. S. (2010). Functional overlap between regions involved in speech perception and in monitoring one's own voice during speech production. *Otolaryngology, 22*(8), 1770–1781. <https://doi.org/10.1162/jocn.2009.21324>. Functional

