The Pennsylvania State University

The Graduate School

College of Engineering

ASSESSING THE AUDITORY AND REWARD RESPONSES TO ROOM ACOUSTICS AND MUSIC USING FUNCTIONAL MAGNETIC RESONANCE IMAGING

A Dissertation in

Acoustics

by

Martin S. Lawless

Submitted in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

May 2018

The dissertation of Martin S. Lawless was reviewed and approved* by the following:

Michelle C. Vigeant
Assistant Professor of Acoustics and Architectural Engineering
Dissertation Advisor
Chair of Committee

Victor W. Sparrow
Professor of Acoustics
Director of the Graduate Program in Acoustics

Daniel A. Russell
Teaching Professor of Acoustics

Nanyin Zhang
Professor of Biomedical Engineering

Pamela M. Cole
Liberal Arts Professor of Psychology and Human Development and Family Studies

Charles J. Limb
Professor of Otolaryngology, University of California San Francisco
Special Member

*Signatures are on file in the Graduate School

ABSTRACT

Room acoustics considerably change how a listener perceives music because spaces introduce temporal and spatial cues into the auditory signals that can alter subjective impression. In particular, the reverberation of a room can influence a listener's overall preference to the musical passage. Reverberation has the potential to affect how the auditory and reward networks in the brain respond to auditory signals, but very little work has studied how room acoustics affects these responses even though much research has examined the effect of music. The overall goal of this doctoral work is to investigate the neural response to room acoustics stimuli using functional magnetic resonance imaging (fMRI). Four studies were conducted with express objectives to: 1) substantiate fMRI as an appropriate tool to examine the reward response to room acoustics, 2) establish a concert hall stimulus set to effectively evoke a reward response, 3) investigate the auditory and reward responses to liked and disliked room acoustic conditions, and 4) further explore the auditory response to coherent and incoherent auditory information presented by room acoustics.

The first experiment, an fMRI pilot study, was first completed with five subjects to ascertain whether a reward response could be detected in the brain due to variations in room acoustic conditions. The individual- and group-level analyses identified that a reward response to preferred concert hall conditions may be significant with the inclusion of more subjects and a set of concert hall stimuli that more readily elicits listener preferences.

The second study, a conventional subjective listening test outside of the MRI scanner with 71 subjects, sought to establish specific musical motifs and room acoustic conditions to improve the stimulus selection process for neuroimaging experiments. Within that objective, the conventional listening study also examined the effect of participant listening expertise on preference to identify suitable candidates for recruitment based on the ability to discriminate between stimuli and the reliability of their preference ratings of reverberant stimuli. A k-means clustering analysis revealed that five groups exhibited starkly different preferences, prompting the selection of individualized stimulus sets for MRI experiments and the recruitment of only expert listeners, defined as musicians.

The third and fourth experiments evaluated the auditory and reward responses to the room acoustics stimuli with fMRI. In the third study, auralizations of each individual's most liked and most disliked room acoustics conditions as well as anechoic musical motifs were presented to 18 participants in an MRI scanner. The results showed that the left and right primary auditory cortices (PAC) activated more so for clearer, simpler stimuli than for the more reverberant, complex stimuli. Specifically, the superior temporal gyri (STG) and Heschl's Gyri (HG) exhibited higher activations in the presence of the anechoic musical motifs (no room acoustics effects) than the most-liked and most-disliked stimuli. Furthermore, a region-of-interest (ROI) analysis of the caudate nucleus and NAcc revealed significant activations in the contrast of the most-liked and most-disliked stimuli as well as the contrast between the most-liked and anechoic stimuli, which were also disliked by the participants. These activations decisively indicated that room acoustics can evoke a reward response.

The fourth study aimed to narrow down the cause of the auditory cortex sensitivity to reverberation, defined as incoherent acoustic information, by presenting all 16 subjects with the same room conditions. The experiment also examined the auditory cortical response to coherent acoustic harmonic content by contrasting the neural responses to the same melody comprised of pure sine tones and trumpet tones, respectively. The results showed that only the presence of reverberation and not the level of the incoherent acoustic information degraded the response of the auditory cortex. Conversely, the presence of coherent harmonic information bolstered the response of the auditory cortex. Since there were no significant interaction effects between the presence of coherent and incoherent acoustic information, these two neural mechanisms were determined to be independent.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ABR	–	Auditory Brainstem Response
ANOVA	–	Analysis of Variance
ANCOVA	–	Analysis of Covariance
BOLD	–	Blood-Oxygen-Level Dependent
BPM	–	Beats Per Minute
BQI	–	Binaural Quality Index
BRIR	–	Binaural Room Impulse Response
BSH	–	Boston Symphony Hall
C_{80}	–	Clarity Index
CNR	–	Contrast-to-Noise Ratio
DLPFC	–	Dorsal Lateral Prefrontal Cortex
EDT	–	Early Decay Time
EEG	–	Electroencephalography
EPI	–	Echo Planar Imaging
FFR	–	Frequency-Following Response
FFT	–	Fast Fourier Transform
fMRI	–	Functional Magnetic Resonance Imaging
FOV	–	Field of View
FWE	–	Family-Wise Error
FWHM	–	Full-Width at Half Maximum
GLM	–	General Linear Model
GUI	–	Graphical User Interface
HG	–	Heschl's Gyrus
HRF	–	Hemodynamic Response Function
HRTF	–	Head-Related Transfer Function
IACC	–	Interaural Correlation Coefficient
IR	–	Impulse Response
IRB	–	Institutional Review Board
J_{LF}	–	Lateral Fraction
JND	–	Just Noticeable Difference

LEV	–	Listener Envelopment
LGP	–	Lateral Globus Pallidus
L_J	–	Late-Lateral Energy Level
MANOVA	–	Multivariate Analysis of Variance
MEG	–	Magnetoencephalography
MNI	–	Montreal Neurological Institute
MUSHRA	–	Multiple Stimulus Hidden Reference and Anchor
NAcc	–	Nucleus Accumbens
PAC	–	Primary Auditory Cortex
PET	–	Positron Emission Tomography
PP	–	Planum Porale
PPI	–	Psychophysiological Interaction
PT	–	Planum Temporale
RIR	–	Room Impulse Response
ROI	–	Region of Interest
RT	–	Reverberation Time
s.d.	–	Standard Deviation
s.e.	–	Standard Error
SLEIC	–	Social, Life, and Engineering Sciences Imaging Center
STG	–	Superior Temporal Gyrus
STS	–	Superior Temporal Sulcus
SVR	–	Slow Vertex Response
T_{30}	–	Reverberation Time, as calculated with 30 dB decay
T_{60}	–	Reverberation Time, as calculated with 60 dB decay
TE	–	Echo Time
TFL	–	Turboflash
TR	–	Repetition Time
VTA	–	Ventral Tegmental Area
WIP	–	Work-In-Progress
Δt_1	–	Initial Time Delay Gap
τ_e	–	Effective Duration of Autocorrelation Function

ACKNOWLEDGEMENTS

I would like to acknowledge the many, many people that have been supportive of my pursuit of a PhD. I will never be able to fully express my love and appreciation in words, but here goes:

First, I would like to thank my academic advisor, Dr. Michelle Vigeant. Michelle, you have been the most wonderful advisor and mentor, and I cannot thank you enough for all of the things, both big and small, that you have done. I appreciate everything that you have taught me in regard to conducting research, and am especially thankful for you sharing your ideas on mentorship, teaching, academia, professional networks, service, and so much more. I enjoyed exploring acoustics and neuroscience with you. I truly appreciate the open communication that we have had about the project and life in general. I look forward to our continued professional relationship, collaboration, and most importantly friendship.

I would also like to thank the members of our research team, the Sound Perception and Room Acoustics Laboratory (SPRAL), both past and present. Thank you all for being there to talk about research or about nothing. I would especially like to acknowledge Matthew Neal and David Dick. Matthew, thank you for the talks about research ideas, the rides home, the dinners we cooked together, and putting up with all of my complaining (and I guess just me in general). Dave, thank you for providing your sage (or not-so-sage) wisdom, humor, and also for putting up with my complaining. I would also like to thank Peter Moriarty for providing an outlet for my need to 3D print brains.

I would like to acknowledge the members of my doctoral committee, Dr. Victor Sparrow, Dr. Daniel Russell, Dr. Pamela Cole, Dr. Nanyin Zhang, and Dr. Charles Limb. Thank you for all of your helpful comments, questions, and critiques. You have all helped this work become the best that it could be. Thank you also to the Social, Life, and Engineering Sciences Imaging Center (SLEIC), especially Dr. Michele Diaz, Debra Weston, Scott Rougeux, Dr. Xiaoxiao Bai, and Dr. Thomas Neuberger.

Throughout the PhD, I would not have remained sane without my family. To everyone in this paragraph and the next, thank you for providing me with a place to call home. Thank you, Mom and Dad, for supporting me in everything that I have sought and giving me a leg to stand on when I needed it. Steven, thank you for giving me the material to start off this dissertation. Thank you Matt, Maggie, Erica, Laura, Paras, Peter, Dave, Jake, Jenn, Alyki, and Jess for being my friends despite the distance.

I could not have made it through this experience at all without my wonderful girlfriend/fiancé/wife (depending on how far in the future we are reading this), Therese Miranda. Therese, you have really held me together through this whole ordeal. Thank you for always being there for me no matter what, no matter when, no matter where. We have endured the almost 5 years of a long-distance relationship. I cannot wait to start our lives together as doctors.

CHAPTER 1

GRAND INTRODUCTION

About four years into my Ph.D. research, my brother asked me, "Why are you mapping the human brain? Won't that lead to artificial intelligence that will destroy humanity?" The answer to the second question is "maybe," but that is not why this research is important. The truth is that there are countless reasons why we, as scientists, must study the brain, the first and most basic being that there are many things that we still do not know about it. Considering how long scientists have been studying the human body, only recently (in the last two decades) have the necessary tools been developed to safely, accurately, and precisely map the behavior and processes of the brain. In the end, this information will help humanity by aiding in the development of more effective treatments for physiological and neurological disorders.

From an acoustician's point of view, mapping the auditory pathway in the brain can provide insight on the perception of sound. Knowing how this pathway works can lead to the development of better hearing aids that work in reverberant spaces or virtual auditory devices that can better exploit human binaural hearing, among others. Without this information, one may have to rely on the subjective experiences of test subjects. For the architectural acoustician in particular, neuroimaging may be used to improve the measurement of these subjective experiences, which are crucial in the design of pleasurable rooms and spaces.

Subjective testing is a useful tool for architectural acousticians, specifically in the design of concert halls, because it allows the acoustician to receive direct feedback from the benefactors of their research. In a typical test setting, subjects rate halls through surveys, formal listening tests, or sensory analysis techniques (Refs. [1], [2], and [3] as an example of each, respectively). The obtained ratings can be used to inform the design of objective metrics within a concert hall setting to improve the overall quality of a space.[4]

Though the subjective testing approach provides good insight into the perceptions of the listeners, reliability can pose a problem if not adequately controlled. A number of factors, including guessing, mistakes in submitted responses, and misunderstanding of constructs or scales, contribute to poor test-retest reliability.[5,6] Even when these factors are minimized, responses of individuals to explicit questions may not be entirely accurate. Inconsistencies have been reported between explicit and implicit cognitive (conscious decision),[7] attitudinal,[8,9] and emotional[10] responses. Moreover, implicit affective processes (i.e. feelings and emotions) have been shown to be independent of explicit perceptual and cognitive responses.[11,12] These results suggest that conventional subjective testing using questions assessing explicit responses may not truly reflect how a subject perceives differences in room acoustics stimuli.

Neuroimaging techniques offer an evaluation of implicit responses to stimuli by measuring physiological responses. In an ideal world, these techniques would allow us to peer directly into a participant's brain (not unlike Fig. 1.1). Although this allowance has yet to become reality, neuroimaging can provide an objective measure to the subjective experience by quantifying the neural activity in the brain. The present work uses functional magnetic resonance imaging (fMRI) methods to reveal the cortical and subcortical response to room acoustic stimuli of simulated concert halls. This particular technique has helped uncover the auditory cortical response and subcortical reward response to music,[13] but until now has not been used to measure these responses to music played in different room conditions.

The goal of the present work is to investigate the neural correlates to room acoustics stimuli using fMRI. To achieve this goal, this dissertation over the course of four experiments aims to validate fMRI as an appropriate tool, establish appropriate concert hall stimuli and participant types, and determine the auditory and reward responses to reverberant stimuli.

Fig. 1.1: Peering into the lying brain. Adapted from Brett Ryder.[14]

1.1 ROOM ACOUSTICS BACKGROUND

1.1.1 Characterization of rooms: the impulse response

The formal, academic study of room acoustics dates back to 1895 when Wallace Clement Sabine attempted to improve the acoustics of a lecture hall by characterizing its *reverberation time (RT),* the time it takes for the sound level to fall 60 dB after an excitation in a room.[15] Since then, rooms have been described with many more acoustical metrics, most of which derive from the impulse response (IR). In any system, the IR is the characteristic response to an excitation. For rooms, the room IR (RIR) corresponds to the sound level recorded by a microphone after an impulsive excitation, such as a balloon burst, though it may be measured or simulated with a variety of different techniques.[4]

Classically, the RIR is divided into direct, early, and late sound (see Fig. 1.2). The direct sound, typically represented by the first peak in the RIR, is the initial sound level at the receiver from the source. For the early part of the sound (before 80 ms), the sharp peaks in the RIR typically represent distinct reflections from surfaces around the receiver, while the late part contains overlapping higher-order reflections which constitute the diffuse sound field.[4]

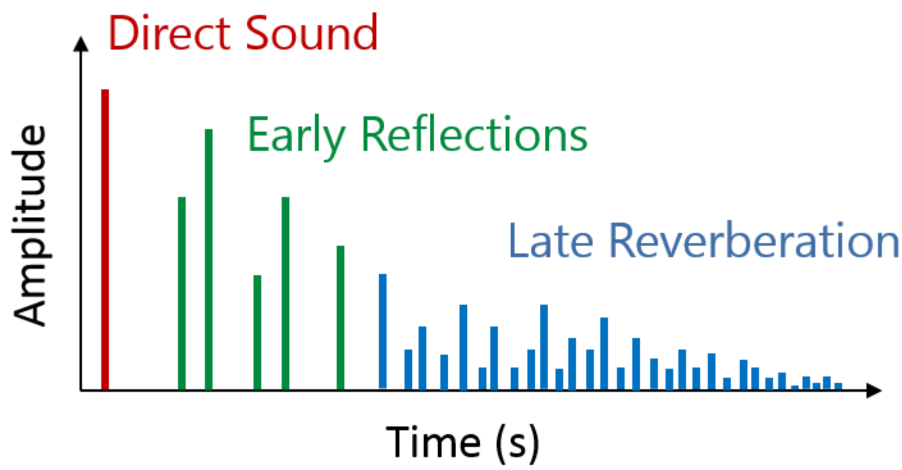


Fig. 1.2: Graphical representation of an idealized room impulse response divided into the direct, early, and late sound by color. From Neal (2015), Figure 4-2.[16]

1.1.1.1 Simulation of the RIR

All of the room acoustics stimuli in this dissertation were generated by convolving a binaural room impulse response (BRIR) with an anechoic recording of music. The resulting auralizations (like visualizations, but with sound) presented a realistic reproduction of the musical passages playing in the simulated rooms. The BRIRs were created in Odeon, a room acoustics software package.[17] Odeon generates BRIRs with a hybrid method that considers the early and late sound, separately.

The early part of the BRIR is calculated with the image source method. With this approach, reflections are modeled as images of the source mirrored across flat surfaces (see Fig. 1.3). After the direct sound, the reflections from these image sources arrive at different times due to the distances between the image sources and the receiver and with different amplitudes due to the absorption and scattering properties of the surfaces. Higher-order reflections derive from reflecting the image sources across additional surfaces. Since the number of image sources increases exponentially with the order of the reflections, it quickly becomes computationally inefficient to generate the entire impulse response with the image source method. Consequently, Odeon limits the order of the reflections, and uses a ray-tracing algorithm to ensure that only "visible" image sources (i.e. those in a direct ray, not obscured by another surface) from the receiver contribute to the BRIR.

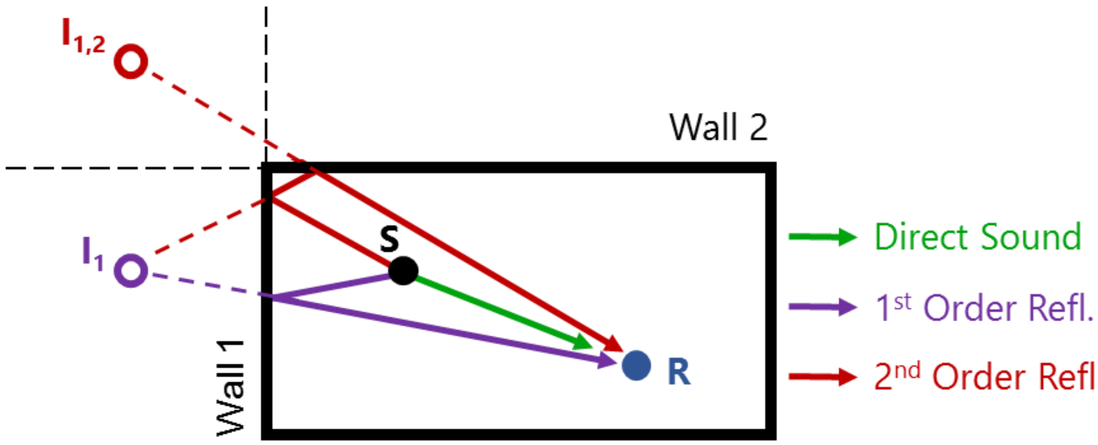


Fig. 1.3: Illustration of the image source method. Only one first-order and one second-order reflections are shown.

The ray-tracing method produces the late part of the BRIR in Odeon. Thousands of rays proceed in random directions from the source and account for specular, scattered, and diffracted energy in the room. A spherical grid placed at the location of the receiver detects ray intersections. The information from the rays is used to help shape a stochastically generated reverberant tail in time, frequency, and space to match the overall behavior of the late energy of the particular room geometry and material properties.

Several user-inputs explicitly affect the resulting BRIR. Since the image source and ray tracing methods depend on the locations of reflections, the geometry of the room model, including the placement of the source and receiver positions, greatly impact the BRIR. These features determine the timing and direction of the direct and early sound. Secondly, the absorption and scattering properties of the surfaces influence the amplitude of the early reflections and late reverberation. Each of the experiments in this dissertation utilized the same source/receiver positions in the same room geometry model that varied only by the absorption and scattering properties of the surfaces. Therefore, all of the BRIRs generated for the listening tests had the same number of early reflections that arrived at the receiver at the same time. However, the level of the early reflections and late reverberation differed between the simulated rooms and enabled the differences in the listener perceptions.

1.1.2 Reverberance and reverberation time

Many attributes can describe an individual's perceptual experience in a concert hall. Examples of attributes include: *loudness*; *intimacy*, the feeling of being close to a source; *listener envelopment (LEV)*, the sense of being immersed in a sound field; and *reverberance*, the perception of how sound lingers in a room. While each of these attributes contribute in some manner to the overall impression of a space, research has shown that intimacy, LEV, and reverberance greatly impact overall preference.[2,18-20] Many studies have demonstrated that reverberance in particular is an important factor that influences a listener's preference.[1,18,21] Listeners generally prefer a moderate amount of reverberance for musical performances – too little does not adequately support the music, but too much causes the notes to muddle together.[22]

Reverberance significantly correlates with the metrics of reverberation time, such as RT, T_{30}, and early decay time (EDT).[23-25] As mentioned above, RT (or T_{60}) denotes the amount of time that it takes for the IR to decay by 60 dB. Reverberation time measured by T_{30} is the amount of time for the IR to decay by 30 dB, and then multiplied by 2 to be comparable to T_{60} times. Neither T_{60} nor T_{30} consider the direct sound as part of the calculation. Instead, the calculations of these metrics on the smoothed IR start at 5 dB below the direct sound to ensure that the metrics accurately characterize the late reverberation. The computation of EDT however begins at the direct sound level and determines the amount of time for the IR to decay 10 dB, which is then multiplied by 6 to again compare to traditional T_{60} metrics. In this manner, EDT characterizes the contribution of the early part of the IR.

Since the room models used in this dissertation altered the absorption and scattering properties of the surfaces while keeping the same source and receiver positions, the levels of the early reflections and overall late reverberation changed proportionately. Although T_{30} and EDT describe different aspects of the IR, the manner in which the BRIRs were constructed dictates that T_{30} and EDT characterize the rooms in the same way. In fact, T_{30} and EDT significantly correlated (Pearson's correlation coefficient, $R = 0.998$). T_{30} was the metric used to denote differences between the room acoustics stimuli since it was more established in the literature. In regards to preference, several studies have shown that highly rated halls have T_{30} between 1.6 and 2.1 seconds, and can vary based on the type of music played (e.g. chamber or symphony).[4]

1.1.3 Conventional listening test methods

A variety of listening test methods exist that may be used in many different applications. For example, researchers in the audio industry can utilize listeners to determine preferences to samples or the appropriate bit rate needed to accurately reproduce a signal. Concert hall acousticians typically rely on listeners to ascertain the subjective hall characteristics, such as the perceptions of reverberance, clarity, and listener envelopment, and correlation the extent of the impressions with objective acoustic metrics, such as reverberation time (T_{30}), clarity index (C_{80}), and late-lateral energy level (L_J), respectively.[26]

In concert hall acoustics, listening tests may occur in a laboratory or *in situ*, i.e. in the concert hall itself. Each of these scenarios has tradeoffs. An *in situ* test, like that reported in Ref. [18], has the advantage of giving the participants direct access to the spaces. Since a concert hall experience involves a multitude of modal stimuli, such as auditory and visual, having the listeners rate their perceptions to an actual performance allows them to truly have a sense of the hall.

Unfortunately with *in situ* listening tests, the experimenter loses control over many experimental aspects, including those concerning subjects and sound sources. The following discussion will focus solely on the loss of control over subject variabilities. If an experiment only takes place during one performance, each of the subjects will listen from different locations, which can drastically affect their perceptions of a space since the distances to each of the sources will be different and reflections will come from different directions. For example, a subject sitting much closer to the orchestra may have stronger impression of source clarity than another may further away in the reverberant field. If the research concerns multiple halls, it is extremely unlikely for the same participants to hear each one (as in Ref. [22]), which makes for a statistically unbalanced design. Even if one were able to get the exact same participants in each concert hall, humans have poor auditory memory,[27] causing the comparisons between halls to be inaccurate.

Laboratory listening tests give researchers control over such aspects by sacrificing the real experience of the concert hall. Typically, the participants in these tests listen to the stimuli over a loudspeaker array or headphones. Loudspeaker arrays provide a natural, out-of-head experience, and can accurately replicate the rooms using spatial reproduction techniques, such as Ambisonics.[28] However, the use of loudspeaker arrays is not always practical, especially for experiments that take place inside the small, constrained environment of an MRI machine, such as the present work. Headphones afford a portable alternative, but an individual's head-related transfer function (HRTF), the response that dictates how anatomical features affect the sound arriving at the ear, is required to provide accurate spatial impressions of the room (or at least a good estimation of the HRTF is necessary).[29] The MRI environment in particular necessitates a careful selection of headphones since one cannot bring ferrous or magnetic objects into the magnetic field of the scanner.

Once reproduction methods have been established, the researcher must decide which specific listening test method will best gather the subjective data (refer to Fig. 1.4 for examples). Many listening tests utilize the successive method of testing (T1 in Fig. 1.4) where each stimulus is consecutively presented to the listener one at a time.[5] This method most emulates the *in situ* testing style since listeners typically cannot repeat stimulus playback and must rate after receiving the entire stimulus. However, reliability and accuracy may be improved by permitting the participants to directly compare two or more stimuli. The paired comparison method (T3 in Fig. 1.4) presents two stimuli to the subjects on the same screen and forces them to choose one based on a specific question, such as which one is more preferred. The multi-stimulus comparative method (T2 in Fig. 1.4) takes this a step further, and presents all of the stimuli in question, along with a rating scale for each, on the screen at the same time. While the comparative method does not directly promote a one-to-one comparison of all the stimuli since the subjects do not necessarily compare each stimulus to every other one like in the pair-wise method, subjects tend to perform faster on the former test, which either reduces listening fatigue and/or allows for longer testing periods.[30] (A fuller description of the successive and comparative methods is included in Chapter 3; the conventional listening tests in the present dissertation utilize these methods).

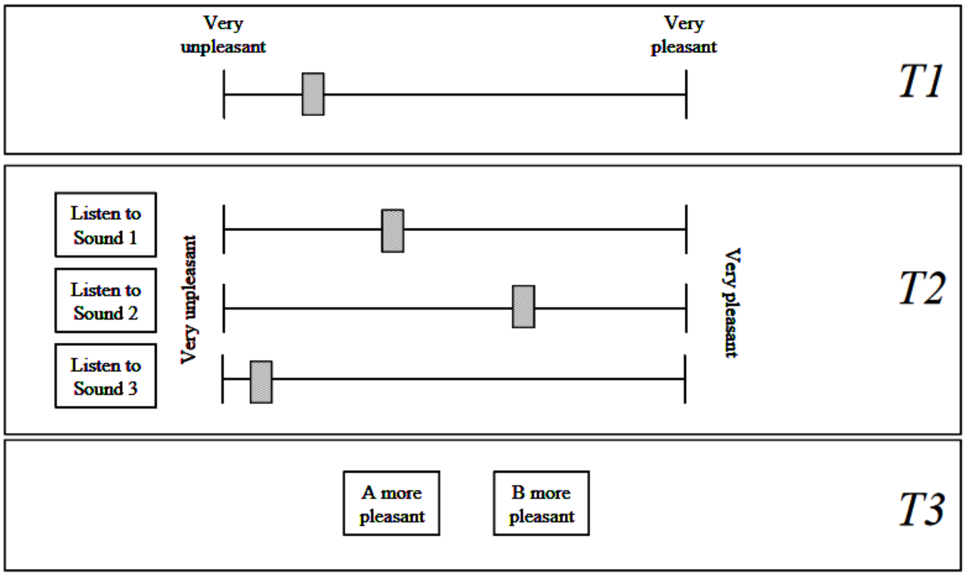


Fig. 1.4: Examples of the successive (T1), comparative (T2), and paired comparisons (T3) listening test methods. Image from Figure 1, Parizet et al. (2005).[30]

Whether *in situ* or laboratory tests are performed, conventional listening test methods commonly analyze the results with a repeated-measures analysis of variance (ANOVA).[31] An ANOVA statistically assesses the differences between group averages with respect to the amount of variance associated with each mean. When a test measures the response to the same dependent variable over multiple subjects (i.e. repeatedly), samples obtained from the same subject (within-subject) may not necessarily be independent of one another. A repeated-measures ANOVA accounts for the covariance between ratings and removes the variability between subjects when calculating the means. All of the statistically analyses conducted in the present work, including the fMRI analyses, use repeated-measures ANOVAs.

1.2 NEUROSCIENCE BACKGROUND

1.2.1 Introduction to fMRI methods

Functional neuroimaging techniques aim to measure localized neuronal activity in order to associate behaviors and stimuli with different brain regions or activity patterns. Functional MRI indirectly determines neural activity in the brain by acquiring images of the blood-oxygen-level dependent (BOLD) response, a measure of the hemodynamic response of neurons.[32] The MRI machine contains a strong, permanent magnetic field (typically 1.5 or

3 T) that causes the hydrogen nuclei (protons) in the body to align. By applying a gradient magnetic field throughout the permanent one, the hydrogen nuclei are spatially segregated. A radiofrequency (RF) pulse perturbs the nuclei from the magnetic alignment and the machine measures the localized energy emitted as the nuclei return to equilibrium.

Because oxygenated and deoxygenated hemoglobin have different magnetic properties, fMRI sequences can resolve the location of the oxygen required by the neurons as they activate, i.e. the hemodynamic response (blue line in Fig. 1.5a). The BOLD measurement corresponds to the amount of deoxygenated hemoglobin in a given region since deoxygenated hemoglobin distorts the MR signal. When a neuron activates, blood surges to the localized area, leading to an increase in oxygenated hemoglobin and the BOLD signal. After the neuron receives the oxygen that it needs, the BOLD signal returns to baseline.[33]

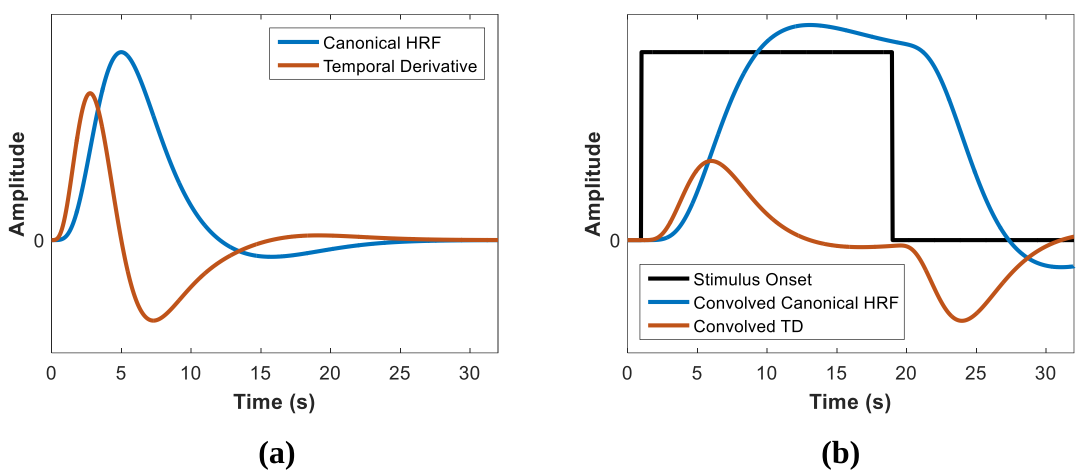


Fig. 1.5: (a) The canonical hemodynamic response function (HRF, in blue) and its first derivative, called the temporal derivative (in orange). (b) The canonical HRF and temporal derivative convolved with the step function corresponding to the block stimulus presentation (in black).

The statistical analysis of fMRI data compares the measured BOLD signal to a general linear model of the hemodynamic response.[34] To model the response over time, the hemodynamic response function (HRF) is convolved with a step function that has the length of the stimulus presentation. The resulting model, shown in Fig. 1.5b, has an extended length. Note that the BOLD response in blue lags behind the stimulus presentation because it does not necessarily correspond to the neuron firing, but rather the body's reaction to the activation. The peak of the response occurs several seconds after the stimulus onset and the signal continues after the stimulus has ended.

While the canonical HRF is good at modeling the BOLD response across most cortical regions, it does not work well for some. In fact, the hemodynamic response can vary substantially across both regions and people. Researchers can use a basis set that expands the capability of the HRF to account for the variability.[35] By taking the first and second derivatives of the HRF, a temporal basis set that can model the small differences in the latency and duration of the peak response. Including these in the statistical analysis effectively reduces noise in the fMRI data. The first derivative, called the temporal derivative, and the convolved temporal HRF are shown in Fig. 1.5 in orange.

The presentation of a stimulus induces small, local changes in the BOLD signal. However, the brain is always functionally active, which makes it difficult to detect these small signal changes during the period when a specific stimulus is presented. The local signals may be isolated by contrasting test images with images acquired during the presentation of a control stimulus.[34] In this method, which could be thought of as a "subtraction method," activated areas common to both the test and control stimuli disappear, while activations or deactivations caused by the test stimulus are emphasized. Statistical tests then determine whether the differences between the test and control images are significant. Effective control stimuli are designed to be similar enough to the test stimuli to activate the same regions of the brain of no interest to the experimenter, but differ in a key aspect of processing that may be isolated. An example of auditory or acoustical control stimulus may be a monaural representation of an auditory scene when one wishes to uncover binaural processes with the test stimuli.

The fMRI technique allows for a temporal resolution on the order of seconds and a high degree of spatial resolution of about 1-3 mm.[32] The temporal resolution appertains to the present work since the musical stimuli are 15-20 seconds long. The excellent spatial resolution supports the study of small regions of interest. Additionally, fMRI is able to detect the BOLD response of both cortical and subcortical regions, while other neuroimaging methods such as electroencephalography (EEG) are limited to cortical activations and inferior spatial resolution. (An advantage of EEG is the high temporal resolution on the order of milliseconds, which can be beneficial in short event-related experimental design).

The fMRI image contains tens of thousands of data points, called voxels (volume pixel). Since the statistical analysis tests each of these data points concurrently, corrections for multiple comparisons are made to reduce the number of false negative results, also known as type I error. Several correction methods exist to lessen the amount of type I errors including the family-wise error (FWE) rate and false discovery rate (FDR).[34] FWE rates, such as Bonferroni corrections,[31] are more stringent than FDR, and tend to be used more frequently. However, the corrections can be overly conservative, especially when the hypotheses of the experiments only concern certain regions in the brain. Region-of-interest (ROI) analyses focus the statistical tests on specific areas.[34] In these analyses, the fMRI data is averaged over an ROI, which is typically a predetermined anatomical region based on the hypothesis. This method reduces the number of statistical tests to the number of ROIs considered, and is useful for uncovering significant activations that may not necessarily appear in the whole brain analysis.

1.2.1.1 Overview of fMRI acquisition parameters

To collect MRI data, the experimenter must choose a particular imaging protocol based on the type of images they wish to acquire.[34] The MRI experiments in this dissertation used echo planar imaging (EPI) sequences to obtain functional data. EPI sequences exploit the gradient magnetic field to obtain image slices in a matter of 20-100 ms. In this manner, entire brain volumes can be collected in a couple of seconds instead of minutes. Other sequences, such as Turboflash (TFL) used to acquire structural data, sacrifice this temporal resolution for higher spatial resolution. The imaging sequence is defined by specifying values for several acquisition parameters, outlined in Table 1.1 below. The parameters affect both the spatial resolution and the contrast of the image, which emphasizes certain types of brain matter (e.g. gray matter, white matter).

Table 1.1: fMRI acquisition parameter descriptions and typical ranges for echo-planar imaging (EPI) and Turboflash (TFL) sequences, which collect functional and structural data, respectively.

Parameter	**Description**	**Typical Range**	
		EPI	*TFL*
Repetition Time (TR)	The time between two RF excitation pulses used to obtain images.	1-3 seconds (one brain volume)	1-3 seconds (one brain slice)
Echo Time (TE)	The time between an RF excitation pulse and the peak of the measured MR signal.	< 30 ms	< 5 ms
Matrix Size	The spatial resolution of each 2D slice determined by the frequency- and phase-encoded gradient sequence. Multiple of 2.	64, 128	> 128
Field of View (FOV)	The physical extent of the acquired volume determined by the number of slices and slice thickness.	210-250 mm	
Number of Slices	The number of acquired slices per whole volumes.	< 50	> 100
Slice Thickness	The spatial resolution of each individual slice.	2-3.5 mm	1 mm
Flip Angle	The amount of rotation, in degrees, from the z-axis caused by the RF pulse that determines signal intensity.	70-110°	< 10°

1.2.1.2 How to read fMRI data

This section will provide an overview of how to interpret the results reported on brain maps in this dissertation for those who are unfamiliar with the analysis of fMRI data. Fig. 1.6 contains example functional images from two stimuli and the activation map from the method of contrasts, which was detailed in Section 1.2.1. The "greater than" symbol (">") denotes that the functional data obtained during the presentation of stimulus 2 were subtracted from the functional data of stimulus 1. The resulting activation map has areas of "positive activation" and "negative activation", indicated in Fig. 1.6 in red and blue, respectively. Positive activation means that the measured BOLD response of that voxel was larger during the presentation of stimulus 1, while negative activations, which could also be referred to as *deactivations*, express that the measured BOLD response was greater during stimulus 2. The statistical significances of these activations are determined by *t*-

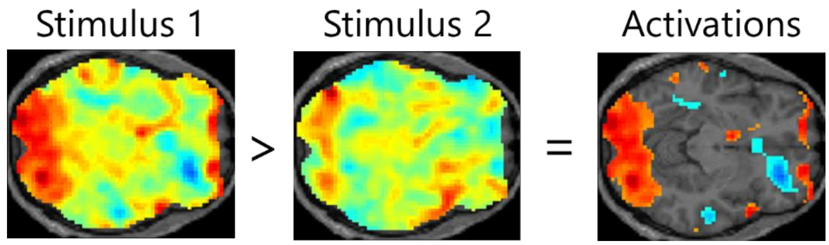


Fig. 1.6: Example contrast demonstrating how activations are isolated. In this example, positive activations are colored red, while negative activations are colored blue.

tests at each individual voxel (and corrected for multiple comparisons), and are reported in accompanying tables. Positive *t* values represent positive activations, and vice versa.

For each contrast, the implicated regions are determined by the x, y, z coordinates of the activated or deactivated voxels.[34] Each subject's functional and structural MRI data is normalized to fit a prescribed atlas, or coordinate system. One of the most commonly used atlases is the Montreal Neurological Institute (MNI) stereotaxic coordinate system, which was developed using an average of structural MRI scans from 152 normal, right-handed participants. The names or descriptions of the implicated regions are typically included in the accompanying statistics tables.

1.2.1.3 Acoustic MRI challenges and considerations

The MRI environment provides quite a number of challenges for presenting auditory stimuli. As mentioned above, one cannot introduce magnetic or ferrous materials into the magnetic field for safety reasons, and therefore must choose a playback system that operates without a magnet, such as electrostatic headphones. Additionally, the MRI sequences may be loud due to large, rapid changes in the electrical current through the gradient coils. While functional scans tend to be quieter than structural scans, measured sound levels of these protocols can be up to 96 dB (ref: 20 μPa). One must provide sound attenuation not only for the subjects to hear the auditory stimuli, but also to protect their hearing.

The sounds emitted from the MRI machine can change depending on the specifications of the acquisition parameters. The frequency and loudness of the MRI noise rely on the designated repetition time (TR) and echo time (TE). Also, the timing of the collection of the brain data can be adjusted to lower the scanner noise during the presentation of auditory stimuli with sparse temporal sampling or interleaved silent steady state imaging sequences.[36,37] Although these methods can reduce the interference of scanner noise during listening periods,[38] they sacrifice signal intensity[39] and likely miss the peak and transient aspects of the hemodynamic response for musical passages that are multiple seconds in duration.

1.2.2 Brief overview of the auditory cortex

Auditory information transduces from sound waves to electrical impulses in the cochlea, processes through the midbrain and thalmus, and ultimately ends in the auditory cortex as the last stop of the official auditory neural pathway (see Fig. 1.7). It is in the auditory cortex that the basic cognitive perception of acoustic signals takes place. Without cortical auditory areas, humans would most likely not have any awareness of sound.[40]

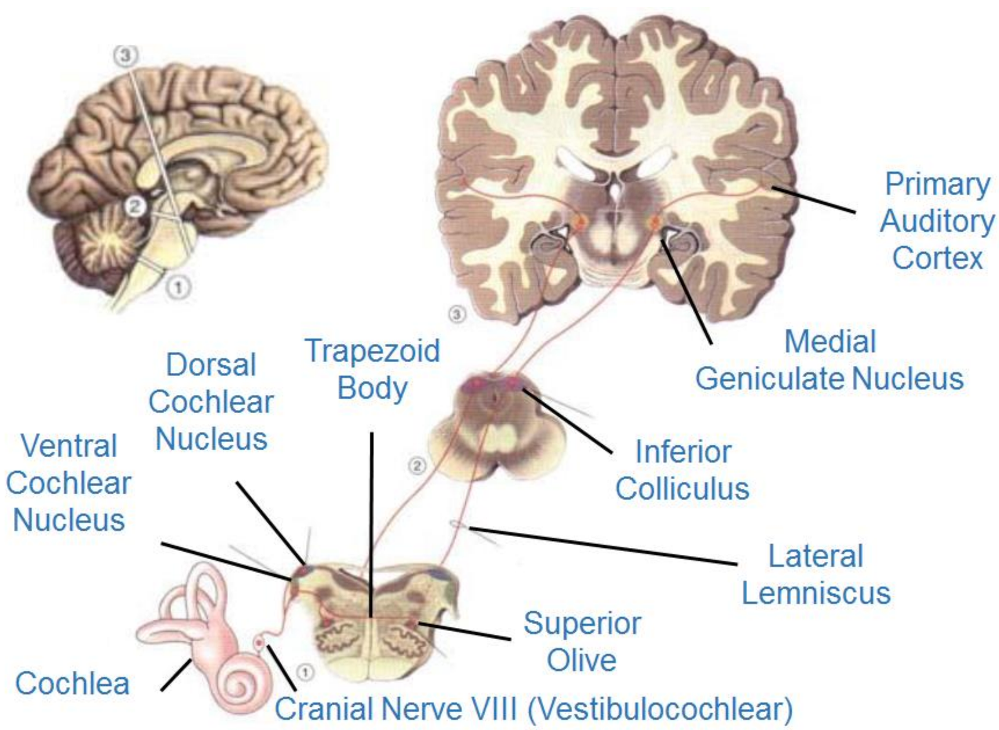


Fig. 1.7: The auditory neural pathway. Auditory signals travel from the cochlea to the cochlear nuclei via the eighth cranial nerve. The cochlea nuclei send bilateral efferents to the superior olivary complexes. The information then travels up the lateral leminiscus to the inferior colliculi. The auditory signal is modulated in the medial geniculate nucleus of the thalamus, and then finally processed in the primary auditory cortex, which lies along the superior temporal gyrus (STG). Image adapted from Fall 2015 NEURO 512 course lecture notes by Dr. Kevin Alloway at Penn State University.

The human auditory cortex lies along the bilaterally superior temporal gyri (STG) and Heschl's gyri (HG), and receives information from both the left and right ears.[41] It is generally divided into primary and secondary areas. The primary auditory cortex (PAC, also known as area Te1), located mainly on HG, contains a tonotopic map for the processing of frequency information preserved throughout the auditory pathway. Architectonic research has shown that the PAC comprises three distinct regions, Te 1.0, 1.1, and 1.2 (Fig. 1.8),[42] that may each serve unique processes. Area Te 1.0 is the most primary, core region, while Te 1.1 and 1.2 cover more medial and lateral areas, respectively.

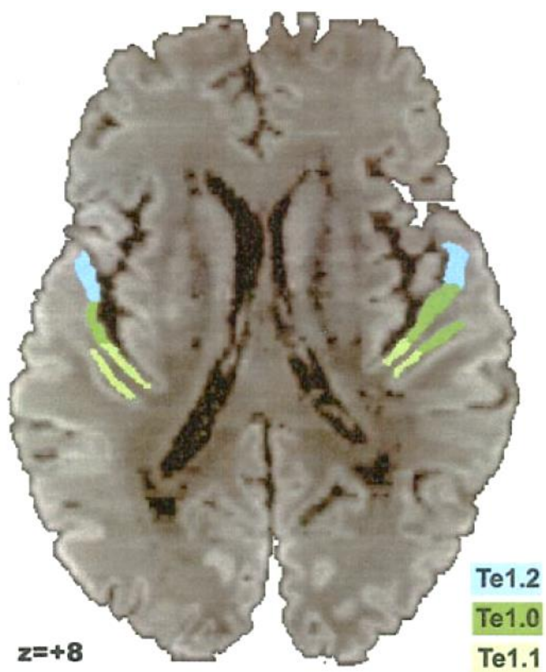


Fig. 1.8: Map of areas Te 1.0, 1.1, and 1.2 on a horizontal histological section. Adapted from Fig. 11 from Morosan et al. (2001).[42]

Beyond the PAC, secondary auditory areas analyze specific characteristics of the auditory signal.[41] These "belt" areas specialize in sound recognition and localization, as well as speech and music perception. Some of these processes may occur asymmetrically in the brain. For example, speech perception and production is lateralized* to the left side of the

* The term "lateralization" has different meanings in acoustics and neuroimaging. Acoustic, or auditory, lateralization is the perception or localization of sounds that appear to originate within the head when reproduced over headphones.[167] Lateralization in neuroimaging studies signifies that neural activations are biased, or specialized, to one hemisphere or the other.[168] This dissertation adheres to the latter definition in any mention of lateralization.

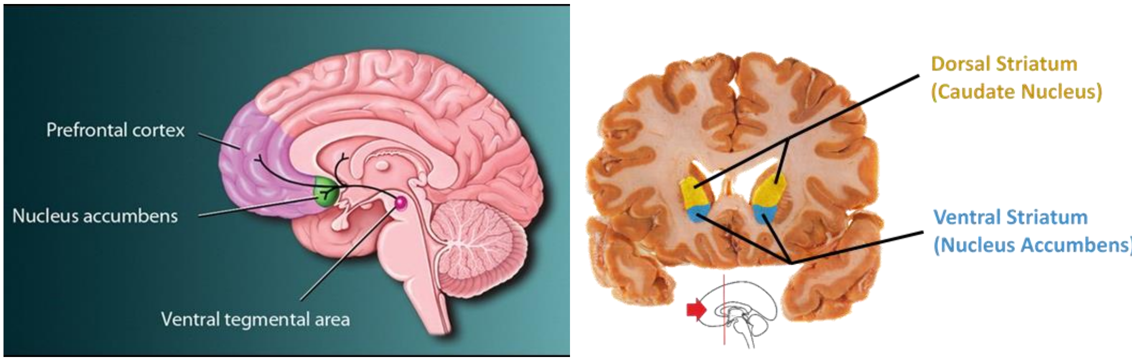
brain in Wernicke and Broca's areas, respectively.[43] On the other hand, music perception such as melodic processing has been associated with activations in the right hemisphere. Along those lines, temporal cues may be processed more effectively by the left auditory cortical areas, while the right regions manage spectral cues.[44]

1.2.3 Functional correlates between music and reward

The study of music's effect on human emotion is extraordinarily multi-disciplinary. It spans studies of musical culture, anthropology, sociology, psychology, and neurobiology. Each discipline has its own perspective and interest in how music production and listening influences emotions and development.[45] Although preference is not classically considered an emotion, research has often included preference as an arousal to pleasurable stimuli (e.g. Ref. [46]). The following discussion focuses on how the brain processes preferred stimuli through reward.

Although pleasure is colloquially thought to be a solely subjective experience, humans exhibit physiological and neurological responses that can serve as objective measures.[47] One such measureable response occurs in the mesolimbic reward pathway (Fig. 1.9a). This response involves an increased dopaminergic release in the ventral striatum of the basal ganglia from the ventral tegmental area (VTA).[40,48] Several different measurement techniques have detected the reward response, including fMRI.

Though rewarding stimuli are typically thought to be tangible incentives, such as food, sex, drugs, and money, music can also activate reward regions of the brain despite being a more abstract stimulus. The reward response to music involves two main regions in the basal ganglia: the caudate nucleus and the nucleus accumbens (NAcc), shown in Fig. 1.9b. The NAcc, located in the ventral striatum of the basal ganglia, activates during the presence of rewarding music.[46,49,50] Other studies have shown that pleasurable music stimulates the caudate in the dorsal striatum during the anticipation of reward.[51]



(a) **(b)**

Fig. 1.9: (a) The mesolimbic pathway includes a dopaminergic connection between the ventral tegmental area (VTA) and the nucleus accumbens (NAcc).[52] (b) The caudate nucleus and the NAcc have been shown to activate during the anticipation and experience of a reward, respectively. Adapted from Ref. [40].

The experiments referenced above typically tested unique musical passages, sometimes specifically chosen by the participants themselves as their favorite and least favorite compositions (e.g. Ref. [49]). The present studies propose to measure musical passages differing by the much more subtle effect of various acoustic conditions. While the room conditions may be diverse from anechoic to extremely reverberant, the underlying musical passages are the same. The neural response differences between liked and disliked room conditions would ostensibly be much smaller than the differences between liked and disliked musical passages. Thus, it was important to validate that fMRI would be able to detect such smaller changes in the reward network.

1.2.4 Electroencephalography correlates of room acoustics stimuli

Very few neuroimaging experiments exist that investigate the effects of room acoustics on the subjective impression of auditory signals. In fact, the body of the work comprises merely two sets of EEG experiments involving speech and music stimuli, respectively, that examined evoked potentials in the auditory brainstem and cerebral cortex. Bidelman and Krishnan (2010) studied the effects of reverberation on the brainstem frequency-following responses (FFRs) to speech signals.[53] Reverberation caused a reduction in the magnitude of the FFR encoding of formant-related harmonics in both musicians and non-musicians. However, the neural encoding of pitch was not affected by the amount of reverberation in the signals. Another interesting finding of the study was that musicians exhibited stronger responses to the speech stimuli across almost all of the room conditions, which suggests that musicians are more equipped to process acoustic features.

Ando et al. focused on the auditory brainstem response to examine the perceptual effects of room acoustics on music, specifically looking at how subjective preference influences the slow vertex response (SVR).[54] The latency of the SVR of the auditory evoked potential correlated to a listener's preference of loudness, delay time of the first early reflection (Δt_1), and interaural cross-correlation coefficient (IACC), which is a measure of the disparity between left and right ear signals.[55] However, this effect was not observed for stimuli with variations in reverberance; in fact, no significant correlations were found between SVR and RT.

When the brainstem response did not correlate with reverberance preference, Ando et al. examined the gross cortical response to the room acoustic stimuli to investigate listener preferences to reverberation.[56] Analysis of the α-wave rhythms (8-13 Hz) in the EEG signal did reveal that the effective duration, τ_e, of the autocorrelation functions of the signal in the left hemisphere was greater in an RT-preferred stimulus than a non-preferred condition, while τ_e averaged over the right hemisphere was lower in the preferred case. While this result seems promising, the opposite effect occurred, i.e. right hemisphere τ_e was greater for the preferred case, when IACC varied between stimuli. The discrepancy between the RT and IACC results suggest that preference was not truly being measured with τ_e, especially since the metric was averaged over entire cortical hemispheres and the measurement system could not detect activations in subcortical regions involved in reward or preference.

After this review, many questions remained unanswered concerning the cortical auditory response to reverberation and the preference response. Room acoustics, specifically reverberation, could potentially affect other auditory mechanisms, including the response of the auditory cortex. Moreover, the reward or preference response to reverberant stimuli was yet to be truly characterized. The present studies provide some insight on these unanswered questions.

1.3 DISSERTATION OUTLINE

The remainder of the dissertation follows:

Chapter 2 discusses a pilot study conducted to determine if fMRI would be a suitable technique for measuring a reward response to room acoustics stimuli. Note, this chapter is a reproduction of a published article, entitled M. S. Lawless and M. C. Vigeant, "Investigating the emotional response to room acoustics: a functional magnetic resonance imaging study," *J. Acoust. Soc. Am.*, **138,** EL417-EL423, 2015. Footnotes have been added to provide additional clarification.

Chapter 3 contains a reproduction of M. S. Lawless and M. C. Vigeant, "Effects of test method and participant musical training on preference ratings of stimuli with different reverberation times," *J. Acoust. Soc. Am.*, vol. 142, no. 4, pp. 2258-2272, 2017. This experiment investigated the influence of the musical expertise of participants on room acoustics preferences and the differences between the successive and comparative listening test methods. Footnotes were also added to this chapter to provide additional clarification.

Chapter 4 details an fMRI study examining the auditory cortical response to reverberant stimuli and the reward response to preferred room acoustics stimuli.

Chapter 5 consists of a final fMRI study that further investigates the auditory sensitivity to reverberation as well as the sensitivity to musical instrumental timbre.

Chapter 6 provides summary, conclusion, and recommended future work.

Appendix A contains relevant MATLAB code and further information for creating the control stimuli for the fMRI experiments.

Appendix B includes pictures of the equipment used in the MRI and listening experiments.

CHAPTER 2

INVESTIGATING THE EMOTIONAL RESPONSE TO ROOM ACOUSTICS: A FUNCTIONAL MAGNETIC RESONANCE IMAGING STUDY

The following chapter presents an fMRI pilot study that intended to verify if the neuroimaging technique could detect a reward response to room acoustic conditions. The chapter reproduces a manuscript published in the Journal of the Acoustical Society of America in October 2015, (*J. Acoust. Soc. Am.* **138**, EL417 (2015)). Footnotes have been added to provide additional clarification as recommended by the doctoral committee. The text in its original form may be found at: http://asa.scitation.org/doi/10.1121/1.4933232.

Reprinted in compliance with the Acoustical Society of America "Transfer of Copyright Agreement"* from:

> M. S. Lawless and M. C. Vigeant, "Investigating the emotional response to room acoustics: a functional magnetic resonance imaging study," *J. Acoust. Soc. Am.*, **138,** EL417-EL423, 2015. Copyright 2015, Acoustical Society of America

* Available at: http://asa.scitation.org/pb-assets/files/publications/jas/jascpyrt.pdf

Investigating the emotional response to room acoustics: a functional magnetic resonance imaging study

M. S. Lawless and M. C. Vigeant

Graduate Program in Acoustics, The Pennsylvania State University, University Park, Pennsylvania 16802, USA

(Originally published in: *J. Acoust. Soc. Am.* **138**, EL417 (2015))

Abstract

While previous research has demonstrated the powerful influence of pleasant and unpleasant music on emotions, the present study utilizes functional magnetic resonance imaging (fMRI) to assess the positive and negative emotional responses as demonstrated in the brain when listening to music convolved with varying room acoustic conditions. During fMRI scans, subjects rated auralizations created in a simulated concert hall with varying reverberation times. The analysis detected activations in the dorsal striatum, a region associated with anticipation of reward, for two individuals for the highest rated stimulus, though no activations were found for regions associated with negative emotions in any subject.

2.1 INTRODUCTION

Music can elicit a wide range of emotions, including happiness, sadness, and reward. By indirectly measuring neuronal activations in various regions of the brain, functional neuroimaging methods, such as positron emission tomography (PET) and functional magnetic resonance imaging (fMRI), have shown that music provokes intensely pleasurable experiences,[49,57] as well as strong unpleasant feelings.[46,58] In particular, activations of the nucleus accumbens (NAcc) in the ventral striatum correlate with dopaminergic mechanisms that prompt reward-motivated behavior,[48] and has been shown to occur after the stimulation of pleasurable music.[50,51] The anticipation of reward, associated with activations in the dorsal striatum, specifically the caudate, precludes this reward-like behavior.[51,59]

A room's acoustics affect the perception of music and the preferences of listeners. The room impulse response (IR), typically divided into the early and late sound, can be used to derive objective measures that correlate with a listener's subjective experience. One such measure, reverberation time (RT), quantifies how long sound lingers in a space, and is one of the acoustical factors that contributes to the overall impression of a room.[2,21] The optimal range of RT for a concert hall has been estimated to be between 1.6 and 2.1 seconds, but an individual's preference may differ depending on personal opinion or musical period.[1,4] Listeners fall into several groups according to overall preference[18] and reverberation.[3] Though listeners may have different affinities for their optimal RT preference, individuals should experience reward activity while listening to pleasing stimuli, i.e. stimuli at their individual preferred RT. Electroencephalography (EEG) experiments have shown correlations between measured brain wave patterns and stimuli with preferred room acoustics.[54] However, EEG is limited to measuring neural activity in the cortical regions of the brain, which do not necessarily process emotions. Thus, the present work proposes to use fMRI, which has the capacity to measure the entire brain, to assess the responses to pleasing and unpleasing RT stimuli in emotional regions of the brain.

In this study, fMRI was used to measure the blood-oxygen-level dependent (BOLD) response to stimuli of varying reverberation to incite positive and negative emotions. Participants were asked to rate their preference for each stimulus to determine which one offered the most pleasing acoustics. Stimuli with a moderate amount of reverberation were expected to elicit higher preference ratings than those with no reverberation or excessive reverberation. We hypothesized that stimuli with pleasing room acoustics would trigger activations in the dorsal striatum (caudate) and ventral striatum (NAcc) involved in the response to reward and positive emotions. Additionally, it was hypothesized that activations would occur in the parahippocampal gyrus and amygdala, regions engaged in the processing of negative emotions, during the presentation of displeasing stimuli.

2.2 METHODS

Testing consisted of two sessions over a period of two days for each participant. The first session took place in a mock scanner, the Psychology Software Tools, Inc. MRI Simulator™, while the second occurred in an actual MRI scanner, the Siemens 3T MAGNETOM Trio™ (refer to Appendix B, Fig. B.1). The simulator was used to familiarize the subjects with the small, restricting physical MRI environment, which also presents a loud, tonal noise during operation. The same test was performed by the participants in both the simulator and scanner to allow for the validation of the results obtained in the actual scanner. During the first session, the participants completed a tutorial before beginning the test.

2.2.1 Subjects

Five healthy, right-handed subjects (2 male, 3 female) between the ages of 21 and 39 participated in the study. All participants were native English speakers. Each subject had extensive musical training – an average of 18 years of formal lessons and actively playing in an ensemble or lessons at the time of the testing. Participants were required to have a maximum hearing threshold of 15 dB HL for octave band frequencies from 250-8000 Hz, which was screened before the first session. Consent was obtained from each participant prior to testing, in accordance with the protocol approved by the Internal Review Board at Penn State. A sixth subject participated in the experiment, but was not included in the results due to excessive movements in the scanner.

2.2.2 Music and paired noise stimuli

To generate the auralizations used as stimuli, a single, binaural IR was calculated using a model of the Boston Symphony Hall in Odeon v12.00. The model was artificially modified to simulate an excessively reverberant condition with an RT of 9.0 s. Five IRs with RTs ranging from 0.0 to 5.3 s were derived from this single IR by manipulating the late sound. By fitting curves with various double-slopes over the original IR, the new IRs had the same energy in the direct and early sound, but deviated in the late reverberant tail. This method was executed to maintain the listener's impressions that primarily depend on the early part of the IR.

The auralizations (music stimuli) were created by convolving the five IRs with an anechoic excerpt. The 11.2 second excerpt was extracted from an anechoic recording of Bruckner's 8th Symphony obtained from Aalto University.[60] The extracted section, played in *tutti forte* (meaning the orchestra played in unison at level *forte*), comprised predominantly eighth notes at an *Allegro moderato* tempo of approximately 132 beats per minute.

In order to perform a main contrast of interest analysis with the fMRI data, control stimuli were produced with the purpose of activating the same regions of the brain as the music stimuli except for those regions involved in reward and emotion processing. For each music stimulus, a white noise signal was filtered such that the frequency content of the music and the noise matched closely. The temporal envelope of each music signal was then applied to the matched noise to integrate rhythmic cues. Five control stimuli were made in this manner – one corresponding to each music stimulus.[*]

2.2.3 Stimulus presentation

The binaural stimuli were presented over STAX MR 003-MK2 electrostatic, MRI-compatible earbuds (Fig. B.2). The playback was equalized to comfortable listening levels at 75 dBA. Howard Leight QM24+ Earmuffs attenuated the noise from the simulator and the scanner by approximately 24 dB from measured levels as high as 96 dB (ref: 20 μPa). The scanner noise predominantly consists of 928 and 945 Hz tones pulsing at a rate of 17 Hz. A custom-built Python graphical user interface (GUI) was employed to present the stimuli and collect ratings. The music stimuli were interleaved with noise stimuli to permit a sufficient amount of time for the BOLD response in the reward regions of the brain to decay between music stimuli.

An indicator tone (0.5 s of silence, 1.0 s of 440 Hz sine wave, 0.5 s of silence) preceded each presentation. After listening to a music stimulus, the participants used Nordic Neurolab Response Grips (see Fig. B.3) to interact with the computer program that was being projected into the bore of the MRI machine. The participants rated each stimulus on "Overall Preference". The continuous, 5-point scale ranged from -2 to +2, representing "Strongly Dislike" to "Strongly Like". The midpoint of the scale (0) was labeled "Neither

[*] Refer to A.1 for more information on the construction of the control stimuli.

Like nor Dislike". Subjects did not rate noise stimuli and were presented with silence after each noise stimulus. Ten seconds were allotted after each stimulus either for the rating prompts or silence.

The test consisted of four functional runs containing the 10 stimuli presented twice in random order for a total of 20 stimuli per run. As a result, each stimulus was repeated eight times across all runs. The large number of repetitions was used in order to ensure significant MR signals in the fMRI data analysis. Each run required 8 minutes and 44 seconds to complete. Between the first two runs, 4 minutes were used to perform structural scans. Short intermissions were given between the other runs for the participants to rest. The entire length of the scanning session for each subject was approximately 45 minutes.

2.2.4 fMRI acquisition

The structural and functional data were obtained in the Social, Life, and Engineering Sciences Imaging Center (SLEIC) at Penn State. A 12-channel head coil was used to collect all imaging data. Structural images were acquired with a T1-weighted Turboflash (MPRage) sequence with 160 sagittal slices each 1.0 mm thick. The repetition time (TR) and echo time (TE) for the MPRage sequence were 1650 ms and 2.03 ms, respectively, with a FOV of 256 mm and matrix size of 256 by 256. An echo planar imaging (EPI) sequence (TR = 2000 ms, TE = 28 ms, flip angle = 78° [61]) was used to collect the BOLD functional images. The 260 volumes with physical dimensions of 210 by 210 by 108 mm^3 consisted of 33 axial slices with a slice thickness of 3.3 mm, FOV of 210 mm, and a matrix size of 64 by 64. The total length of the EPI sequence was 8 minutes and 44 seconds, which was matched to the length of the stimulus presentation sets.

2.2.5 Image processing

The structural MRI and fMRI data were processed using statistical parametric mapping via the SPM8 Matlab platform.[62] The fMRI images were realigned using a three-dimensional coordinate reference frame with six degrees of freedom. The three translational and three rotational realignments were checked to ensure that the subjects moved no greater than 1.0 mm in the x-, y-, or z-directions, nor rotated more than 1° in the pitch, roll, and yaw directions. The amounts of realignments in each direction were used as nuisance regressors in the statistical analysis. Slice-timing acquisition differences were corrected for using sinc-interpolation. Spatial normalization was applied to the images, which were resampled into 3-mm isotropic voxels. An 8-mm FWHM Gaussian kernel filter was applied to smooth the data.

2.2.6 Statistical analysis

An analysis of variance (ANOVA)[31] of the subjective preference ratings was conducted using a repeated-measures model in SAS 9.4. The interaction between the ratings and type of session (simulator or scanner) was also modeled to determine if the same ratings were given in both the mock and actual scanners. The highest and lowest rated stimuli, representing the most liked and most disliked reverberance levels, were noted and used in the statistical analysis of the fMRI data.

The functional data were analyzed using a classical GLM mass-univariate approach through SPM8. Eleven conditions were established in the block design matrix for the five reverberance levels, five noise counterparts, and the condition of rating the stimuli. Movement vectors generated in the realignment preprocessing stage contributed as six additional regressors. Linear contrasts between each music condition and corresponding noise condition were evaluated. (A contrast between two conditions is denoted as *Condition 1* > *Condition 2*.) Contrasts were also created between the music stimuli rated the highest and the lowest, as noted above. One-sample t-tests for each aforementioned contrast were conducted in a second-level group analysis. A region of interest (ROI) analysis [34] was performed over a small volume sphere of 25-mm radius centered at (0,14,4) mm MNI coordinates [63] for each contrast listed above. This region included the ventral and dorsal striatum, thalamus, and putamen, which correlate with the anticipation and experience of reward. Due to the large number of voxels being measured simultaneously, a family-wise error rate (FWE) correction for multiple comparisons was applied to the p-values to determine significance.[34]

2.3 RESULTS

2.3.1 Preference ratings

The subjective preference ratings were averaged across all participants for each reverberation level of music stimuli. Fig. 2.1 shows the averaged results for the simulator and MRI scanner sessions. The preference ratings in the simulator were not significantly different from the ratings in the MRI scanner based on the insignificance of the interaction between RT and type of session (p = 0.304). This result indicates that future tests may not require participants to rate the stimuli in the MRI scanner, which would avoid confounding visual stimuli and motor movements after stimulus presentation.

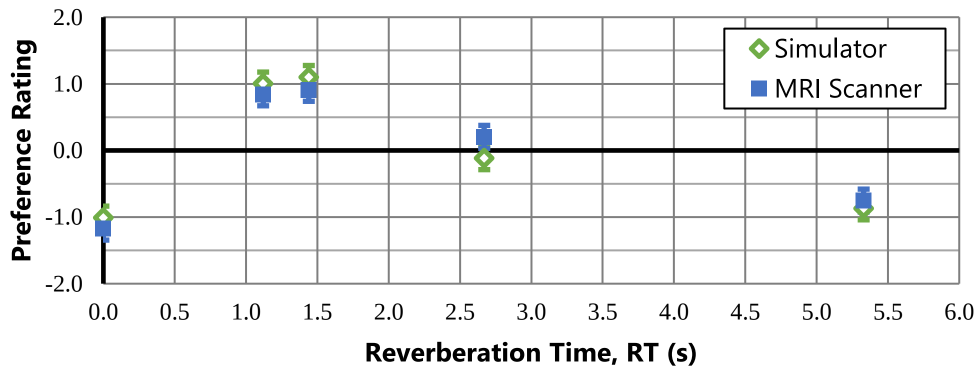


Fig. 2.1: Preference ratings as a function of RT. The 1.2 and 1.4 s stimuli were the "Most Liked," the 0.0 and 5.3 s stimuli were the "Most Disliked," and the 2.7 s stimulus had a neutral rating (p<0.001).

The ANOVA results specified the most preferred and least preferred stimuli using the Tukey method.[31] The anechoic (RT = 0.0 s) and most reverberant stimuli (RT = 5.3 s) were the "Most Disliked" stimuli with average ratings of -1.09 and -0.81, respectively. The two stimuli that rated the highest, the "Most Liked" stimuli, were those with RTs of 1.4 and 1.2 s. These stimuli had average ratings of +1.06 and +0.92, respectively. The stimulus with an RT of 2.7 s was rated neutrally with an average rating of 0.04. From this analysis, neuroimaging contrasts were created between each "Liked" stimulus and each "Disliked" stimulus.

2.3.2 Brain activations

Five contrasts for each of the music and noise pairs were created to analyze the activations (e.g. *RT 1.4 Music > RT 1.4 Noise*). An additional four contrasts were generated to examine differences in brain activations for the two "Most Liked" conditions and the two "Most Disliked" conditions (e.g. *RT 1.4 Music > RT 0.0 Music)*. To determine significant findings, the minimum cluster size was set to 20 voxels, and the threshold p-value was set to 0.001 before the FWE correction for multiple comparisons.

For some of the *Music > Noise* contrasts in three of the five subjects, activations were detected in the insula ($p < 0.05$, corrected), the brain region that handles emotional recall.[64] Insular stimulation likely occurred in the music conditions over the noise conditions because subjects were only prompted to rate the music stimuli. Thus, when the subjects were required to remember their preference to the musical stimuli, the insula was engaged.

The ROI analysis (Table 2.1) determined that left caudate activations occurred while two subjects listened to a pleasing stimulus. Activations in the caudate surpassed the threshold for the *RT 1.4 Music > RT 1.4 Noise* contrast with corrected p-values of 0.001 and 0.033 for each subject, respectively. Fig. 2.2 shows the activations for one of the subjects, denoted "Subject 1." The stimulation of the caudate agrees with previous studies, which have demonstrated caudate activations during the anticipation of a reward from musical stimuli, suggesting that these participants may have experienced an anticipation of reward during the music stimulus and not the noise stimulus. While this inference could be resolved for any of the music and noise stimulus combinations, since the noise stimuli were more unpleasant than the music, caudate activations were only significant during the presentation of the "Most Liked" stimulus. Therefore, the results of these two subjects support the claim that reward anticipation processes correlate with pleasing room acoustics.

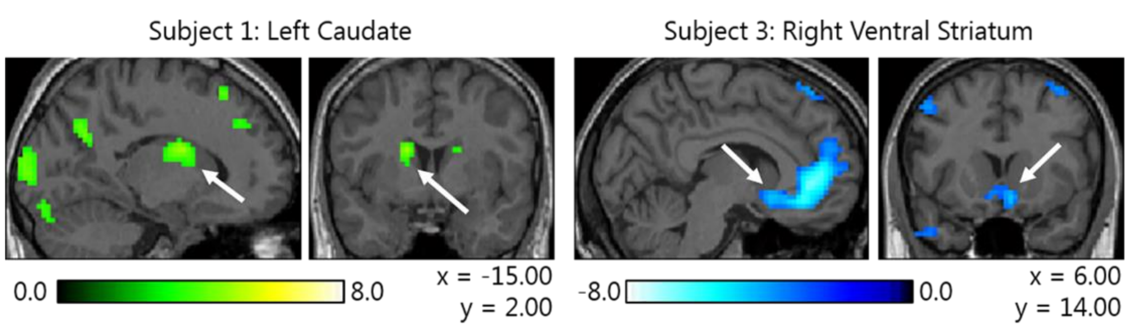


Fig. 2.2: ***RT 1.4 Music > RT 1.4 Noise*** **contrast for two subjects. Subject 1 (left) experienced positive activations in the caudate (p = 0.001, corrected), while Subject 3 (right) exhibited deactivations in the ventral striatum (p = 0.001, corrected).**

Table 2.1: ***RT 1.4 Music > RT 1.4 Noise*** **contrast ROI analyses of a small volume of 25-mm radius centered at (0, 14, 4) MNI coordinates.**

Analysis	Region	Peak-Level			MNI Coor. (mm)		
		$p_{FWE\text{-}corr}$	*T*	*Z*	*x*	*y*	*z*
Subject 1	**Left Caudate**	0.001	4.71	4.68	-15	2	19
	Left Thalamus	0.049	3.75	3.74	-3	-10	10
Subject 2	**Left Caudate**	0.033	3.89	3.88	-18	17	7
Subject 3	**Right Anterior Cingulum**	< 0.001	-6.40	-6.33	9	35	-5
	Right Ventral Striatum	0.001	-4.66	-4.63	6	14	-11
	Left Ventral Striatum	0.066	-3.60	-3.59	-6	14	-11
Group	**Right Caudate**	0.665	9.46	3.39	18	17	10

Contrary to the hypothesis, a third subject demonstrated negative activations in the ventral striatum, including the nucleus accumbens (NAcc), for the *RT 1.4 Music > RT 1.4 Noise* contrast. The strength of the deactivations in Subject 3 (Fig. 2.2) were the same as the activations in Subject 1, both with FWE corrected p-values of 0.001. This result would suggest that Subject 3 preferred the RT 1.4 s noise stimulus to the music counterpart. While that outcome is entirely possible, it is unlikely since Subject 3 rated the RT 1.4 s music stimulus the highest. If the participant had preferred the noise stimuli overall, activations in the NAcc should have also been as significant, if not more, in the contrasts of the disliked stimuli. Since no significant activations or deactivations in the NAcc were detected in the other contrasts, Subject 3 probably did not prefer the noise stimuli. A possible reason for the NAcc deactivation involves the impulsivity of decision making. Previous research has shown that in NAcc inhibited by lesions[65] or chemical receptors,[66,67] impulsive behavior increases compared to uninhibited NAcc. Due to the time constraints imposed on the rating of stimuli, Subject 3 may have felt the need to rush and make quick decisions on preference, resulting in the deactivation of the NAcc and reward response for the RT 1.4 s stimulus. Since the deactivations were not found in the contrasts of the other stimuli, the RT 1.4 s music stimulus, Subject 3 may have found this stimulus to be more difficult to rate than the others.

To find activations in the second-level, group analysis, the minimum cluster size was kept at 20 voxels, but the p-value threshold was relaxed to 0.005 since the analysis only included five participants. No significant activations were found in any of the one-sample t- tests that assessed the nine contrasts between subjects. In the ROI-analysis of the *RT 1.4 Music > RT 1.4 Noise contrast*, the threshold map showed evidence of activations in the right caudate, but these activations were not significant after FWE correction (Fig. 2.3). The lack of significance at the group-level may be a result of the small number of participants.

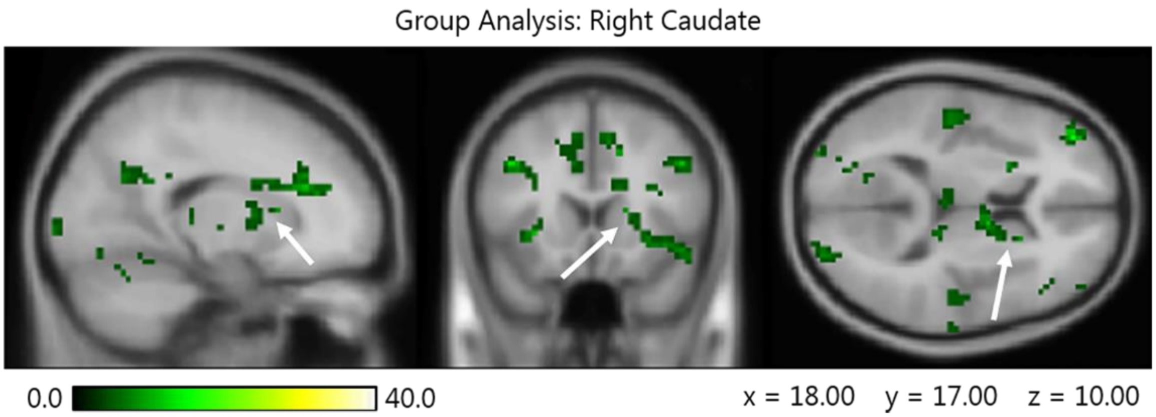


Fig. 2.3: *RT 1.4 Music > RT 1.4 Noise* contrast from the group analysis. Positive activations were found in the right caudate, but was not significant, possibly due to the small number of participants ($p < 0.001$, uncorrected; $p = 0.665$, corrected). Threshold maps superimposed on smoothed template brain (avg152T1 in SPM).

In neither the individual nor group analyses were significant activations exhibited in the parahippocampal gyri, amygdala, or other regions of the brain associated with negative emotions. The absence of activations in these regions was unexpected given that participants disliked two of the stimuli. The stimuli may not have been potent enough to elicit strong negative emotional responses from the subjects, also evidenced by the fact that only two of the five subjects exhibited positive activations in the reward regions. The two testing sessions as well as the large number of repetitions may have saturated the activations. In addition, stimuli may have been too similar to facilitate a high contrast to noise ratio (CNR) in contrasts between two music stimuli.

2.4 CONCLUSION

For the first time, this work has examined individual preferences to room acoustics by objectively measuring activity in emotional regions of the brain with fMRI. Subjectively, the five participants gave the highest preference ratings for the two stimuli with a moderate amount of reverberation. Two of the participants demonstrated positive activations in anticipatory reward regions of the brain for the contrast of the most preferred condition and noise baseline as hypothesized. This result suggests that pleasing room acoustics can trigger the anticipation of reward in some individuals. On the other hand, one participant experienced negative activations in the NAcc in the same contrasted conditions. The deactivations may have been an effect of impulsive decision making, but more evidence is needed to support this claim. Across all individuals, no significant positive activations were found in the NAcc for the "Most Liked" stimuli, nor in the parahippocampal gyri or amygdala for the "Most Disliked" stimuli. A group analysis revealed that the dorsal striatum may activate while a person listens to pleasurable room acoustics, reinforcing the conclusions attained from the first two subjects, but the inclusion of more subjects is needed to strengthen this assertion. Also, the stimuli may not have elicited strong enough responses from the subjects. Future work includes the development and subjective testing of stimuli aimed at provoking stronger emotional responses and a larger range of preference ratings. With more powerful stimuli, additional fMRI data will be obtained that may verify the results of this pilot study.

2.5 ACKNOWLEDGMENTS

We thank the staff of the SLEIC at Penn State, especially Dr. Thomas Neuberger and Dr. Xiaoxiao Bai for help with fMRI data acquisition and stimulus presentation, and Dr. Michele Diaz for reviewing this manuscript. We also thank Dr. Nanyin Zhang for his assistance with experimental design and manuscript review, and Dr. Zhifeng Liang for his support in data analysis. Lastly, we thank Dr. Victor Sparrow and Tae Kwang Yoo for their assistance in obtaining the MRI scanner noise measurement. Work was completed under the Penn State Internal Review Board Protocol #44471.

CHAPTER 3

EFFECTS OF TEST METHOD AND PARTICIPANT MUSICAL TRAINING ON PREFERENCE RATINGS OF STIMULI WITH DIFFERENT REVERBERATION TIMES

The following chapter presents two conventional listening test experiments with the overall goal of building a set of stimuli to use in future fMRI studies. The first experiment also aimed to certify that successive and comparative listening test methods yielded comparable results. The second experiment examined differences between the preference responses of musicians and non-musicians. The chapter reproduces a manuscript published in the Journal of the Acoustical Society of America in October 2017, (*J. Acoust. Soc. Am.* **142**, 2258 (2017)). Footnotes have been added to provide additional clarification as recommended by the doctoral committee. The text in its original form may be found at: http://asa.scitation.org/doi/abs/10.1121/1.5006065.

Reprinted in compliance with the Acoustical Society of America "Transfer of Copyright Agreement"* from:

> M. S. Lawless and M. C. Vigeant, "Effects of test method and participant musical training on preference ratings of stimuli with different reverberation times," *J. Acoust. Soc. Am.*, **142**, 2258-2272, 2017. Copyright 2017, Acoustical Society of America

* Available at: http://asa.scitation.org/pb-assets/files/publications/jas/jascpyrt.pdf

Effects of test method and participant musical training on preference ratings of stimuli with different reverberation times

Martin S. Lawless and Michelle C. Vigeant

Graduate Program in Acoustics, The Pennsylvania State University, University Park, Pennsylvania 16802, USA

(Originally published in *J. Acoust. Soc. Am.* **142**, 2258 (2017))

Abstract

Selecting an appropriate listening test design for concert hall research depends on several factors, including listening test method and participant critical-listening experience. Although expert listeners afford more reliable data, their perceptions may not be broadly representative. The present paper contains two studies that examined the validity and reliability of the data obtained from two listening test methods, a successive and a comparative method, and two types of participants, musicians and non-musicians. Participants rated their overall preference of auralizations generated from eight concert hall conditions with a range of reverberation times (0.0–7.2s). Study 1, with 34 participants, assessed the two methods. The comparative method yielded similar results and reliability as the successive method. Additionally, the comparative method was rated as less difficult and more preferable. For study 2, an additional 37 participants rated the stimuli using the comparative method only. An analysis of variance of the responses from both studies revealed that musicians are better than non-musicians at discerning their preferences across stimuli. This result was confirmed with a k-means clustering analysis on the entire dataset that revealed five preference groups. Four groups exhibited clear preferences to the stimuli, while the fifth group, predominantly comprising non-musicians, demonstrated no clear preference.

3.1 INTRODUCTION

Concert hall research fundamentally involves discerning listener preferences to various room acoustics settings in order to improve the design of performance spaces. An essential component of this research includes subjective testing to directly evaluate the perceptions of listeners. Many subjective studies utilize *in situ* surveys,[18] conventional laboratory listening tests,[2] or sensory analysis techniques[3] to assess participant perceptions of the hall. Research in the field has also begun to explore uncovering the emotional response to room acoustics by measuring physiological responses, such as neural activity[54,68] and skin conductance.[69]

Within the context of traditional laboratory listening tests, a variety of subjecting testing designs exist that may be employed to obtain a listener's subjective preference to auditory stimuli. Researchers in a previous study compared six such test methods investigating the pleasantness of in-car ventilation noises.[30] Out of the six test methods, four were variations on a paired-comparison test, one was a successive method, and the last was a multi-stimulus comparative design. The paired-comparison method presents a forced choice in which a participant must choose the preferred stimulus between two competing excerpts. In the successive method of testing, the subject is prompted to individually rate each stimulus after it has played to completion. The multi-stimulus comparative method presents all the stimuli on the screen at the same time. The subjects are able to listen to each stimulus as many times as necessary and can instantaneously switch back and forth between stimuli. The primary findings of the study established that the paired-comparison designs were much more time intensive than either the successive or comparative methods, most likely because paired-comparisons necessitate $k(k-1)/2$ trials, where k is the number of test stimuli. The authors went on further to recommend the multi-stimulus comparative method for use in subjective studies since the subjects yielded more consistent results than the successive method while maintaining a high level of discrimination between stimuli.[30]

In selecting an appropriate test method for an acoustics subjective study, one must consider other factors in addition to testing time and the reliability of the data. For example, *in situ* studies in concert halls necessitate a successive testing method since one is unable to instantaneously compare two different conditions. Additionally, laboratory studies utilizing physiological measurements require a successive method of testing because correlations must be drawn between each stimulus separately and the physiological activity. In this regard, the time to complete a set of the successive method in such studies is fixed because it depends entirely on the length of the stimuli. In studies without physiological measurements, other listening tests may be more desirable, e.g. the comparative method recommended by Parizet et al.,[30] which may exhibit differences in reliability and the time necessary to complete a set. Whether the comparative method could predict responses of a study that requires the successive method is contingent upon the similarity of results between the two tests. While Parizet et al. found that the two test methods produced similar results for in-car ventilation noise,[30] a study has yet to be completed analyzing these methods for room acoustics stimuli, which are spectrally and perceptually very different from noise.

The second topic of this paper involves selecting an appropriate participant pool for concert hall subjective studies. A literature survey of concert hall subjective studies from 2011-2015 was conducted to determine the types of participants commonly used in concert hall research. Peer-reviewed journals were obtained by searching in the Thomson Reuters Web of Science database with the keyword ‘concert hall’ and the year published. Articles were excluded if the study did not contain a subjective listening test, did not focus on concert hall research, used data that had been previously reported in the literature, or specifically required musical performers. Twenty-one unique principal research institutions contributed to 41 experiments from 35 articles that fit these criteria. The experiments were examined to determine if the researchers recruited expert listeners (e.g. musicians, audio engineers, acousticians) and/or naïve listeners (e.g. non-musicians, general public). The literature survey (see Table 3.1) revealed that 66% of the studies included expert listeners only, 17% included a mix of expert and naïve listeners, 2% used naïve listeners only, and 15% did not state subject experience levels.[3,19,68,70-101]

Table 3.1: The number of experiments that recruited each type of participant (41 total). The literature survey included 35 peer-reviewed journal articles with concert hall subjective testing from 2011-2015.

Types of Participants	Number of Experiments	
Expert Only	27	(66%)
Naïve Only	1	(2%)
Mixed Expert and Naïve	7	(17%)
Unknown	6	(15%)

While the majority of experiments tested expert listeners, only five articles (14%) provided reasons as to why expert listeners were recruited. These articles demonstrated or cited that experts more readily learn the testing procedures[102,103] and may have previous experience with the testing constructs.[104] Other studies have also emphasized that subjects with prior critical listening experience rate stimuli more consistently[105-108] and can better distinguish between stimuli with small differences.[106,108,109] These studies provide strong evidence of the importance of including expert listeners, which may imply that the choice of participant type could have significant impacts on the overall findings of subjective-based research.

Despite the consistency of the ratings, the use of expert listeners as representatives for the entire population is only valid if expert and naïve listeners produce similar results. In a subjective test of loudspeaker preferences of trained and untrained listeners, Olive found that 12 expert listeners demonstrated similar preferences as 256 inexperienced listeners.[107] Although one study has shown that expert and non-expert preferences to the visual aspects of concert halls may be aligned,[84] no such conclusion has been demonstrated concerning the auditory preferences of listeners in concert hall research.

The term "expert" must first be defined, though it may change depending on the objective of each study. For example, in the study by Olive mentioned above, experts were described as a panel of listeners from Harman International Industries Inc. who were highly experienced with loudspeaker subjective tests. In the literature survey, 46% of the studies specified musicians or those with formal musical training as experts, and the present study adopted this definition. The population of concert-goers consists of both musicians and non-musicians,[110] and if musicians and non-musicians exhibit diverse subjective ratings, musicians should not be the only participants tested.

Research in the field of psychology has shown that musicians and non-musicians are similarly able to differentiate between the emotional affinities of several musical excerpts. Bigand et al. and Vieillard et al. separately utilized multidimensional scaling to demonstrate that the subjects consistently grouped musical passages into the same emotional categories regardless of musical experience.[111,112] On the other hand, musicians and non-musicians may not experience concert halls in the same way. Cox and Shield found that musician and non-musician concert-goers perceived the quality of the hall's sound differently.[110] Recent work by Roy et al. has also demonstrated that musicians and non-musicians may exhibit diverse preferences to musical excerpts with variations in reverberation time (RT).[113] However, the stimuli tested had RT values outside of the range of typical concert halls (RT = 3, 5, and 20 s, along with anechoic and RT = 1 s) and were generated using artificial reverberation, which may not have provided a realistic listening experience. The present study reevaluates the recruitment of musicians and non-musicians for concert hall subjective studies based on the consistency and similarity of preference ratings to reverberant stimuli.

The overall goal of this work was to investigate the influence of two key experimental design considerations, test method and participant type in terms of musical training, on preference ratings of stimuli with varying concert hall conditions, while also taking into account test-method and participant-type reliability. The first objective was to compare the successive and comparative listening test methods for reliability and to determine if the ratings from the two methods are statistically indistinct as found by Parizet et al. in their study on sound quality of vehicle ventilation noise. The second objective was to determine the necessity of including non-musicians (inexperienced listeners), in addition to musicians (experienced listeners), due to the concern that the preferences of musicians may not be representative of the entire population. Two studies were conducted to achieve these objectives. The purpose of Study 1 was to evaluate the listening test methods, while both Study 1 and Study 2 were used to assess the similarity of the preference ratings of the two participant types, musicians and non-musicians. Study 1 examined if the listening test method (successive versus comparative) had an effect on the preference ratings of two solo-instrumental motifs simulated in a variety of reverberant room acoustics settings. The possible interaction between testing method and test participant type was also evaluated. Finding no interaction effects and also that the comparative method required less time to complete, Study 2 was conducted to further evaluate if preference differences exist for the two participant types. In addition to the solo-instrumental motifs used in Study 1, two orchestral motifs simulated in the same room acoustics settings were also used to evaluate preference differences as a function of motif type.

3.2 METHODS

3.2.1 Listening test stimuli

Since a major objective of the present work was to determine if musicians and non-musicians have similar preferences to concert hall acoustics, it was important to select a variable that has already been established to correlate with preference. Past studies have identified several room acoustics attributes that correlate with listener preference including *reverberance*, the perception of sound lingering in a space;[1,18-20] *intimacy* (sometimes referred to as *proximity*), the feeling of being close to a source;[1,18,19] and *listener envelopment*, the sense of being immersed in a sound field.[18,19] These three attributes are not necessarily independent of each other in regards to both concert hall design and a listener's perception, and the listener preferences may be attributed to specific room acoustics conditions. However, in the generation of stimuli for an experiment, a variable must be altered in a consistent manner to test the hypothesis. In this regard, the attribute of reverberance was chosen to vary across the stimuli. As such, the models were modified to achieve significant differences in reverberation time (T_{30}), while taking care to control for other room acoustic perceptions. It was expected that most musicians would rate stimuli with a moderate amount of reverberation favorably,[4] while non-musicians and some musicians would prefer drier room conditions.[113]

The stimuli were generated in Odeon v12.0[17] using the model of Boston Symphony Hall (BSH), which is included with the software, in order to best control for the direct sound and early reflections. Seven room acoustics settings were created that ranged from relatively dry (T_{30} = 1.0 s) to extremely reverberant (T_{30} = 7.2 s) to elicit a wide extent of preference. These settings are denoted Conditions B through H in Table 3.2. To simulate each setting, the absorption and scattering properties of the surfaces in the model were artificially altered. The real BSH does not have variable acoustics, but the geometry of the hall provided a sufficient model to simulate a variety of settings.

Table 3.2: T_{30}, EDT, C_{80}, BQI, and J_{LF}* in the 1000 Hz octave band for Conditions A-H. These metrics do not have defined quantities for Condition A, the anechoic setting, or are outside of a realistic range of use. It was assumed that the anechoic impulse response had T_{30} and EDT values of 0.0 s.

Condition	T_{30} (s)	EDT (s)	C_{80} (dB)	BQI	J_{LF}
A	0.0	0.0	-	-	-
B	1.0	0.9	4.8	0.73	0.23
C	1.2	1.0	3.6	0.66	0.22
D	1.7	1.4	1.2	0.75	0.24
E	2.1	1.9	-0.3	0.75	0.26
F	2.8	2.3	-2.1	0.75	0.27
G	4.2	4.4	-5.5	0.79	0.28
H	7.2	7.3	-7.5	0.77	0.28

The auralizations of an eighth, anechoic room acoustics setting, denoted Condition A, were simulated with a custom MATLAB script since the auralizations produced using a model of BSH with completely absorbing surfaces did not accurately depict a room with no reflections. Using anechoic recordings as the "auralizations" of an anechoic room would have resulted in monaural stimuli with no spatial cues. The MATLAB script allowed the anechoic auralizations to contain distance and spatial cues, thereby providing an "out-of-head" experience. The script accounted for differences in source-receiver distances by attenuating the direct sounds due to spherical spreading and air absorption, as well as applying attenuation based on the instrument directivities and source look-directions.

The settings were designed such that the metrics of T_{30}, early decay time (EDT), and clarity index (C_{80}) were altered in a relatively proportional manner across octave bands from 125-4000 Hz (the 1000 Hz octave band results are reported in Table 3.2). T_{30} and EDT have been shown to positively correlate with the perception of reverberance,[23-25] while C_{80} negatively correlates with reverberance.[2] While it is generally accepted that EDT provides

* Definitions of room acoustics metrics in Table 3.2:[26]

1. Reverberation Time (T_{30}) – The time for the response of a room to decay by 60 dB, based on the linear fit between -5 and -35 dB (multiplied by 2) on a backwards-integrated RIR (also known as a Schroeder curve). Used to estimate the perception of reverberance.
2. Early Decay Time (EDT) – The time (multiplied by 6) for the response of a room to decay from 0 to -10 dB based on a linear fit on a Schroeder curve. Used to estimate the perception of reverberance.
3. Clarity Index (C_{80}) – The ratio of early sound energy (before 80 ms) to late sound energy (after 80 ms). Used to estimate the perception of clarity.
4. Binaural Quality Index (BQI) – A measure of the similarity in the signals at the left and right ears. BQI = 1-$IACC_{E3}$, where $IACC_{E3}$ is the interaural cross-correlation coefficient averaged across the 500, 1000, and 2000 Hz octave bands. BQI is typically measured with a head and torso simulator.
5. Lateral fraction (J_{LF}) – The ratio of early sound energy (5-80 ms) arriving laterally over early sound energy arriving from all directions. The lateral sound is typically measured with a dipole microphone.

a more accurate prediction of perception of reverberance,[25] the present work uses T_{30} to represent each of the simulated hall conditions to validate the participants' preferences with the literature.[4] In addition, the T_{30} and EDT values were found to be correlated with a Pearson's correlation coefficient, R = 0.998 ($p < 0.0001$). Thus, for the purposes of this study, each condition was represented with T_{30} values.

Please note that the anechoic (T_{30} = 0.0 s), dry (T_{30} = 1.0 s), and extremely reverberant settings (T_{30} = 4.2 and 7.2 s), were outside the typical range of concert hall conditions, especially for a hall the size of the BSH. The anechoic and dry conditions were included to expressly compare the results of the present work to the findings of Roy et al. (2015), which showed that some participants prefer extremely dry room acoustics.[113] On the other hand, the excessively reverberant stimuli were designed as distinct anchors to encourage the participants to use the full range of the rating scale, as used in other subjective tests such as the Multi Stimulus test with Hidden Reference and Anchor (MUSHRA)[114] and psychometric fits of just noticeable differences (JND).[115]

The source and receiver locations were consistent across all of the conditions to ensure that the timing and direction of the direct sound and early reflections were constant. A single receiver location, as shown in Fig. 3.1, was used for all of the hall settings to control for seat-by-seat variation. A head-related transfer function (HRTF), Subject 21 from the UC Davis CIPIC HRTF database,[116] was used to filter the auralizations to present spatial cues to the listeners in both Odeon and the MATLAB script. Subject 21 represents an average HRTF measured from a KEMAR mannequin, and was deemed appropriate to use for all the subjects of the present study since the perception of reverberance does not rely too heavily on spatial characteristics. To maintain the aforementioned early reflection consistency, the binaural quality index (BQI) and early lateral energy fraction (J_{LF}) were controlled to be comparable across settings, as seen in Table 3.2. The mean value and standard deviation of BQI was 0.74 ± 0.04, and the range of values were within 2 JNDs with the majority of conditions being well within 1 JND.[117] For J_{LF}, the mean and standard deviation was 0.26 ± 0.02, and the range was 1 JND.[26]

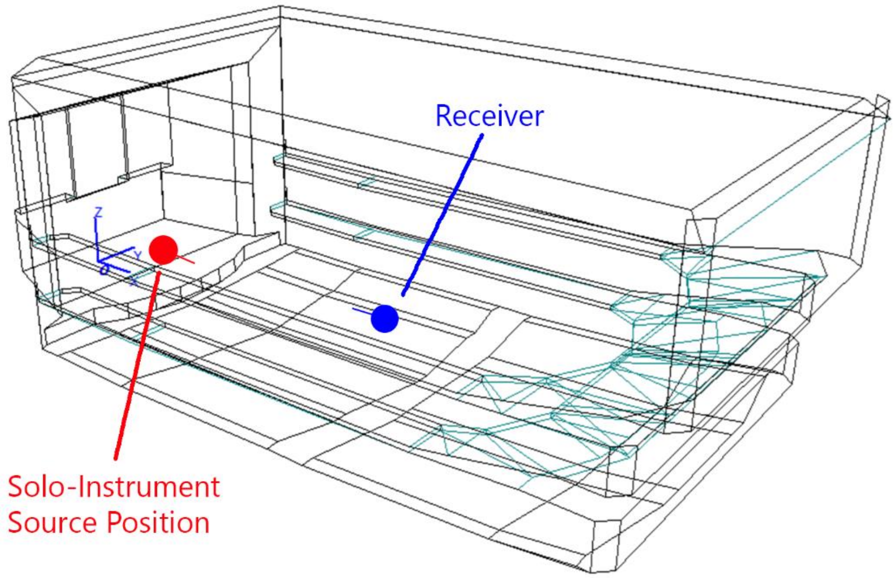


Fig. 3.1: Odeon model of Boston Symphony Hall used to create auralizations with the receiver and solo-instrument source positions shown.

For auralizations with solo-instrument motifs, a single source location was placed where a soloist would stand on the stage. (This same source was used to calculate the room acoustics metrics above.) For orchestral auralizations, multiple sources were used to simulate an entire orchestra, divided into sections according to instrument (see Fig. 3.2). One anechoic motif was used for each specific section of the orchestra. Within any one section, the different distances between each source and receiver, as well as the distances between each source and each surface, resulted in binaural room impulse responses (BRIRs) that contained phase shifts in both the direct and reflected sound. When convolved with each anechoic motif, these phase shifts provoked the impression of multiple musicians playing the same piece together.

The instrument directivity for each source was also considered. Odeon contains a small database of instrument directivities including string, woodwind, and brass instruments. For instrumental sources not contained in Odeon's library, substitute directivities were chosen from similar instruments based on relative size and shape.

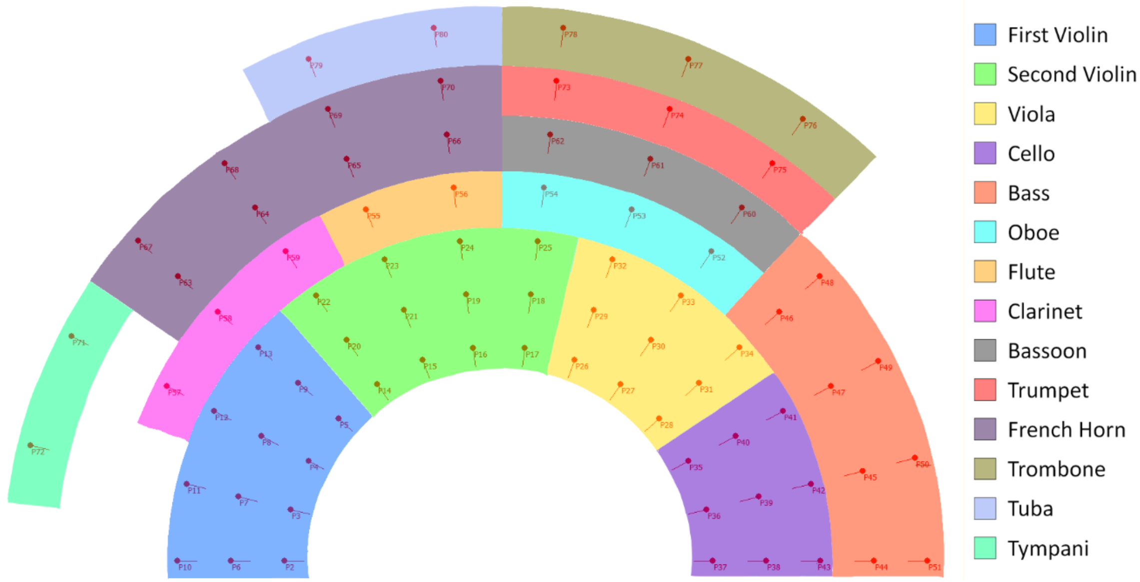


Fig. 3.2: Depiction of how the orchestral sources were distributed on the simulated stage in the Odeon model. The dark dots represent the source locations, while the dark lines indicate the source orientation for the directivities of the instruments.

Four motifs, two solo-instrumental and two orchestral, were used to generate the eight auralizations for each room acoustics setting, resulting in 32 total stimuli. Fast-tempo motifs were chosen to elicit more dynamic preferences to reverberance since in highly reverberant conditions, notes in quick succession tend to muddle together creating an unpleasant sound. The first solo-instrumental motif, the Trumpet Polka from the Archimedes Project CD,[118] contained a solo-trumpet performing a brisk musical passage composed predominantly of sixteenth notes. Over the course of 16 seconds and 33 beats, 110 notes are played.

The three remaining motifs were obtained from Aalto University's collection of anechoic recordings.[60] The second solo-instrumental recording was a selection from the first-violin part of Mahler's Symphony No. 1. Though not originally intended as a solo-instrumental musical phrase, the passage serves as an excellent example of violin performance, consisting of mainly sixteenth notes and *vibrato*. In 15 seconds, approximately 34 beats occur with 126 played notes. The two orchestral motifs were excerpts from Beethoven's Symphony No. 7 and Bruckner's Symphony No. 8, respectively. The Beethoven piece contained 32 beats, while the Bruckner excerpt included 24 beats. Both passages featured the simulated orchestras playing in *tutti forte*. Fig. 3.2 shows the orchestral layout used in

Odeon for the Bruckner motif; the Beethoven motif employed a slightly different arrangement due to a fewer number of French horn parts and the complete absence of trombone and tuba parts.

Once all of the auralizations were generated, the wave files were imported into Adobe Audition CC 2015 to normalize both length and overall level across the excerpts. For each motif, the lengths of the convolved stimuli were adjusted to be one second longer than the anechoic stimulus, and a 0.5-s linear fade-out was applied to remove the reverberant tails. This procedure ensured that the subjects did not rate based on the length of the stimuli or allow them to solely rely on the last note for the perception of reverberance. The lengths and tempos of the auralizations for each motif are shown in Table 3.3. Additionally, the A-weighted levels of all auralizations were equalized by adjusting the overall gains of the stimuli. This equalization was applied to eliminate the possible confounding variable of loudness,[19,54] which may have biased the preference ratings since the sound pressure levels of the models increased by 10 dB from Conditions B to H.

Table 3.3: Length and tempo of the auralizations of each of the four motifs used in the studies.

Motif	Composition Name	Type	Length (s)	Tempo (BPM)	Study
Trumpet	*Trumpet Polka*	Solo instrument	17.0	120	1 & 2
Violin	Mahler's *Symphony No. 1*	Solo instrument	16.5	140	1 & 2
Beethoven	*Symphony No. 7*	Orchestral	16.0	140	2
Bruckner	*Symphony No. 8*	Orchestral	13.0	130	2

3.2.2 Test procedures

3.2.2.1 Preference rating scale

For both studies and both listening test methods, the participants were prompted to rate each stimulus using a slider bar to indicate responses. The instructions dictated that the subjects should rate each stimulus based on their overall preference to the acoustic quality. The five-point, continuous Likert scale[6] ranged from -2 ("Strongly Dislike") to 0 ("Neither Like nor Dislike") to +2 ("Strongly Like"). "Strongly Dislike" was defined as "really disliking the acoustic quality of the track (low preference)" and "Strongly Like" meant that the participant "really liked the acoustic quality of the track (high preference)." These anchors allow the subjects to select a negative number when indicating a negative response, displeasure, and use a positive number to designate a positive response, pleasure. The subjects could place the slider bar anywhere on the scale, and the values were rounded to the nearest hundredth of a point to ensure that the computer's screen resolution did not alter the responses.

3.2.2.2 *Listening test method 1: successive method*

In the successive listening test method, the participants rated the stimuli for overall preference, but could only listen to and rate one stimulus at a time. While subjects were listening to each stimulus, a screen was displayed (Fig. 3.3)* on the testing computer that instructed the participant to "Listen to the excerpt." After each stimulus presentation was complete, the subjects were prompted to rate the stimuli, and the rating scale appeared on the screen. Once a rating was submitted, a three-second countdown was started to announce the beginning of the next stimulus. The participants were not allowed to replay any of the stimuli to better emulate listening tests that restrict such an allowance, such as tests conducted with physiological measures. The stimulus presentations were organized into sets; any given set only contained the eight auralizations of one motif. Both the stimuli order within the sets and set order were randomized.

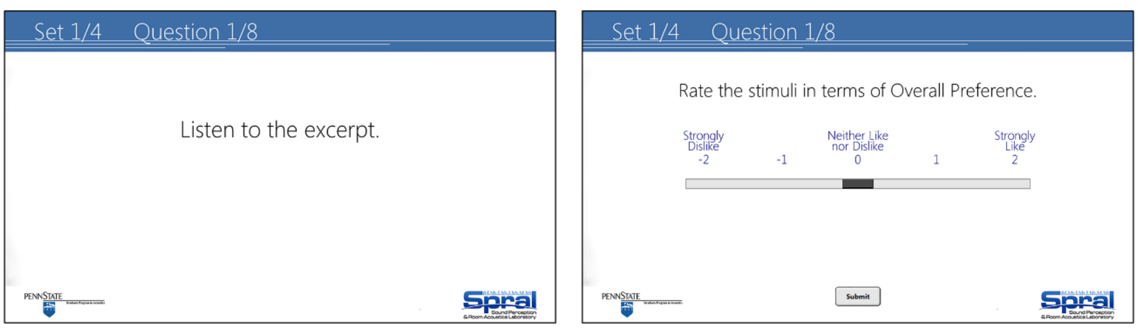


Fig. 3.3: Images of the custom-built Python stimulus presentation program of the successive method.

3.2.2.3 *Listening test method 2: comparative method*

The comparative method allowed participants to immediately compare and rate overall preference of all eight stimuli within each set. The stimuli were represented in a random order on the screen (Fig. 3.4)† by "Play" buttons labeled A through H (these labels did not correspond to the stimuli condition labels). The subjects could start and stop stimuli at will by clicking on the "Play" buttons, or start the stimuli from the beginning by pressing a

* Fig. 3.3 was not included in the original publication of this chapter in the *J. Acoust. Soc. Am.*
† Fig. 3.4 was not included in the original publication of this chapter in the *J. Acoust. Soc. Am.*

“Reset” button. By clicking a “Play” button of a stimulus while another stimulus was being presented, the participants could instantaneously switch between stimuli such that the musical passage appeared to continue seamlessly as the room acoustics setting changed. As such, only the eight auralizations of a single motif were presented on the screen during a set. The switch between stimuli was instantaneous such that there was no latency between the perceptions to the different stimuli, though a click may have been audible. This feature afforded a more direct comparison of the room acoustics variable, and none of the participants reported problems in the audio quality of the musical tracks in the feedback survey. The rating scales for each stimulus appeared underneath each stimulus, respectively. As with the successive method, both the set and stimuli orders were completely randomized to reduce the likelihood of any effects of stimulus order on the results.

The subjects were permitted to play each stimulus as many times as necessary. Additionally, the participants were required to listen to each stimulus before submitting their responses. If one or more of the stimuli were not played when a subject attempted to submit their ratings for a given set, a warning would appear that instructed the participant to listen to and rate each stimulus.

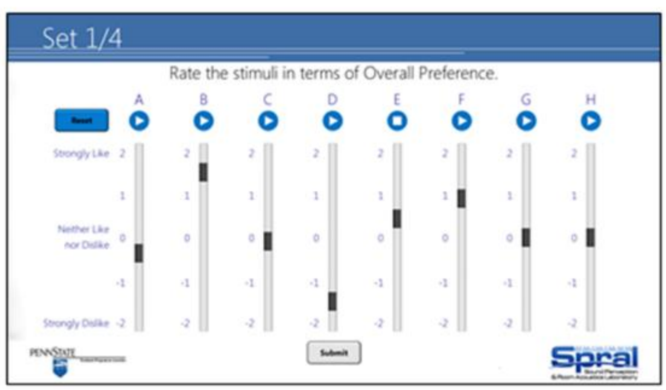


Fig. 3.4: Image of the multi-stimulus comparative method presented by the custom-built Python program. Playback of the musical passages is continuous and the participants can play any stimulus multiple times.

3.2.2.4 Testing protocols

The participants of both Study 1 and Study 2 began the testing session in the same manner. First, a hearing prescreening test was administered to each subject to determine if the participant was eligible to participate in the study. Following the hearing test, the subjects completed a training tutorial. The tutorials introduced the subjects to the construct of preference, the rating scale, and the listening test methods. After the tutorial was finished, the subjects began the testing protocol, which differed depending on whether the subject was participating in Study 1 or 2, as outlined below. For both studies, the stimuli were presented binaurally over STAX SR-207 electrostatic headphones (see Fig. B.4) with a Creative Sound Blaster external sound card. The L_{eq} of the excerpts was 75 dBA, a relatively realistic and comfortable level for the listeners.

The primary purpose of Study 1 was to determine if participants gave similar ratings of the stimuli across the two test methods. Each subject was randomly assigned to start the testing protocol with either the successive method or the comparative method, and completed all of the sets of one method before being tested with the second method. Only the two solo-instrumental excerpts, Trumpet and Violin, were presented to the listeners to reduce testing time since participants had to rate all stimuli twice using both test methods. Within one set, the subject would only listen to the auralizations of one motif. The subjects rated two sets of each motif for each listening test method in order to use the repetitions of ratings for a test-retest reliability analysis. In all, each subject was presented with two repetitions, two motifs, and two listening test methods for a total of eight sets.

To more fully achieve the second objective of finding differences in preference ratings between musicians and non-musicians, Study 2 prompted a new set of participants to rate overall preference for all four motifs—Trumpet, Violin, Beethoven, and Bruckner. Based on the findings of Study 1, only the comparative method was utilized in Study 2 to test the participants. This method was chosen because the testing time was shorter than that of the successive method. Also, the subjects maintained reliable results that were consistent with those from the successive method. Again, the subjects were only presented one motif per set and two sets of each motif. As such, the subjects in Study 2 were given eight sets of stimuli: four motifs and two repetitions of each motif.

3.2.3 Subjects

Thirty-four subjects (16 male and 18 female, 18-47 years, mean 25 years; 17 musicians and 17 non-musicians) rated the two solo-instrument auralizations in both the successive and comparative listening test methods in Study 1. In Study 2, an additional 37 new participants (18 male and 19 female, 18-47 years, mean 24 years; 18 musicians and 19 non-musicians) were tested with only the comparative test method, rating both the solo-instrument and orchestral stimuli. To classify as a musician, a subject must have had at least five years of formal music lessons from a professional musician and was currently active in lessons and/or an ensemble. Non-musicians, on the other hand, must have had no more than two years of formal musical training and must not have been taking lessons or in an ensemble. Each participant also provided information on the instruments they played, type of ensemble experience, higher education music courses, and critical listening experience, summaries of which may be found in Table 3.4.

Each participant was pre-screened before testing to ensure they fit the criteria of a maximum hearing threshold of 15 dB HL to improve reproducibility.[106] In addition to the 71 participants above, 12 subjects (14.5%) were excluded prior to testing for not meeting the hearing threshold requirement. Out of the 12 subjects, 2 were classified as musicians (5.7% of the total musician participants) and 10 as non-musicians (27.8% of this subject pool). Past studies conducted by the authors' research group that solely contain musicians typically have hearing screen failure rates between 5 and 12%.

3.2.4 Statistical analysis

3.2.4.1 Comparing the listening test methods in Study 1

Several factors were considered in evaluating the differences between the successive and comparative listening test methods. In particular, the similarity between the preference ratings and the internal reliability of the methods were assessed to directly compare the results obtained by each method. Secondary factors included overall testing time, perceived test difficulty, and test method preference.

Table 3.4: Summary of musical experience of the musician and non-musician participants.

	Musicians	Non-musicians
Instrument[a]		
Brass	13	2
Guitar	6	1
Percussion	3	1
Piano	14	2
Viol Family	12	1
Voice	12	1
Woodwind	10	3
None	0	26
Ensemble Type[b]		
Band	23	2
Orchestra	18	0
Choir	17	7
Jazz Band	4	0
Rock Band	1	0
None	0	27
Years of Experience		
Formal Musical Training	15 ± 5	1 ± 1
Ensemble Participation	9 ± 4	1 ± 1
College Music Courses	2 ± 2	0 ± 0.2
Concerts Attended per Year		
0	2	17
1-4	16	18
5-9	2	1
10-50	6	1
50+	9	0
Critical Listening (Hours/Week)		
0	1	24
1-3	13	5
4-6	11	3
7-9	2	0
10+	8	4

[a]83% of the musicians studied more than one instrument.
[b]60% of the musicians participated in more than one ensemble.

The preference rating responses between the two methods were compared to determine if the two tests produce similar results for each reverberant setting. A repeated measures analysis of variance (ANOVA)[31] was used to investigate the statistical similarity of preference ratings. The factors included in the general linear model (GLM) were Setting, Motif, Participant Type (e.g. musician/non-musician), Test Method, Test Order, and the interactions between these factors. Mean comparisons were drawn for this and subsequent ANOVAs with a Tukey-Kramer adjustment.[31]

The test-retest reliability of each testing method was evaluated using Cronbach's α coefficient.[119] The α coefficient is the ratio of the sum of covariances among individual items to the total variance of responses, indicating the correlation between the items. In the case of test-retest reliability, α is used as a measure of internal consistency between the repeated responses. In other words, each repetition of the stimuli serves as an item in a test-retest reliability test. Larger values of α represent a more reliable test, where values above 0.7 are considered reliable, between 0.2 and 0.7 are somewhat reliable, and below 0.2 are not reliable.[5] A limitation of Cronbach's α is that the coefficient tends to increase with the addition of items. In the present study, only two repetitions of responses were collected for each stimulus, the minimum number of items for such a reliability test, meaning that Cronbach's α remained a suitable metric for reliability.

The time required for each participant to complete each set was recorded and assessed using a repeated measures ANOVA. The ANOVA incorporated the same factors listed above except for Setting and Motif. The interaction effects between the test method, test order, and participant type were examined to determine if familiarization with the stimuli due to receiving the successive method first and/or musical expertise affected the subjects' performance on the comparative method.

In addition to the similarity of preference ratings, reliability, and testing time, perceived test difficulty and test preference were also considered. For perceived test difficulty, the participants were asked to rate the difficulty of each test method. Since the subjects were not privy to the names of the tests, they were prompted to rate the difficulty of "Test 1" and "Test 2" (which were randomly assigned for each subject) on a continuous scale from "Very Difficult" to "Very Easy." Participants were also asked which test method they preferred. For this question, a continuous scale was presented with “Test 1” on the left and “Test 2” on the right. These responses were evaluated with two additional repeated measures ANOVA with test difficulty and test preference as the response variables, respectively.

3.2.4.2 Comparing musicians and non-musicians in Study 1 and Study 2

The results from the comparative listening test method in both studies were examined to draw conclusions about differences in preference ratings and test performance between musicians and non-musicians. A repeated measures ANOVA was used to evaluate the ratings between the two groups of participants to determine if musicians and non-musicians have diverse preferences to reverberant stimuli. Setting, Motif, and Participant Type were chosen as the factors to include in the GLM, as well as the statistical interactions between them. Test Method and Test Order were not included in this analysis since only the participants in Study 1 were exposed to both test methods. Cronbach’s α coefficient was used to assess test-retest reliability between the two subject types. Once again, the two repetitions of responses for each stimulus were evaluated as the items in the analysis.

To keep the test length to a maximum of one hour, the participants were not asked to rate the stimuli in terms of perceived reverberance. The scope of the present study was limited to examining the differences between the preferences of the two participant types without knowing explicitly how they perceived the reverberance differences. However, the stimuli had T_{30} values that span 37 JND limens based on the accepted JND of 5%.[26] The minimum difference between any two stimuli was 4 JNDs. Therefore, it was expected that the participants would be able to hear clear differences between the reverberant stimuli.

3.2.4.3 Clustering analysis

Since several concert hall studies have demonstrated that listeners either prefer clear, intimate sound or loud, reverberant sound,[2,18,19,113] a *k-means clustering* analysis[120] was conducted to determine if musicians and non-musicians in this current study could be categorized into these two preference groups or possibly further subdivided into more than two groups. The k-means algorithm collects data into groups by applying a least squares fit between the data points and a pseudo-randomized cluster center. Since the cluster centers are pseudo-random, the process iterates in an attempt to minimize the total within-cluster variance. This process is highly dependent on the number of cluster centers. A *scree test* was used to determine the appropriate number of cluster centers by evaluating the total amount of variance attributed to clustering the data,[121] and an *agglomerative hierarchical clustering* analysis was performed using *Ward's method*[122] to verify the number of determined clusters. If an inappropriate number of clusters was used in the k-means analysis, the cluster centers may have been unique in regards to algorithm, but may have not been statistically significantly different from one another. A repeated measures ANOVA was run replacing Participant Type with Participant Preference Group to ensure that the cluster centers were significantly different.

3.3 RESULTS AND DISCUSSION

3.3.1 Comparisons between the comparative and successive testing methods in Study 1

3.3.1.1 Effect of test method on reverberation time preference ratings

A repeated measures ANOVA was performed to determine if the preference ratings obtained by the comparative and successive testing methods were in agreement. For this dataset and subsequent datasets, the residuals of the data are normally distributed. The preference ratings for each setting were adjusted for multiple comparisons using the Tukey method.[31] The main effect of Setting on the preference ratings was significant ($F(7,2080) = 140.00$, $p < 0.0001$) across both methods, indicating that the participants reported preference differences between the stimuli, as shown in Fig. 3.5. On average across test method, motif, participant type, and test order, the subjects preferred the stimuli with a moderate amount of reverberation (T_{30} = 1.2-2.8 s) and disliked the anechoic stimulus (T_{30} = 1.0 s) and extremely reverberant stimulus (T_{30} = 7.2 s).

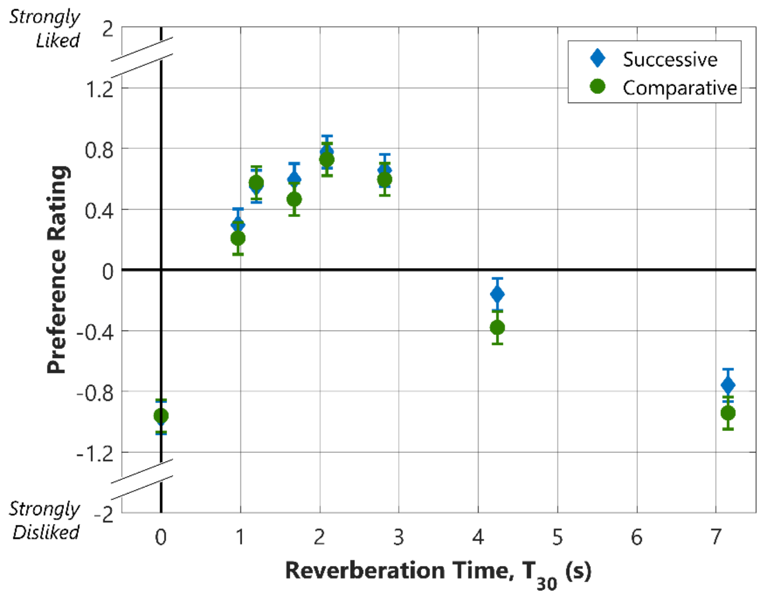


Fig. 3.5: Preference ratings as a function of T_{30} and Test Method. The bars indicate the standard error of the mean responses. The auralizations from the conditions with T_{30} from 1.2 to 2.8 s received the highest ratings, while the anechoic (T_{30} = 0.0 s) and extremely reverberant (T_{30} = 7.2 s) were rated the lowest. The successive and comparative methods produced statistically indistinct results for each setting ($p > 0.05$).

The main effect of Test Method was determined to be statistically significant ($F(1,2080) = 4.44$, $p < 0.05$) indicating that there was a significant difference in the average ratings between methods. On average across all settings, the successive method yielded higher preference ratings than the comparative method with mean values and standard deviations of 0.12 ± 0.07 and 0.04 ± 0.07, for each test method respectively. This difference in average preference ratings between the two methods is a small effect, only 2% of the total range of the rating scale, as determined by the small F-statistic (F = 4.44) compared to the large number of observations (N = 2080).[31] Furthermore, when separate pair-wise comparisons were drawn between the two test methods for each of the conditions, no significant differences existed between the successive and comparative methods (p > 0.05 for each condition; see Fig. 3.5). Additionally, the interaction effect between Test Method and Setting is not significant ($F(7,2080) = 0.56$, $p = 0.79$), which suggests that the responses to both methods exhibited the same trend across the room acoustic conditions. Therefore, it was determined that the successive and comparative methods yielded the same preference results despite the successive method having a slightly higher average rating across settings.

Similarly, the main effect of Motif was significant ($F(1,2080) = 23.58$, $p < 0.0001$). When averaged across all settings, the rating of the Violin motif was higher than that of the Trumpet motif with mean values and standard deviations of 0.18 ± 0.07 and -0.02 ± 0.07, respectively. However, the interaction effect between Setting and Motif was found to be insignificant ($F(7,2080) = 1.83$, $p = 0.08$), and the preferences ratings for each condition were within the standard error between motifs ($p > 0.05$). The main effects of Test Order and Participant Type on the mean preference ratings, as well as the higher-order interaction effects, were all insignificant at $p > 0.05$.

3.3.1.2 Reliability of preference ratings as a function of test method

The reliability of the data obtained as a function of the two test methods, successive and comparative, were evaluated using the method of test-retest reliability.[119] The Cronbach's α coefficients for the successive and comparative methods were found to be 0.839 and 0.836, respectively. The α coefficients were also calculated for each stimulus in each test method, and a paired t-test was conducted to determine if the reliability between the successive and comparative methods were statistically different. Cronbach's α differed between stimuli with the ratings of the distinct anchors of the anechoic and excessively reverberant conditions tending to have higher α values. However, the differences between the α coefficients between the two tests were not significantly different ($p = 0.88$).

Since the test-retest reliability of the anchor stimuli was greater than those conditions in the typical range of concert halls, an additional analysis was conducted to test if the anchor stimuli inflated the test-retest reliability estimates. Excluding the stimuli generated with T_{30} = 0.0, 1.0, 4.2, and 7.2 s, Cronbach's α for the successive method decreased from 0.839 to 0.821, which indicates that the anchors unfairly biased the reliability for the successive test. However, the α coefficient of the comparative method increased from 0.836 to 0.860 in the reduced model, suggesting that the anchors did not unfairly inflate the reliability estimates for the comparative test.

3.3.1.3 Effect of test order on testing time

The time to complete the sets of each test method was recorded and compared using an ANOVA and the Tukey method of comparisons. The ANOVA evaluated the effects of Test Method, Test Order, and Participant Type on the dependent variable of time. It must be noted that the successive method generally had a fixed testing time since the subjects were required to listen to the entirety of each stimulus before they could submit their rating. The testing time of the successive method was constrained more so by the length of the stimuli than the subjects' performance. Thus, the primary purpose of the statistical analysis of the time data was to compare the test length of the comparative method to this fixed time.

It was found that the order in which the participants received the test methods influenced the amount of time spent on each method, as evidenced by the significance of the Test Method by Test Order interaction effect ($F(1,30) = 4.36$, $p < 0.05$). The lengths of the successive method sets only varied slightly based on the response times of the participants and were statistically the same regardless of test order. For the participants who received the comparative method first, the time to complete a set was the same as that of the successive method. However, for the participants who received the comparative method second, the time to complete a set was an average of 60 ± 20 seconds shorter than a set using the successive method ($p < 0.005$). This result indicates that the comparative test was quicker than a successive test when the subjects were primed, or familiar with the test stimuli. The participants that were not familiar with the stimuli took longer using the comparative method than those who had heard the stimuli previously in a successive method set. By training the subjects prior to the listening tests, extra time may be afforded by the comparative method that could either allow for additional experimental sets or potentially reduce subject fatigue.

3.3.1.4 Subjective evaluation of test methods: test method difficulty and preference ratings

The repeated measures ANOVA of the participant's feedback examined the effects of Test Method, Test Order, Participant Type, and interaction effects on the ratings of test difficulty. The only significant effect on test difficulty ratings was Test Method ($F(1,30) = 4.77$, $p < 0.05$), shown in Fig. 3.6. Though both methods were rated positively denoting an "easy" difficulty level, the comparative method earned a difficulty rating of 1.4 ± 0.1, while the successive method was given a rating of 1.1 ± 0.1. Musicians and non-musicians rated the difficulty of both the successive and comparative methods similarly ($p > 0.05$).

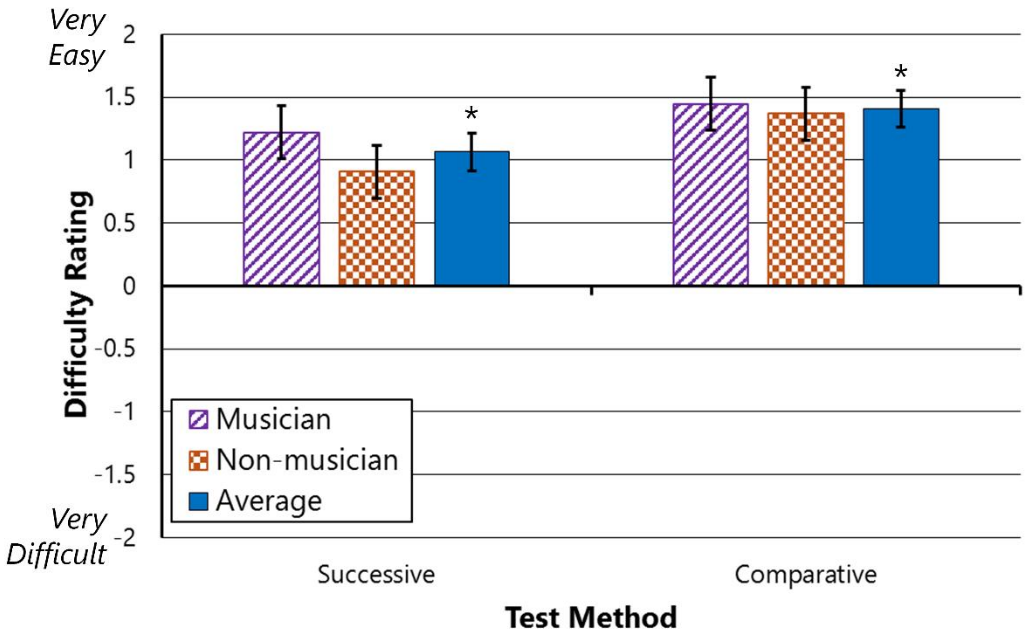


Fig. 3.6: Difficulty ratings for each test method. Differences between the two average ratings of +1.06 for the successive method and +1.41 for the comparative method indicated by the asterisks are significant with a p-value of less than 0.05. While both methods were rated positively, meaning that the participants found the tests to be easy, the comparative method was rated as the easier test.

A paired-comparison test was used to assess the participants' preferences for the two test methods. Participants were asked to rate their preference using a continuous scale ranging from "Test 1" to "Test 2". The repeated measures ANOVA assessed the effects of Test Order and Participant Type on whether the participants preferred the comparative method or the successive method. In this context, positive preference ratings indicate a preference for the comparative method, while negative ratings signify that the participants preferred the successive method (see Fig. 3.7). Overall, the participants preferred the comparative method over the successive method with a test preference rating of +1.0 ± 0.5. However, the amount the subjects preferred the comparative test depended on the order in which the tests were received and the type of participant, as demonstrated by the significance of the interaction effect between Test Order and Participant Type ($F(1,30) = 4.64$, $p < 0.05$). As shown in Fig. 3.7, the non-musicians preferred the comparative method regardless of which test was completed first. The ratings that the non-musicians provided were +1.2 ± 0.5 and +1.0 ± 0.5 for the two test orders, respectively. On the other hand, musicians preferred the comparative method more if they received the successive test first (test preference rating of +1.9 ± 0.5), and rated neutrally (test preference rating of 0.0 ± 0.5) if they completed the comparative method first ($p < 0.01$).

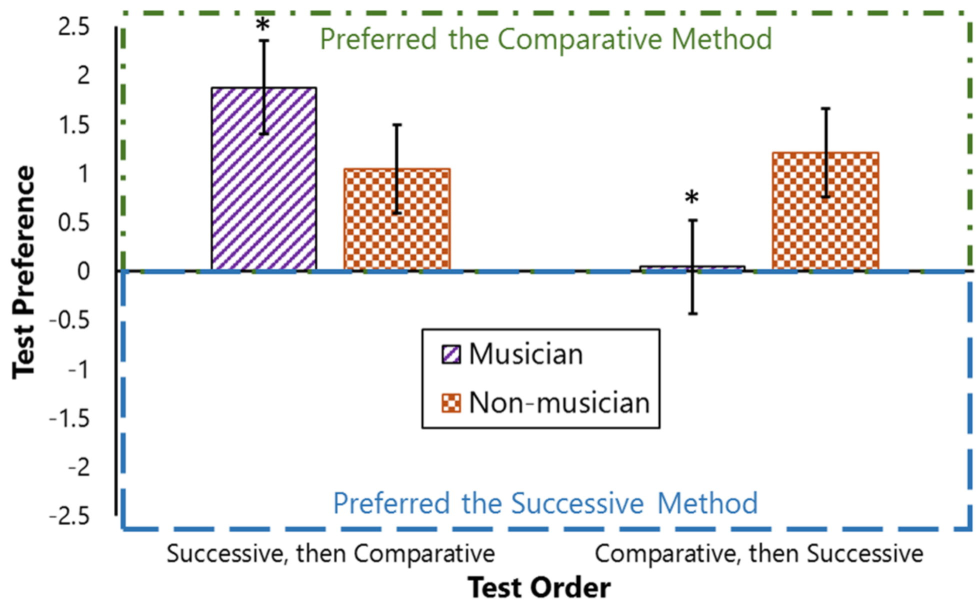


Fig. 3.7: Test preference ratings for each test order where positive values represent preference for the comparative method and negative values represent preference for the successive method. The musicians greatly preferred the comparative method (rating of +1.9 ± 0.5) if they completed the successive method first, but did not prefer one method over the other (rating of 0.0 ± 0.5) if they received the comparative method first. The non-musicians preferred the comparative method equally regardless of test order (average ratings of +1.1 ± 0.4). The asterisks mark the only significant Tukey comparison ($p < 0.01$).

3.3.1.5 Summary of the comparison of the two test methods

The results of the study comparing the two test methods agree with previous literature in that the comparative method was more robust to test order than the successive method. Parizet et al. suggest that the successive method requires a training period to produce equally accurate results as the comparative method.[30] Along those lines, Bech argued that a training period was necessary to ensure that the ratings and performance of the subjects were consistent.[106] In the present work, the participants rated the stimuli similarly with the same level of reliability for both listening test methods regardless of test order. However, the significant effect of test order was found on test duration for the comparative sets, where the participants who received the comparative method second completed the set in less time than those who had that set first. Furthermore, the comparative test was rated as the easier and more preferred method of rating the room acoustics stimuli used in this study.

In experiments aimed at establishing stimuli to use in experiments that require a successive method of testing, the results of Study 1 indicate that the comparative method can quickly and reliably predict the preference responses to reverberant stimuli presented using the successive method. The speed and reliability of the comparative method can be beneficial when prompting participants to rate very similar acoustic stimuli such as in-car ventilation noise, concert hall acoustics, and likely other acoustic stimuli. Based on these results, only the comparative method was used to subsequently compare the preference ratings and reliability of musicians and non-musicians in Study 2.

3.3.2 Comparisons between musicians and non-musicians across Studies 1 and 2

3.3.2.1 Differences in preference ratings between participant types

Combining the data from both studies, an ANOVA examined the effects of Setting, Motif, and Participant Type on preference. (The factors of Test Method and Test Order were not included in the analysis because the participants only rated the stimuli using the comparative method.) The main effects of Setting, Motif, and Participant Type were all significant at $p < 0.005$. The significance of the Setting by Motif interaction effect ($F(21,3322) = 5.47$, $p < 0.0001$) suggests that the preference ratings for each of the motifs exhibited different trends across settings. Specifically, most reverberant conditions (T_{30} = 4.2 and 7.2 s) were rated positively for the Beethoven motif and negatively for the other three motifs. Likewise, the significance of the Participant Type by Motif effect ($F(3,3357) = 3.33$, $p < 0.05$) indicates that the difference between RT preferences between musicians and non-musicians differed for each motif. However, the lack of significance of the higher-order Setting by Motif by Participant Type interaction ($F(21,3322) = 1.44$, $p = 0.09$) and Tukey comparisons ($p > 0.05$) indicate that the previously mentioned interaction effect across Motif and Participant Type is consistent across settings. Consequently, the immediate discussion will focus on the Setting by Participant Type interaction.

The significance of the Setting by Participant Type interaction effect ($F(7,3322) = 11.51$, $p < 0.0001$) indicates that musicians and non-musicians rated the stimuli differently, as shown in Fig. 3.8. The differences in the preference ratings between the two groups lie in the ratings for the excessively reverberant stimuli, T_{30} = 4.2 and 7.2 s. The musicians disliked these stimuli, responding with average ratings of -0.41 ± 0.09 and -1.0 ± 0.09, respectively. The non-musicians, on the other hand, gave neutral ratings of 0.15 ± 0.09 and -0.21 ± 0.09 for the T_{30} = 4.2 and 7.2 s conditions, respectively. Another critical difference between the two groups of subjects is that the musicians on average provided clearer distinctions on which stimuli they preferred according to the Tukey means comparisons with $p < 0.05$. In the Tukey means comparison analysis, the musicians rated two conditions, T_{30} = 1.7 and 2.1 s, as the most preferred group of stimuli, while the non-musicians rated four of the eight conditions (T_{30} values ranging from 1.2 to 2.8 s) as the most preferred. As the non-musicians had statistically similar ratings for half of the stimuli, it was determined that the non-musicians were not able to distinguish their preferences towards the settings as effectively as the musicians.

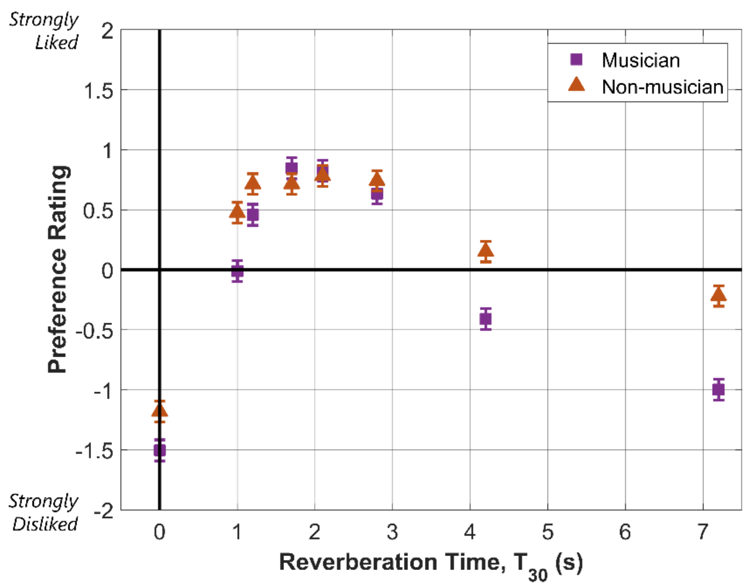


Fig. 3.8: Preference ratings as a function of T_{30} averaged over motifs. The bars indicate the standard error of the responses. The differences between the trends exhibited by the two participant types was statistically significant (*p* < 0.0001) and are most apparent in the preference ratings for the two excessively reverberant conditions with T_{30} values of 4.2 and 7.2 s, respectively. The Tukey means comparison showed that the musicians rated two conditions as the most preferred (T_{30} = 1.7 and 2.1 s), while the non-musicians rated four stimuli as the most preferred (T_{30} values from 1.2-2.8 s) since these stimuli were statistically similar (*p* > 0.05 for both groups of stimuli, respectively). All of the other comparisons with the most preferred stimuli were statistically significant at p < 0.05.

3.3.2.2 Reliability of preference ratings as a function of participant type

Again using the method of test-retest reliability, the reliability of the data obtained as a function of the two participant types, musicians and non-musicians, were evaluated. On average, the test-retest reliability of the comparative method for all of the subjects in Studies 1 and 2 was similar to the reliability reported in Section 3.3.1.2 with an α coefficient of 0.833. However, in examining the test-retest reliability for both groups separately, the α coefficients were found to be 0.875 and 0.782 for the musicians and the non-musicians, respectively. A paired t-test between the participant types was conducted on the Cronbach's α for each stimulus to determine if these two values were statistically different. The t-test revealed that musicians were statistically more reliable than non-musicians ($p < 0.05$). Nonetheless, the non-musicians still did demonstrate reliability based on the commonly-accepted α reliability threshold of 0.7.[5]

As in Section 3.3.1.2, the reliability analysis was rerun excluding the conditions that were outside the range of typical concert halls to verify that the ratings of these conditions did not inflate Cronbach's α. The α coefficient of both the musicians and non-musicians increased in the new analysis, which showed that the ratings for the anchor stimuli did not exaggerate reliability estimates.

3.3.2.3 Summary of the participant type comparison

The analysis of the preference ratings of the musicians and non-musicians recruited in this study demonstrated that the two types of participants experience differences in their preferences to reverberant stimuli. Also, the musicians in the study displayed a higher level of reliability than the non-musicians and were able to distinguish their preference ratings for nearly all of the stimuli. Despite the statistical differences between the musicians and non-musicians, the broad-stroke trends exhibited by both groups are similar: stimuli with a moderate amount of reverberation were significantly preferred over the anechoic and excessively reverberant stimuli (Fig. 3.8). This result contradicts the conclusions of Roy et al. that non-musicians only preferred anechoic or drier conditions[113] because the non-musicians in the present study demonstrated a dislike for the anechoic stimuli, on average. The disagreement between the two studies may have occurred because of the disparity between the two sets of stimuli: the majority of the stimuli tested by Roy et al. were outside the range of typical concert halls,[113] while the present study included more realistic conditions. However, testing non-musicians may still not be appropriate because the broad-strokes trends and reliability analysis suggest that musicians are more discriminating and reliable.

3.3.3 Reverberation time preference groups

3.3.3.1 Preference rating clustering results from all 71 participants

The purpose of the clustering analysis was to establish if there were preference groups based on the reverberant conditions among the 71 participants. Using the scree test and verified with Ward's method of agglomerative hierarchical clustering, five clusters were determined to be appropriate for this data set. The ANOVA was re-run as described in Section 3.2.4.2, but the independent variable of Participant Type was replaced with Participant Preference Group based on the clustering analysis. The interaction between Setting and Participant Preference Group was determined to be significant ($F(28,3194) = 17.06$, $p < 0.0001$), which indicates that each of the five clusters rated the stimuli with unique trends.

The groups uncovered by the k-means clustering analysis with five cluster centers (Fig. 3.9) will be referred to according to the condition rated the highest within the group. If two stimuli received high ratings but were statistically indistinct from one another, both conditions were included in the name of that particular group. As such, four subject categories were revealed, denoted "RT 1.0 s," "RT 1.2 & 1.7 s," "RT 2.1 s," and "RT 2.1 & 2.8 s" that preferred dry to moderately reverberant stimuli, respectively. Specifically, the RT 1.0 s group preferred the relatively dry condition, RT 1.2 & 1.7 s preferred the lower range of typical concert hall reverberation times , and groups RT 2.1 s and RT 2.1 & 2.8 s preferred the upper range of typical concert halls.

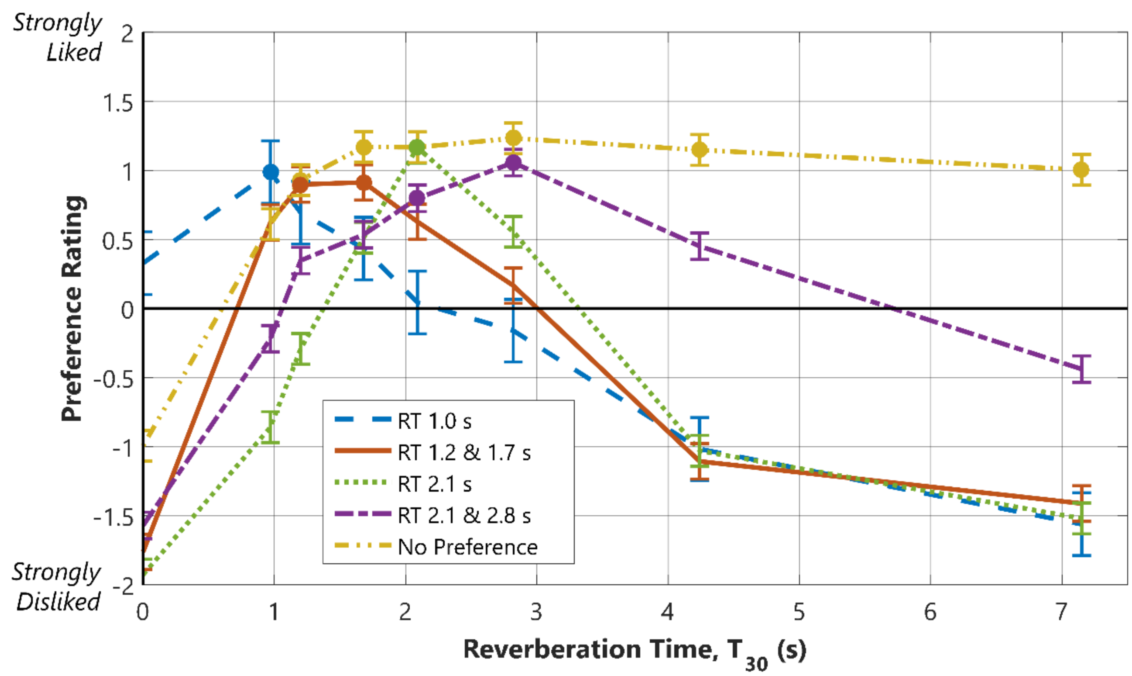


Fig. 3.9: The five cluster centers for preference ratings as a function of T_{30} ($p < 0.0001$). The groups exhibit a range of preferred T_{30} from 1.0 to 2.8 s. The particular T_{30} value(s) that was/were rated the highest in each group are indicated with round data markers.

The fifth group had statistically indistinct ratings for the conditions with T_{30} values between 1.2 and 7.2 s ($p > 0.05$ for all Tukey comparisons). This category was named the "No Preference" group since the participants provided ratings that were statistically similar for six of the eight conditions, thus not indicating clear preference differences. It is not possible to confirm that the participants could hear differences between the conditions since concert hall attribute ratings were not included in the study. However, it is unlikely that the "No Preference" participants were not capable of differentiating between the reverberant stimuli. The range of T_{30} for these conditions spans approximately 37 JND limens,[26] which suggests that the differences in perceived reverberance were vast. Nonetheless, the results indicate that these participants were unable to distinguish their preferences to these room acoustics conditions.

The analysis result of the five cluster centers, or preference groups, differs with the outcomes of the prior research of two prevalent preference groups that favor clear, intimate halls and reverberant halls, respectively.[18,19] Namely, the five groups represent a larger range of concert hall preferences than previously found with two of the groups preferring less reverberant conditions and two preferring more reverberant conditions. More groups may have emerged from the present study's clustering analysis because a greater number of participants were tested, which allowed the unique preference trends to be recognized and not clustered with another group. This result is further confirmed by the fact that the participants were relatively evenly distributed between groups in that no group had less than 10% nor more than 25% of the subjects.

3.3.3.2 Comparison of preference groups as a function of participant type

Comparing the Setting by Participant Preference Group interactions in Fig. 3.8 and Fig. 3.9, the differences between cluster trends were much more defined than the difference between the trends of musicians and non-musicians as demonstrated by the noticeable differences in the shape of the curves. By examining how musicians and non-musicians are allocated between the cluster groups (Table 3.5), rating trend differences between the two subject types become clearer. While only 11% and 22% of non-musicians were grouped into the No Preference group for the violin and trumpet solo-instrumental auralizations, respectively, the No Preference group had a large number of non-musicians for the orchestral auralizations at 42% and 79% for the Bruckner and Beethoven motifs, respectively. In comparison, the No Preference group contained 0-6% of the musicians for the solo-instrumental motifs and 6-28% for the orchestral motifs (see Fig. 3.10*).

Table 3.5: Distribution of musicians and non-musicians into the five preference groups for each of the four motifs.

Group	*Musicians / Non-musicians*			
	Trumpet	**Violin**	**Beethoven**	**Bruckner**
RT 1.0 s	5 / 3	8 / 9	1 / 0	1 / 0
RT 1.2 & 1.7 s	10 / 9	12 / 11	1 / 1	5 / 4
RT 2.1 s	9 / 5	3 / 2	7 / 1	7 / 2
RT 2.1 & 2.8 s	9 / 11	12 / 10	4 / 2	4 / 5
No Pref.	2 / 8	0 / 4	5 / 15	1 / 8

* Fig. 3.10 was not included in the original publication in *J. Acoust. Soc. Am.* **142**, 2258 (2017).

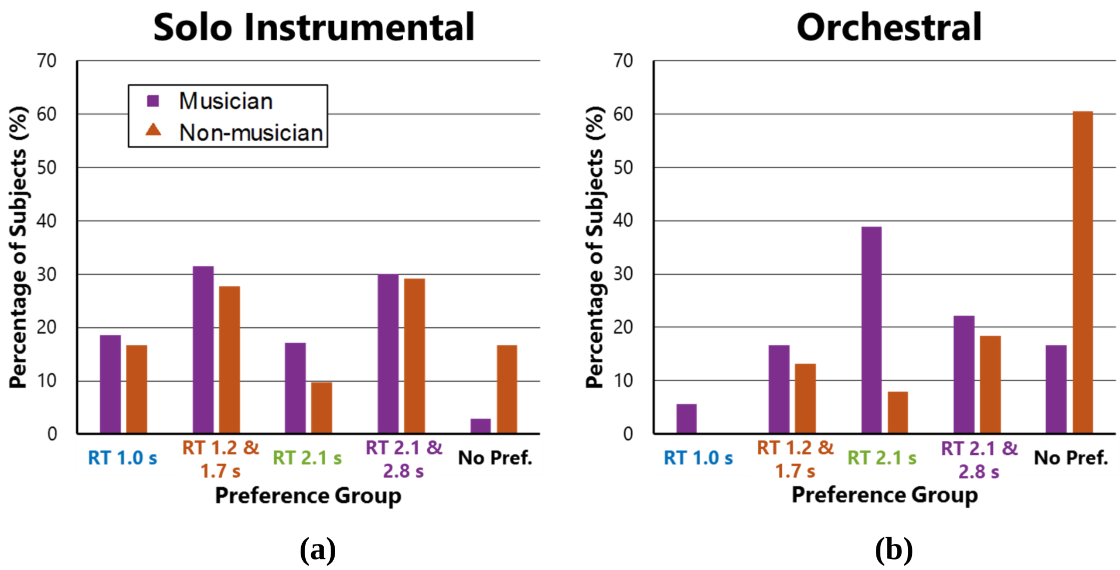


Fig. 3.10: The percentages of musicians and non-musicians present in each preference group for the (a) solo-instrumental and (b) orchestral motifs.

It is important to note that the majority of musicians were clustered in the groups with preferred T_{30} values from 1.2-2.8 s. In fact, almost 80% of musicians were distributed in these three categories, which agrees with the range for "ideal" reverberation times for concert halls according to past studies.[2,22,24] It is also interesting that a larger percentage of musicians preferred drier conditions (T_{30} = 1.0, 1.2, and 1.7 s) for the solo-instrumental motifs. Conversely, more musicians preferred the more reverberant conditions (T_{30} = 2.1 and 2.8 s) for the orchestral motifs. This finding agrees with past research that has shown that the optimal reverberation time depends on the type of music being played or reproduced.[4]

For each of the four motifs, the No Preference group predominantly comprised a large number of non-musicians, which suggests that many of the non-musicians could not recognize their preferences for most of the auralizations with differences in reverberance. This result agrees with the effect demonstrated in the Section 3.3.3.3.2 that on average, non-musicians were unable to distinguish their preferences towards many of the stimuli.

The reason that non-musicians may not exhibit a clear preference was not explicitly tested in this study. However, a possible explanation may be that non-musicians do not have the same or any musical expectations as musicians. The musicians may have disliked anechoic and extremely reverberant stimuli in accordance with dissonance theory[123,124] if they had expected more moderate reverberation for the solo-instrumental and orchestral motifs. Conversely, the musicians would have liked the stimuli that matched their expectations. Without these expectations, the non-musicians may not have disliked any of the stimuli. These expectations could have derived from the musical experiences of each participant type, summarized in Table 3.4. On average, the musicians boasted 15 years of musical experience, while the non-musicians reported only 1 year of experience. When asked about the amount of time spent actively listening to music without performing another task, i.e. critical listening, the musicians and non-musicians professed 7 hours and 3 hours per week, respectively. Similarly, on average, the musicians attended 24 concerts (i.e. any unamplified, live performance) per year and the non-musicians only attended approximately 1 with some of them never having attended a classical performance.

Irrespective of the reason, non-musicians may not be appropriate participants for evaluating room acoustics stimuli because they do not demonstrate clear preferences. However, the non-musicians that placed in the other preference groups, i.e. participants that could discriminate their preferences for the reverberant stimuli, demonstrated similar preference trends as the majority of the musicians. Either these non-musician participants had similar expectations as the musicians or they had developed similar tastes through some other means. Therefore, musicians and non-musicians can share preferences to concert hall stimuli. It is possible that non-musicians may be able to discern their preferences with more musical training, but musicians are more desirable test participants because they can readily demonstrate clear preferences with minimal training.

The k-means clustering analysis was rerun with the data from the musicians and from the non-musicians, separately. These groups are named with sequential numbers to avoid confusion with the groups from the previous analysis. For the musician data, four cluster groups emerged from the scree test and Ward's method. These four groups resemble the groups of the initial k-means clustering analysis. Musician Groups 1 through 3 were similar to the RT 1.0 s, RT 1.2 & 1.7 s, and RT 2.1 s clusters, respectively, while Musician Group 4 appeared to be the combination of the RT = 2.1 & 2.8 s and No Preference groups. On the other hand, the scree test and k-means clustering analysis showed that the non-musicians only fell into one of two preference groups, which resemble the average non-musician trend (see Fig. 3.8) and No Preference group, respectively. The interaction effect of Setting and Participant Preference Group for the musician data ($F(21,1505) = 17.96$, $p < 0.0001$) and the non-musician data ($F(7,1612) = 16.89$, $p < 0.0001$) in the new ANOVAs were both significant.

These separate analyses confirm the results shown in Table 3.5. The musicians exhibited a diverse range of preferences while the non-musicians were largely unable to discriminate between their preferences. Even the non-musicians that did not place into Non-musician Group 2 (~No Preference) rated four of the stimuli as statistically indistinct. Furthermore, 22% of the musicians' responses were placed in Musician Group 4 (~No Preference), which suggests that a prescreening test may be necessary to filter out participants that do not demonstrate clear preferences to the room acoustics stimuli.

3.4 SUMMARY AND CONCLUSIONS

The present work had the overall goal of examining the influence of two important subjective listening test experimental design considerations, test method and participant musical expertise, on preference ratings of concert hall stimuli. For the first objective, preference ratings of a variety of concert hall conditions were used to compare the successive and comparative listening test methods, since it may be more desirable to use one method over another depending on testing design constraints. For example, while the comparative method may provide more consistent results by promoting the direct comparison of stimuli,[30] some studies, such as those involving physiological correlates to auditory signals, require the use of a successive method. At the same time, it is imperative that the successive and comparative methods yield similar results such that the participant preference ratings do not rely on the implemented test method. Additional factors, including reliability of the data obtained, were also considered when comparing the two methods. The second objective of the present work was to examine preference rating and consistency differences between musicians and non-musicians. A literature survey on concert hall subjective studies from 2011-2015 revealed that these two participant types should be examined since 46% of the studies recruited musician participants as experts. However, recent research has shown that the opinions of musicians may not be the same as non-musicians,[110,113] and thus may not be representative of the general population.

The investigation was divided into two studies, which included a total of 71 participants.

In Study 1, the 34 participants rated stimuli simulated with varying amounts of reverberation on overall preference using both the successive and comparative test methods. The listening tests produced similar ratings with an equivalent level of reproducibility. However, the comparative method was a more desirable method based on test time and participant feedback. The participants completed the comparative test faster when primed by receiving the successive method first. Additionally, the participants provided more positive feedback for the comparative test than the successive test both in terms of test difficulty and test preference. While these two test methods had been compared in a previous study,[30] the present work provides the similarities and differences of the two testing methods in the context of reverberant stimuli, which have clear perceptual differences from the in-car ventilation stimuli of the literature. Note that these results have been validated only for the reverberant conditions and motifs used in this study, thus future work is needed to verify these findings for other room acoustics attributes and motifs.

Study 2 prompted an additional 37 subjects of both musicians and non-musicians to rate preference of the same concert hall conditions of Study 1 using the comparative listening test method only. Two orchestral motifs were added to the two solo-instrument motifs used in Study 1. Comparing the results from musicians and non-musicians from both Studies 1 and 2, the present work found that the two participant groups exhibited different reverberance preference trends on average. The non-musicians on average were less able to elect a most preferred room condition than the musicians, possibly because the less experienced listeners lacked explicit expectations of room acoustic conditions. Additionally, the non-musicians demonstrated a lower test-retest reliability than the musicians, albeit one that would still be considered reliable based on general psychometric acceptance levels.[5]

Further examination of participant type differences using a k-means clustering analysis revealed five emergent preference groups. Four of these groups had the highest preference ratings for conditions with T_{30} values ranging between 1.0 s to 2.8 s. The majority of musicians were placed into these preference groups. A fifth group also emerged in which participants did not exhibit clear preferences for any of the reverberant conditions. Many non-musicians were placed into this fifth group. This clustering result agrees with the result from the ratings averaged over all of the non-musicians in that non-musicians were less able to discern their preferences to the concert hall conditions. The ratings from these participants would not help to inform concert hall design for the benefit of others that have stronger opinions about reverberance. While the study suggests that non-musicians should not be recruited to evaluate listener preference of reverberation, further work must be completed to extend this conclusion to other room acoustics characteristics.

Studies that have also shown that musicians and non-musicians exhibit different auditory processes in the brain.* Musicians have demonstrated improved brainstem and cortical auditory evoked potentials to acoustical features of speech[125,126] and an enhanced neural response to pitch.[127] Additionally, the structural and functional plasticity of musicians' brains differ from non-musicians.[128,129] Based on these results and the results of the current study, only musicians were considered as participants in the following MRI experiments.

3.5 ACKNOWLEDGMENTS

Recruitment and consent of the participants were obtained in accordance with Penn State's Internal Review Board (IRB) Protocol PRAMS44471. We would like to thank Evan Savage, an undergraduate research assistant, for his help in data acquisition, and Penn State's Statistical Consulting Center for their advice on data analysis.

* This paragraph was not included in the original publication in *J. Acoust. Soc. Am.* **142**, 2258 (2017).

CHAPTER 4

SENSITIVITY OF THE HUMAN AUDITORY CORTEX AND REWARD NETWORK TO REVERBERANT STIMULI

The following chapter explores the auditory and reward response to room acoustic stimuli with fMRI. The chapter was prepared as a manuscript and will be submitted to a peer-reviewed neuroscience journal.

Sensitivity of the human auditory cortex and reward network to reverberant stimuli

Martin S. Lawless and Michelle C. Vigeant

Graduate Program in Acoustics, The Pennsylvania State University, University Park, Pennsylvania 16802, USA

Abstract

The room acoustics of a space can alter subjective impressions of music, including a listener's preference, though not much research has characterized the brain's response to room conditions. Functional magnetic resonance imaging was used to uncover the cortical auditory and subcortical reward responses to concert hall stimuli. Outside the scanner, 18 participants heard a solo-instrumental musical passage and an orchestral motif simulated in eight different room acoustic conditions. The subjects rated the stimuli for preference to identify their most liked and disliked conditions. The liked and disliked stimuli had reverberation times of 1.0-2.8 s and 7.2 s, respectively. In the MRI, the moderately reverberant most-liked and extremely reverberant disliked stimuli were presented to the participants along with the anechoic stimuli and scrambled versions of the anechoic musical passages. The auditory cortex was found to be sensitive to the acoustic complexity of the stimuli as it activated more strongly for the clear, simple stimuli, the solo-instrumental and anechoic conditions, than for the acoustically complex stimuli, which were the orchestral and reverberant conditions. Contrasting liked and disliked reverberant stimuli, a reward response in the basal ganglia was detected in a region of interest analysis using a temporal derivative model of the canonical hemodynamic response function. Being small and transient, this response represents the modest differences between preferential responses to stimuli with subtle variations.

4.1 INTRODUCTION

Room acoustics research has shown that the spaces in which music is heard affect how music is perceived. Rooms, or more specifically concert halls, can add impressions of lingering sound (defined as *reverberance*), closeness to a source or musician (*intimacy* or *proximity*), and being immersed in a sound field (*listener envelopment*), among others.[1] Simply speaking, the characteristics of a concert hall alter the auditory information presented to the listener, thereby affecting the auditory response and subsequent subjective impressions of listeners.[18,19]

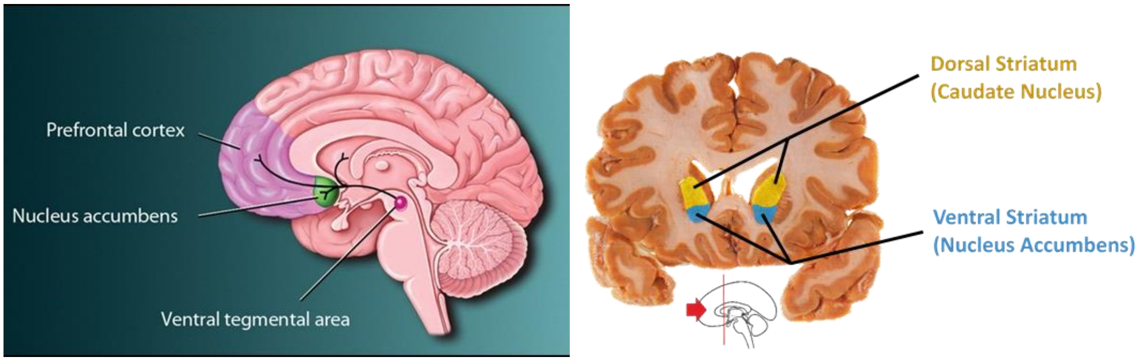
Despite the extensive functional neuroimaging research involving music, very little work has been conducted studying how room acoustics may affect the brain's response. The body of the work, conducted by Ando et al., demonstrated electroencephalographic (EEG) correlates to changes in several concert hall parameters. The research investigated the relationship between the auditory brainstem response (ABR) and slow vertex responses (SVR) and changes in sound level and the difference in signals presented to the left and right ears as quantified by the interaural cross-correlation coefficient (IACC). The latencies of the ABR and SVR tended to decrease as sound level increased and as the correlation between the left and right signals increased.[55]

For the present work, the auditory response to room acoustics stimuli was measured using functional magnetic resonance imaging (fMRI). Specifically, the present work intended to study the effect of increasing reverberation time (RT) on the extent of activations in the auditory cortex. The reverberation time (RT) of a hall correlates with the subjective impression of reverberance,[23] which in turn contributes considerably to a listener's overall impression of a hall.[18,19] Thus, the present study uses RT to predict a listener's perceived reverberance.

The auditory cortex is sensitive to the coherence of signals[130] in that increased coherence has evoked increased activations in Heschl's Gyri (HG) and the planum temporale (PT). Similarly, the addition of auditory objects (i.e. different sound sources) into a scene decreases the response along the superior temporal sulcus (STS).[131] Prior research has also shown that the degradation of the acoustic signal, whether by scrambling music,[132] speech,[133] or by including incoherent noise into speech signals,[134] reduces the activation extent in the primary auditory cortex (PAC). Since adding reverberation to a musical passage effectively inserts additional, incoherent auditory objects or noise into the stimulus, it was hypothesized that areas in the PAC would experience less activations as the RT of a room increased. The addition of reverberation to speech signals has also been shown to obstruct the neural encoding of formants in the brainstem, though it is not clear whether the same effect would occur during pitch-dominated music signals.[53]

Ando et al. also examined EEG correlates in the left and right hemispheres to differences in the subjective preferences of concert hall parameters; however, the results were contradictory.[56] The effective duration of the autocorrelation function (τ_e) for α-wave signals (8-13 Hz) were analyzed with respect to acoustic signals that the participants liked and disliked. In the analysis of the subjective preference as a function of RT, τ_e exhibited a lateralization towards the left hemisphere such that the value of τ_e was greater in the left hemisphere for the preferred RT of 1.2 seconds. The τ_e value in the left hemisphere was also greater for the preferred RT than the values for the disliked RTs of 0.2 and 6.4 seconds. On the other hand, the τ_e in the right hemisphere was lower for the preferred RT than the disliked RTs. When testing signals with different IACC, the exact opposite effect was found. The subjects exhibited τ_e lateralization favoring the right hemisphere, and the τ_e value was greater in the right hemisphere for the preferred IACC than the disliked IACC.

The ambiguity of the preference results call into question the exact neural mechanisms that Ando et al. were measuring. Typically, the response to musical preferences has concentrated in the reward network in the brain. In particular, the nucleus accumbens (NAcc) in the ventral striatum of the basal ganglia has been shown to activate in the presence of a rewarding, preferred musical stimulus (Fig. 4.1).[49-51,57] Additionally, research has also demonstrated that the caudate nucleus in the dorsal striatum activates during the anticipation of a musical reward.[51] Ando et al. were unable to measure signals from this subcortical system due to the limitations of EEG. Thus, the goal of a pilot study by the authors[68] and the present work was to examine the influence of reverberation time on the reward response in the basal ganglia using fMRI to see if the subtle changes in room acoustics could induce a similar effect as liked and disliked musical passages.



(a) **(b)**

Fig. 4.1: (a) The mesolimbic pathway includes a dopaminergic connection between the ventral tegmental area (VTA) and the nucleus accumbens (NAcc).[52] (b) The caudate nucleus and the NAcc have been shown to activate during the anticipation and experience of a reward, respectively. Adapted from Ref. [40]. (This figure also appeared in Ch. 1 as Fig. 1.9.)

The current study aimed to use functional magnetic resonance imaging to objectively quantify the response to concert hall stimuli. The first objective of this study was to measure the cortical auditory response to various concert hall conditions. To achieve this objective, participants were presented two musical passages simulated with different reverberation times ranging from anechoic (i.e. no reverberation) to extremely reverberant (RT = 7.2 s). This design intended to isolate the areas of the brain that are sensitive to changes in perceived reverberance with the hypothesis that the activations of these regions would decrease with increasing reverberation. The second objective was to determine if fMRI is a viable tool to assess preferences to room acoustics stimuli. The participants rated each of the reverberant stimuli in terms of overall preference, and the differences in responses in the mesolimbic reward network in the basal ganglia was observed between the most liked and most disliked stimuli.

4.2 MATERIALS AND METHODS

4.2.1 Experimental procedure

Testing occurred in two sessions on separate days. During Session 1 in an office, the participants were prescreened (see Section 4.2.2 for details) and asked to rate their preferences to the stimuli to identify their most-liked and most-disliked stimuli. The subjects were then placed in a mock scanner, the Psychology Software Tools, Inc. MRI Simulator (Fig. B.1a), to familiarize them with the MRI environment. While in the mock scanner, the participants were also prompted to rate the stimuli for preference to assess whether they provided the same responses both outside and inside of the scanner. (A previous study by the authors found that listening tests in the mock and actual MRI scanners yielded statistically similar preference ratings of the room acoustic stimuli.[68])

Session 2 occurred in the actual MRI scanner, a Siemens 3T Prisma Fit. The participants were presented their most liked and most disliked concert hall stimuli as determined by the ratings in the mock scanner. Two sets of control stimuli, anechoic recordings and scrambled versions of the anechoic recordings of the two musical passages, were also played to the listeners. The control stimuli will be described in 4.2.3. The subjects did not rate the stimuli during the MRI scans. While it is unclear if the act of rating influences the reward network response, all of the subjects rated the stimuli during Session 1 and were already aware that the concept of preference was being tested. Additionally, it was expected that the subjects experienced similar preferences in the mock and MRI scanners since the participants in a previous study provided similar ratings in both environments (see Chapter 2.3.1).[68]

4.2.2 Subjects

Eighteen (9 male and 9 female, 18-35 years, mean 23 years) normal, healthy musicians participated in the experiment. The subjects had between 8 and 25 years of formal music lessons and were currently active in lessons or a musical ensemble. All participants were right-handed, verified on the first day of testing with the Edinburgh Handedness Inventory.[135] The participants were also screened to have a maximum hearing threshold of 15 dB HL to ensure reproducibility in their ratings.[106] Three subjects beyond the 18 included in the study did not pass the hearing screening. The Major Depression Inventory[136] was used to screen the subjects for depression to ensure that they could accurately express preferences.

In a previous subjective study by the authors, 71 subjects rated the reverberant stimuli (see Section 4.2.3) in terms of overall preference.[137] From a k-means clustering analysis, five groups emerged that demonstrated different preference trends to the room acoustic conditions (see Fig. 3.9). One of the groups, denoted No Preference, was unable to distinguish between their preferences to the room acoustics stimuli, and rated six of the eight conditions as the most liked stimuli. It was expected that participants unable to report clear preferences to the stimuli would not exhibit different responses in the reward regions of the brain. As such, participants that placed in the No Preference group were excluded from the study (four total in addition to the included 18 participants).

4.2.3 Reverberant stimuli

Concert hall perceptual studies have identified several room acoustics attributes that influence a listener's preference. One such attribute, *reverberance* or the perception of sound lingering in a space, has been shown to correlate with preference.[18,19] The stimuli were designed to evoke a large range of perceived reverberances by simulating musical passages in different room acoustic conditions that ranged from anechoic to extremely reverberant. Based on the literature and the results from a previous study,[137] it was expected that the listeners would prefer the simulated musical passages, termed auralizations, with a moderate amount of reverberation and would dislike the anechoic and extremely reverberant auralizations.

4.2.3.1 Stimulus set for Session 1 preference ratings

The stimulus set for Session 1 consisted of 16 auralizations: two musical motifs in eight room acoustic settings. The two motifs were a solo-instrumental excerpt, 16 s from the Trumpet Polka from the Bang & Olufsen "Music from Archimedes" CD,[118] and an 11 s orchestral excerpt from Bruckner's 8th Symphony, obtained from Aalto University's collection of anechoic recordings.[60]

The eight room acoustic settings were applied using room models of the Boston Symphony Hall in Odeon v12.0[17] (see Fig. 3.1). Odeon's algorithm for predicting room acoustics divides the processing of the room's response into two time regions: early and late. First, the early reflections are added in a room using an image-source method (see Fig. 1.3). The late reverberation is then included via a stochastic generation of diffuse sound. The models used in the present study varied reverberation times (T_{30}) from 0.0 to 7.2 s by changing the absorption and scattering properties of the surfaces. Each model simulated the same source and receiver locations to ensure that the early reflections of the sound arrived from the same locations. In this manner, raising T_{30} increases the sound levels of the image sources as well as the level and time of decay of the late reverberation, but keeps the number of early reflections constant.

A head-related transfer function (HRTF) filter from Subject 21 of the CIPIC database[116] was applied to the auralizations to create binaural stimuli with spatial cues. The loudness of the stimuli was equalized based on A-weighted levels. (A full description of the stimulus preparation can be found in Chapter 3.)

The spectrograms in Fig. 4.2 depict the spectral content of the auralizations. With increasing T_{30}, the amplitudes of the reflections increase, the sound becomes "muddier," and the perceived reverberance of the room increases.[4] This range of T_{30} values for the present work spanned 37 just noticeable difference (JND) limens according to the accepted JND of 5%.[26] The minimum number of JNDs between two stimuli was four.

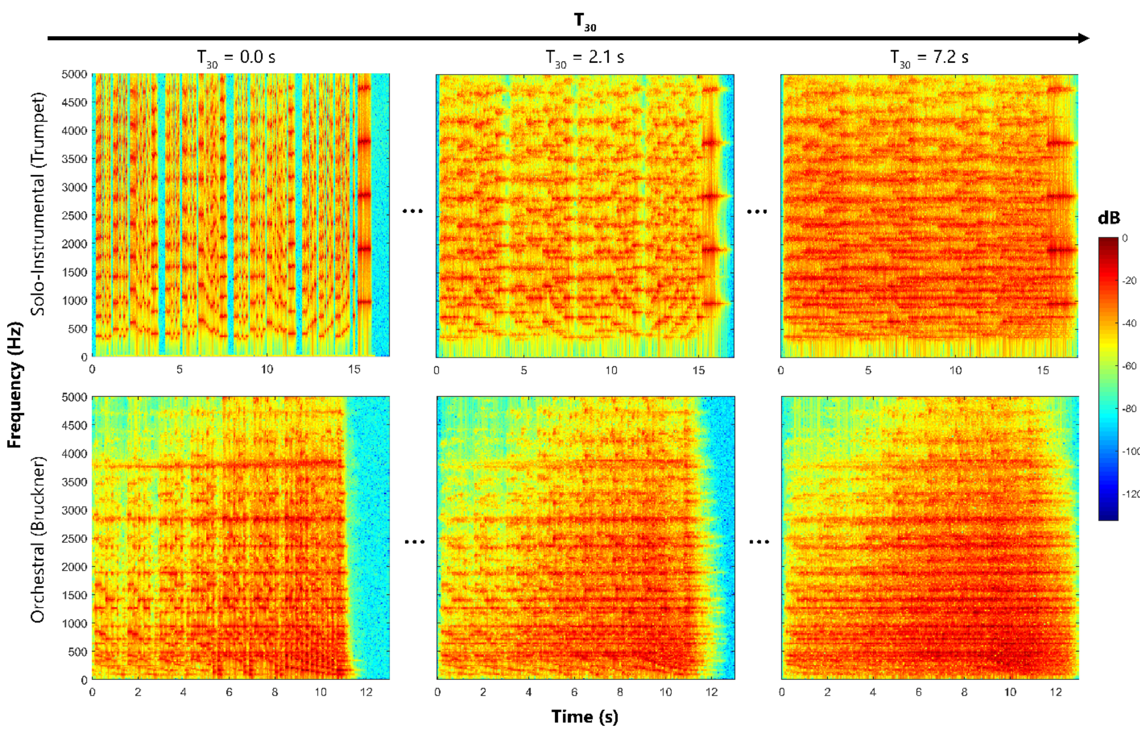


Fig. 4.2: Spectrograms of a subset of the solo-Instrumental and orchestral stimuli from 0 to 5 kHz. Each spectrogram shows the frequency as a function of time with the amplitude (normalized to the peak amplitude across all of the stimuli) denoted by the color scale. As T_{30} increases (left to right), the energy at each frequency spreads in time since the energy decay time of the modeled rooms becomes longer. Note that the solo-instrumental motif exhibits a clear harmonic structure, while the fundamental frequencies and harmonics of the multiple instruments of the orchestral motif are overlaid on top of one another.

4.2.3.2 Stimulus set for Session 2 fMRI scans

The first objective of the present work was to observe the brain's auditory response to room acoustics using fMRI. Thus, the test stimuli in Session 2 comprised auditory signals that contained room acoustic information and the control stimuli consisted of signals with no such information. The control stimuli were the two anechoic (T_{30} = 0.0 s) room acoustic stimuli of the solo-instrumental and orchestral motifs. To investigate if any of the effects were solely due to music or musical structure, two additional "Scrambled Music" stimuli were created by randomizing the phase of the anechoic recordings of both motifs and filtering the resulting waveform with the temporal envelope of the respective recordings. This process ensured that the overall frequency and rhythmic content of the Scrambled Music remained the same as the anechoic recordings. The Scrambled Music wavefiles were then auralized in the anechoic room acoustic condition. The resulting control stimuli (Fig. 4.3) maintained the temporal distinctness present in the anechoic musical passages, but blurred the frequency content. Refer to Appendix A.2 for more information on the construction of the control stimuli.

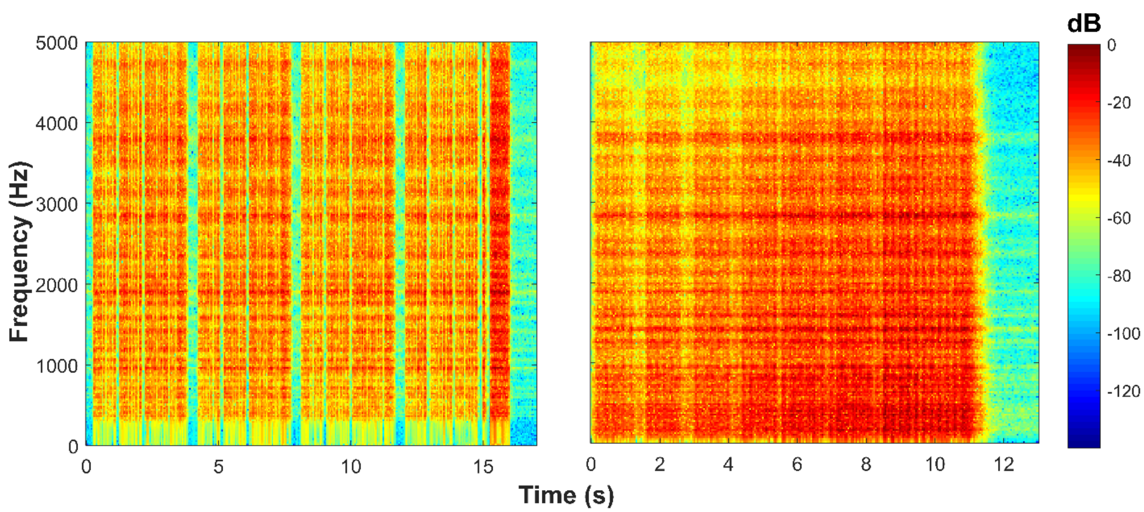


Fig. 4.3: Spectrograms of the scrambled music of the solo-instrumental (left) and orchestral (right) motifs from 0 to 5 kHz. . Each spectrogram shows the frequency as a function of time with the amplitude (normalized to the peak amplitude across all of the stimuli) denoted by the color scale.

Each participant's most-liked and most-disliked room acoustic conditions for both the solo instrumental and orchestral motifs were the test stimuli with applied room acoustics. These stimuli were determined by comparing each subject's preference ratings with the five preference groups found in Ref. [137]. The preference group of each participant was determined by minimizing the least means square distance between their preference ratings and the cluster centers (see Fig. 4.4 for comparisons between the averaged participant ratings and cluster centers for each group). As mentioned, the "No Preference" group was excluded from the study, which meant that the most-liked and most-disliked stimuli were those rated the highest and lowest in each of the remaining four preference groups. As shown in Fig. 4.4, the most-liked stimuli had T_{30} values of 1.0, 1.7, 2.1, and 2.8 s. The most-disliked stimuli for three of the five preference groups were the 7.2 s stimuli, while the fourth group disliked the anechoic stimuli. Since each participant would receive the anechoic stimuli anyway, the second-most-disliked stimuli, the 7.2 s stimuli, were chosen for participants that placed in the fourth group. Therefore, each participant's most-disliked stimuli had a T_{30} of 7.2 s.

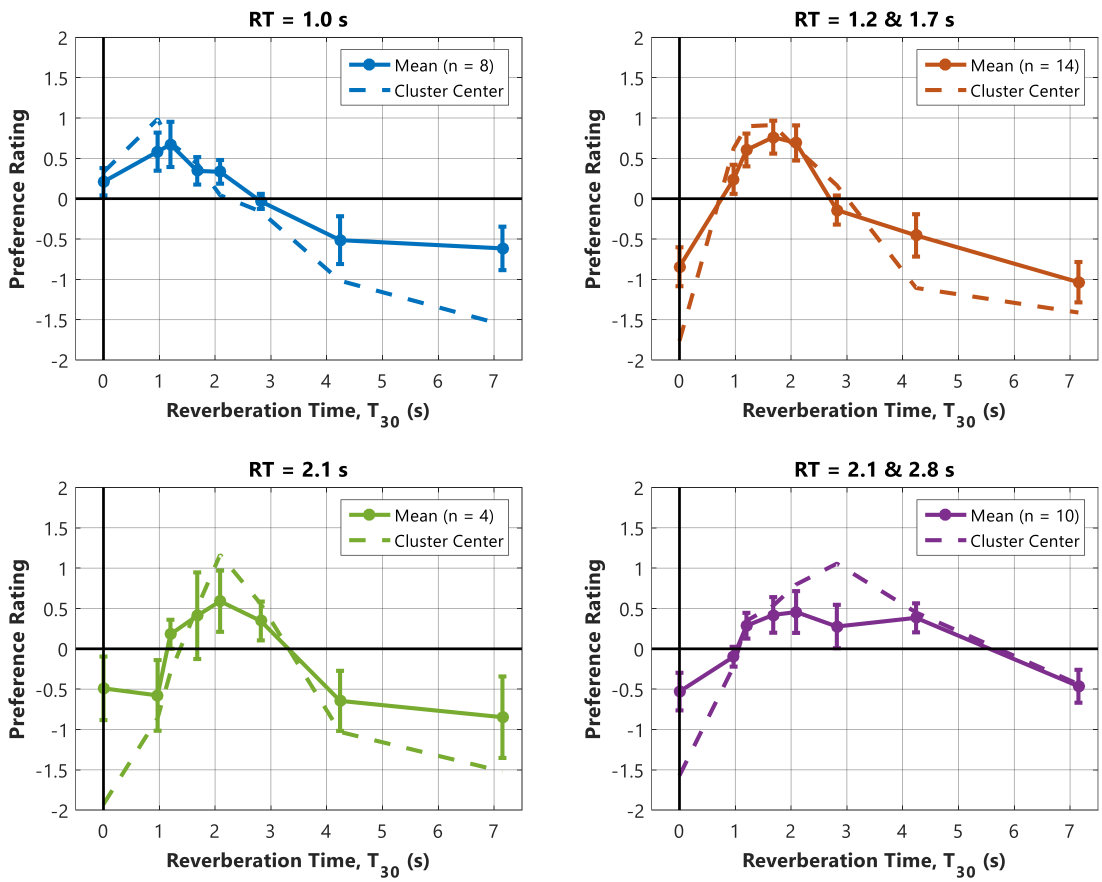


Fig. 4.4: Average preference ratings for each of the reverberation time preference groups determined by the least square fit to the cluster centers found in Ref. [137] (Chapter 3). The average data visually matches each cluster trend fairly well. Note that each participant was categorized individually, which is why some of the mean values deviate from the cluster center values.

4.2.4 Listening test methods

In the aforementioned previous study by the authors, two testing methods, a successive and a comparative test, were compared to determine whether both produced similar preference results and reliability. The study found that the two methods yielded similar preferences and reproducibility.[137] However, the subjects tended to perform quicker on the comparative method once they had been familiarized with the stimuli. Consequently, the present work utilized both methods, described below. Outside of the mock and MRI scanners, the comparative method was used after each participant listened to each stimulus. The successive method was employed inside the mock scanner because the ability to interface with the test program was limited and the block design of the MRI scans required successive presentation of the stimuli. The subjects did not rate the stimuli while in the MRI scanner.

4.2.4.1 Session 1: ratings with the comparative and successive method in the mock scanner

In the first session, each subject rated the stimuli in terms of overall preference in an office and inside the mock scanner. In both cases, the stimuli were presented with a custom-built Python graphical user interface (GUI) through STAX MR 003-MK2 electrostatic earbuds (see Fig. B.2). Each stimulus was rated on a continuous, five-point Likert scale.[6] The scale ranged from -2 ("Strongly Dislike") to 0 ("Neither Like nor Dislike") to +2 ("Strongly Like"). To rate a stimulus, the participants positioned a slider bar anywhere on the continuous scale. The preference ratings were rounded to the nearest hundredth's place.

Outside of the scanner, the subjects were prompted to rate the stimuli with the comparative method of testing. In the comparative method, all eight room acoustics conditions for each motif were concurrently tested to promote the direct comparison of stimuli. The stimuli were presented on the screen at the same time in a random order. "Play" buttons for each stimulus, labeled A through H, would start or pause the musical passage (see Fig. 3.4). When switching between stimuli, the musical passage would continuously play while the room acoustics setting changed. Each stimulus could be repeated as many times as necessary. All of the rating scales for each stimulus were also present on the screen underneath the buttons.

The successive method of testing (Fig. 3.3) was used to obtain the preference ratings while the subjects were in the mock scanner. Each stimulus was presented one at a time to completion during which the participants were instructed to record the tempo (i.e. count each beat) of the musical passage using the right thumb button of Nordic Neurolab Response Grips (Fig. B.3). After the stimulus finished playing, the rating scale appeared, and the subjects used the Response Grips to move the slider bar and submit their answer. To better emulate the format and length of Session 2 in the MRI scanner, any individual presentation of the stimuli could not be repeated. Each stimulus was only presented once. Howard Leight QM24+ earmuffs were used to attenuate the simulated scanner noise.

4.2.4.2 Session 2: successive presentation of stimuli in the MRI scanner

Session 2 utilized a successive presentation method similar to that employed in the mock scanner. Each stimulus was played one at a time followed by an inter-stimulus interval jittered between 8 and 12 seconds with a 10 second average. The order of the stimuli was pseudo-randomized such that two consecutive stimuli were not auralizations of the same musical motif nor the same room condition (T_{30} value). Again, the participants recorded the tempo of the musical passages with the Response Grips in an attempt to ensure that the participant actively listened to the stimuli. No other input was received from the subjects. The stimuli were presented over Sensimetric Model S14 earbuds (see Fig. B.5) with sound attenuation applied via Resonance Technology, Inc. VisuaStim over-ear headphones (see Fig. B.6). All of the stimuli were pre-filtered to account for the response of the Sensimetric earbuds. Each stimulus was presented a total of nine times over the course of five functional runs.

4.2.5 fMRI data acquisition

Structural and functional MRI scans were acquired with the Siemens 3T Prisma Fit (Fig. B.1b) at the Social, Life, and Engineering Sciences Imaging Center (SLEIC) at Penn State. The imaging data were obtained with a 20-channel head coil. A T1-weighted Turboflash (MPRage) sequence with 192 sagittal slices and 0.9-mm voxel size collected the structural images. For this sequence, the parameters were as follows: repetition time (TR) – 2300 ms, echo time (TE) – 2.32 ms, field of view (FOV) – 240 mm, matrix size – 256, and flip angle – 8^{o}.

There were five functional runs that collected a total of 1008 volumes with 33 axial slices each (slice thickness = 3.0 mm). The echo planar imaging (EPI) sequence that acquired the data had TR = 2000 ms, TE = 28 ms, FOV = 210 mm, matrix size = 64, and flip angle = 78°. The total length of the testing protocol was approximately 45 min; the length of five EPI sequences was 33 min and 36 s. The EPI protocol was a work-in-progress (WIP) sequence that allowed for the simultaneous acquisition of each participant's respiration using the Siemens Respiration Belt. Due to technical failure, the respiration data for three subjects was unable to be acquired.

4.2.6 Statistical analysis on behavioral data

In addition to analyzing each participant's preference ratings to determine the most-liked and most-disliked stimuli, the preference ratings from outside and inside of the mock scanner were compared across subjects to establish whether the subjects provided the same results in both scenarios. Using a repeated measures analysis of variance (ANOVA)[31] in SAS 9.4, the statistical similarity between to the two sets of preference ratings was evaluated. The general linear model (GLM) included the following treatments: T_{30} (8 levels), Motif (2 levels), and Scenario (2 levels), as well as the interactions between these three main effects.

4.2.7 fMRI image processing and statistical analysis

Both the structural and function MRI data were processed with statistical parametric mapping using the SPM8 program in MATLAB.[62] The fMRI images were first realigned to the first acquired volume. The six degrees of freedom (x-, y-, and z- translations; pitch, roll, and yaw rotations) were used as nuisance regressors in the statistical analysis. If the subjects had moved more than 1.0 mm in translation or 1° in rotation during an EPI run, the scans from that set were omitted from the analysis. A sinc-interpolation routine was used to correct for slice-timing acquisition differences. The structural image was coregistered to the mean of the functional images. Spatial normalization to the MNI avg152 standard template was applied to both the structural and functional images with resampled isotropic voxel sizes of 1 mm and 3 mm, respectively. Finally, spatial smoothing was applied with an 8-mm full-width at half maximum (FWHM) Gaussian kernel filter.

Respiration data was processed with the Physio Toolbox,[138] which removed clipped samples, normalized the data, and generated eight RETROICOR (RETROspective Image CORection) regressors to reduce noise in functional analysis.[139] An analysis was also performed to determine the relationship between respiration rate and the stimuli. The processed respiration data was divided into chunks based on the onsets and durations of the stimuli and inter-stimulus silences. To calculate the respiration rate for each time chunk, the processed respiration data was first low-pass filtered (0-1 Hz). The respiration rate was then calculated by finding the peak of the FFT of the low-passed respiration data. A repeated measures analysis of variance (ANOVA) was used to evaluate if the stimuli evoked different respiration rates across the participants during presentations and silences.

At the first level, a classical GLM mass-univariate model in SPM8 was used to analyze each subject's functional data. A factorial design was established with Motif (Solo-Instrumental or Orchestral) and Stimulus Type (Anechoic, Most-Liked, Most-Disliked, or Scrambled Music) as the main effects. This design resulted in eight separate conditions modeled with the canonical hemodynamic response function (HRF). Past studies have suggested that the hemodynamic response can vary across different brain regions,[140] including the basal ganglia.[141] As such, the time-derivative of the canonical HRF was included in the model for each condition, as outlined by Friston et al. (1998),[35] specifically to model the hemodynamic response in the basal ganglia. Linear contrasts were created between the treatment levels of the main effects and the interaction between the main effects. The second-level, group analysis evaluated these linear contrasts with one-sample *t* tests in a random-effects model. Activations are reported at peak voxel locations that are statistically significant after correcting for multiple comparisons using family-wise error (FWE) correction at $\alpha = 0.05$.

Region-of-interest (ROI) maps of the auditory cortices and basal ganglia were evaluated for all contrasts. The ROIs for the bilateral auditory cortices included the core cortical areas Te 1.0, 1.1, 1.2[42] that have been shown to correlate with primary auditory functions in primates.[142] Area Te 3 was also included to examine gross secondary function. These probabilistic cytoarchitectural maps were obtained from the SPM Anatomy library[143] transformed to MNI space. Similarly, the basal ganglia probabilistic maps of the dorsal and ventral striata were acquired from a 7T MRI study of 30 subjects.[144] Upon visual inspection of the ROIs overlaid on the average images from the 18 subjects of the present work (Fig. 4.5), these maps appeared to be correct. The ROI analysis was conducted using the MarsBar v0.42 toolkit,[145] and a Bonferroni correction was applied to account for the five separate ROIs evaluated.[31]

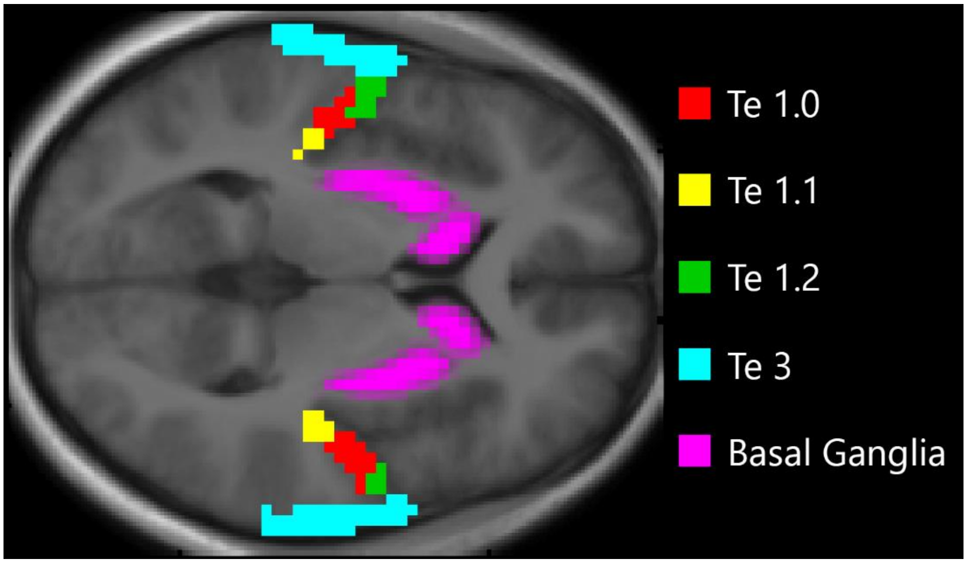


Fig. 4.5: The five ROIs overlaid on the average T1-weighted anatomical scan of the 18 participants at the horizontal plane z = 4.

4.3 RESULTS

4.3.1 Behavioral results: preference ratings, tempo task, and respiration

The preference ratings in the office and mock scanner were compared with a repeated measures ANOVA to test if the participants produced the same results in both scenarios. To satisfy a criteria of the ANOVA, the residuals of the data were checked for normal distribution. The main effects of T_{30} and Scenario as well as the interaction between Motif and T_{30} were significant at $p < 0.05$. No other effects showed any statistical significance. It was expected that the subjects would have preferred some stimuli over others. The Tukey-adjusted means comparison for the T_{30} effect showed that the 1.2, 1.7, and 2.1 s stimuli were the most-liked, while the anechoic (0.0 s) and 7.2 s stimuli were the most-disliked.

The significance of the Scenario main effect suggests that the mean preference ratings between the office and the mock scanner were statistically different. The office scenario produced an average preference rating of -0.1 ± 0.1 (s.d.), while the mean preference rating in the mock scanner was 0.1 ± 0.1 (s.d.), a difference of 0.2 ± 0.1 (s.d.). However, neither of these averages are statistically different from 0, which provides an indicator that the subjects tended to balance their preference ratings on the scale in both scenarios. Furthermore, the lack of significance of the interaction between T_{30} and Scenario indicates that the overall trends of preference over T_{30} are similar (Fig. 4.6). In fact, the only main difference demonstrated in the Tukey means comparison was for the 4.2 s stimuli. Since the 4.2 s stimuli were not the most-liked or most-disliked, this difference had no effect on the results of this study. Additionally, the algorithm to predict the most-liked and most-disliked stimuli produced the same stimuli for each subject regardless of which scenario's preference ratings were used.

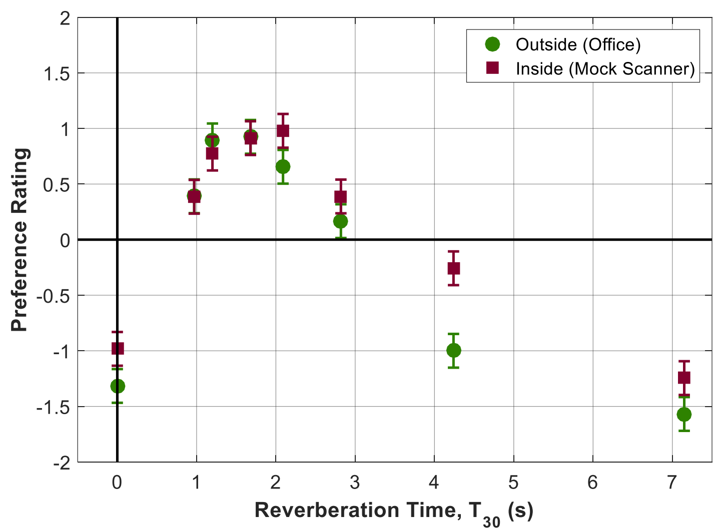


Fig. 4.6: Preference rating as a function of T_{30} for the scenarios outside and inside the mock scanner. The results indicate that the overall trends across T_{30} are similar, though the average preference ratings in the mock scanner were slightly higher than those in the office.

For the tempo task, almost all of the participants recorded the tempo accurately and consistently, approximately 120 BPM and 130 BPM for the solo-instrumental and orchestral motifs, respectively. Thus, these subjects were judged to be attentive throughout the entire MRI session. There was one subject that recorded slower tempos, specifically half of the rate for both motifs, which means that they were counting every other beat. Since this subject was consistent in this regard, this subject was also determined to have been attending to the stimuli.

The difference between the respiration rates of the stimulus presentations was evaluated with a repeated measures ANOVA. The subjects breathed faster when a stimulus was being presented than silence, a respiration rate difference of 0.015 ± 0.003 Hz ($p < 0.0001$). However, no two stimuli evoked different respiration rates either during the presentation or the immediately followed silence ($p > 0.05$).

4.3.2 Functional results: whole-brain analysis

One sample *t*-tests were used to evaluate the contrasts for the effects of stimulus type (i.e. anechoic, liked [moderate reverberation], disliked [extreme reverberation], and scrambled music) and motif (Table 4.1). On the second-level, the contrasts between the room acoustics and motif conditions revealed large clusters of activations bilaterally located in the superior temporal gyrus (STG). Several of the clusters also contained peak activations in the left and right Heschl's Gyri.

Across the contrasts, the STG and Heschl's Gyri activated more so for clearer, simpler stimuli than for the more reverberant, complex stimuli (Fig. 4.7). Specifically, for contrasts of the main effect of stimulus type, the anechoic stimuli elicited greater activations in the L and R STG than the extremely reverberant stimuli. The *Anechoic* > *Liked* contrast also revealed significant activations in the R STG. Similarly, bilateral STG activations were greater during the presentations of the solo-instrumental motif than those of the orchestral motif. Overlap of the activations for the contrasts of these two main effects occurred primarily in the primary auditory cortex (PAC), including areas in the bilateral STG and Heschl's Gyri (see Fig. 4.8). The left postcentral gyrus only exhibited significant activations in the *Solo-Instrumental* > *Orchestral* contrasts, most likely because the tempo of the two motifs that the participants were required to record with their right hands differed by 10 beats per minute (BPMs).

Table 4.1: Whole-brain group-level one sample *t*-test results. Peak levels are reported for significant p_{FWE} < 0.05. The differences in activations between conditions are primarily located in the primary auditory cortex, including the Right (R)/Left (L) superior temporal gyrus (STG) and Heschl's Gyrus (G). Implicated regions also include the R dorsolateral prefrontal cortex (DLPFC), R lateral globus pallidus (LGP), L postcentral G, and L temporal pole.

Contrast	Region	Peak-Level p_{FWE}	Peak-Level T	Voxels in Cluster	MNI Coordinates (mm) x	y	z
Anechoic > Liked	**R STG**	**0.042**	**6.68**	**315**	**66**	**-22**	**7**
Anechoic > Disliked	**L STG**	**< 0.001**	**11.22**	**702**	**-57**	**-28**	**7**
	L Heschl's G	< 0.001	10.11		-33	-31	10
	L STG	0.002	8.75		-54	-16	1
	R Heschl's G	**0.001**	**9.09**	**565**	**42**	**-25**	**13**
	R STG	0.002	8.93		66	-25	7
	R STG	0.019	7.42		57	-10	1
Scrambled Music > Liked	**R STG**	**0.007**	**7.96**	**376**	**54**	**-31**	**13**
	R STG	0.016	7.40		63	-31	10
	R STG	0.021	7.17		45	-34	13
	R DLPFC	**0.080**	**6.24**	**238**	**48**	**17**	**25**
Scrambled Music > Disliked	**R STG**	**0.021**	**7.29**	**366**	**60**	**-31**	**10**
	R LGP	**0.027**	**7.09**	**100**	**18**	**-4**	**7**
Solo-Instr. > Orchestral	**L Heschl's G**	**< 0.001**	**14.61**	**1206**	**-39**	**-31**	**13**
	L STG	< 0.001	11.95		-60	-19	7
	L STG	< 0.001	10.23		-48	-22	10
	R STG	**< 0.001**	**12.17**	**1079**	**48**	**-16**	**1**
	R Heschl's G	< 0.001	11.10		42	-28	13
	R STG	< 0.001	11.03		63	-25	4
	L Postcentral G	**0.002**	**8.90**	**258**	**-36**	**-22**	**49**
	L Temporal Pole	**0.029**	**7.07**	**67**	**-45**	**11**	**-29**
Motif x Type Interaction: Scrambled Music > Disliked	**L STG**	**0.008**	**8.07**	**257**	**-39**	**-31**	**10**
	R STG	**0.049**	**6.79**	**162**	**39**	**-22**	**1**

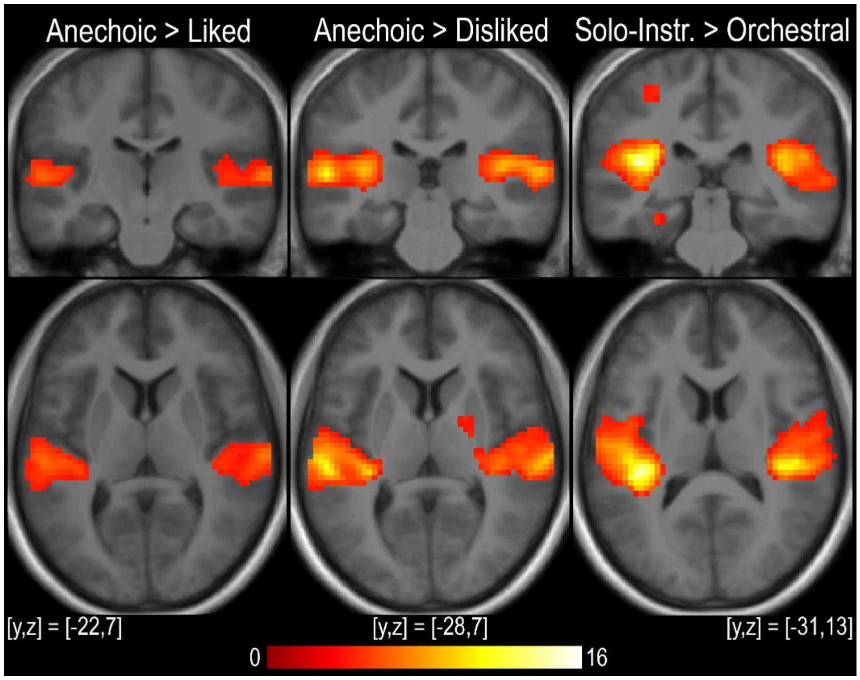


Fig. 4.7: Significant activations for the *Anechoic > Liked*, *Anechoic > Disliked*, and *Solo-Instrumental > Orchestral* contrasts. The three activation maps are determined based on uncorrected threshold of p < 0.001, with peak-voxel threshold at p < 0.05, corrected. The colorbar indicates the *T* value associated with each voxel.

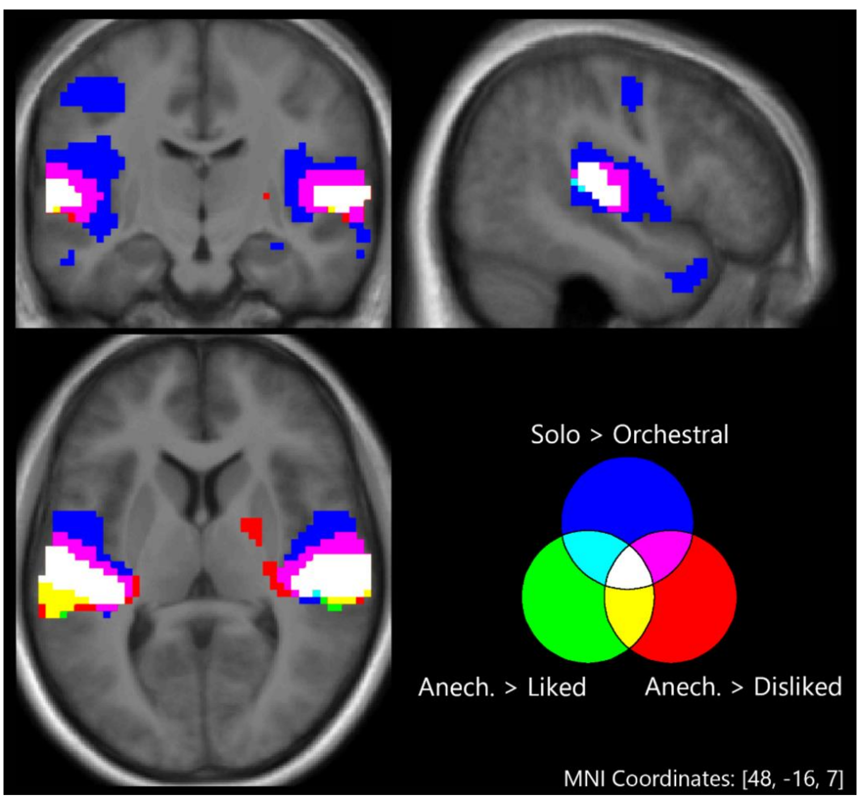


Fig. 4.8: Overlapping activation regions for each of the contrasts listed in Fig. 4.7 as specified by the color Venn diagram. Each of the three contrasts revealed significant activations in the primary auditory cortex (indicated in white), and other areas of the bilateral STG.

Over the whole brain, there were no significant findings for the *Anechoic > Scrambled Music* contrast, which suggests that the scrambled music stimuli well matched the frequency and rhythmic content of the anechoic recordings. Moreover, the R STG exhibited higher activations for the scrambled music than for either the liked or disliked stimuli. This result may be analogous to the results demonstrated by the contrasts between the anechoic and liked/disliked stimuli, in that adding reverberation to the binaural stimuli caused the STG to activate less. As seen in Fig. 4.3, the scrambled music stimuli had the same temporal clearness of the anechoic stimuli, which means that the significant activations in the *Scrambled Music > Liked/Disliked* contrasts were most likely due to the differences in reverberation in the stimuli.

The interaction between Motif and Stimulus Type for the *Scrambled Music > Disliked* contrast also proved to be significant. This result means that the solo-instrumental motif evoked a larger neural response in the left and right STG in this contrast than the orchestral motif. The interaction was not significant for any of the other contrasts.

4.3.3 Region of Interest (ROI) results

For the ROI analysis of the auditory response, data was extracted from the left and right auditory cortices using probabilistic cytoarchitectonic maps of areas Te 1.0, 1.1, 1.2, and 3. The analysis was performed on unsmoothed functional images since the regions border each other. Each ROI was evaluated separately for each contrast, and the p-values were FWE corrected.

Similar to the results of the whole brain analysis, the ROI analysis revealed that the auditory regions exhibited significant activations for the effects of room acoustics and motif (Table 4.2). Area Te 1.0, the most primary-like region of the auditory cortex, activated more strongly during the presentations of the anechoic stimuli than for the more reverberant Liked and Disliked stimuli. This region also demonstrated stronger activations for the solo-instrumental motif than the orchestral. Areas Te 1.1, 1.2, and 3 presented significant activations for these contrasts as well, although the *Anechoic > Liked* contrast had no significant activations in area Te 1.2. These results confirm that the primary auditory cortex has a greater response to clear, simple stimuli than for more reverberant, complex stimuli.

Table 4.2: ROI analysis results in the auditory cortex for each of the main effect contrasts of stimulus type and motif. Only significant activations (p_{FWE} < 0.05) are reported.

ROI	Anechoic > Liked		Anechoic > Disliked		Solo Instr. > Orchestral		Anechoic > Scrambled		Scrambled > Liked		Scrambled > Disliked	
	p_{FWE}	T	p_{FWE}	T	p_{FWE}	T	p_{FWE}	T	p_{FWE}	T	p_{FWE}	T
Te 1.0	0.003	4.20	< 0.001	6.69	< 0.001	10.12	0.003	4.19	-	-	-	-
Te 1.1	< 0.001	5.88	< 0.001	8.07	< 0.001	12.68	0.005	3.94	0.002	4.47	< 0.001	5.24
Te 1.2	-	-	< 0.001	5.70	< 0.001	7.49	< 0.001	5.04	-	-	-	-
Te 3	0.009	3.68	< 0.001	6.43	< 0.001	4.82	< 0.001	5.50	-	-	-	-

Lateralization of the auditory response was evaluated by conducting a paired-*t* test between the mean contrast values in the left and right ROIs. The left auditory area Te 1.1 exhibited stronger activations for the *Solo-Instrumental* > *Orchestral* contrast ($t(17) = -5.02$, $p_{FWE} < 0.001$), while the right auditory area Te 1.1 had stronger activations in the *Anechoic* > *Liked* ($t(17) = 5.22$, $p_{FWE} < 0.001$) and *Anechoic* > *Disliked* ($t(17) = 3.17$, $p_{FWE} = 0.02$) contrasts. No other ROI showed a significant effect of hemisphere.

Significant activations were only found in area Te 1.1 for each of the Motif by Stimulus Type interaction contrasts (Table 4.3). This result signifies that the solo-instrumental motif caused stronger activations in area Te 1.1 than the orchestral motif over each of the room acoustic conditions. Consequently, the significances of the activations in area Te 1.1 in the main effect contrasts may have been primarily due to the differences between the two motifs, namely frequency content. However, areas Te 1.0, 1.2, and 3 did not have any significant activations for the interaction contrasts, meaning that these regions responded separately to differences between effects of motif and, more importantly to the purposes of this study, room acoustics condition.

Table 4.3: ROI analysis results in the auditory cortex for the interactive effects between motif and stimulus type.

ROI	Motif x Type: Anechoic > Liked		Motif x Type: Anechoic > Disliked		Motif x Type: Scrambled > Liked		Motif x Type: Scrambled > Disliked	
	p_{FWE}	T	p_{FWE}	T	p_{FWE}	T	p_{FWE}	T
Te 1.0	-	-	-	-	-	-	0.002	4.39
Te 1.1	0.005	4.02	0.001	4.68	0.004	3.15	< 0.001	8.17
Te 1.2	-	-	-	-	-	-	-	-
Te 3.0	-	-	-	-	-	-	-	-

A fifth ROI consisting of the probabilistic cytoarchitectonic maps of the left and right striatum of the basal ganglia was evaluated once again on unsmoothed functional data. The dorsal and ventral striatum comprise the caudate nucleus and putamen, and the nucleus accumbens (NAcc), respectively. As mentioned in Section 4.2.7, it was expected that the HRF of the basal ganglia was the temporal derivative of the canonical HRF. True to the hypothesis, the ROI analysis using the model of canonical HRF did not demonstrate any significant activations of the basal ganglia. However, with the assumption of the temporal derivative HRF model for the basal ganglia, significant activations were found in the contrasts of the main effect of room acoustics condition. Specifically, the striatum activated more during the presentation of the Liked room acoustics condition than the Disliked ($t(17) = 3.40$, $p_{FWE} = 0.017$). Similarly, the Liked conditions evoked higher activations in basal ganglia than the anechoic stimuli in the *Liked* > *Anechoic* contrast ($t(17) = 3.44$, $p_{FWE} = 0.016$). This result agrees with the behavioral task, where both the Disliked and anechoic stimuli were both rated as “disliked.” The basal ganglia ROI did not exhibit any significant activations for any other contrast in the canonical and temporal derivative HRF models.

The temporal derivative contrasts were also examined for the auditory ROIs, and most were found to be significant at $p_{FWE} < 0.05$. However, these results will not be discussed further in the text to avoid any circular analysis and because the auditory cortex was not considered to exhibit a hemodynamic response that differed from the canonical model.

4.4 DISCUSSION

4.4.1 Auditory cortex response to changes in reverberance, motif, and musical structure

The results of the whole-brain and ROI analyses show evidence of an auditory response to changes in the reverberation of a music stimulus. Specifically, the addition of reverberation in the musical signals degraded the activations in the auditory cortex as compared to the anechoic passage. At the whole-brain level, the influence of reverberation extended throughout large clusters in the left and right STG, while at the ROI-level, the effect manifested in all of the bilateral auditory ROIs evaluated: areas Te 1.0, 1.1, 1.2, and 3.

The auralizations of the reverberant musical stimuli were created by convolving the anechoic recordings of each of the instruments in the motifs with the binaural room impulse responses (BRIRs) of the eight room conditions that simulated the early reflections and late reverberation. The early reflections were modeled with the image source method, while the late reverberation was stochastically generated. As reverberation time increased in the auralizations, both the sound levels of the image sources and the level of the late reverberation increased. The sensitivity of the auditory cortex to reverberation could have been due to the early reflections, late reverberation, or some combination of both. The present study did not explicitly test for this effect, and future work is needed to reach any firm conclusions.

Past research suggests that both the addition of early reflections and late reverberation contributes to the decrease of the cortical auditory response.[130,131,134] The early reflections of a room add images of the source around the listener. The sounds from these images have an inherent phase lag based on the differences between the source-receiver and image-receiver distances. Images with more than 50-80 ms of lag, as determined by the Precedence Effect,[146] essentially act as incoherent source objects. Thinking of the room acoustic signals in this regard, the results of the present study compare with those of past research of the effect of auditory objects. As incoherent auditory objects were added into a scene, the extent of activations in the auditory cortex decreased.[130,131] A similar effect was observed when testing the difference between the solo-instrumental and orchestral motifs. The auditory cortex was sensitive to the number of instruments (incoherent auditory objects) present in the auditory signal. However, it is important to note that the frequency content and musical structure of these motifs were different from each other, which may have contributed to the large differences of activations between the two conditions.

The sensitivity of the auditory cortex to the addition of incoherent noise has been established with respect to the clarity of speech signals.[134] Adding late, random reverberation into the musical signals may be equivalent to adding noise to the music signals (see Fig. 4.9). The present study demonstrated a similar sensitivity with respect to musical stimuli since the late reverberation caused a degradation of activations in the auditory cortex. In fact, visual comparisons between Fig. 4.8 and Fig. 4.9 show analogous regions of activation. In the present study, that more reverberation was more detrimental to the auditory response; however, this trend was not significant in the *Liked* > *Disliked* contrast.

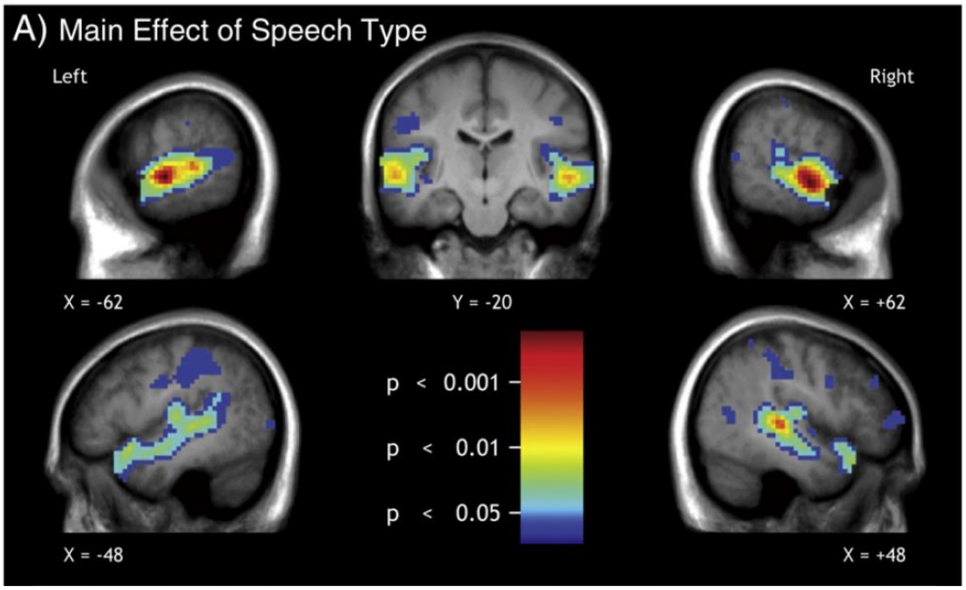


Fig. 4.9: Voxels that exhibited a sensitivity to the clarity of speech in noise. These auditory regions demonstrated stronger activations in the presence of clear speech than noise-vocoded speech. Image from Fig. 3(A) in Wild et al. (2012).[134]

To compare the effects of reverberance and musical structure on the auditory cortex, contrasts were established between the room acoustics stimuli and a scrambled version of the anechoic stimulus. The generation of the scrambled music intended to create a stimulus that contained the same rhythmic content, while changing the phase of the frequency content to eradicate the melodic structure. This procedure for scrambling the music differed from past studies (e.g. Refs. [132] and [147]), but was expected to produce similar effects in the auditory cortex. While the *Anechoic > Scrambled Music* contrast was not revealed to be a significant simple effect in the whole-brain analysis (Table 4.1), all four auditory ROIs presented significant activations for this contrast (Table 4.2). This outcome demonstrates that the removal of the melodic structure may produce a small effect on the primary and secondary regions of the auditory cortex.

Interestingly, the contrasts between the Scrambled Music and the reverberant room conditions (Liked and Disliked) did exhibit significant activations on the whole-brain level (see Table 4.1). The ROI analysis revealed that these activations manifested in area Te 1.1, while no other auditory ROI had significant activations (see Table 4.2). The lack of activations in the majority of the auditory ROIs suggests that the scrambling of music produced a similar effect as the addition of reverberation for both primary and secondary regions of the auditory cortex since areas Te 1.0, 1.2, and 3 were all activated at least to a small extent for the *Anechoic* > *Scrambled Music* contrast. Furthermore, area Te 1.1 exhibited higher activation for the scrambled music than the reverberant conditions, which blurred the rhythmic content of the motifs. The strong activations in area Te 1.1 as displayed in the whole-brain and ROI analyses indicate that this region may be the more sensitive to the clarity of musical structure.

This conclusion is supported by the significant activations in the Te 1.1 ROI for each of the simple effects of the interaction between motif and room acoustics condition (Table 4.3). The positive activations imply that area Te 1.1 responded more strongly for the solo-instrumental motif, i.e. the motif with the clearest musical structure, for each of these conditions. While area Te 1.0 is considered to be the primary "core" region of the auditory cortex, area Te 1.1 serves as an anterior and medial "belt" region that may serve as a pathway to secondary auditory areas for the purposes of music processing.

4.4.2 Basal ganglia response to differences in preference

The activation maps of the *Liked* > *Disliked* and *Liked* > *Anechoic* contrasts in the temporal derivative model (Fig. 4.10) both show the locations of the activations in the basal ganglia. Although the ROI analysis of the bilateral striata revealed similar levels of activation for these contrasts, the exact locations of the activations were not the same. Activations in the caudate nucleus were prevalent in the *Liked* > *Disliked* contrast, while the putamen and NAcc were active in the *Liked* > *Anechoic* contrast.

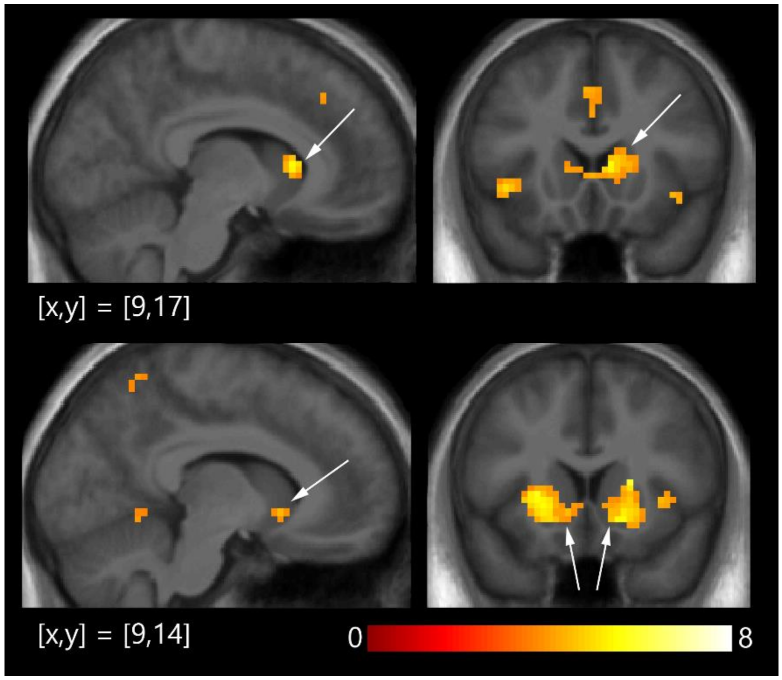


Fig. 4.10: (*top*) Activation map of the *Liked* > *Disliked* contrast in the temporal derivative HRF model, showing the caudate nucleus, indicated by the arrow, as the area in the ROI that was primarily activated. (*bottom*) Activation map of the *Liked* > *Anechoic* contrast in the temporal derivative HRF model. In this map, the putamen and the ventral striatum of the basal ganglia, including the NAcc (indicated by arrows), were the activated areas.

The ventral striatum has been related to the experience of the reward,[48,51,148] including during the entire presentation of musical passages.[50,57] Past research has also implicated the dorsal striatum in processing of an anticipation of a reward.[51,48] Research has also shown activation of the caudate nucleus in contrasts involving the entire duration of musical passages[149] when the activation of the NAcc was not present. The aforementioned music studies utilized the canonical HRF model to demonstrate the activations of the NAcc and caudate in contrasts of two completely different musical passages. In the present study, the activations of the dorsal and ventral striata were only found in a contrast of the temporal derivative model, meaning that the activations in this region were transient and did not last for the duration of the music. Also, these basal ganglia activations were found in an ROI analysis and not at the whole-brain level, which implies that the activations were not as prominent as in the past studies. As one may expect, the reward response is much stronger for preferred music than preferred room conditions, especially when the music is subject-selected. However, the present study shows that a reward still occurs in response to the more subtle effect of room acoustics on the same musical passage, though to a lesser extent and duration.

The temporal derivative model may be more typical of the "everyday" reward response in the brain. Individuals are not constantly being presented with monetary incentives or their favorite pieces of music that would trigger large, extended responses in the basal ganglia. Instead, an individual's reward response to nominally liked stimuli may be transient and to a lesser extent, but nevertheless present.

4.4.3 Considering auditory activity as an input to the reward network

The auditory cortex may act as an input to the reward network since the reward network responds to the perception of sound. A psychophysiological interaction (PPI) analysis was conducted to examine the connectivity of the auditory and reward networks.[34] PPI analyses use regression models to test if the connectivity between two distant brain regions is modulated by experimental context. The experimenter chooses the first region, called the "seed," and uses the PPI analysis to determine which regions may be affected by the context. Fig. 4.11 shows a schematic of how the results of a PPI may be interpreted in one of two ways provided the activity in the seed region k and the experimental factor. Either the experimental factor modulates the influence of the seed region or the activity in the seed region modulates the experience of the experimental factor. A limitation of PPI analyses is that it is difficult to decipher which interpretation is correct.

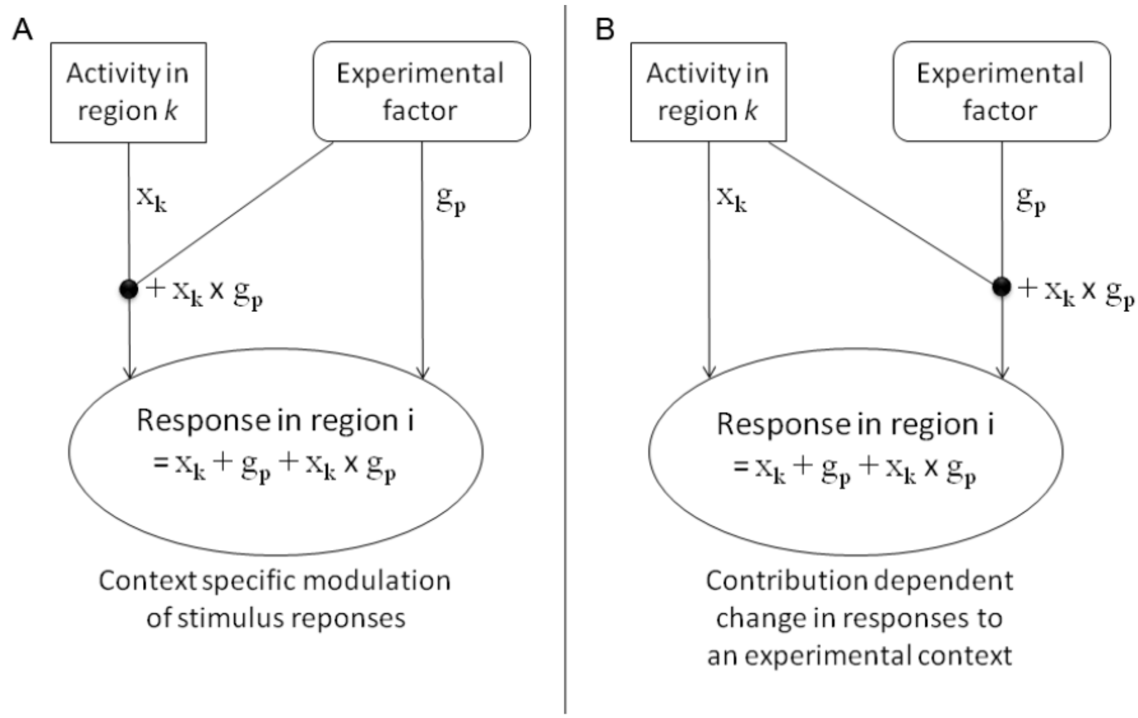


Fig. 4.11: Two alternative interpretations of PPI effects. A) The contribution of one area (k) to another (i) is altered by the experimental (psychological) context. B) The response of an area (i) to an experimental (psychological) context due to the contribution of region (k). Image and caption from SPM8 Manual, Fig. 33.2.[62]

For the present study, the auditory cortex was considered as the seed region, while the preference level of the stimuli (Anechoic, Liked, Disliked) was the experimental factor. The PPI analysis examined whether the preference level of the auditory stimuli interacted with the activity of the auditory cortex. There were no significant activations at the group level in any region of the brain, which suggests that the connectivity of the auditory cortex to other brain regions was not modulated by preference, or vice versa. In other words, the amount of correlation between the responses of the auditory cortex and other brain regions, including the reward network, did not change depending on the participants' preferences to the reverberant stimuli.

An analysis of covariance (ANCOVA)[31] was run to confirm this result specifically for the basal ganglia (reward network). The ANCOVA was designed to test if the correlation between the measured response of the basal ganglia and the measured activity in the PAC changed according to the preference level of the stimuli. For the ANCOVA, the raw BOLD response of the basal ganglia was regressed against the raw BOLD activity in the auditory cortex for the three preference conditions. If the relationship between the responses in the auditory cortex and basal ganglia (i.e. the slope of the regression) proved to be significantly different over the levels of preference, it is likely that the activity of the auditory cortex was modulated by preference.

The results demonstrated that preference most likely does not interact with the activity of the auditory cortex (see Fig. 4.12). While there was a significant interaction between the PAC activity and the preference condition ($p < 0.001$), only the Anechoic condition produced a significantly different slope than the other two conditions. The differences in slope may be more accurately accounted for by the presence of reverberation since the Liked and Disliked conditions yielded statistically similar slopes ($p = 0.85$). Additionally, the PAC experienced higher amounts of activation for the anechoic stimuli than either the liked or disliked stimuli. This effect was demonstrably more pronounced than the measured reward response and may outweigh any modulation effect.

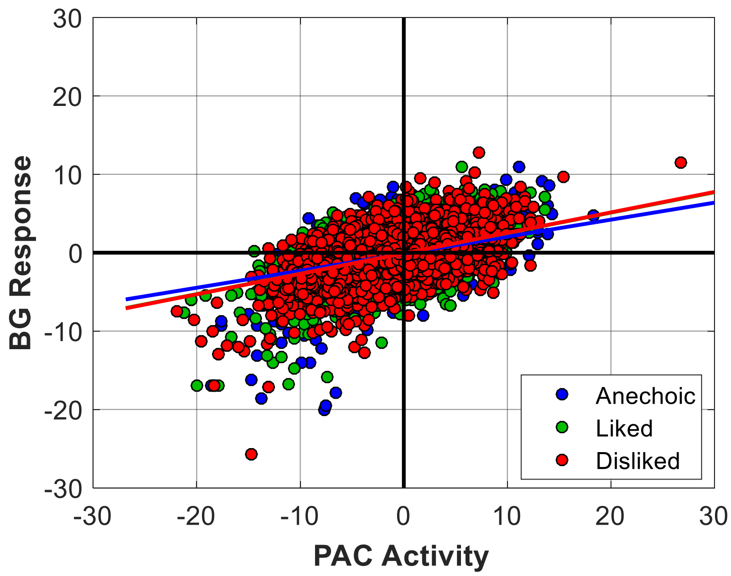


Fig. 4.12: Scatterplot of measured Basal Ganglia (BG) BOLD Response vs. Primary Auditory Cortex (PAC) BOLD Response for the Anechoic, Liked, and Disliked conditions (the points of the Disliked condition obscure the view of the other points). For all three conditions, the auditory activity and reward responses were correlated ($p < 0.001$). The slope of the response to the Anechoic condition differed from the others ($p < 0.001$), but the slopes of the Liked and Disliked conditions were not significantly different ($p = 0.85$).

4.4.4 Discussion of possible conflation between reverberation and preference

Several experimental design choices were made to study the effects of both reverberation and preference. One major decision involved presenting each participant's most liked stimulus as the moderately reverberant condition in order to reduce the total number of stimuli presented in the MRI scanner. As a result, not all participants received the same moderately reverberant stimuli. This choice considerably lowered the total number of contrasts analyzed (from 43[*] to 13[†] for both the canonical and temporal derivative HRF models) as well as the amount of time that each participant remained in the scanner. Consequently, the variables of reverberation and preference could potentially be confounding. Specifically, the measured auditory response could have been due to differences in preference rather than reverberation. However, reverberation and preference were both considered to be inherent properties of sound in the context of the present study. Although preference is not a physical property of the test signals, preference was ascribed to each condition for each participant through preference ratings, which were obtained in the first testing session outside the MRI. Additionally, there were no significant differences between the Liked and Disliked conditions in the auditory cortex. Thus, it was presumed that the conflation between reverberation and preference had a limited effect.[‡]

4.5 SUMMARY AND CONCLUSION

Before now, the brain's response to room acoustic stimuli had not been explicitly established using functional magnetic resonance imaging. Building off the research of the response to auditory objects, noise, and music, the findings of the present study provide a basis for the auditory and preference response to room acoustics stimuli.

The auditory cortex displayed a sensitivity to the reverberation of a room similar to the previously reported sensitivity to incoherent auditory objects and noise. While not explicitly tested in this study, it is likely that both the addition of auditory objects by the early reflections of the room and incoherent noise from the late reverberation contribute to the degradation of the auditory response. This conclusion was confirmed by the similar

---

[*] This number is based on the inclusion of all 4 most-liked stimuli, T_{30} = 1.0, 1.7, 2.1, and 2.8 s, in addition to the anechoic, most-disliked (T_{30} = 7.2 s), and scrambled music conditions for both motifs.

[†] This number is based on the inclusion of only one most-liked stimulus, in addition to the anechoic, most-disliked (T_{30} = 7.2 s), and scrambled music conditions for both motifs.

[‡] The follow-up experiment (Chapter 5) eliminated the confounder of preference in the examination of the effect of reverberation by presenting the same four T_{30} conditions to each of the subjects. However, this decision potentially sacrificed the extent of the reward response since the participants were not necessarily listening to their individualized most-preferred conditions.

sensitivity of the auditory cortex in the presence of additional musical instruments.

Furthermore, the present study compared the effects of room acoustics and the scrambling of music structure. While both reverberation and the scrambling of music may grossly affect the auditory cortex by impairing musical structure in a similar manner, the region most likely responsible for music processing, area Te 1.1, still exhibited activation differences between the two effects. This result suggests that area Te 1.1 may be involved in higher-level music processing.

The reward response to room acoustic stimuli was also examined. The present study demonstrated activations in the ROI of the dorsal and ventral striatum of the basal ganglia, indicative of the experience of a reward, in contrasts between the most preferred room acoustic conditions and the most disliked room acoustic conditions using a temporal derivative model of the HRF. As the differences between a listener's preferences to room acoustic may be more subtle than their preferences to different subject-selected musical passages, the measured effect was small and transient. Nevertheless, the observed reward response shows that fMRI is a viable tool for distinguishing listener preferences to room acoustics stimuli.

4.6 ACKNOWLEDGMENTS

We would like to thank the Social, Life, and Engineering Sciences Imaging Center (SLEIC) at Penn State for the use of the facilities, and Dr. Nanyin Zhang for his input on the data analysis and discussion. We would like to express special thanks to Dr. Thomas Beck and Dr. Josef Pfeuffer, Siemens Healthcare, Erlangen for providing us with the Advanced fMRI WIP package.

CHAPTER 5

EVALUATION OF THE HUMAN AUDITORY CORTICAL RESPONSE TO COHERENT HARMONIC CONTENT AND INCOHERENT REVERBERATION IN MUSIC

The following chapter presents a further examination of the auditory and reward response to room acoustic stimuli with fMRI by considering stimuli with and without coherent harmonic content simulated in various room acoustic conditions. The chapter was prepared as a manuscript and will be submitted to a peer-reviewed neuroscience journal.

Evaluation of the human auditory cortical response to coherent harmonic content and incoherent reverberation in music

Martin S. Lawless and Michelle C. Vigeant

Graduate Program in Acoustics, The Pennsylvania State University, University Park, Pennsylvania 16802, USA

Abstract

Reverberation degrades the cortical auditory response to music. However, it is yet unclear whether early reflections, which add incoherent auditory objects, or late reverberation, which adds incoherent noise, contributes to the sensitivity of the auditory cortex. Studies have also shown that timbre influences the response of the auditory cortex, but have not fully demonstrated the differences in responses to a melodic passage with and without coherent harmonic content with functional magnetic resonance imaging. The present study intends to investigate the auditory response to reverberation and timbre. A sine-wave stimulus with the same notes as a trumpet anechoic musical passage was created by removing the harmonic content and applying the attack, decay, and rhythm of the trumpet. Outside of the MRI, sixteen participants listened to the two musical motifs simulated in eight room models with reverberation times from 0.0 to 7.2 s, and rated overall preference, perceived reverberance, and perceived clarity. A subset of the room conditions for both motifs were presented to the subjects in the MRI. Coherent acoustic information bolstered the response of the auditory cortex. Adding incoherent information reduced the response as predicted, but the amount of reverberation did not correlate with the amount of degradation. No interaction between reverberation and timbre was observed; thus, the two effects are independent.

5.1 INTRODUCTION

Sound, and more specifically music, consists of several perceptual attributes, including pitch, loudness, timbre, perceived duration, spatial location, and acoustic environment.[150] While a large body of work exists on the human auditory processing of pitch, loudness, duration, and spatial localization, considerably less work has studied the brain's response to room acoustics. The acoustic environment, or room, can alter a listener's perception of *reverberance*, the sense of sound lingering in a space; *clarity*, the impression of the clearness of the sound; and even *preference* to the heard sound. Some neuroimaging studies have studied the effects of rooms on music and speech,[53,54] but only recently has functional magnetic resonance imaging (fMRI) been used to uncover the cortical auditory response to reverberant stimuli.

Reverberation, which adds both spatial, auditory objects and random, diffuse noise, lowers the response of the auditory cortex to musical stimuli (see Chapter 4). This finding agrees with research that has shown that the auditory cortex is sensitive to incoherent frequency content. In particular, auditory cortical activation tends to decrease with increasing incoherent modulation,[130] the addition of extraneous auditory objects,[131] or the inclusion of random noise[134] on acoustic signals. However, it is yet unclear if the amount of reverberation in a signal affects the extent of degradation of the auditory response. The level and length of the reverberation, characterized by the reverberation time (RT), may correlate with the sensitivity of the auditory response.

If reverberation adds incoherent acoustic information into a musical signal, it is also interesting to consider the effects of adding coherent acoustic information. For musical instruments, coherent acoustic information comprises the harmonic structure that influences the perception of *timbre*, the quality of sound that allows listeners to distinguish between different sound sources. Several functional neuroimaging experiments have implicated the superior temporal gyrus (STG) and superior temporal sulcus (STS) as the centers for processing timbre.[151-156] These studies contrasted timbre profiles that differed in either harmonic structure, temporal cues, or both, but did not examine the contrast between tones or a melodic passage with and without harmonic content. A magnetoencephalography (MEG) study by Pantev et al. (2001) investigated the differences between violin, trumpet, and pure sine tones.[157] The study determined that the auditory cortical representations of the tones were greater for musical timbres associated with the musician's principal instrument. For example, the measured MEG auditory response for the violinists was larger when listening to the violin tones than for the trumpet or pure sine tones. However, the musicians experienced larger auditory responses for the musical stimuli in general, making it unclear if the measured effect was due to familiarization with the instrumental sources or the presence of coherent acoustic information.

Most acoustic scenes involve both changes in reverberation and timbre, and the amount of reverberation may affect the perception of timbre, or vice versa; yet, very little research has been conducted on the interaction between reverberation and timbre. The ability to discriminate between musical instruments may change in a concert hall based on the type of instrument being played,[158] which suggests that reverberation interacts with the perception of timbre. However, reverberation has also been shown to have a significant effect on the emotional perception of sustained instrumental tones regardless of instrument type,[159,160] indicating that reverberation and timbre may not have an interactive relationship.

Even fewer neuroimaging experiments have tested the relationship between reverberation and timbre. One electroencephalography (EEG) study found that reverberation severely degraded the auditory brainstem representation of formant-related harmonics required to distinguished vowel sounds.[53] The finding suggests that more degradation occurred because the reverberation was able to distort more acoustic information. However, these results may only be true of speech stimuli as opposed to music signals. Clarity ratings for pure tones in reverberation are typically lower than instrumental tones in the same conditions because there is less acoustic information for the listener to interpret. Consequently, it is possible the auditory response would be degraded further by reverberation on pure tones than instrumental tones. With this in mind, testing the interaction between reverberation and timbre can further explore the sensitivity of the auditory cortex by testing whether the clarity of the signals or merely the presence of incoherent acoustic content contribute to the degradation of the auditory cortex response.

The overall goal of the present work is to characterize the auditory response to coherent and incoherent acoustic information using fMRI. The exact objectives were to determine the auditory responses to differences in harmonic content (coherent acoustic information) and reverberation (incoherent acoustic information). Two musical motifs comprising a trumpet and sine-wave tones playing the same melodic passage were simulated in various room acoustic conditions with reverberation times (RT) from 0.0 to 7.2 s. Subjects were first asked to rate the stimuli based on perceived preference, reverberance, and clarity before listening to a subset of the stimuli in an MRI scanner. Examining the differences between the brain's functional responses to the musical stimuli was expected to reveal larger auditory activations in the STG and STS for the trumpet motif, which contained more coherent acoustic information than pure sine tones. Conversely, in accordance with the previous study by the authors, it was hypothesized that as reverberation time increased, the response of the auditory cortex would decrease. The last objective was to determine how the auditory response to coherent harmonic structure interacted with the response to incoherent reverberation to assess whether the response to reverberation would degrade further when harmonic content was present.

5.2 MATERIALS AND METHODS

5.2.1 Experimental procedure

The fMRI experiment consisted of two sessions on two separate days. In the first session, the subjects were asked to listen to each of the stimuli and rate them on each of the concert hall attributes of interest: overall preference, reverberance, and clarity. To prepare them for the second testing session in the MRI scanner, the subjects were placed in a Psychology Software Tools, Inc. MRI simulator (see Fig. B.1a) after rating the stimuli. The simulation aimed to familiarize the participants with the loud and confining MRI environment and the experimental protocol. In the second, MRI session, the participants were presented the stimuli while in a Siemens Prisma Fit 3T MRI scanner (see Fig. B.1b).

5.2.2 Subjects

Sixteen normal, healthy musicians (11 male and 5 female, 19-40 years, mean 27 years) participated in the experiment. Each participant had at least five years of formal musical training (i.e. private lessons from a professional instructor) and was currently active in a musical ensemble. A hearing screening was conducted to ensure that each of the subjects had a minimum hearing threshold of 15 dB HL, which has been shown to correlate with a subject's reliability.[106] The participants were also screened for right-handedness using the Edinburgh Handedness Inventory[135] to avoid the confounder of hand dominance. Since preference was one of the primary constructs being tested, the subjects were screened for depression with the Major Depression Index.[136] An additional 2 subjects were unable to participate in the experiment based on the screening tests.

5.2.3 Stimuli

The main objective of the current study was to examine the auditory response to coherent and incoherent acoustic information. Two musical signals were used to evaluate the auditory sensitivity to coherent, harmonic frequency components. The first signal (see spectrogram in Fig. 5.1A), the Trumpet Polka from the Bang & Olufsen Archimedes CD, comprised a solo-instrumental motif with 33 beats and 110 notes played at 120 BPM over 16.5 seconds.[118] The second musical signal, which will be denoted "sine-trumpet", was the same melody as the trumpet, but with the harmonic content removed. In order to test for the effects of reverberation on the auditory response of the brain, both musical motifs were simulated in eight different room acoustic conditions that perceptually varied from low to high reverberance.

5.2.3.1 Generation of sine-trumpet stimulus

The primary aim of the creation of the sine-trumpet motif (Fig. 5.1B) was to preserve the melody, rhythm, and instrumentality of the trumpet motif and only remove the harmonic frequency content. To determine the fundamental frequencies of the notes in the trumpet motif, peaks in the frequency content over time were found using a peak detection algorithm. Since the notes of the musical passage were all below 1000 Hz, the recording was first low-pass filtered at that frequency to ensure that the detected peaks would belong to the fundamental frequencies. The algorithm then found the start and end times of each note based on changes in peak frequency. Using the determined frequencies and start/end times, sine waves were generated that matched the melody of the trumpet passage.

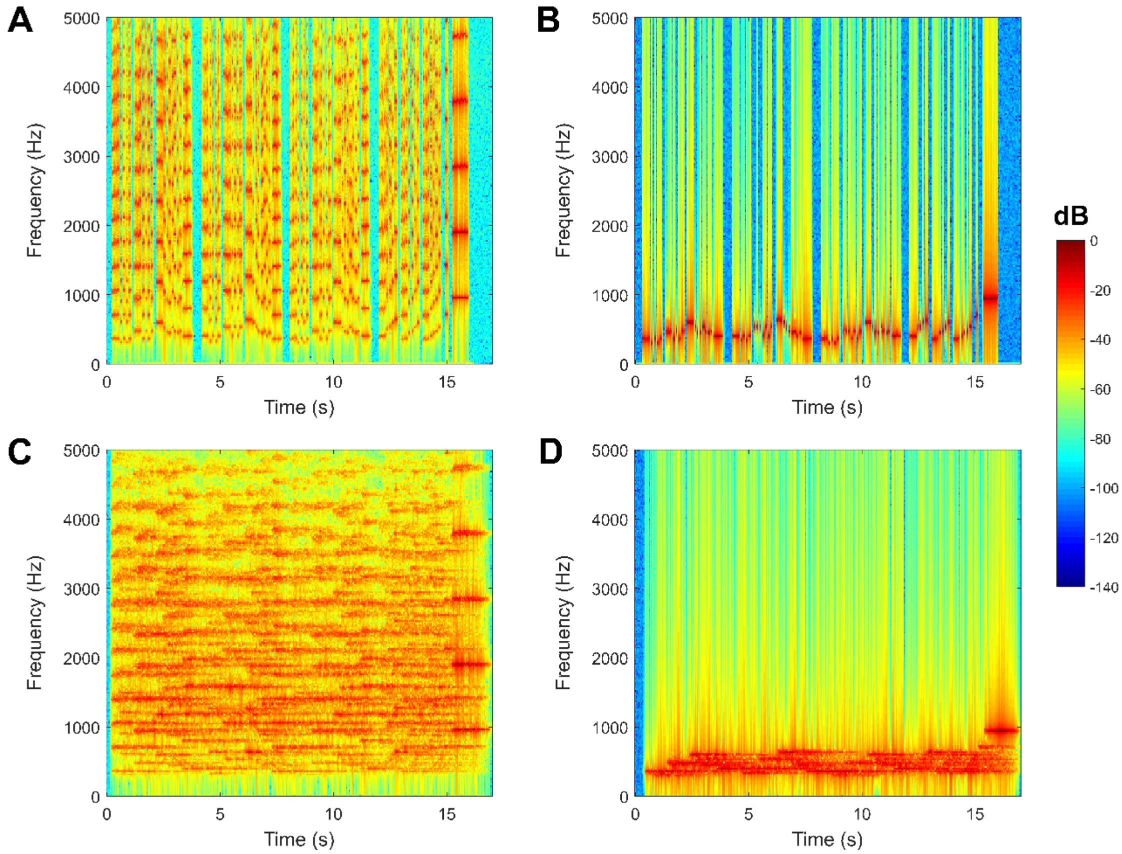


Fig. 5.1: Spectrograms from 0 to 5 kHz of the auralizations of the (A) anechoic trumpet, (B) anechoic sine-trumpet, (C) trumpet at T_{30} = 7.2 s, and (D) sine-trumpet at T_{30} = 7.2 s. The amplitudes of each spectrogram are normalized to the peak amplitude across all stimuli.

Next, the rhythm and instrumentality of the trumpet motif was applied to the generated sine waves. The amplitude envelope of the trumpet recording was obtained using a low-pass filter. By applying the amplitude envelope to the sine waves, the attack, sustain, and decay of each note were maintained in the sine-trumpet motif, providing both a sense of rhythm and maintaining the timbral perceptions that do not rely on harmonic content. Refer to Appendix A.3 for more information on the construction of the sine-trumpet stimuli.

As intended, this process removed coherent harmonic content. At any point in the presentation of the original trumpet excerpt, the listener received sound in which the fundamental frequency and harmonics arrived at the same time and from the same location. However, for the sine-trumpet musical signal, only the fundamental frequency of the melody was included. It was hypothesized that the inclusion of coherent harmonic content would increase the activations in the auditory cortex.

5.2.3.2 *Room acoustic auralizations*

The auralizations of both musical motifs were created using eight room acoustic models with varying room acoustic conditions. The first model for the anechoic room condition (reverberation time, T_{30} = 0.0s) was generated in a custom MATLAB script that provided an “out-of-head” auralization by attenuating the direct sound according to the spherical spreading of sound from a point source and the absorptive properties of air. The seven remaining models were created in Odeon v12.0. The material absorption and scattering properties of the Boston Symphony Hall model were artificially altered to establish conditions with T_{30} ranging from 1.0 to 7.2 s. Across all eight of the room acoustic conditions, the source and receiver locations remained constant to account for the timing and direction of arrival of the direct sound and early reflections. A head-related transfer function (HRTF) from Subject 21 of the CIPIC database[116] was used to create a binaural room impulse response (BRIR). The eight BRIRs were convolved with the anechoic motifs to produce 16 total auralizations. The loudness of all of the stimuli were equalized based on A-weighted sound pressure levels. A full description of the anechoic and Odeon room models may be found in Chapter 3.

Fig. 5.1C and D show spectrograms of the trumpet and sine-trumpet motifs auralized in the most reverberant condition (T_{30} = 7.2 s), illustrating the interaction effect of reverberation and timbre on the spectral content of the stimuli. As the room effects are added to each motif, the auditory stimuli become more complex over time. The reverberation causes any particular note to be less distinctive and clear by spatially introducing incoherent auditory objects and temporally increasing the length of each note.

5.2.4 Session 1: ratings of overall preference, reverberance, and clarity

In the first session, the participants were asked to rate their perceptions of overall preference, reverberance, and clarity for each of the 16 stimuli. The perceptual data was collected to evaluate whether the subjective ratings or objective room acoustics metrics (such as T_{30} or clarity index, C_{80}) correlated more with the sensitivity of the auditory cortex. To prevent bias in the preference ratings, the subjects were prompted to rate preference first, followed by reverberance and clarity in a balanced, random order.

Each of the concert hall attributes were rated on a five-point, continuous Likert scale.[6] The scale for preference ranged from -2 to +2, "Strongly Dislike" to "Strongly Like." For both the reverberance and clarity scales, the subjects rated the extent of the perceived attribute in the stimuli from being "Not at All" present (0%) to "Moderately" present (50%) to "Completely" present (100%).

A comparative listening test method was employed to gather the subjective ratings. In the comparative method, multiple stimuli are presented on the screen at the same time (see Fig. 3.4) in an effort to promote direct comparisons between each stimulus. The participants may choose when to listen to each stimulus and may switch between them as many times as needed. For a given set in the current study, all eight auralizations of one particular motif were presented simultaneously, which allowed the participants to instantaneously switch between the different room acoustic conditions while the musical passage played continuously. The stimuli in the first session were presented over STAX MR 003-MK2, MR-compatible electrostatic earbuds (Fig. B.2).

5.2.5 Statistical analysis on room acoustic perceptual data

A multivariate analysis of variance (MANOVA)[31] was conducted in SAS 9.4 to evaluate the preference, reverberance, and clarity ratings. The general linear model (GLM) contained two treatments, T_{30} and Motif, representing the eight room acoustic conditions and two musical motifs, respectively. The interaction effect between T_{30} and Motif was also examined to determine if the participants' perceptions were independent of motif type. To interpret the results, three univariate repeated measures analysis of variance (ANOVA) tests were performed to evaluate the ratings for preference, reverberance, and clarity, separately.

5.2.6 Session 2: presentation of stimuli in the MRI

Due to the need for several repetitions of the stimuli and time constraints, the second session in the MRI scanner only included a subset of the 16 auralizations. Four room acoustics conditions, T_{30} = 0.0, 1.2, 2.1, and 7.2 s, were chosen to present a wide range of reverberant conditions to the participants, giving eight total stimuli (4 room acoustic conditions x 2 motifs) for the MRI experiment. Each stimulus was presented nine times over the course of the entire session to improve the MR signal-to-noise ratio.

The stimuli were presented one at a time in a pseudo-randomized order such that two consecutive auralizations were neither the same motif nor the same room acoustic condition. The inter-stimulus interval varied between 8 and 12 seconds, averaging at 10 seconds between stimuli. Sensimetric Model S14 earbuds (see Fig. B.5) were used to present the stimuli, which were pre-filtered with the equalization filter of the headphones. Resonance Technologies Inc. VisuaStim over-ear headphones (Fig. B.6) provided sound attenuation. During each presentation, the subjects recorded the tempo of each stimulus using the right-thumb button on a pair of Nordic Neurolab Response Grips (see Fig. B.3). The tempo data was collected to ensure that the participants were actively listening to the auralizations.

5.2.7 fMRI data acquisition

Functional and structural MRI data were obtained with a 20-ch head coil on a Siemens 3T Prisma Fit scanner. For the fMRI data, five runs of an echo-planar imaging (EPI) sequence collected a total of 1080 volumes with 33 axial slices and a slice thickness of 3.0 mm. This sequence had a repetition time (TR) of 2000 ms, echo time (TE) of 28 ms, field-of-view (FOV) of 210 mm, matrix size of 64, and flip angle of 78°. The length of all five functional runs was 36 minutes and 20 seconds, and allowed for nine repetitions of each stimulus. During each run, a Siemens Respiration Belt simultaneously acquired each subject's respiration via a work-in-progress (WIP) EPI sequence. After the third functional run, a T1-weighted MPRage sequence (TR = 2300 ms, TE = 2.32 ms, FOV = 240 mm, matrix size = 256, flip angle = 8°) collected structural data. The structural scan contained 192 sagittal slices with an isotropic voxel size of 0.9 mm. The total length of the entire testing protocol was approximately 45 minutes.

5.2.8 fMRI image processing and statistical analysis

The processing of the functional and structural MRI scans was completed with SPM8 in MATLAB.[62] First, the functional scans were realigned with the first acquired volume, and the amount of realignment along the x-, y-, and z- directions, as well as the pitch, roll, and yaw of rotation, was examined to ensure that each subject had not moved more than 1.5 mm in translation or 1° in rotation within an EPI run. If a subject's movement was in excess of either or both of these criteria, all the scans from that particular set were discarded from the analysis. Slice-timing acquisition differences were then corrected with a sinc-interpolation routine, followed by the coregistration of the structural image to the mean of the functional images. Both the functional and structural data were spatially normalized to the MNI standard template and resampled with 3-mm and 1-mm isotropic voxel sizes, respectively. For the last pre-processing step, the fMRI images were smoothed spatially with a 8-mm full-width at half maximum (FWHM) Gaussian kernel filter.

The Physio Toolbox in MATLAB[138] was used to processed the respiration data. The toolbox removed clipped samples, normalized the data, and produced eight RETROICOR (RETROspective Image CORrection)[139] regressors. Both the RETROICOR and the six realignment vectors were used as nuisance regressors to diminish the noise in the statistical analysis.

In SPM8, a classical GLM mass-univariate model was constructed to analyze each participant's functional data at the first, individual level. Similar to the analysis of the perceptual data, the factorial design included treatments for the main effects of T_{30} and Motif, as well as the interaction between these effects. The T_{30} treatment had four levels, while the Motif treatment contained two, resulting in eight separate conditions. The block events in these conditions were modeled with the canonical hemodynamic response function (HRF).

The present study also included the study of preference and reward. The caudate nucleus and the nucleus accumbens (NAcc) of the basal ganglia have been shown to activate with the anticipation and experience of a reward to musical stimuli.[46,49-51] A past study by the authors (see Chapter 4) demonstrated that a similar reward could be detected in response to stimuli with preferred room conditions in the basal ganglia. Since previous studies have shown that the hemodynamic response of the basal ganglia may be more transient,[141] the time-derivative of the canonical HRF[35] was also included in the model (see Fig. 1.5b). Within each modeled HRF, linear contrasts representing differences in the main and interaction effects were created between the conditions.

With a random-effects model, one-sample t tests evaluated the linear contrasts of all the participants for the second-level, group analysis. Both a whole-brain and region-of-interest (ROI) analysis is reported. For the whole-brain analysis, statistically significant activations (p < 0.05) at peak voxel locations were determined based on a family-wise error (FWE) correction for multiple comparisons. The ROI analysis on the unsmoothed functional data included areas in the primary and secondary auditory cortex, as well as the basal ganglia. The probabilistic cytoarchitectonic maps of the auditory cortical areas Te 1.0, 1.1, and 1.2, as well as Te 3,[42] were acquired from the SPM Anatomy library.[143] The Te 1.x regions have been associated with primary auditory function.[142] The basal ganglia probabilistic maps were obtained from Keuken et al. (2014),[144] a 7T MRI study of 30 subjects. The ROIs of both the auditory and reward areas were overlaid on the average structural images of the 16 subjects to check for viability (Fig. 5.2). The MarsBar v0.42 tookit[145] was used to perform the ROI analysis. Significant activations were determined after a Bonferroni correction to account for the multiple comparisons of the five ROIs.[31]

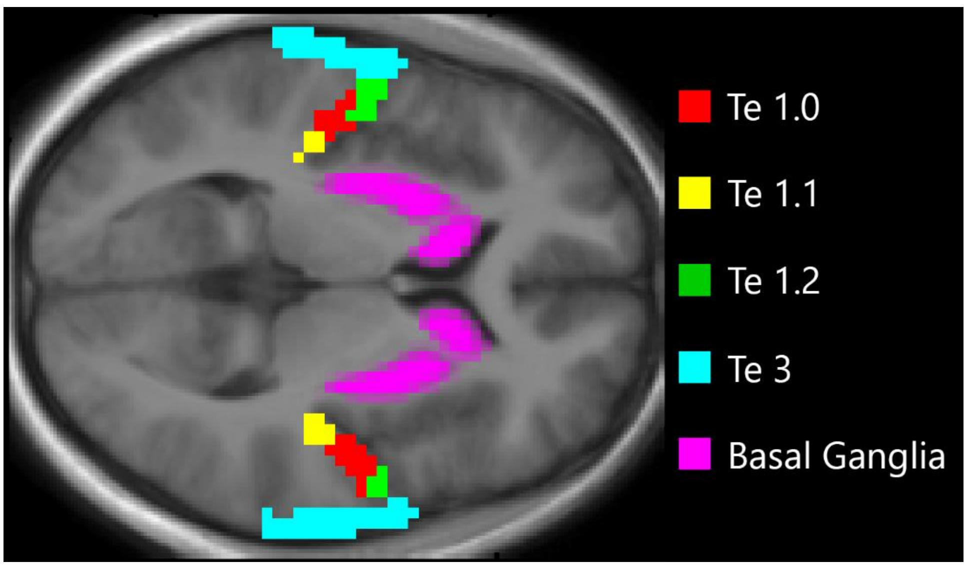


Fig. 5.2: The five ROIs overlaid on the average T1-weighted anatomical scan of 16 subjects at horizontal plane z = 4.

5.3 RESULTS

5.3.1 Behavioral results: preference, reverberance, and clarity ratings

The perceptual data were evaluated with a two-way repeated-measures MANOVA. The statistical analysis tested the effects of reverberation time (T_{30}) and motif on subjective ratings of preference, reverberance, and clarity. The residuals of the rating data of all three tests were visually inspected to ensure a multivariate normal distribution to satisfy one of the assumptions of ANOVA.

In the MANOVA using Wilk's lambda, there was a significant effect of T_{30} ($\Lambda = 0.13$, $F(14, 448) = 56.69$, $p < 0.0001$), Motif ($\Lambda = 0.75$, $F(2, 224) = 38.04$, $p < 0.0001$), and the interaction effect ($\Lambda = 0.87$, $F(14, 448) = 56.69$, $p = 0.004$). These results signify the ratings of at least one of the dependent variables differed for each one of these treatments. Three separate univariate ANOVA were conducted to interpret the results of the MANOVA.

For the subjective preference ratings, the main effect of (T_{30}) was statistically significant at $p < 0.0001$, meaning that the participants demonstrated different preference levels across the eight room acoustic conditions. Since the ANOVA did not show that the main effect of motif was significant ($p = 0.15$), the preference ratings between the trumpet and sine-trumpet motifs were not statistically different averaged over all of the room acoustic conditions. The interaction effect between motif and T_{30} had a significant p-value of 0.002, which shows that the two motifs were rated differently across T_{30}, as seen in Fig. 5.3. Specifically, the most-liked stimuli for the trumpet motif were the room conditions with T_{30} values of 1.7 and 2.1 s, while the most-disliked stimuli were the 0.0 and 7.2 s conditions. This outcome agrees with the preferences of the subjects in the previous studies by the authors (Chapters 3 and 4). For the sine-trumpet motif, the room condition with T_{30} = 1.0 s was the most liked and the one with T_{30} = 7.2 s was the most disliked.

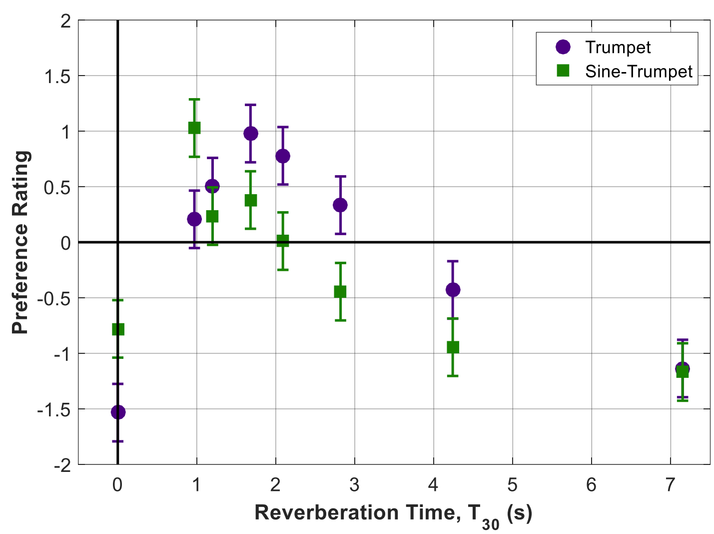


Fig. 5.3: Preference ratings as a function of reverberation time (T_{30}) for the trumpet and sine-trumpet motifs. For the trumpet motif, the T_{30} = 1.7 and 2.1 s stimuli were rated as the most-liked, while the 0.0 and 7.2 s stimuli were rated as the most-disliked. The most-liked and most-disliked sine-trumpet stimuli had T_{30} values of 1.0 and 7.2 s, respectively.

All of the main and interaction effects were significant when examining the responses of perceived reverberance ($p < 0.05$). For both motifs, as reverberation time increased, the ratings of perceived reverberance increased (see Fig. 5.4). The reverberance ratings of the sine-trumpet motif were higher than those of the trumpet motif ($p < 0.0001$) at a difference of 9 ± 2 (s.e.) percentage points. This difference was largely demonstrated for the stimuli in the typical range of concert halls with T_{30} = 1.0 – 2.1 s stimuli ($p < 0.05$). The stimuli with T_{30} values of 0.0, 2.8, 4.2, and 7.2 s received statistically comparable reverberance ratings between the two motifs ($p > 0.05$).

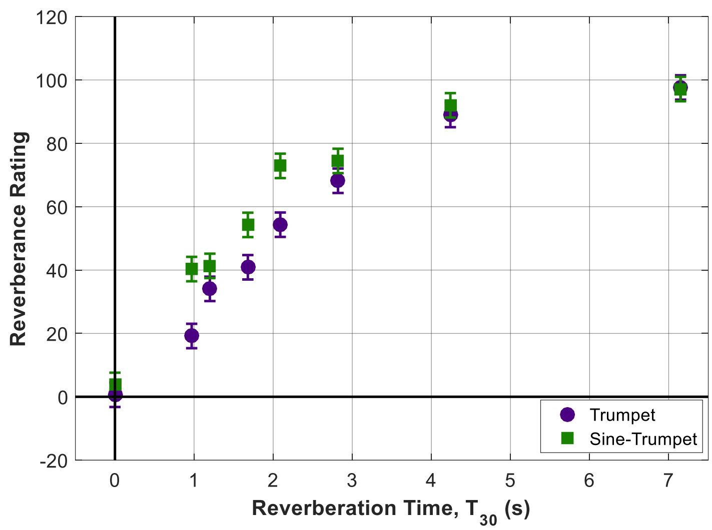


Fig. 5.4: Perceived reverberance ratings as a function of T_{30} for the trumpet and sine-trumpet motifs. Overall, the sine-trumpet motif was rated as more reverberant than the trumpet motif (p < 0.0001). For both motifs, perceived reverberance increased with reverberation time.

The perceived clarity ratings exhibited similar findings. Again, the two main effects of T_{30} and motif, as well as the interaction effect between them, were determined to be significant at $\alpha = 0.05$. As expected in relation to the reverberance ratings, the perceived clarity ratings decreased as T_{30} increased for both the trumpet and sine-trumpet motifs (Fig. 5.5). Also, in agreement with the reverberance ratings, the trumpet motif was rated clearer than the sine-trumpet motif by an average of 17 ± 2 (s.e.) percentage points ($p < 0.05$). This trend was true across all of the room acoustic conditions except for the anechoic setting, where the two motifs were rated to have the same clarity.

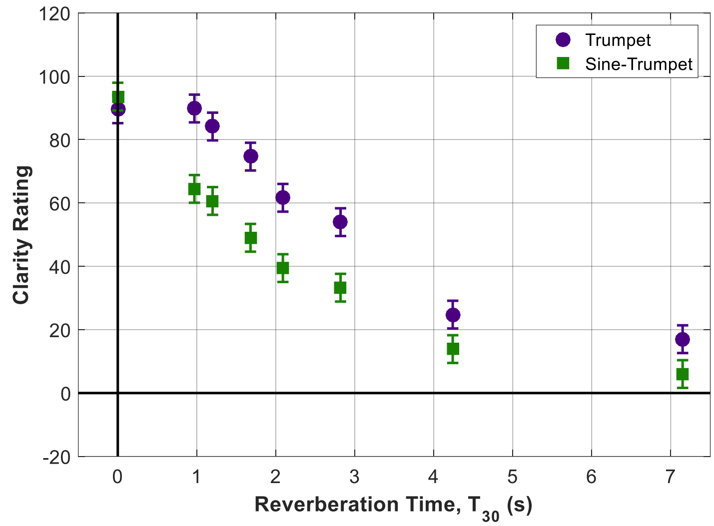


Fig. 5.5: Perceived clarity ratings as a function of T_{30} for the Trumpet and Sine-Trumpet motifs. Overall, the sine-trumpet motif was rated as less clear than the trumpet motif ($p < 0.0001$), which agrees with the perceived reverberance ratings. As reverberation time increased, perceived clarity decreased.

The reverberance rating of the most-liked sine-trumpet stimulus, T_{30} = 1.0 s, statistically matched the rating of the one of the most-liked trumpet stimuli, T_{30} = 1.7 s ($p > 0.05$). In much the same way, the clarity rating of the most-liked sine-trumpet stimulus equaled the clarity rating of the other most-liked trumpet stimulus, T_{30} = 2.1 s ($p > 0.05$). Furthermore, the most-disliked stimuli for both motifs (T_{30} = 0.0 and 7.2 s) were respectively rated similarly for both the perceptions of reverberance and clarity. These results indicate that the participants considered the perceived reverberance and clarity of the stimuli in their preferences ratings despite the fact that they were unaware that these two concert hall attributes would be tested while they were rating the stimuli for preference.

5.3.2 Functional results: whole-brain analysis

At the second-level, the linear contrasts of the main and interaction effects were evaluated with one-sample *t* tests. The peak activations (see Table 5.1) occurred primarily in the right superior temporal gyrus (STG) spanning the primary auditory cortex (PAC). These activations were only reported in contrasts of the main effects of interest.

Table 5.1: Results from the whole-brain group-level analysis. The significant peak activations ($p_{FWE} < 0.05$) are reported along with the number of voxels in the cluster and MNI coordinates. For all of the contrasts of main effects, the activations or deactivations were located in the right (R) superior temporal gyrus (STG). Other significant regions include the left (L) Rolandic Operculum and R Heschl's gyrus (G).

Contrast	Region	Peak-Level		Voxels in Cluster	MNI Coordinates (mm)		
		p_{FWE}	T		x	y	z
T_{30} 1.2 > T_{30} 0.0	**R STG**	**< 0.001**	**-10.90**	**377**	**54**	**-25**	**4**
T_{30} 2.1 > T_{30} 0.0	**R STG**	**0.002**	**-9.45**	**702**	**57**	**-22**	**4**
T_{30} 7.2 > T_{30} 0.0	**R STG**	**0.03**	**-7.71**	**263**	**60**	**-25**	**7**
	R Heschl's G	0.05	-7.10		42	-19	7
	L Rolandic Oper.	**0.05**	**-7.19**	**199**	**-36**	**-31**	**16**
Trumpet > SineTrumpet	**R STG**	**0.05**	**7.08**	**254**	**48**	**-28**	**10**

Out of the six possible linear contrasts for the main effect of T_{30}, three contrasts yielded significant findings at the second-level. When contrasted with the anechoic room condition, each of the reverberant room conditions exhibited a deactivation in the right STG. The negative *T* denote that the auditory cortex responded more for the anechoic stimuli than for the reverberant stimuli. The activation maps of these contrasts (Fig. 5.6) show that these deactivations covered the majority of the right PAC for all three comparisons. None of the contrasts between two reverberant room conditions (T_{30} = 1.2, 2.1, or 7.2 s) demonstrated any significant activations at the whole-brain level.

The main effect of motif showed significant activations in the PAC. The contrast between the trumpet and sine-trumpet motif (see activation map in Fig. 5.7) displayed peak activations in the right STG. While the region that exhibited the activation for the motif contrast was similar to that implicated in the T_{30} contrasts, the extent of the activations were opposite. In fact, the auditory cortex experienced a larger response to the trumpet stimuli than the sine-trumpet stimuli.

There were six linear contrasts to test the interaction effect between T_{30} and motif. However, none of these contrasts had any significant activations over the whole brain.

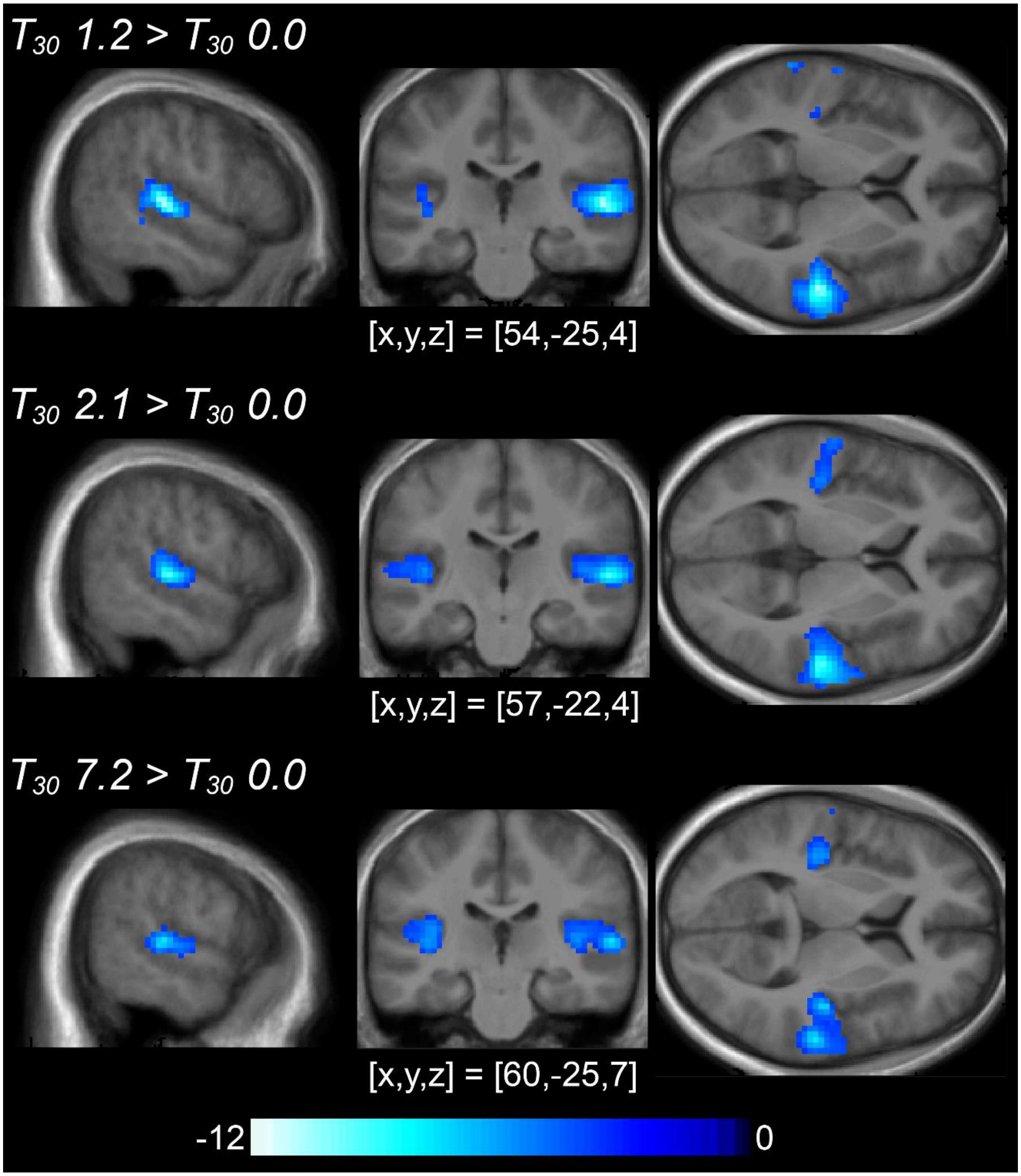


Fig. 5.6: (*From Top to Bottom*) Activation maps for the *T_{30} 1.2 > T_{30} 0.0*, *T_{30} 2.1 > T_{30} 0.0*, and *T_{30} 7.2 > T_{30} 0.0* contrasts superimposed on the average T1-weighted anatomical image of the 16 subjects. The three activations maps were constructed based on an uncorrected threshold of $p < 0.001$ with a peak-vowel threshold of $p_{FWE} < 0.05$. The colorbar indicates the *T* value associated with each voxel. While not significant at the whole-brain level, the clusters in the left STG encompasses regions that were significant in the ROI analysis. Activations of the right STG were significant in both the whole-brain and ROI analyses.

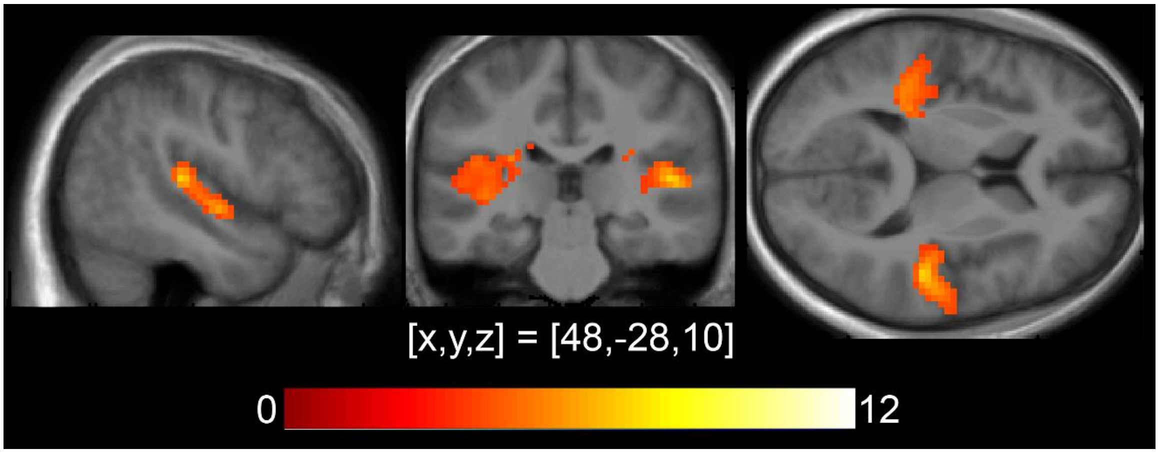


Fig. 5.7: Activation map for the *Trumpet* > *Sine-Trumpet* contrast superimposed on the average T1-weighted anatomical image of the 16 subjects. The map was determined based on uncorrected threshold of $p < 0.001$ and peak-voxel threshold at $p < 0.05$, corrected. The colorbar indicates the *T* value associated with each voxel. While not significant at the whole-brain level, the cluster in the left STG encompasses regions that were significant in the ROI analysis. Activations of the right STG were significant in both the whole-brain and ROI analyses.

5.3.3 Region of Interest (ROI) results

Probabilistic cytoarchitectonic maps of the left and right auditory areas Te 1.0, 1.1, 1.2, and 3 were examined for the ROI analysis. Since these areas lie next to each other, the analysis only included unsmoothed functional data. The fifth ROI was the dorsal and ventral striata of the basal ganglia, which was also evaluated with unsmooth data. The significance levels reported for each ROI and each contrast were corrected for the multiple comparisons of the five ROIs.

The ROI analysis of the auditory regions agreed with the findings of the whole-brain analysis. In particular, both of the main effects of T_{30} and motif showed significant results, while the interaction effect was not significant. As in the whole-brain analysis, the auditory cortex activated more strongly for the anechoic stimuli than for the reverberant stimuli (see Table 5.2). Area Te 1.1 was significantly deactivated for all three contrasts with the anechoic stimulus. Both areas Te 1.0 and 3 also yielded significant deactivations for the contrasts with T_{30} = 2.1 and 7.2 s stimuli. Also in agreement with the whole-brain analysis, the ROI analysis showed that the auditory cortex experienced a stronger response for the trumpet than the sine-trumpet motif. Areas Te 1.0, 1.1, and 1.2 were all positively activated for this contrast, while area Te 3 did not exhibit significant activations.

Table 5.2: Results of the ROI analysis for each of the main effect contrasts of T_{30} and motif. Activations were considered significant at $p < 0.05$, corrected for multiple comparisons.

ROI	T_{30} 1.2 > T_{30} 0.0		T_{30} 2.1 > T_{30} 0.0		T_{30} 7.2 > T_{30} 0.0		Trumpet > SineTrumpet	
	p_{FWE}	T	p_{FWE}	T	p_{FWE}	T	p_{FWE}	T
Te 1.0	-	-	0.004	-4.25	0.006	-3.95	0.04	3.08
Te 1.1	< 0.001	-4.94	< 0.001	-5.56	< 0.001	-6.44	< 0.001	5.31
Te 1.2	-	-	-	-	-	-	0.04	3.08
Te 3	-	-	0.01	-3.65	0.04	-3.09	-	-

The left and right auditory cortices were also considered as separate ROIs to determine if the auditory response was lateralized to one hemisphere. A paired *t*-test evaluated the difference between the mean contrast values of the left and right ROIs of Te 1.0, 1.1, 1.2, and 3. None of the auditory regions exhibited lateralization in that the activations of the left and right cortical areas were not statistically different. Furthermore, the left and right ROIs were separately significantly activated after correction for eight multiple comparisons (Fig. 5.6 and Fig. 5.7).

The ROI analysis for the basal ganglia ROI was conducted on both the canonical HRF and time-derivative HRF models. It was expected that basal ganglia would activate for the more preferred stimuli in the time-derivative model. However, neither model demonstrated any significant activations in the basal ganglia.

5.4 DISCUSSION

Continuing from the authors' previous work (Chapter 4), the current study further examined the sensitivity of the auditory cortex to acoustic information present in auralizations of musical passages in reverberant room conditions. The past study demonstrated that reverberant room effects reduce the activation of the PAC compared to an anechoic stimulus; however, each subject was presented an individualized stimulus set, and no activation differences were found between reverberant stimuli. In the current MRI experiment, each subject listened to each reverberant condition to determine if the amount of reverberation affected the degradation of the auditory response.

Based on both the whole-brain and ROI analyses, including reverberation decreases the response of the auditory cortex. Three of the contrasts of interest for the main effect of T_{30}

reported significant deactivations with reverberant stimuli. However, these contrasts all involved the anechoic stimuli, and no contrast between two reverberant conditions exhibited any significant deactivation. These findings suggest that simply the presence of incoherent, spatial auditory content degrades the response of the auditory cortex, and the level of the incoherent content is inconsequential.

The reverberation of a room is composed of two parts. Early reflections add source images around the receiver, which act as incoherent auditory objects if the phase lag is beyond the influence of the Precedence Effect (50-80 ms).[146] Late reverberation subsequently inserts incoherent noise around the listener. The perception of reverberance increases when the level of both the early reflections and late reverberation increases. Although T_{30} is a time-based metric, the different room conditions were created by adjusting the level of the reverberation. Increasing the level of the incoherent sources and stochastic reverberation did not further the degradation of the auditory response as seen by the insignificant difference in auditory response between any two reverberant stimuli, even between auralizations with starkly different reverberation times, T_{30} 1.2 and T_{30} 7.2. However, prior research has demonstrated that the number of auditory objects negatively correlates with the extent of the response in the auditory cortex.[131] Therefore, the auditory cortex may be more sensitive to the number of early reflections than the level of the incoherent objects or noise.

The present study also showed with fMRI methods that adding harmonic content, i.e. coherent acoustic information, increases the response in the PAC. This result agrees with previous work that found that including timbral frequency content heightened the MEG auditory cortical representation of tones.[157] The ROI analysis specifically showed that timbre affected the response of the primary auditory areas and not the gross secondary area of Te 3. The current study applied the attack and decay of the trumpet to the pure-sine melody, maintaining the temporal timbre cues across motifs. Therefore, the differences between the two motifs focused chiefly on the spectral timbre cues, which may explain why the activations occurred in the PAC.

Another possible reason that the auditory cortex responded stronger for the trumpet than the sine-trumpet could be because of the perceived clarity of the stimuli. The sine-trumpet motif was rated as more reverberant by 9 ± 2 (s.e.) reverberance rating points ($p < 0.0001$), and the trumpet motif was rated as more clear by 17 ± 2 (s.e.) clarity rating points ($p < 0.0001$). However, upon examination of the clarity ratings and contrasts of the reverberant stimuli, it becomes apparent that the auditory cortex is not responding to the clarity of the motifs. The differences in clarity ratings for the T_{30} = 0.0 s and T_{30} = 1.2 s stimuli were 33 ± 6 (s.e., $p < 0.0001$) and 5 ± 6 (s.e., $p > 0.05$) points for the sine-trumpet and trumpet motifs, respectively. If the brain was responding according to the perceived clarity of the stimuli, one would expect to see significant activations for the $T_{30}\ 1.2 > T_{30}\ 0.0$ contrast for the sine-trumpet motif, which demonstrated large clarity differences, and not for the trumpet motif. Although there was no significant interaction effect, when examined separately, the trumpet $T_{30}\ 1.2 > T_{30}\ 0.0$ contrast exhibited significant activations in the left STG (p_{FWE} = 0.01, T = -8.17), while the same contrast for the sine-trumpet motif did not have activations above threshold. Thus, the perceived clarity differences between the two motifs does not explain the auditory response differences, and the reason that the two motifs evoked different responses in the auditory cortex most likely agrees with the initial hypothesis that the addition of coherent frequency content enhances the response.

Some auditory perception studies of timbre have reported lateralization of timbre processing,[152,154-156] while others have not.[151,153] While the whole-brain analysis primarily implicated the right auditory cortex, the ROI analysis did not discriminate between left or right sides of the brain, meaning that no lateralized processing transpired. This finding suggests that the processing of musical spectral content may occur bilaterally, and perhaps the auditory response is lateralized for some other aspect of timbre, such as temporal cues.

Contrary to the initial hypothesis, the addition of more harmonic information did not conflate the sensitivity of the auditory cortex to incoherent reverberation. There was no significant interaction effect between reverberation and harmonic content, which shows that the two effects may be independent of one another. Independent of motif, adding reverberation, i.e. adding incoherent auditory objects, degraded the auditory cortical response. Conversely, adding coherent frequency content (harmonic content) universally increased the response of the auditory cortex to the musical stimuli. This result does not necessarily contend with the findings of Bidelman & Krishnan (2010).[53] The ability to distinguish between speech signals inherently requires interpreting formants, which can be severely temporally distorted by reverberation. The interpretation of the harmonics of speech stimuli may be a primary process, but may become a more associative process for music stimuli, which may not be as affected by reverberation.

A secondary goal of the present work was to determine if the preferred stimuli could evoke a reward response. The participants reported significant differences in their preferences to the stimuli included in the MRI study, and activations in the basal ganglia in the temporal derivative model were expected based on the results of the previous study by the authors (Chapter 4). However, no such activations were significant in either the whole-brain or the ROI analysis. The difference between the two studies is that the stimulus sets in the present study were not individualized to each particular subject. We suspect that the participants who did not strictly prefer the 1.2 or 2.1 s stimuli were unable to experience a reward response. With the inclusion of more participants, a significant reward response may emerge, but probably not for the sine-trumpet stimulus, since the preference ratings for these two room acoustic conditions for this motif were close to zero.

5.5 CONCLUSION

Processing sound involves neural mechanisms that have yet to be fully studied in the human auditory cortex.[41] The present study sheds some light on the auditory cortical response to complex acoustic phenonmena, namely timbre and reverberation, by considering them as coherent and incoherent acoustic information, respectively. Previous research has indicated that the extent of the auditory response increases with coherency.[130,131,134,157] The present work agrees with this conclusion, demonstrating that adding coherent harmonic content bolstered the auditory response and including incoherent reverberation caused a degradation of said response. Though unexpected, inserting coherent acoustic information did not influence the effect of the incoherent content. Further research is needed to confirm the impact of the number and timing of early reflections on the sensitivity of the auditory response.

5.6 ACKNOWLEDGMENTS

We would like to thank the Social, Life, and Engineering Sciences Imaging Center (SLEIC) at Penn State for the use of the facilities. We would like to express special thanks to Dr.

Thomas Beck and Dr. Josef Pfeuffer, Siemens Healthcare, Erlangen for providing us with the Advanced fMRI WIP package.

CHAPTER 6

GRAND CONCLUSION

6.1 RESULTS SUMMARY

The main objective of the current work was to use fMRI to assess how room acoustics affects the neural response to music. Four studies aimed to achieve this goal. The first study involved a pilot experiment of five participants to see if fMRI was an appropriate tool to investigate the reward response to room conditions. Based on the results of the first study, the second study was a conventional listening test to establish a set of reward-eliciting stimuli. This study also compared two listening test methods and participant expertise. The third study returned to the MRI scanner and examined both the cortical auditory response and mesolimbic reward response to room acoustics. The final study focused on the cortical auditory response to coherent and incoherent acoustic information by investigating the relationship between coherent harmonic content and incoherent early reflections and late reverberation.

In the pilot study (Chapter 2), five participants listened to and rated five reverberant musical stimuli while in an MRI machine. While the subjective ratings demonstrated that the participants on average preferred two moderately reverberant room conditions and disliked the anechoic and extremely reverberant room conditions, only two of the participants exhibited a reward response, as shown by activations in the caudate nucleus. The detection of a reward response in these participants, as well as group-level analysis, provided evidence that fMRI could measure the subtle differences between room acoustic conditions. However, the absence of reward response in the other participants either presented counter-evidence or suggested that the chosen stimuli did not elicit powerful preferences. Future neuroimaging work would require a careful selection of room conditions and musical motifs in order to confirm the reward response fMRI measurement.

The primary objective of the second study (Chapter 3) in the context of this dissertation was to create such a set of rewarding room acoustics conditions by prompting subjects to rate a large set of stimuli consisting of eight simulated rooms and four musical motifs and eight simulated rooms with T_{30} ranging from 0.0 s (anechoic) to 7.2 s. Because many subjects participated and the rating range for each room conditions was large, the subjective preference data was scrutinized for specific trends using a k-means clustering algorithm. The clustering analysis found that the subjects could be categorized into five distinct preference groups. One set of room acoustics stimuli would most likely not be able to elicit a reward response in all of the participants. Instead, the MRI stimulus sets of room acoustics stimuli need to conform to individual room preferences.

The second study also examined differences between conventional listening test methods. The comparison of the listening test methods aimed to determine if the results between listening tests inside and outside of the MRI scanner, the successive and comparative methods, respectively, were analogous in terms of the ratings results and reliability. The experiment demonstrated that both methods yielded statistically similar preference results and reliability, though the participants preferred and performed quicker on the comparative method. This outcome agreed that the comparative method was a better listening test with a past comparison of multiple listening tests on the preferences of in-car ventilation noise.[30]

The final objective of the conventional listening test study was to determine if inexperienced listeners could display similar preferences as experienced listeners. For concert hall acoustics, musicians with at least five years of formal musical training and currently active in lessons or an ensemble were considered to have extensive critical listening experiences. Conversely, the non-experienced listeners were non-musicians with less than two years of musical training. The comparisons between the two groups showed that musicians were better than non-musicians at distinguishing between their preferences to the room acoustics stimuli, and produced more reliable results. This result lends to the fact that these room acoustics stimuli exhibit subtle differences to the untrained ear despite tremendously large disparity in the amount of reverberation. Given that non-musicians in general were unable to self-report their preferences to the room conditions, it was unclear if the fMRI would be able to detect neural response differences, especially since non-musicians have diminished responses to auditory stimuli compared to musicians. [126,127] Therefore, only musicians were recruited as participants in the MRI experiments. However, since some musicians could not discriminate between their preferences akin to the non-musicians, these participants would need to be pre-screened.

Following the conventional listening test, the third study (Chapter 4) tested the individualized room acoustic preferences sets for 18 participants in the MRI. Each of the participants preferred a moderate amount of reverberation (T_{30} between 1.0 and 2.8 s) and disliked the anechoic and extremely reverberant (T_{30} = 7.2 s) conditions. On the group level, the auditory cortex exhibited a sensitivity to the complexity of the musical stimuli. In agreement with a previous study on auditory objects,[131] a solo-instrumental motif prompted a stronger response in the auditory cortex than an orchestral motif with multiple auditory objects, although the musical melody and structure of the two passages differed. Similarly, the addition of reverberation also degraded the response of the auditory cortex. However, it was unclear if this sensitivity was due to the presence of early reflections forming auditory objects[130] or the late reverberation introducing incoherent noise in the signals, as shown in other studies.[134]

The third experiment also demonstrated a reward response to pleasurable room acoustics stimuli. Contrasts between the most-liked and the most-disliked stimuli exhibited activations in the caudate nucleus and NAcc. These activations were only found in a ROI analysis of the basal ganglia using a temporal derivative model as the basis for the HRF. This model was used to account for potential differences in the hemodynamic response of the reward network.[141] The reward result in the third study indicates that while fMRI can sense a reward response to subtle differences in room acoustic stimuli, the response itself is smaller and more transient than had been found in previous studies comparing distinct musical motifs.

The last study (Chapter 5) aimed to investigate if a correlation exists between the auditory cortex response and the amount of reverberation as well as examine the cortical auditory response to musical timbre, specifically harmonic content. In regards to the previous study, the reverberation and harmonic content were considered incoherent and coherent acoustic information, respectively. While in the MRI, 16 participants were presented two motifs, one with harmonic content (trumpet) and one without (sine-trumpet), simulated in four room conditions ranging from anechoic to extremely reverberant. Because the final study was not concerned with preference, all of the participants received the same room conditions.

As before, the presence of reverberation caused the auditory response to decrease, but higher levels of reverberation did not necessarily induce a stronger degradation. Since the room conditions were created by adjusting the levels of the early reflections and late sound, not the number or direction of the early reflections, the sensitivity of the auditory cortex could be explained by the addition of incoherent auditory objects in the form of early reflections. While the level of the incoherent noise of the reverberant tail could influence the response of the auditory cortex, it is most likely not the dominant effect.

Conversely, the presence of harmonic content bolstered the response of the auditory cortex as hypothesized. For each of the room conditions, the trumpet motif elicited a stronger response in the auditory cortex than the sine-trumpet motif. Additionally, no interaction effect between motif and room condition was observed, which suggests that the cortical auditory responses to coherent and incoherent acoustic information are independent of one another.

In summary, the contributions and impact of this dissertation are as follows:

1. The auditory cortex demonstrates a sensitivity to the presence of reverberation that is consistent with the degradation of auditory response caused by incoherent auditory objects and noise.
2. The auditory cortical responses to coherent and incoherent auditory information comprise two independent processes.
3. Despite the subtle differences between the same musical passages simulated in varying room conditions, the reward network exhibits activations for preferred room acoustic stimuli, which confirms the importance of room acoustics for the perception of music.

4. The comparison of the preference results of experienced and non-experienced listeners in conventional listening test methods for concert hall studies encourages the exclusion of non-experienced listeners based on the inability to discriminate preferences.
5. While the comparative and successive conventional listening test methods yield similar preference results with equivalent reliability, participants tend to prefer and perform faster on the comparative method in testing the overall preference of concert hall stimuli.

6.2 FUTURE WORK

The experiments in this dissertation present for the first time an investigation into how room acoustics affect the perception of music using fMRI. There is much more work that may be accomplished to further understand the role of the auditory cortex and reward network in the processing of spatial sound. The proposed future work includes: (1) further evaluation of the sensitivity of the auditory cortex to early reflections, and (2) an investigation of the effect of the timing and spatial distribution of early reflections on the auditory cortical response.

The results of the last two MRI studies in the present work provided the first steps in understanding the sensitivity of the auditory cortex to reverberation. While the presence of reverberation degrades the response of the auditory cortex, increasing the sound level of the early reflections and late reverberation does not contribute to additional deactivation. Therefore, the number of incoherence source objects in the form of early reflections arriving after 50-80 ms most likely facilitates the reduction in the auditory response. This effect may be tested explicitly in an fMRI experiment by presenting subjects with musical stimuli auralized in room conditions with varying amounts of early reflections that arrive from a range of spatial locations. In the simulation of these room conditions, the source-receiver distance should remain constant, as well as the level of the reflections and late reverberant tail. The BRIRs of these stimuli may be easily constructed using a program that places individual image sources to model the early reflections, while not necessarily needing to worry about the specific geometry of the room. The late reverberation in this case may be generated with a random Poisson process.[161] In this manner, the perception of reverberance would change according to the number of early reflections (which affects the EDT) and not the RT of the rooms.

Future work into the effect of early reflections may also focus on how the Precedence Effect manifests in the auditory cortex. The Precedence Effect, also known as the Haas Effect, defines the time needed between the arrivals of sound from two identical sources for the listener to perceive two distinct auditory objects.[146] Before this time point, the two identical sources are perceived as a single, fused source, providing impressions of loudness and spaciousness. The Precedence Effect also specifies that a listener may perceive a fused auditory event even if the level of the identical source is 10 dB louder as long as the wavefront arrives within 5 to 30 ms after the first-arriving sound. For speech signals, the time needed to perceive two distinct objects is considered to be around 50 ms, while for music, it is around 80 ms. (This disparity between the Haas Effect for speech and music is the reason that clarity index is calculated from these time points for each signal type, respectively.)

Although EEG experiments have studied the neural correlates to fused and non-fused images in the auditory cortex,[162] the Precedence Effect has yet to be characterized in the auditory brain using fMRI. Doing so may help explain the sensitivity of the auditory cortex to early reflections. Two experiments may be conducted that evaluating the timing and spatial distribution of early reflections, respectively. The first experiment may involve stimuli simulated in rooms with only the direct sound and one early reflection. Between stimuli, the distance from the image source to the receiver would change such that the reflection arrives from the same direction, but later in time. Air attenuation may also be applied to adjust for the level of the early reflections, but both constant level and varying level stimuli should be included in the experiment. The second experiment to test the spatial distribution of the Precedence Effect would control for the timing of the first reflection, but separate the sources in space at a fixed radius such that the image source to receiver distance was constant. Both experiments could use stimuli comprising clicks, pure-tones, and instrumental tones with short durations.

APPENDIX A

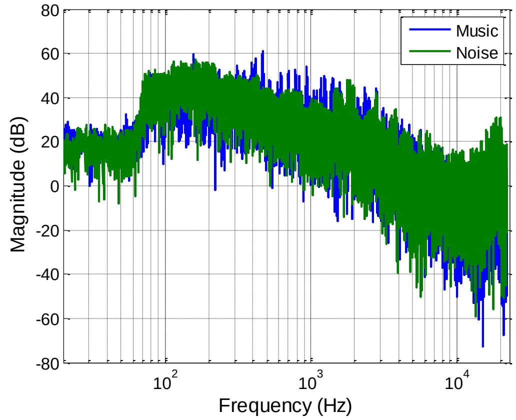
MATLAB Code for Generating fMRI Control Stimuli

A.1 RHYTHMIC NOISE CONTROL STIMULI

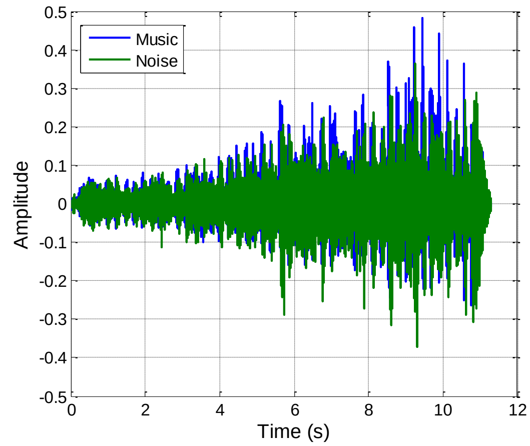
The purpose of the rhythmic noise control stimuli were to replicate the magnitude of the frequency content and rhythm of the music passages in a noise-like stimulus. The resulting auditory signals were white noise “notes,” and were described by the participants to sound like a snare drum.

```
%% Rhythmic Noise

clc
clear all
close all


%% Input the file name and load the file.
RT = 4.00;
rname = 'Auralizations/Bruckner - %.2f.wav';
rname = sprintf(rname,RT);
[aural,fs] = wavread(rname);
[rows columns] = size(aural);
if mod(rows,2) == 1
    aural(end,:) = [];
end

%% Compute necessary constants.
dt = 1/fs;
N = length(aural);
t = (0:N-1)*dt; t = t.';
T = N*dt;
df = 1/T;

%% Create white noise with a random number generator.
noise = zeros(N,1);
for ii = 1:length(aural)
    number = rand;
    if number > 0.5
        number = number-1;
    end
    noise(ii) = number;
end

%% Take the FFT for the noise time series.
noise_fft = fft_move(noise);

f = (0:N/2)*df;
f = f.';

%% (Optional) Create pink noise
```

```
pinknoise_fft = noise_fft./f;
pinknoise = ifft_half(pinknoise_fft);
pinknoise = real(pinknoise);

%% Take the FFT of the auralization and filter with noise.
auralfft1 = fft_move(aural(:,1));
auralfft2 = fft_move(aural(:,2));

noisefft1 = auralfft1*noise_fft;
noisefft2 = auralfft2*noise_fft;

%% Obtain the time series of the noise with inverse FFT.
noise(:,1) = ifft_move(noisefft1);
noise(:,2) = ifft_move(noisefft2);

%% The envelope (env_lp) of auralization is determined by low-pass
%% filtering the absolute value of the time series. The envelope is then
%% applied to the noise time series. The noise time series is then scaled
%% to match the root mean square of the music time series.
[b,a] = butter(4,10/(fs/2),'low');
% env_lp = filter(b,a,env);
env_lp = filter(b,a,abs(aural(:,1)));
noise = noise.*env_lp;

noisesum = sum(noise.^2);
auralsum = sum(aural(:,1).^2);
noise = sqrt(auralsum/noisesum)*noise;
noise_fft = fft_move(noise(:,1));

%% Display the results.
figure(2)
plot(t,aural(:,1),t,noise(:,1));
xlabel('Time (s)');
ylabel('Amplitude');


figure(1)
semilogx(f,db(abs(auralfft1)),f,db(abs(noise_fft(:,1))));
xlabel('Frequency (Hz)');
ylabel('Magnitude (dB)');
```



(a)



(b)

Fig. A.1: Example test and matched noise control stimulus in the frequency domain (a) and time domain (b) showing the similar frequency and rhythmic content.

A.2 SCRAMBLED MUSIC CONTROL STIMULI

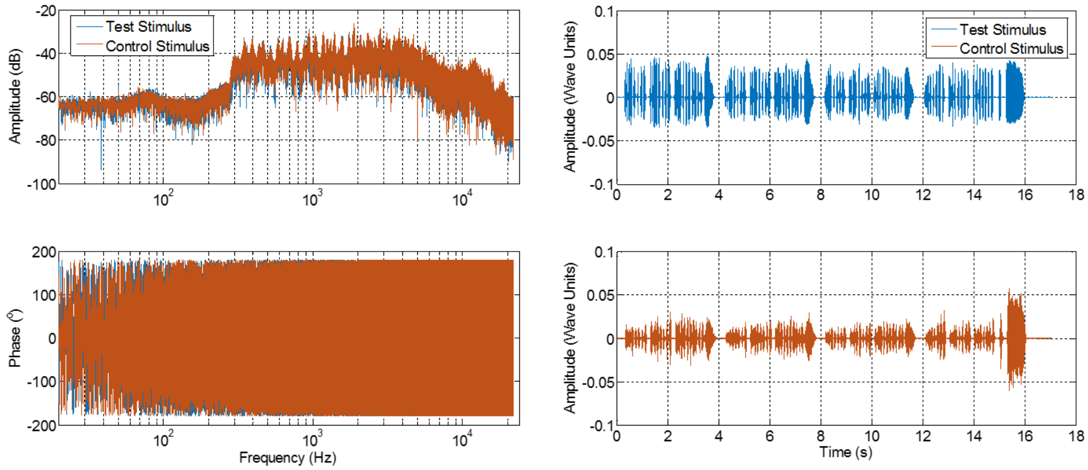
The purpose of the scrambled music control stimuli were to produce controls that maintained the same frequency and rhythmic content as the musical passages, while eliminating all semblance of musical structure. This objective was completed by randomizing the phase of the musical passages so that the resulting auditory signal sounded as if all notes were played concurrently.

```
%% Scrambled Music

clc
clear all
close all

%% Load the original files. There are 8 RT times and 3 different musical
%% motifs.
RTs = [0.00,0.97,1.20,1.68,2.09,2.82,4.24,7.15];
composers = {'Bruckner','Trumpet'};
for vv = 1:length(composers)
    eval([composers{vv} '=struct();']);
end

total_comp = length(composers)*length(RTs);
count = 0;
sums = zeros(length(RTs),length(composers));
for jj = 1:length(composers)
    composer = composers{jj};
    for ii = 1:length(RTs)
        count = count+1;
        clc;
        disp(sprintf('Percent Complete:
%d',round(100*count/total_comp)));
        RT = RTs(ii);
```

```
        if strcmp(composer,composers(1))
            fstring = '%s - Boston - Multiple - %0.2f.wav';
        elseif RT ~= 0.00;
            fstring = '%s - Boston - %0.2f.wav';
        else
            fstring = '%s - Anechoic - %0.2f.wav';
        end
        filename = sprintf(fstring,composer,RT);
        [music, fs] = audioread(filename);

        N = length(music);
        dt = 1/fs;
        t = (0:N-1)*dt;

%% Takes the FFT of the newly import file (music), and sets it as the fft
%% of new control stimulus (scrambled_fft). The phase of the scrambled_fft
%% is then randomized, while the magnitude remains the same.
        [music_fft f] = fft_move(music,fs);
        scrambled_fft = music_fft;
        if mod(N,2)==0
            randoms = rand(N/2+1,1)*2*pi;
            scalers = [cos(randoms), (1i)*sin(randoms)];
            index = find(f==0);
            scrambled_fft(index:end,:) =
abs(scrambled_fft(index:end,:)).*[(scalers(:,1)+scalers(:,2)),(scalers(
:,1)+scalers(:,2))];
            scrambled_fft(1:index-1,:) = abs(scrambled_fft(1:index-
1,:)).*[flipud((scalers(2:end-1,1)-scalers(2:end-
1,2))),flipud((scalers(2:end-1,1)-scalers(2:end-1,2)))];
            scrambled = ifft_move(scrambled_fft,fs);
            scrambled = real(scrambled);
        else
            randoms = rand(floor(N/2)+1,1)*2*pi;
            scalers = [cos(randoms), (1i)*sin(randoms)];
            index = find(f==0);
            scrambled_fft(index:end,:) =
abs(scrambled_fft(index:end,:)).*[(scalers(:,1)+scalers(:,2)),(scalers(
:,1)+scalers(:,2))];
            scrambled_fft(1:index-1,:) = abs(scrambled_fft(1:index-
1,:)).*[flipud((scalers(2:end,1)-
scalers(2:end,2))),flipud((scalers(2:end,1)-scalers(2:end,2)))];
            scrambled = ifft_move(scrambled_fft,fs);
            scrambled = real(scrambled);
        end

%% The envelope (env_lp) of music is determined by low-pass filtering the
%% absolute value of the time series. The envelope is then applied to the
%% scrambled time series. The scrambled time series is then scaled to
%% match the root mean square of the music time series.
        [b,a] = butter(4,10/(fs/2),'low');
        % env_lp = filter(b,a,env);
        env_lp = filter(b,a,abs(music));
        scrambled_env = scrambled.*env_lp;
        scrambledsum = sum(scrambled_env.^2);
        musicsum = sum(music.^2);
%         musicsums(ii,jj) = musicsum;
        scrambled_eq = sqrt(musicsum/scrambledsum)*scrambled_env;
```

```
%% Each file is saved appropriately.
        field = sprintf('RT%d',ii);
        eval([composer '=setfield(' composer ',field,music);']);
        field = sprintf('scrambled%d',ii);
        eval([composer '=setfield(' composer ',field,scrambled_eq);']);
        % soundsc(music,fs);
        % input('');
        % soundsc(scrambled,fs);
        ostring = '%s - Scrambled - %0.2f.wav';
        outname = sprintf(ostring,composer,RT);
        audiowrite(outname,scrambled_eq,fs);
    end
end

%% Displays the results.
music = Trumpet.RT1;
scrambled = Trumpet.scrambled1;
[music_fft fm] = fft_move(music,fs);
[scrambled_fft fn] = fft_move(scrambled,fs);

Nm = length(music);
Nn = length(scrambled);
dt = 1/fs;
tm = (0:Nm-1)*dt;tm=tm.';
tn = (0:Nn-1)*dt;tn=tn.';

figure(4)
subplot(2,1,1);
set(gca,'FontSize',16);
semilogx(fm,10*log10(abs(music_fft(:,2))),fn,10*log10(abs(scrambled_fft
(:,2))));
xlim([20 24000]);
ylabel('Amplitude (dB)');
legend('Test Stimulus','Control Stimulus');
grid on;
subplot(2,1,2);
set(gca,'FontSize',16);
semilogx(fm,(angle(music_fft(:,2)))*180/pi,fn,(angle(scrambled_fft(:,2)
))*180/pi);
xlim([20 24000]);
ylabel('Phase (^o)');
xlabel('Frequency (Hz)');
grid on;

figure(6)
subplot(2,1,1);
set(gca,'FontSize',16);
plot(tm,music(:,2));
ylim([-.1 .1]);
ylabel('Amplitude (Wave Units)');
grid on;
subplot(2,1,2);
set(gca,'FontSize',16);
plot(tm,music(:,2)/10000);
hold on;
plot(tn,scrambled(:,2),'Color',[0.7500,0.3250,0.0980]);
```

```
hold off;
ylim([-.1 .1]);
ylabel('Amplitude (Wave Units)');
xlabel('Time (s)')
legend('Test Stimulus','Control Stimulus');
grid on;
```



(a) (b)

Fig. A.2: Example test and matched scrambled music control stimulus in the frequency domain (a) and time domain (b) showing the similar frequency and rhythmic content.

A.3 SINE-TRUMPET CONTROL STIMULUS

The purpose of the sine-trumpet control stimulus was to remove the harmonic content of the trumpet so that the same musical passage could be heard without the timbral elements of the trumpet. The following code was created in a generic way that allows any anechoic musical passage to be translated into sine-waves, including orchestral pieces with each instrument in separate files.

```
%% Sine Stimuli

clc
clear all
close all

%% Get original wave files
stim_dir = 'C:\Users\msl224\Documents\MATLAB\4 MRI Retim\Stimuli';
cd(stim_dir);
%%
files = dir(stim_dir);
files = files(3:end);
data = cell(length(files),1);
for ii = 1:length(files)
    [new_data,fs] = audioread(files(ii).name);
    data{ii} = new_data;
end
```

```
dt = 1/fs;

%% Find the f0 of each individual file. First, the power spectral
density (PSD) of the time series is found over time. Each PSD is low-
pass filtered at a specified frequency depending on the instrument. For
the trumpet motif, this frequency was 1000 Hz.
total_melody = zeros(length(data{1}),1);
for data_ind = 9:9
    new_data = data{data_ind};
    N = length(new_data);
    dt = 1/fs;
    t = (0:N-1)*dt;t = t.';
    N_rec = 2048;
    [average_num sot] = create_sot(new_data(:,1),N_rec,0)
    [b1,a1] = butter(3,1000/fs,'low');
    [Gxx_av, f_half, Gxx_set, rms] =
spectral_average(filter(b1,a1,new_data(:,1)),fs,average_num,N_rec);
    % [Gxx_av, f_half, Gxx_set, rms] =
spectral_average(new_data(:,1),fs,average_num,N_rec);

    figure(data_ind)
    h = plot_spect(t,f_half,Gxx_set,max(max(Gxx_set)));
    colormap(jet);
    ylim([0 6000]);
    % xlim([0 4]);

% Next, the peak of the PSD was found at each time point. If the
amplitude of the peak did not reach threshold, no note was considered
to be played.
    f0s = [];
    onsets = [];
    zero_onsets=[];
    [rows,columns] = size(Gxx_set);
    for ii = 1:columns
        [pks,locs] = findpeaks(Gxx_set(1:48,ii));
        if max(pks)>2e-9
    %           plot(f_half(locs),pks);
            [~,f0_ind] = max(pks);
            f0_ind = locs(f0_ind);
            f0 = f_half(f0_ind);

            f0s(end+1) = f0;
            onsets(end+1) = sot(ii);
        else
            zero_onsets(end+1) = sot(ii);
        end
    end
    plot(onsets,f0s,'LineStyle','none','Marker','o');
    ylim([0,1000]);

% Sine-waves were created at each of the prescribed time points and
fundamental frequencies. To avoid clipping, a half-Hann window was used
to fade in and out if there was a change in frequency or no sequenctial
note.
    all_onsets = [onsets,zero_onsets];
    all_onsets = sort(all_onsets);
    melody = zeros(N,1);
```

```
    f0_counter = 1;
    fader1 = hann(1024);
    fade_in1 = fader1(1:length(fader1)/2);
    fade_out1 = fader1(length(fader1)/2+1:end);
    fader2 = hann(512);
    fade_in2 = fader2(1:length(fader2)/2);
    fade_out2 = fader2(length(fader2)/2+1:end);
    for ii = 2:length(all_onsets)
        if any(all_onsets(ii-1)==onsets)
            melody(all_onsets(ii-1):all_onsets(ii)) =
sin(2*pi*f0s(f0_counter)*t(all_onsets(ii-1):all_onsets(ii)));
            if any(all_onsets(ii)==zero_onsets)
                melody(all_onsets(ii)-
length(fader1)/2+1:all_onsets(ii)) = melody(all_onsets(ii)-
length(fader1)/2+1:all_onsets(ii)).*fade_out1;
            elseif f0s(f0_counter)~=f0s(f0_counter+1)
                melody(all_onsets(ii)-
length(fader2)/2+1:all_onsets(ii)) = melody(all_onsets(ii)-
length(fader2)/2+1:all_onsets(ii)).*fade_out2;
            end
            if any(all_onsets(ii-2)==zero_onsets)
                melody(all_onsets(ii-1):all_onsets(ii-
1)+length(fader1)/2-1) = melody(all_onsets(ii-1):all_onsets(ii-
1)+length(fader1)/2-1).*fade_in1;
            elseif f0s(f0_counter)~=f0s(f0_counter-1)
                melody(all_onsets(ii-1):all_onsets(ii-
1)+length(fader2)/2-1) = melody(all_onsets(ii-1):all_onsets(ii-
1)+length(fader2)/2-1).*fade_in2;
            end
            f0_counter=f0_counter+1;
        end
    end

% The envelope of the original wave file was found and applied to the
new sine waves to maintain the attack and decay of each note.
    [b,a] = butter(4,40/(fs/2),'low');
    env_lp = filter(b,a,abs(new_data(:,1)));
    melody_env = melody.*env_lp;

% If multiple wave files were used, this allowed the user to listen to
the mixed resulting signals.
    total_melody = total_melody+melody_env;
end
soundsc(total_melody,fs);

%% Display the results. (see Fig. 5.1)
[Gxx_av, f_half, Gxx_set, rms] =
spectral_average(new_data(:,1),fs,average_num,N_rec,sot,1);
figure(9)
h = plot_spect(t,f_half,Gxx_set,max(max(Gxx_set)));
colormap(jet);
ylim([0 6000]);
xlim([0 4]);

[Gxx_av, f_half, Gxx_set, rms] =
spectral_average(melody_env,fs,average_num,N_rec,sot,1);
figure(2)
```

```
h = plot_spect(t,f_half,Gxx_set,max(max(Gxx_set)));
colormap(jet);
ylim([0 6000]);
xlim([0 4]);
```

APPENDIX B

MRI and Listening Testing Equipment

(a) (b)

Fig. B.1: (a) Psychology Software Tools, Inc. MRI Simulator[163] and (b) Siemens 3T Prisma Fit, MRI Scanner.

Fig. B.2: STAX MR 003-MK2 electrostatic, MRI-compatible earbuds. Used in the pilot study (Chapter 2) and outside of the MRI in the two other MRI experiments (Chapters 4 and 5).

Fig. B.3: Nordic Neurolab Response Grips.[164]

Fig. B.4: STAX SR-207 headphones.[165] Used in the conventional listening study (Chapter 3).

Fig. B.5: Sensimetrics Model S14 ear buds.[166] Used in the MRI experiments (Chapters 4 and 5).

Fig. B.6: Resonance Technology, Inc. VisuaStim Digital.[167]

REFERENCES

[1] L. L. Beranek, "Concert hall acoustics," *J. ASA*, vol. 92, no. 1, pp. 1-39, 1992.

[2] M. R. Schroeder, D. Gottlob and K. F. Siebrasse, "Comparative study of European concert halls: correlation of subjective preference with geometric and acoustic parameters," *J. Acoust. Soc. Am.,* vol. 56, no. 4, pp. 1195-1201, 1974.

[3] T. Lokki, J. Pätynen, A. Kuusinen, V. Heikki and S. Tervo, "Concert hall acoustics assessment with individually elicited attributes," *J. ASA,* vol. 130, no. 2, pp. 835-849, 2011.

[4] H. Kuttruff, Room Acoustics, Fourth Edition, London: Spon Press, 2000.

[5] J. C. Nunnally and I. H. Bernstein, Psychometric Theory, Third ed., New York: McGraw-Hill, 1994, pp. 251-252, 271-276, 304.

[6] R. F. DeVellis, Scale Development: Theory and Applications, Newbury Park, CA: Sage, 1991, pp. 68-70.

[7] R. H. Fazio and M. A. Olson, "Implicit measures in social cognition research: their meaning and use," *Annu. Rev. Psychol.,* vol. 54, pp. 297-327, 2003.

[8] P. G. Devine, "Stereotypes and prejudice: their automatic and controlled components," *J. Pers. Soc. Psychol.,* vol. 56, no. 1, pp. 5-18, 1989.

[9] W. Hofmann, T. Gschwendner and M. Schmitt, "On implicit-explicit consistency: the moderating role of individual differences in awareness and adjustment," *Eur. J. Pers.,* vol. 19, pp. 25-49, 2005.

[10] A. Gyurak, J. J. Gross and A. Etikin, "Explicit and implicit emotion regulation: a dual-process framework," *Cognition Emotion,* vol. 25, no. 3, pp. 400-415, 2011.

[11] R. B. Zajonc, "Feeling and thinking - preferences need no inferences," *Am. Psychol.,* vol. 35, no. 2, pp. 151-175, 1980.

[12] K. R. Scherer, "Neuroscience projections to current debates in emotion psychology," *Cognition Emotion,* vol. 7, no. 1, pp. 1-41, 1993.

[13] S. Koelsch, Brain and music, Hoboken, NJ: Wiley-Blackwell, 2012.

[14] B. Ryder, "A tissue of lies," The Economist, 9 June 2012. [Online]. Available: http://www.economist.com/node/21556548. [Accessed 13 April 2017].

[15] W. C. Sabine, Collected papers on acoustics, Cambridge: Harvard University Press, 1922, pp. 1-68.

[16] M. T. Neal, Investigating the sense of listener envelopment in concert halls using third-order ambisonic reproduction over a loudspeaker array and a hybrid room acoustics simulation method, University Park: The Pennsylvania State University, 2015.

[17] C. L. Christensen, *Odeon Room Acoustics Program: Version 12.0 User Manual,* Lyngby: Odon A/S, 2013.

[18] M. Barron, "Subjective study of British symphony concert halls," *Acustica*, vol. 66, no. 1, pp. 1-14, 1988.

[19] T. Lokki, J. Pätynen, A. Kuusinen and S. Tervo, "Disentangling preference ratings of concert hall acoustics using subjective sensory profiles," *J. Acoust. Soc. Am.*, vol. 132, no. 5, pp. 3148-3161, November 2012.

[20] K. Yamaguchi, "Multivariate analysis of subjective and physical measures of hall acoustics," *J. Acoust. Soc. Am.*, vol. 52, no. 5, pp. 1271-1279, 1972.

[21] Y. Ando, Concert hall acoustics, Heidelberg: Spring-Verlag, 1985.

[22] L. L. Beranek, Concert halls and opera houses: music, acoustics, and architecture, New York: Springer, 2004, p. 536.

[23] R. J. Hawkes and H. Douglas, "Subjective acoustic experience in concert auditoria," *Acustica,* vol. 24, pp. 235-250, 1971.

[24] S. Hase, A. Takatsu, S. Sato, H. Sakai and Y. Ando, "Reverberance of an existing hall in relation to both subsequent reverberation time and SPL," *J. Sound. Vib.*, vol. 232, no. 1, pp. 149-155, 2000.

[25] J. S. Bradley, "Review of objective room acoustics measures and future needs," *Appl. Acoust.*, vol. 72, pp. 713-720, 2011.

[26] "ISO 3382-1:2009: Acoustics – Measurement of room acoustic parameters – Part 1: Performance spaces".

[27] J. Bigelow and A. Poremba, "Achilles' Ear? Inferior human short-term and recognition memory in the auditory modality," *PLOS One*, vol. 9, pp. 1-8, 2014.

[28] M. Gerzon, "Periphony: with-height sound reproduction," *J. Audio Eng. Soc.,* vol. 21, pp. 2-10, 1973.

[29] B. Xie, Head-Related Transfer Function and Virtual Auditory Display, 2nd ed., Plantation, FL: J. Ross Publishing, 2013, pp. 1-42, 233-264, 351-370.

[30] E. Parizet, N. Hamzaoui and G. Sabatié, "Comparison of some listening test methods: a case study," *Acta Acust. United Ac.,* vol. 91, pp. 356-364, 2005.

[31] M. H. Kutner, C. J. Nachtsheim, J. Neter and W. Li, Applied Linear Statistical Models, Boston: McGraw-Hill Irwin, 2005, pp. 641-1208.

[32] R. B. Buxton, Introduction to Funcitonal Magnetic Resonance Imaging: Principles and Techniques, Cambridge: Cambridge University Press, 2009, pp. 114-118.

[33] S. M. Breedlove, M. R. Rosenzweig and N. V. Watson, Biological Psychology: An Introduction to Behavioral, Cognitive, and Clinical Neuroscience, 5th ed., Sunderland, MA: Sinauer Associates, 2007, pp. 23-58.

[34] R. A. Poldrack, J. A. Mumford and T. E. Nichols, Handbook of Functional MRI Data Analysis, New York: Cambridge UP, 2011.

[35] K. J. Friston, P. Fletcher, O. Josephs, A. Holmes, M. D. Rugg and R. Turner, "Event-related fMRI: Characterizing differential responses," *Neuroimage,* vol. 7, pp. 30-40, 1998.

[36] D. A. Hall, M. P. Haggard, M. A. Akeroyd, A. R. Palmer, A. Q. Summerfield, M. R. Elliott, E. M. Gurney and R. W. Bowtell, ""Sparse" temporal sampling in auditory fMRI," *Hum. Brain Mapp.,* vol. 7, pp. 213-223, 1999.

[37] C. Schwarzbauer, M. H. Davis, J. M. Rodd and I. Johnsrude, "Interleaved silent steady state (ISSS) imaging: A new sparse imaging method applied to auditory fMRI," *NeuroImage,* vol. 29, pp. 774-782, 2006.

[38] S. Skouras, M. Gray, H. Critchley and S. Koelsch, "fMRI scanner noise interaction with affective neural processes," *PLoS One,* vol. 8, no. 11, p. e80564, 2013.

[39] K. Mueller, T. Mildner, T. Fritz, J. Lepsien, C. Schwarzbauer, M. L. Schroeter and H. E. Möller, "Investigating brain response to music: A comparison of different fMRI acquisition schemes," *NeuroImage,* vol. 54, pp. 337-343, 2011.

[40] J. Nolte, The Human Brain - An introduction to its Functional Anatomy, Sixth ed., Philadelphia: Mosby, 2009, p. 602.

[41] S. Clarke and P. Morosan, "Architecture, Connectivity, and Transmitter Receptors of Human Auditory Cortex," in *The Human Auditory Cortex*, D. Poeppel, T. Overath, A. N. Popper and R. R. Fay, Eds., New York, Springer-Verlag, 2012, pp. 11-38.

[42] P. Morosan, J. Rademacher, A. Schleicher, K. Amunts, T. Schormann and K. Zilles, "Human primary auditory cortex: cytoarchitectonic subdivisions and mapping into a spatial reference system," *Neuroimage*, vol. 13, pp. 684-701, 2001.

[43] A. Giraud and D. Poeppel, "Speech Perception from a Neurophysiological Perspective," in *The Human Auditory Cortex*, D. Poeppel, T. Overath, A. N. Popper and R. R. Fay, Eds., New York, Springer-Verlag, 2012, pp. 225-260.

[44] R. J. Zatorre, P. Belin and V. B. Penhune, "Structure and function of auditory cortex: music and speech," *Trends Cogn. Sci.*, vol. 6, pp. 37-46, 2002.

[45] P. N. Juslin and J. A. Sloboda, "Introduction: Aims, organization, and terminology," in *Handbook of Music and Emotions: Theory, Research, Applications*, Oxford, Oxford University Press, 2011.

[46] S. Koelsch, "Investigating emotion with music: neuroscientific approaches," *Annals of the New York Academy of Sciences*, vol. 1060, pp. 412-418, 2005.

[47] K. C. Berridge and M. L. Kringelbach, "Pleasure systems in the brain," *Neuron*, vol. 86, pp. 646-664, 2015.

[48] B. H. Schott, L. Minuzzi, R. M. Krebs, D. Elmenhorst, M. Lang, O. H. Winz, C. I. Seidenbecher, H. H. Coene, H. Henize, K. Zilles, E. Duzel and A. Bauer, "Mesolimbic functional magnetic resonance imaging activations during reward anticipation correlate with reward-related ventral striatal dopamine release," *Journal of Neuroscience*, vol. 28, no. 52, pp. 14311-14319, 2008.

[49] A. J. Blood and R. J. Zatorre, "Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion," *PNAS*, vol. 98, no. 20, pp. 11818-11823, 2001.

[50] V. Menon and D. J. Levitin, "The rewards of music listening: Response and physiological connectivity of the mesolimbic system," *NeuroImage,* vol. 28, pp. 175-184, 2005.

[51] V. N. Salimpoor, M. Benovoy, K. Larcher, A. Dagher and R. J. Zatorre, "Anatomically distinct dopamine release during anticipation and experience of peak emotion to music," *Nature Neuroscience,* vol. 14, no. 2, pp. 257-U355, 2011.

[52] L. Kibiuk, "Mesolimbic Pathway," Soc. for Neurosci., 22 Mar. 2009.

[53] G. M. Bidelman and A. Krishnan, "Effects of reverberation on brainstem representation of speech in musicians and non-musicians," *Brain Res.,* vol. 1355, pp. 112-125, 2010.

[54] Y. Ando and P. Cariani, Auditory and visual sensations, New York: Springer, 2009.

[55] Y. Ando, K. Yamamoto, H. Nagamastu and S. H. Kang, "Auditory brainstem response (ABR) in relation to the horizontal angle of sound incidence," *Acoustic Letters,* vol. 15, pp. 57-64, 1991.

[56] Y. Ando, "Evoked potentials relation to the subjective preference of sound fields," *Acustica,* vol. 76, pp. 292-296, 1992.

[57] S. Koelsch, T. Fritz, D. v. Cramon, K. Muller and A. Friederici, "Investigating emotion with music: an fMRI study," *Human Brain Mapping,* vol. 27, pp. 239-250, 2006.

[58] A. J. Blood, R. J. Zatorre, P. Bermudez and A. C. Evans, "Emotional responses to pleasant and unpleasant music correlate with activity in paralimbic brain regions," *Nature Neuroscience,* vol. 2, no. 4, pp. 382-387, 1999.

[59] V. N. Salimpoor and R. J. Zatorre, "Neural interactions that give rise to musical pleasure," *Psychology of Aesthetics, Creativity, and the Arts,* vol. 7, no. 1, pp. 62-75, 2013.

[60] J. Pätynen, V. Pulkki and T. Lokki, "Anechoic Recording System for Symphony Orchestra," *Acta Acustica united with Acustica,* vol. 94, no. 6, pp. 856-865, 2008.

[61] R. R. Ernst and W. A. Anderson, "Application of Fourier Transform Spectroscopy to Magnetic Resonance," *Rev. Sci. Instruments,* vol. 37, no. 1, pp. 93-102, 1966.

[62] J. Ashburner, G. Barnes, C. Chen, J. Daunizeau, G. Flandin, K. Friston, D. Gitelman, S. Kiebel, J. Kilner, V. Litvak, R. Moran, W. Penny, K. Stephan, R. Henson, C. Hutton, V. Glauche, J. Mattout and C. Phillips, "SPM8 Manual," 21 July 2010. [Online]. Available: http://www.fil.ion.ucl.ac.uk/spm/software/spm8/. [Accessed 6 March 2015].

[63] J. Mazziotta, A. Toga, A. Evans, P. Fox, J. Lancaster, K. Zilles, R. Woods, T. Paus, G. Simpson, B. Pike, C. Holmes, L. Collins, P. Thompson, D. MacDonald, M. Iacoboni, T. Schormann, K. Amunts, N. Palomero-Gallagher, S. Geyer, L. Parsons, K. Narr, N. Kabani, G. Le Goualher, D. Boomsma, T. Cannon, R. Kawashima and B. Mazoyer, "A probabilistic atlas and reference system for the human brain: International Consortium for Brain Mapping (ICBM)," *Phil. Trans. R. Soc. Lond. B,* vol. 356, no. 1412, pp. 1293-1322, 2001.

[64] K. Phan, T. Wager, S. F. Taylor and I. Liberzon, "Functional neuroanatomy of emotion: a meta-analysis of emotion activation studies in PET and fMRI," *NeuroImage,* vol. 16, pp. 331-348, 2002.

[65] R. N. Cardinal, D. R. Pennicott, C. L. Sugathapala, T. W. Robbins and B. J. Everitt, "Impulsive choice induced in rats by lesions of the nucleus accumbens core," *Science,* vol. 292, pp. 2499-2501, 2001.

[66] S. Peciña and K. C. Berridge, "Hedonic hot spot in nucleus accumbens shell: where do μ-opioids cause increased hedonic impact of sweetness?," *J. Neurosci,* vol. 25, no. 50, pp. 11777-11786, 2005.

[67] M. Feja, L. Hayn and M. Koch, "Nucleus accumbens core and shell inactivation differentially affects impulsive behaviours in rats," *Prog. in Neuro-Psychopharmacology & Biol. Psychiatry,* vol. 54, pp. 31-42, 2014.

[68] M. S. Lawless and M. C. Vigeant, "Investigating the emotional response to room acoustics: a functional magnetic resonance imaging study," *J. Acoust. Soc. Am.,* vol. 138, no. 4, pp. EL417-EL423, 2015.

[69] J. Pätynen and T. Lokki, "Concert halls with strong and lateral sound increase the emotional impact of orchestra music," *J. Acoust. Soc. Am.,* vol. 139, no. 3, pp. 1214-1224, 2016.

[70] A. Giménez, R. M. Cibrián, S. Girón, T. Zammarreño, J. J. Sendra, A. Vela and F. Daumal, "Questionnaire survey to qualify the acoustics of Spanish concert halls," *Acta Acoust. United Ac.,* vol. 97, pp. 949-965, 2011.

[71] T. Lokki, J. Pätynen, S. Tervo, S. Siltanen and L. Savioja, "Engaging concert hall acoustics is made up of temporal envelope preserving reflections," *J. Acoust. Soc. Am.,* vol. 129, no. 6, pp. EL223-EL228, 2011.

[72] A. M. Sarroff and J. P. Bello, "Toward a computational model of perceived spaciousness in recorded music," *J. Audio Eng. Soc.,* vol. 59, no. 7/8, pp. 498-513, 2011.

[73] R. Shimokura, L. Tronchin, A. Cocchi and Y. Soeta, "Subjective diffuseness of music signals convolved with binaural impulse responses," *J. Sound Vib.,* vol. 330, pp. 3526-3537, 2011.

[74] J. Yoo, J. Seo, H. Shim, H. Chung, K. Sung and K. Kang, "Subjective listening experiments on a front and rear array-based WFS system," *ETRI J.,* vol. 33, no. 6, pp. 977-980, 2011.

[75] M. Galiana, C. Llinares and Á. Page, "Subjective evaluation of music hall acoustics: response of expert and non-expert users," *Build. Environ.,* vol. 58, pp. 1-13, 2012.

[76] A. Giménez, R. M. Cibrián and S. Cerdá, "Subjective assessment of concert halls: a common vocabulary for music lovers and acousticians," *Arch. Acoust.,* vol. 37, no. 3, pp. 331-340, 2012.

[77] A. Miśkiewicz, T. Rogala, T. Rościszewska, T. Rudzki and T. Fidecki, "Concert hall sound clarity: a comparison of auditory judgments and objective measures," *Arch. Acoust.,* vol. 37, no. 1, pp. 41-46, 2012.

[78] N. Nishihara and T. Hidaka, "Loudness perception of low tones undergoing partial masking by higher tones in orchestral music in concert halls," *J. Acoust. Soc. Am.,* vol. 132, no. 2, pp. 799-803, 2012.

[79] S. Sato, S. Wang, Y. Zhao, S. Wu, H. Sun, N. Prodi, C. Visentin and R. Pompoli, "Effects of acoustic and visual stimuli on subjective preferences for different

seating positions in an Italian style theater," *Acta Acoust. United Ac.*, vol. 98, no. 5, pp. 749-759, 2012.

[80] D. L. Valente, J. Braasch and S. A. Myrbeck, "Comparing perceived auditory width to the visual image of a performing ensemble in contrasting bi-modal environments," *J. Acoust. Soc. Am.*, vol. 131, no. 1, pp. 205-217, 2012.

[81] Y. Xiong, S. Yan and S. W. Sheng, "Experiment research on apparent source width (ASW)," *Adv. Mat. Res.*, Vols. 374-377, pp. 1155-1158, 2012.

[82] A. Avni, J. Ahrens, M. Geier, S. Spors, H. Wierstorf and B. Rafaely, "Spatial perception of sound fields recorded by spherical microphone arrays with varying spatial resolution," *J. Acoust. Soc. Am.*, vol. 133, no. 5, pp. 2711-2721, 2013.

[83] M. Frank, "Source width of frontal phantom sources: perception, measurement, and modeling," *Arch. Acoust.*, vol. 38, no. 3, pp. 311-319, 2013.

[84] S. Kawase, "Factors influencing audience seat selection in a concert hall: a comparison between music majors and nonmusic majors," *J. Environ. Psychol.*, vol. 36, pp. 305-315, 2013.

[85] P. W. Robinson, J. Pätynen and T. Lokki, "The effect of diffuse reflections on spatial discrimination in a simulated concert hall," *J. Acoust. Soc. Am.*, vol. 133, no. 5, pp. EL370-EL376, 2013.

[86] P. W. Robinson, J. Pätynen, T. Lokki, H. S. Jang, J. Y. Jeon and N. Xiang, "The role of diffusive architectural surfaces on auditory spatial discrimination in performance venues," *J. Acoust. Soc. Am.*, vol. 133, no. 6, pp. 3940-3950, 2013.

[87] P. W. Robinson, A. Walther, C. Faller and J. Braasch, "Echo thresholds for reflections from acoustically diffusive architectural surfaces," *J. Acoust. Soc. Am.*, vol. 134, no. 4, pp. 2755-2764, 2013.

[88] S. Tervo, J. Pätynen, A. Kuusinen and T. Lokki, "Spatial decomposition method for room impulse responses," *J. Audio Eng. Soc.*, vol. 61, no. 1/2, pp. 17-28, 2013.

[89] J. van Dorp Schuitman, D. de Vries and A. Lindau, "Deriving content-specific measures of room acoustic perception using a binaural, nonlinear auditory model," *J. Acoust. Soc. Am.*, vol. 133, no. 3, pp. 1572-1585, 2013.

[90] Y. Xiong, "The influence of early sound direction on apparent source width (ASW)," *Appl. Mech. Mater.,* Vols. 357-360, pp. 160-163, 2013.

[91] A. Giménez, R. M. Cibrián, S. Cerdá, S. Girón and T. Zammareño, "Mismatches between objective parameters and measured perception assessment in room acoustics: a holistic approach," *Build. Environ.,* vol. 74, pp. 119-131, 2014.

[92] A. Haapaniemi and T. Lokki, "Identifying concert halls from source presence vs room presence," *J. Acoust. Soc. Am.,* vol. 135, no. 6, pp. EL311-EL317, 2014.

[93] T. Lee, Y. Baek, Y. Park and D. Youn, "Stereo upmix-based binaural auralization for mobile devices," *IEEE T. Consum. Electr.,* vol. 60, no. 3, pp. 411-419, 2014.

[94] S. Merchel and M. E. Altinsoy, "The influence of vibrations on musical experience," *J. Audio Eng. Soc.,* vol. 62, no. 4, pp. 220-234, 2014.

[95] S. Tervo, P. Laukkanen, J. Pätynen and T. Lokki, "Preferences of critical listening environments among sound engineers," *J. Audio Eng. Soc.,* vol. 62, no. 5, pp. 300-314, 2014.

[96] S. Cerdá, A. Giménez, R. Cibrián, S. Girón and T. Zamarreño, "Subjective ranking of concert halls substantiated through orthogonal objective parameters," *J. Acoust. Soc. Am.,* vol. 137, no. 2, pp. 580-584, 2015.

[97] J. Grosse and S. van de Par, "Perceptually accurate reproduction of recorded sound fields in a reverberant room using spatially distributed loudspeakers," *IEEE J. Sel. Top. Signa.,* vol. 9, no. 5, pp. 867-880, 2015.

[98] A. Kuusinen and T. Lokki, "Investigation of auditory distance perception and preferences in concert halls by using virtual acoustics," *J. Acoust. Soc. Am.,* vol. 138, no. 5, pp. 3148-3159, 2015.

[99] H. Lee, "2D-to-3D ambience upmixing based on perceptual band allocation," *J. Audio Eng. Soc.,* vol. 63, no. 10, pp. 811-821, 2015.

[100] L. Shtrepi, A. Astolfi, S. Pelzer, R. Vitale and M. Rychtáriková, "Objective and perceptual assessment of the scattered sound field in a simulated concert hall," *J. Acoust. Soc. Am.,* vol. 138, no. 3, pp. 1485-1497, 2015.

[101] M. C. Vigeant, R. D. Celmer, C. M. Jasinski, M. J. Ahearn, M. J. Schaeffler, C. B. Giacomoni, A. P. Wells and C. I. Ormsbee, "The effects of different test methods

on the just noticeable difference of clarity index for music," *J. Acoust. Soc. Am.*, vol. 138, no. 1, pp. 476-491, 2015.

[102] A. J. Oxenham, "Informational masking and musical training," *J. Acoust. Soc. Am.*, vol. 114, no. 3, pp. 1543-1549, 2003.

[103] C. Micheyl, K. Delhommeau, X. Perrot and A. J. Oxenham, "Influence of musical and psychoacoustical training on pitch discrimination," *Hearing Res.*, vol. 219, pp. 36-47, 2006.

[104] E. Bigand and B. Poulin-Charronnat, "Are we "experience listeners"? A review of the musical capacities that do not depend on formal musical training.," *Cognition*, vol. 100, pp. 100-130, 2006.

[105] A. Gabrielsson, U. Rosenberg and H. Sjögren, "Judgments and dimension analyses of perceived sound quality of sound-reproducing systems," *J. Acoust. Soc. Am.*, vol. 55, no. 4, pp. 854-561, 1974.

[106] S. Bech, "Selection and training of subjects for listening tests on sound-reproducing equipment," *J. Audio Eng. Soc.*, vol. 40, no. 7/8, pp. 590-610, 1992.

[107] S. E. Olive, "Differences in performance and preference of trained versus untrained listeners in loudspeaker tests: a case study," *J. Audio Eng. Soc.*, vol. 51, no. 9, pp. 806-825, 2003.

[108] N. Zacharov and G. Lorho, "What are the requirements of a listening panel for evaluating spatial audio quality?," in *Proceedings of the Spatial Audio and Sensory Evaluation Techniques Workshop*, University of Surrey, UK, 2006.

[109] R. E. Kirk, "Learning, a major factor influencing preferences for high-fidelity reproducing systems.," *J. Acoust. Soc. Am.*, vol. 28, no. 6, pp. 1113-1116, 1956.

[110] T. J. Cox and B. M. Shield, "Audience questionnaire survey of the acoustics of the Royal Festival Hall, London, England," *Ac. United Acta Acust.*, vol. 85, no. 4, pp. 547-559, 1999.

[111] E. Bigand, S. Vieillard, F. Madurell, J. Marozeau and A. Dacquet, "Multidimensional scaling of emotional responses to music: the effect of musical expertise and of the duration of the excerpts," *Cognition Emotion*, vol. 19, no. 8, pp. 1113-1139, 2005.

[112] S. Vieillard, I. Peretz, N. Gosselin, S. Khalfa, L. Gagnon and B. Bouchard, "Happy, sad, scary and peaceful musical excerpts for research on emotions," *Cognition Emotion,* vol. 22, no. 4, pp. 720-752, 2008.

[113] A. T. Roy, M. Vigeant, T. Munjal, C. Carver, P. Jiradejvong and C. J. Limb, "Reverberation negatively impacts musical sound quality for cochlear implant users," *Cochlear Implants International,* vol. 16, 2015.

[114] "ITU-R BS.1534-1 -- Method for the subjective assessment of intermediate quality level of coding systems," 2001.

[115] F. A. Wichmann and N. J. Hill, "The psychometric function: II. Bootstrap-based confidence intervals and sampling," *Percept. Psychophys.,* vol. 63, no. 8, pp. 1314-1329, 2001.

[116] V. R. Algazi, R. O. Duda, D. M. Thompson and C. Avendano, "The CIPIC HRTF database," in *Proc. of the 2001 IEEE Workshop on Applications of Signal Processing to Audio and Electroacoustics*, New Paltz, New York, 2001.

[117] T. Okano, "Judgments of noticeable differences in sound fields of concert halls caused by intensity variations in early reflections," *J. Acoust. Soc. Am.,* vol. 111, pp. 217-229, 2002.

[118] B. &. Olufsen, "Music from Archimedes," 1992.

[119] L. J. Cronbach, "Coefficient alpha and the internal structure of tests," *Psychometrika,* vol. 16, no. 3, pp. 297-334, 1951.

[120] D. Pollard, "Strong consistency of k-means clustering," *Annals of Stat.,* vol. 9, no. 1, pp. 135-140, 1981.

[121] R. B. Cattell, "The scree test for number of factors," *Multivar. Behav. Res.,* vol. 1, no. 2, pp. 245-276, 1966.

[122] J. H. Ward, "Hierarchical grouping to optimize an objective function," *J. Am. Stat. Assoc.,* vol. 58, no. 301, pp. 236-244, 1963.

[123] L. Berkowitz and P. G. Devine, "Research traditions, analysis, and synthesis in social psychological theories: the case of dissonance theory," *Pers. Soc. Psychol. B.,* vol. 15, no. 4, pp. 493-507, 1989.

[124] P. N. Juslin and D. Västfjäll, "Emotional responses to music: the need to consider underlying mechanisms," *Behav. Brain Sci.,* vol. 31, pp. 559-621, 2008.

[125] C. Faßhauer, A. Frese and S. Evers, "Musical ability is associated with enhanced auditory and visual cognitive processing," *BMC Neurosci.,* vol. 16, p. 59, 2015.

[126] H. K. Sanju and P. Kumar, "Enhanced auditory evoked potentials in musicians: a review of recent findings," *J. Otol.,* vol. 11, pp. 63-72, 2016.

[127] F. Bianchi, J. Hjortkjær, S. Santurette, R. J. Zatorre, H. R. Siebner and T. Dau, "Subcortical and cortical correlates of pitch discrimination: Evidence for two levels of neuroplasticity in musicians," *NeuroImage,* vol. 163, pp. 398-412, 2017.

[128] C. Gaser and G. Schlaug, "Brain structures differ between musicians and non-musicians," *J. Neurosci.,* vol. 23, pp. 9240-9245, 2003.

[129] C. Pantev and S. C. Herholz, "Plasticity of the human auditory cortex related to musical training," *Neurosci. Biobehav. Rev.,* vol. 35, pp. 2140-2154, 2011.

[130] T. Overath, S. Kumar, L. Stewart, K. von Kriegstein, R. Cusack, A. Rees and T. D. Griffiths, "Cortical mechanisms for the segregation and representation of acoustics textures," *J. Neurosci.,* vol. 30, no. 6, pp. 2070-2076, 2010.

[131] R. J. Zatorre, M. Bouffard and P. Belin, "Sensitivity to auditory object features in human temporal neocortex," *J. Neurosci.,* vol. 24, pp. 3637-3642, 2004.

[132] E. Fedorenko, J. H. McDermott, S. Norman-Haignere and N. Kanwisher, "Sensitivity to musical structure in the human brain," *J. Neurophysiol.,* vol. 108, pp. 3289-3300, 2012.

[133] J. Obleser, R. J. S. Wise, M. A. D and S. K. Scott, "Functional integration across brain regions improves speech perception under adverse listening conditions," *J. Neurosci.,* vol. 27, no. 9, pp. 2283-2289, 2007.

[134] C. J. Wild, M. H. Davis and I. S. Johnsrude, "Human auditory cortex is sensitive to the perceived clarity of speech," *NeuroImage,* vol. 60, pp. 1490-1502, 2012.

[135] R. C. Oldfield, "The assessment and analysis of handedness: the Edinburgh Inventory," *Neuropsychologia,* vol. 9, pp. 97-113, 1971.

[136] P. Bech, N.-A. Rasmussen, L. Olsen, V. Noerholm and W. Abildgaard, "The sensitivity and specificity of the Major Depression Inventory, using the Present State Examination oas the index of diagnostic validity," *J. Affect. Disorders,* vol. 66, pp. 159-164, 2001.

[137] M. S. Lawless and M. C. Vigeant, "Effects of test method and participant musical training on preference ratings of stimuli with different reverberation times," *J. Acoust. Soc. Am.,* vol. 142, no. 4, pp. 2258-2272, 2017.

[138] L. Kasper, S. Bollmann, A. O. Diaconescu, C. Hutton, J. Heinzle, S. Iglesias, T. U. Hauser, M. Sebold, Z. Manjaly, K. P. Pruessmann and K. E. Stephan, "The PhysIO Toolbox for modeling physiological noise in fMRI data," *J. Neurosci. Meth.,* vol. 276, pp. 56-72, 2017.

[139] T. B. Jones, P. A. Bandettini and R. M. Birn, "Integration of motion correction and physiological noise regression in fMRI," *NeuroImage,* vol. 42, pp. 582-590, 2008.

[140] D. A. Handwerker, J. M. Ollinger and M. D'Esposito, "Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses," *NeuroImage,* vol. 21, pp. 1639-1651, 2004.

[141] S. Maleki-Balajoo, G. Hossein-Zadeh, H. Soltanian-Zadeh and H. Ekhtiari, "Locally estimated hemodynamic response function and activation detection sensitivity in heroin-cue reactivity study," *Basic Clin. Neurosci.,* vol. 7, no. 4, pp. 299-314, 2016.

[142] T. A. Hackett, T. M. Preuss and J. H. Kaas, "Architectonic identification of the core region in auditory cortex of macaques, chimpanzees, and humans," *J. Comp. Neurol.,* vol. 441, pp. 197-222, 2001.

[143] S. B. Eickoff, K. E. Stephan, H. Mohlberg, C. Grefkes, G. R. Fink, K. Amunts and K. Zilles, "A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data," *Neuroimage,* vol. 25, pp. 1325-1335, 2005.

[144] M. C. Keuken, P. L. Bazin, L. Crown, J. Hootsmans, A. Laufer, C. Müller-Axt, R. Sier, E. J. van der Putten, A. Schäfer, R. Turner and B. U. Forstmann, "Quantifying inter-individual anatomical variabilityin the subcortex using 7 T structural MRI," *Neuroimage,* vol. 94, pp. 40-46, 2014.

[145] M. Brett, J. Anton, R. Valabregue and J. Poline, "Region of interest analysis using an SPM toolbox [abstract]," *Presented at the 8th International Conference on Functional Mapping of the Human Brain,* June 2-6 2002.

[146] H. Haas, "On the influence of a single echo on the audibility of speech," *Acustica,* vol. 1, pp. 48-58, 1951.

[147] D. J. Levitin and V. Menon, "Musical structure is processed in "language" areas of the brain: a possible role for Brodmann area 47 in temporal coherence," *Neuroimage,* vol. 20, pp. 2142-2152, 2003.

[148] A. J. Blood and R. J. Zatorre, "Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion," *PNAS,* vol. 98, no. 20, pp. 11818-11823, 2001.

[149] G. S. Berns, C. M. Capra, S. Moore and C. Noussair, "Neural mechanisms of the influence of popularity on adolescent ratings of music," *Neuroimage,* vol. 49, pp. 2687-2696, 2010.

[150] D. J. Levitin, "Memory for musical attributes," in *Music, Cognition, and Computerized Sound: An Introduction to Psychoacoustics,* P. Cook, Ed., Cambrige, MA, MIT Press, 1999, pp. 209-227.

[151] V. Menon, D. J. Levitin, B. K. Smith, A. Lembke, B. D. Krasnow, D. Glazer, G. H. Glober and S. McAdams, "Neural correlates of timbre change in harmonic sounds," *NeuroImage,* vol. 17, pp. 1742-1754, 2002.

[152] J. D. Warren, A. R. Jennings and T. D. Griffiths, "Analysis of the spectral envelope of sounds by the human brain," *NeuroImage,* vol. 24, pp. 1052-1057, 2005.

[153] E. J. Allen, P. C. Burton, C. A. Olman and A. J. Oxenham, "Representations of pitch and timbre variation in human auditory cortex," *J. Neurosci.,* vol. 37, pp. 1284-1293, 2017.

[154] S. Deike, B. Gaschler-Markefski, A. Brechmann and H. Scheich, "Auditory stream segregation relying on timbre involves left auditory cortex," *Neuroreport,* vol. 15, pp. 1511-1514, 2004.

[155] A. R. Halpern, R. J. Zatorre, M. Bouffard and J. A. Johnson, "Behavioral and neural correlates of perceived and imagined musical timbre," *Neuropsychologia,* vol. 42, pp. 1281-1292, 2004.

[156] S. Reiterer, M. Erb and W. Grodd, "Cerebral processing of timbre and loudness: fMRI evidence for a contribution of Broca's area to basic auditory discrimination," *Brain Imaging Behav.,* vol. 2, pp. 1-10, 2008.

[157] C. Pantev, L. E. Roberts, M. Schulz, A. Engelien and B. Ross, "Timbre-specific enhancement of auditory cortical representations in musicians," *Neuroreport,* vol. 12, pp. 169-174, 2001.

[158] P. J. Goad and D. H. Keefe, "Timbre discrimination of musical instruments in a concert hall," *Music Percept.,* vol. 10, pp. 43-62, 1992.

[159] R. Mo, B. Wu and A. Horner, "The effects of reverberation on the emotional characteristics of musical instruments," *J. Audio Eng. Soc.,* vol. 63, pp. 966-979, 2015.

[160] R. Mo, R. H. Y. So and A. Horner, "An investigation into how reverberation effects the space of instrument emotional characteristics," *J. Audio Eng. Soc.,* vol. 64, pp. 988-1002, 2016.

[161] R. Heinz, "Binaural room simulation based on an image source model with addition of statistical methods to include the diffuse sound scattering of walls and to predict the reverberant tail," *Appl. Acoust.,* vol. 38, pp. 145-159, 1993.

[162] A. D. Brown, G. C. Stecker and D. J. Tollin, "The precedence effect in sound localization," *J. Assoc. Res. Otolaryngol.,* vol. 16, pp. 1-28, 2015.

[163] "MRI Simulator," SLEIC, Penn State, 11 February 2018. [Online]. Available: http://imaging.psu.edu/facilities/3t-mri/equipment/mri-simulator. [Accessed 11 February 2018].

[164] "Response Grip," NordicNeuroLab, 20 November 2017. [Online]. Available: http://www.nordicneurolab.com/products/ResponseGrip.html. [Accessed 11 February 2018].

[165] "STAX SR-207," Altei, 11 February 2018. [Online]. Available: https://www.altei.cz/stax-sr207/. [Accessed 11 February 2018].

[166] "Sensimetrics Model S14 earbuds," Sensimetrics, 11 February 2018. [Online]. Available: http://www.sens.com/products/model-s14/. [Accessed 11 February 2018].

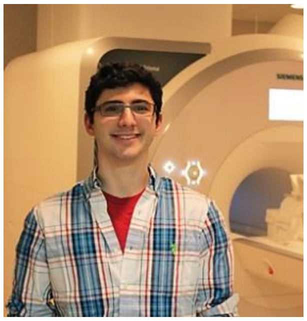
[167] "VisuaStimDigital," Resonance Technology, Inc., 11 February 2018. [Online]. Available: http://www.mrivideo.com/visuastimdigital.php. [Accessed 11 February 2018].

[168] G. Plenge, "On the differences between localizaton and lateralization," *J. Acoust. Soc. Am.,* vol. 56, no. 3, pp. 944-951, 1974.

[169] M. E. Halpern, O. Güntürkün, W. D. Hopkins and L. J. Rogers, "Lateralization of the vertebrate brain: taking the side of model systems," *J. Neurosci.,* vol. 45, pp. 10351-10357, 2005.

VITA

Martin S. Lawless



Martin pursued his undergraduate education in New York City where he received a B.E. in Mechanical Engineering *Summa Cum Laude* from the Cooper Union for the Advancement of Science and Art. While there, he developed and taught a Musical Instrument Design class for his senior design project. The goal of the course was to provide an interdisciplinary educational experience for engineers, artists, and architects with the three diverse perspectives being united by music and acoustics. During undergrad, Martin twice interned at Pratt & Whitney, a jet engine company located in East Hartford, CT. At the start of graduate school in 2013, Martin pursued a Ph.D. in Acoustics from Penn State, researching the brain's response to concert hall acoustics using functional magnetic resonance imaging (fMRI). He also studied the effects of annoying and distracting office noise on cognitive performance and physiological response. During graduate school, Martin was an active member of the Acoustical Society of America (ASA), where he was the Student Council representative for the Technical Committee on Musical Acoustics (TCMU). He also participated as a member of the Students Meet Members for Lunch (SMMfL) subcommittee of the Committee for Education in Acoustics and on the Strategic Task Force 2 committee on Membership Engagement and Diversity. Martin's technical interests include psychoacoustics, functional neuroimaging, music, education, physics, architectural acoustics, noise, and vibration.