INVESTIGATION OF THE EFFECTS OF ROOM ACOUSTICS STIMULI ON REWARD REGIONS IN THE BRAIN.

MS Lawless The Graduate Program in Acoustics, The Pennsylvania State University, University Park, Pennsylvania, USA
MC Vigeant The Graduate Program in Acoustics, The Pennsylvania State University, University Park, Pennsylvania, USA

1 INTRODUCTION

Typical room acoustic studies rely on subjective perceptions to recommend architectural design through the use of surveys,[1] formal listening tests,[2] or sensory analysis techniques.[3] Subjective data heavily depend on the construct,[4] i.e. the conceptual theory being tested. With an abstract or hard-to-define construct, this type of data can fail to be both valid and reliable. Functional neuroimaging techniques, such as functional magnetic resonance imaging (fMRI), may serve as a tool to obtain objective data that may provide more reliable data than subjective results. The brain images may potentially reveal perceptions and cognition that would not be apparent in the conscious, subjective ratings.

Two studies that explore how individual preferences to room acoustics relate to the emotional response in the brain are described in this proceedings paper. An fMRI pilot study was conducted to determine if pleasing or unpleasant simulated room acoustics stimuli engage emotional regions in the brain, where the stimuli are binaural room impulse responses with a range of reverberation times that were convolved with a musical excerpt. The goal of the second study is to establish a set of room acoustics stimuli using conventional listening test methods that would effectively activate such regions for a future fMRI study.

2 NEUROIMAGING (BRAIN IMAGING) PILOT STUDY

2.1 Introduction to Neuroimaging Pilot Study

The goal of the neuroimaging pilot study was to examine individual emotional responses to concert hall stimuli. Previous work has been conducted using EEG, and demonstrated neural correlates to pleasurable room acoustics.[5] However, the limitations of EEG include the inability to measure brain activity beyond the outer cortex, whereas fMRI can be used to image the entire brain and has the spatial resolution to identify activations in specific regions. To the authors' knowledge, no previous studies have been conducted using fMRI. Accordingly, the simple, well-defined construct of preference was considered. Since reverberation time (RT) has been shown to be a major component of preference (e.g. Ando (1985),[6] Beranek (1992),[1] Kuttruff (2000)[7]), it was chosen as the independent variable to alter in the simulations to influence preference.

The authors hypothesized that the stimuli rated as highly preferable would engage regions of the brain associated with positive emotions and reward. Such regions include the nucleus accumbens (NAcc) in the ventral striatum,[8-10] and the caudate of the dorsal striatum.[11,12] Conversely, the stimuli that the participants disliked were expected to activate negative emotional regions such as the amygdala and parahippocampal gyri.[9,13] The activations were quantified using fMRI, which measures the blood-oxygen-level-dependent (BOLD) response.[14]

Testing took place over two one-hour sessions. An MRI simulator made by Psychology Software Tools, Inc. was used in the first session to acquaint subjects with the unusual, loud, and physically restricting MRI environment (see Fig. 1a). The second session occurred in a Siemens 3T MAGNETOM Trio™ MRI scanner to acquire the brain images (see Fig. 1b). The same stimuli and

(a) (b)

Figure 1: (a) The MRI simulator by Psychology Software Tools Inc. artificially reproduces the environment of an MRI scanner, including the loud, tonal noise. (b) The Siemens 3T MAGNETOM Trio™ MRI scanner located at the Social, Life, and Engineering Sciences Imaging Center (SLEIC).

procedure, outlined in Section 2.2, were presented during both sessions of testing in order to validate the preference ratings obtained during the MRI session.

2.2 Description of the Stimuli and Presentation Method in the MRI Scanner

Stimuli consisted of five auralizations and five noise-matched stimuli, with the latter serving as controls for contrasts to isolate the emotional activations in the brain. The auralizations were generated by convolving an 11.2 s anechoic excerpt of Bruckner's 8th Symphony[15] with five different binaural impulse responses (IRs) with RTs of 0.0 s, 1.2 s, 1.4 s, 2.7 s, and 5.3 s. The IRs were derived from one source-receiver combination in the Odeon v12.00[16] model of the Boston Symphony Hall (BSH), shown in Figure 2. The five noise stimuli were created by shaping the spectral content of white noise to match the spectral content of each auralization (or "music stimulus"). The rhythm of each noise stimulus was then fitted to the rhythm of the corresponding music stimulus using a temporal envelope of the music stimulus. The noise stimuli matched the music stimuli in terms of frequency content and rhythm with the expectation that the fMRI contrasts would remove activations in regions of the brain involved in the processing of these sound characteristics.†

The participants were presented with one stimulus at a time and asked to rate each music stimulus in terms of overall preference. This successive method was used because the fMRI analysis requires that each stimulus be played to completion in order to draw correlations between stimuli and brain images. Also, the measured BOLD signal needs time to fully develop and decay before the presentation of the next test stimulus. The music stimuli presentations were alternated with noise stimuli to allow adequate time for the BOLD signal due to the response of the auralizations to decay. For this study, the participants were allotted 10 seconds to rate each stimulus so that the brain images could be aligned with the onset of stimuli presentation. A custom-built Python program was used to present the stimuli; example screens of the graphical user interface are shown in Figure 3.

A 5-point continuous Likert scale was used and was anchored from "Strongly Dislike" (-2) to "Strongly Like" (+2), with the middle of the scale representing "Neither Like nor Dislike" (0). The test included 4 sets of 20 stimuli: each music and noise stimuli were presented twice in a random order. STAX MR 003-MK2 MRI-compatible electrostatic earbuds were used to present the binaural stimuli.

---

† fMRI contrasts are the subtraction of the brain images resulting from two separate stimuli. Since the brain is continuously active, even during rest, non-contrasted images do not reveal significant activations correlating with the stimuli. Control stimuli are used to engage regions of the brain similar to those activated by the test stimuli. The contrast analysis would then eliminate the activations in the analogous regions and potentially isolate activations in regions of interest.

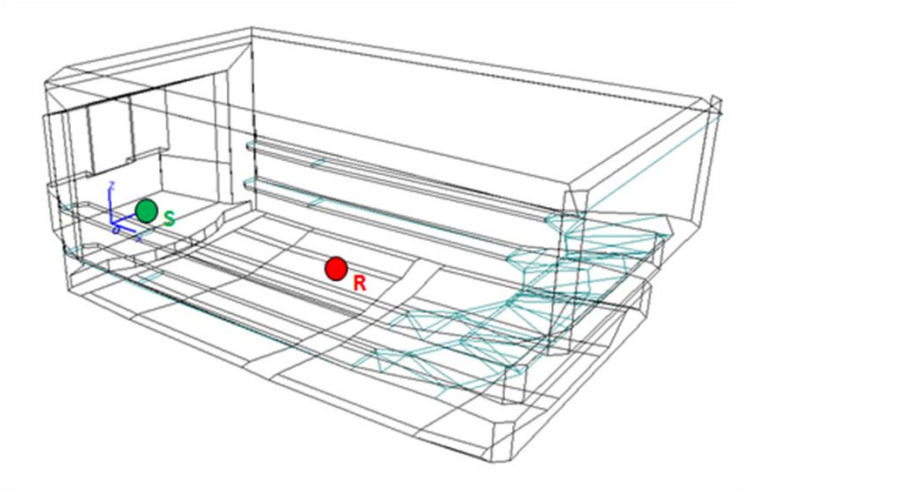

Figure 2: Boston Symphony Hall (BSH) model in Odeon v12.00.[16] The source and receiver positions used for the auralizations are noted in green and red, respectively.

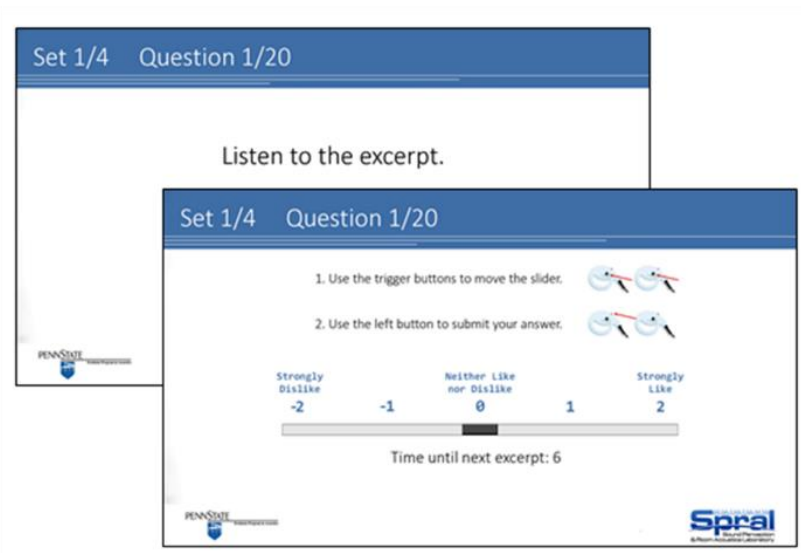

Figure 3: Images of the custom-built Python stimulus presentation program of the successive method. The subjects used Nordic Neurolab Response Grips (illustrated above) to move the cursor and submit a preference rating response. Stimulus presentation time was 16 seconds, and reply time was 10 seconds.

2.3 Subjects

Five healthy, right-handed subjects (2 male and 3 female, 21-39 years) participated to the study. All subjects were musicians actively participating in an ensemble or lessons with an average of 18 years of formal training. Prior to the testing, the participants were screened to have a maximum hearing threshold of 15 dB HL for octave band frequencies from 250-8000 Hz. The testing protocol (#44471) was approved by the Internal Review Board at The Pennsylvania State University.

2.4 Functional Magnetic Resonance Imaging (fMRI) Acquisition

The neuroimaging was performed at the Social, Life, and Engineering Sciences Imaging Center (SLEIC) at The Pennsylvania State University. The MRI scanner utilized a 12-channel head coil to

collect data. A turbo-flash MPRage sequence (Table 1) was used to obtain the structural aspects of the brain, which are used to identify the areas of activations detected in the functional data. The functional images were obtained using an echo-planar imaging (EPI) protocol (see Table 1). The presentation of the stimuli was time-aligned with the EPI sequence, and both had a total length of 8 minutes and 44 seconds.

2.5 Image Processing and Statistical Analysis

The brain images were processed using the SPM8[18] platform in MATLAB. A three-dimensional coordinate reference system with six degrees of freedom was used first to realign the functional images. Nuisance regressors, factors such as subject motion that unpreventably confound the data, were derived from the realignment process and were examined to ensure that each subject moved no more than 1 mm in the x-, y-, and z-directions, nor more than 1° in the pitch, yaw, and roll rotations. (A sixth subject's data was excluded from the study due to movements exceeding these requirements.) Slice-timing via sinc-interpolation was then applied to the functional images. The structural images were spatially normalized into 1-mm isotropic voxels, and the functional images were normalized into 3-mm isotropic voxels. Lastly, the functional images were smoothed with an 8-mm FWHM Gaussian kernel filter.

The preference ratings across the five subjects were evaluated using a repeated-measures analysis of variance (ANOVA)[19] in SAS 9.4. To validate the data obtained during the fMRI scans in the actual scanner, the interactions between the preference ratings obtained in the scanner and in the mock scanner were included in the model. A classical general linear model mass-univariate approach in SPM8 was used to analyse the functional data. The block-design matrix consisted of 11 conditions: the five music stimuli, the corresponding noise stimuli, and the state of rating the stimuli. The nuisance regressors of movement vectors were also included in the model to account for the confounding stimulation of motor regions. Nine linear contrasts comparing the response of one stimuli to a second stimuli, denoted as *Condition 1* > *Condition 2*, were statistically evaluated using t-tests: one between each of the music and noise counterparts (five total) and the two highest-rated stimuli against the two lowest-rated stimuli (four total) as determined by the ANOVA. In the first-level, analysis, a region of interest (ROI) analysis[20] was conducted using a 25-mm sphere centred around (0,14,4) mm MNI coordinates.[21] This sphere encompasses the structures associated with the anticipation and experience of reward.

2.6 Results

2.6.1 ANOVA of Preference Ratings

In the ANOVA, the preference ratings from the five subjects were averaged for each RT. The interaction between ratings and testing session (simulator or actual scanner) had a p-value of 0.304, indicating that participants rated the stimuli similarly in the two sessions. Thus, the results obtained in the scanner were validated with those of the mock scanner, which suggests that preference ratings

Table 1: Specifications for the turbo-flash and EPI sequences used to obtain structural and functional data, respectively. The flip angle for the EPI sequence was calculated to be the Ernst angle for the maximization of signal based on the TR.[17]

	Turbo-flash	*EPI*
Number of Whole-brain Images	1	260
Number of Slices per Image	160	33
Slice Thickness (mm)	1.0	3.3
Repetition Time (TR, ms)	1650	2000
Echo Time (TE, ms)	2.03	28.0
Field of View (mm)	256	210
Matrix Size	256 x 256	64 x 64
Flip Angle (°)	9	78

from mock scanner sessions are sufficient to characterize the subjective response of individuals. Future brain imaging acquisitions may omit the rating stage of stimuli presentation to prevent additional visual and movement confounders.

Tukey comparisons of the RT levels were drawn ($p < 0.001$ from the ANOVA). The 1.2 s and 1.4 s stimuli had statistically indistinct ratings, and were considered the "Most Liked" stimuli with average ratings of 0.9 and 1.0, respectively. Conversely, the lowest-rated stimuli with average ratings of -1.1 and -0.8, constituting the "Most Disliked," had RTs of 0.0 s and 5.3 s, as hypothesized. Lastly, the 2.7 s stimulus was given a neutral average rating of 0.0.

2.6.2 Brain Activations of Preferred Stimuli

The activations in the first-level individual analysis were found by searching for voxels (three-dimensional pixels) with a maximum p-value threshold of 0.001 and minimum cluster size of 20 voxels. After applying a family-wise error rate (FWE) correction for multiple comparisons, significant activations were determined to be those clusters with $p < 0.05$.

For two of the five individuals, positive activations were found for the contrast between one of the "Most Liked" stimulus and noise counterpart. The ROI analysis for the *RT 1.4 Music* > *RT 1.4 Noise* contrast for these subjects indicated that these subjects experienced activations in the caudate, located in the dorsal striatum. The results for these subjects are shown in Figure 4. These activations, in accordance with previous studies, suggest that these two individuals experienced an anticipation to reward during the presentation of the RT 1.4 s stimulus over the presentation of the corresponding noise stimulus. It was hypothesized that all of the music stimuli would have produced a reward-like response in contrast to the noise stimuli, but only the response to the highest-rated stimulus demonstrated significant activations. This result shows that pleasing room acoustics, not simply music as shown in prior studies,[8-13] can elicit reward anticipation. One participant did experience deactivations in the ventral striatum for the same contrast in the ROI analysis, which was unexpected. The inclusion of more subjects in future studies may help to determine if this result is valid or an outlier.

None of the participants exhibited activations in the negative emotional regions for the "Most Disliked" stimuli contrasts, nor any other. Subjects may not have had a strong emotional response due to the inefficacy of the stimuli to elicit a strong response. Previous studies have used subject-chosen music to ensure that the subjects experienced visceral responses.[8,11] Since the stimuli were not selected by each subject individually, the stimuli in the present study may have been unable to elicit strong emotional responses. The goal of a follow-on study is to use a set of stimuli with the capacity to incite both stronger positive emotions than found in this study and also stimulate a negative emotional response. Establishing such a set of experimenter-chosen stimuli is the purpose of the study detailed in the next section.

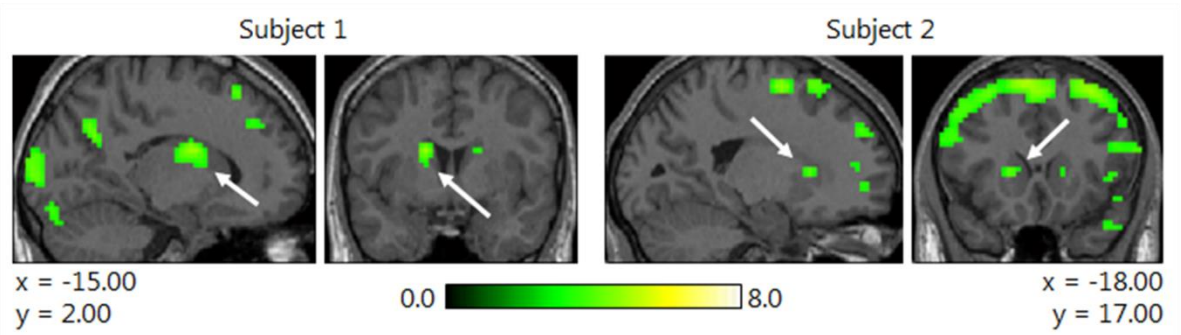

Figure 4: Positive activations in the caudate were found in Subjects 1 ($p = 0.001$, corrected) and 2 ($p = 0.033$, corrected) for the *RT 1.4 Music* > *RT 1.4 Noise* contrast.

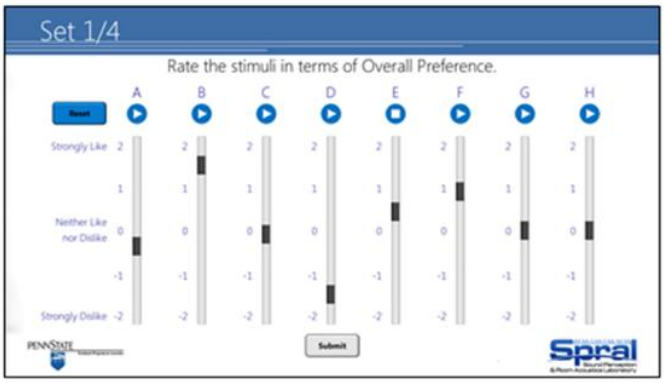

Figure 5: Image of the multi-stimulus comparative method presented by the custom-built Python program. Playback of the musical passages was continuous and the participants could play any particular stimulus multiple times.

3 SUBJECTIVE LISTENING STUDY

The purpose of the subjective listening study was to test a larger number of subjects to accurately characterize preferences to several room acoustics stimuli with a range of RTs and two motifs. A set of stimuli that effectively engages emotional regions of the brain can be created based on the preference ratings to the room acoustics settings and differences between the motifs. The stimuli in this study were simulations generated by varying the material properties in the Odeon model of the BSH (see Fig. 2), which were adjusted to achieve seven RTs ranging from 1.0 to 7.2 s. The simulated binaural room impulse responses were convolved with two solo-instrument motifs: the Trumpet Polka from the Archimedes project CD,[22] and a selection from the first violin part only of Mahler's Symphony No. 1 from Aalto University.[15] In addition, the anechoic excerpt was used as a stimulus (RT=0.0s), resulting in a total of eight stimuli per musical motif.

Thirty-six participants (18 male and 18 female, 18-47 years; 18 musicians and 18 non-musicians) rated four sets of eight stimuli in a multi-stimulus comparative test. In this method, all of the stimuli in a set were presented on the screen at the same time (Fig. 5). The participants were allowed to switch between stimuli in a continuous manner such that the musical passage kept playing, but the room acoustics changed. The rating scales for each of the stimuli were also present on the screen to promote immediate, comparative evaluation of stimuli. As in the previous study, the scales ranged from -2 to +2, representing "Strongly Dislike" to "Strongly Like." Each set contained the eight auralizations from each musical excerpt exclusively, which were presented over STAX SR-207 electrostatic earspeakers.

A two-factor, repeated-measures ANOVA was performed on the subjective ratings using SAS 9.4. The statistical model included RT, motif, and the interaction between these two factors. Using a Tukey comparisons method on the data from both motifs, the stimuli with RTs ranging from 1.2 to 2.8 s were found to be the "Most Liked" stimuli, with average ratings of approximately +0.6 as shown in Figure 6. These ratings were not observed to be statistically different on a 95% confidence interval. Conversely, the 0.0 s and 7.2 s stimuli were determined to be the "Most Disliked" stimuli with statistically indistinct ratings of -0.75 and -0.85, respectively. The interaction between RT and motif was not found to be significant. Minimal differences in preference ratings were found between the two motifs, with the exception of the 1.7 s stimuli, which was found to be more favorable for the violin motif, and the 7.2 s stimuli, which was more preferable for the violin motif. In other words, it is not apparent which motif might induce a stronger emotional response.

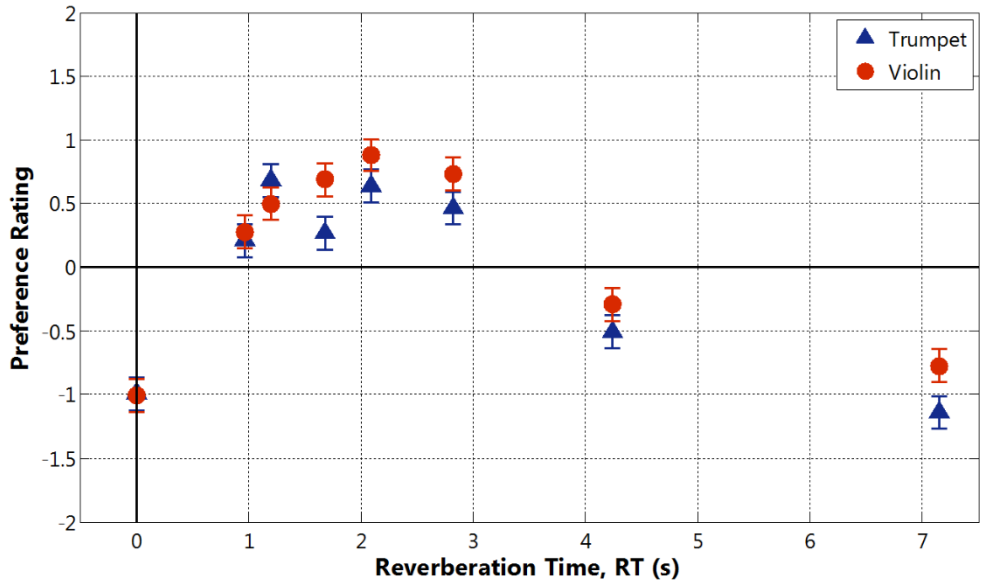

Figure 6: Preference ratings as a function of RT for trumpet and violin motifs. On average, the participants favored stimuli with a moderate amount of reverberation, and disliked dry and excessive reverberant stimuli.

Minimal differences in ratings between the two motifs may have been a result of averaging the ratings across all participants, which effectively suppresses differences in individual preferences. Patterns emerged within participants that shared similar affinities for room settings. A k-means clustering technique[23] was used to distribute the 36 participants into 4 unique groups as shown in Figure 7.The clustering analysis revealed that most participants prefer concert halls with RTs between 1.6 s and 2.1 s, which agrees with a previous review of subjective studies.[7] Two cluster groups, denoted by "RT 1.5 s" and "RT 2.5 s" in Figure 7, reflected this prediction with preferences for stimuli with moderate RTs. The "RT 0.5 s" group contained individuals who preferred dry, clear stimuli. Lastly, participants that did not demonstrate a statistically significant difference between "Liked" stimuli were gathered into the "No Preference" cluster. As seen in Table 2, for the trumpet motif, almost half of the non-musicians fell into the "RT 1.5 s" group and five subjects exhibited no preference. Non-musicians were more evenly grouped for the violin motif. On the other hand, musicians more clearly demonstrated their preferences for either dry or reverberant stimuli. A recent study found that normal hearing (NH) non-musicians strictly preferred anechoic stimuli, whereas NH musicians either preferred anechoic stimuli or moderately reverberant stimuli.[24] These findings differ from those found in the present study, which show that non-musicians fall into a number of preference groups and do not solely favor anechoic stimuli.

The results of the clustering analysis suggest that one set of stimuli cannot elicit potent emotional responses across all participants. In order to increase the likelihood of evoking both strong reward and negative emotional responses in a future fMRI test, the stimuli sets could be personalized to each subject. As a result, for future fMRI studies, it may be beneficial to pre-screen subjects to determine individual preferences.

Table 2: Cluster analysis groupings subdivided by motif and musicianship.

Group	*Musicians / Non-musicians*	
	Trumpet	**Violin**
RT 0.5 s	6 / 2	3 / 3
RT 1.5 s	3 / 8	6 / 4
RT 2.5 s	9 / 3	7 / 6
No Pref.	0 / 5	2 / 5

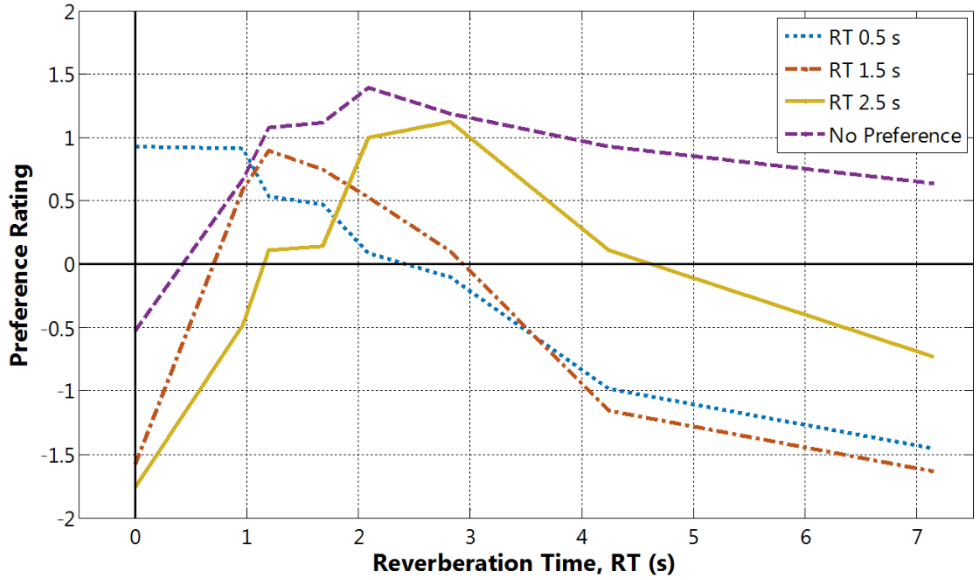

Figure 7: Four groups identified using k-means clustering with affinities for little to no reverberation, moderate reverberation, or no preference. The curves represent the average preference ratings at the respective RT for each group.

4 CONCLUSION

Two studies have been presented that examined the individual emotional response to room acoustics using objective and subjective measures. In the first experiment, a pilot study was conducted using fMRI to investigate the correlation between stimuli with pleasing and unpleasing room acoustics and activations in emotional regions of the brain. Of the five participants, two demonstrated significant activations in the dorsal striatum, specifically the caudate, during the presentation of a preferred stimulus. This region has been shown to correlate with the anticipation of reward. Since only two of the participants exhibited these activations and no negative emotional region activations were observed in any participant, it was determined that more powerful stimuli may be needed to engage emotional response.

The second study aimed to build such a set of stimuli based on the responses from a large number of participants. Two solo-instrumental motifs under several stimulated acoustical conditions were rated by participants. Averaging the whole group of participants together did not delineate the two motifs, but a k-means clustering analysis revealed four groups of participants with different affinities for room acoustics. To more fully develop the stimuli set, orchestral motifs will be considered in additional subjective testing.

A subsequent neuroimaging study will utilize the developed reverberant stimuli sets to elicit strong responses in positive and negative emotional regions. A pre-screening process will determine each participant's room acoustics and possibly motif preference group. Further fMRI experiments will be conducted examining less well-defined variables such as listener envelopment in an attempt to gain objective data on a subjective experience.

5 ACKNOWLEDGEMENTS

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.

6 REFERENCES

1. L. Beranek, "Concert hall acoustics," *J. Acoust. Soc. Am.*, **92**(1), 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.*, **56**(4), 1195-1201 (1974).
3. T. Lokki, J. Pätynen, A. Kuusinen, V. Heikki and S. Tervo, "Concert hall acoustics assessment with individually elicited attributes," *J. Acoust. Soc. Am.*, **130**(2), 835-849 (2011).
4. R. DeVellis, *Scale Development: Theory and Applications*, (Sage, Newbury Park, CA, 1991).
5. Y. Ando and P. Cariani, *Auditory and visual sensations*, (Springer, New York, 2009).
6. Y. Ando, *Concert hall acoustics*, (Springer-Verlag, Heidelberg, 1985).
7. H. Kuttruff, *Room Acoustics*, Fourth Edition, (Spon Press, London, 2000).
8. A. Blood and R. Zatorre, "Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion," *PNAS*, **98**(20), 11818-11823 (2001).
9. S. Koelsch, T. Fritz, D. v. Cramon, K. Muller and A. Friederici, "Investigating emotion with music: an fMRI study," *Human Brain Mapping*, **27**, 239-250 (2006).
10. V. Menon and D. Levitin, "The rewards of music listening: Response and physiological connectivity of the mesolimbic system," *NeuroImage*, **28**, 175-184 (2005).
11. V. Salimpoor, M. Benovoy, K. Larcher, A. Dagher and R. Zatorre, "Anatomically distinct dopamine release during anticipation and experience of peak emotion to music," *Nat. Neurosci.*, **14**(2), 257-262 (2011).
12. V. Salimpoor and R. Zatorre, "Neural interactions that give rise to musical pleasure," *Psychology of Aesthetics, Creativity, and the Arts*, **7**(1), 62-75 (2013).
13. A. Blood, R. Zatorre, P. Bermudez and A. Evans, "Emotional responses to pleasant and unpleasant music correlate with activity in paralimbic brain regions," *Nat. Neurosci.*, **2**(4), 382-387 (1999).
14. R. Buxton, *Introduction to Funcitonal Magnetic Resonance Imaging: Principles and Techniques*, (Cambridge UP, Cambridge, 2009), pp. 114-118.
15. J. Pätynen, V. Pulkki and T. Lokki, "Anechoic Recording System for Symphony Orchestra," *Acta Acustica united with Acustica*, **94**(6), 856-865 (2008).
16. C. Christensen, *Odeon Room Acoustics Program: Version 12.0 User Manual*, (Odon A/S, Lyngby, 2007).
17. R. Ernst and W. Anderson, "Application of Fourier Transform Spectroscopy to Magnetic Resonance," *Rev. Sci. Instruments*, **37**(1), 93-102 (1966).
18. Wellcome Trust Centre for Neuroimaging, "SPM8," 2014. [Online]. Available: http://www.fil.ion.ucl.ac.uk/spm/software/spm8/. [Accessed 6 March 2015].
19. M. Kutner, C. Nachtsheim, J. Neter and W. Li, *Applied Linear Statistical Models*, (McGraw-Hill Irwin, Boston 2005).
20. R. Poldrack, J. Mumford and T. Nichols, *Handbook of Functional MRI Data Analysis*, (Cambridge UP, New York, 2011).
21. 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.
22. "Bang & Olufsen: Music from Archimedes. Audio CD 101," 1992.
23. D. Pollard, "Strong consistency of k-means clustering," *Annals of Stat.,* **9**(1), 135-140 (1981).
24. A. Roy, M. Vigeant, T. Munjal, C. Carver, P. Jiradejvong and C. Limb, "Reverberation negatively impacts musical sound quality for cochlear implant users," *Cochlear Implants International*, **16** (2015).