12-1-2022

Dimensionality of Natural Auditory Scene Perception: A Factor Analysis Study

Margaret A. Mcmullin

DIMENSIONALITY OF NATURAL AUDITORY SCENE PERCEPTION: A FACTOR ANALYSIS STUDY

By

Margaret A. McMullin

Bachelor of Science – Psychology
University of Wisconsin, Parkside
2017

A thesis submitted in partial fulfillment
of the requirements for the

Master of Arts – Psychology

Department of Psychology
College of Liberal Arts
The Graduate College

University of Nevada, Las Vegas
December 2022

UNLV | GRADUATE COLLEGE

Thesis Approval

The Graduate College
The University of Nevada, Las Vegas

December 12, 2022

This thesis prepared by

Margaret A. McMullin

entitled

Dimensionality of Natural Auditory Scene Perception: A Factor Analysis Study

is approved in partial fulfillment of the requirements for the degree of

Master of Arts – Psychology
Department of Psychology

Joel Snyder, Ph.D.
Examination Committee Chair

Erin Hannon, Ph.D.
Examination Committee Member

Colleen Parks, Ph.D.
Examination Committee Member

Beiyu Lin, Ph.D.
Graduate College Faculty Representative

Alyssa Crittenden, Ph.D.
Vice Provost for Graduate Education &
Dean of the Graduate College

Abstract

Theories of auditory and visual scene analysis suggest the perception of scenes relies on the identification and segregation of objects within it, resembling a detail-oriented processing style, but it is possible that a global-oriented process also occurs while evaluating auditory scenes. There is evidence for global properties that enable rapid recognition of visual scenes, even without recognizing the individual objects comprising the scene. It is our understanding that a similar line of research has not been explored in the auditory domain; therefore, we evaluated the contributions of high-level global and low-level acoustic information to auditory scene perception. A secondary aim is to increase the field's ecological validity by utilizing our collection of high-quality auditory scenes. Participants rated scenes on 8 global properties (e.g., open vs. enclosed) and an acoustic analysis evaluated which low-level features predicted the ratings. We submitted the acoustic measures and average ratings of the global properties to separate exploratory factor analyses (EFAs). The EFA of the acoustic measures revealed a seven-factor structure explaining 57% of the variance in the data, while the EFA of the global property measures revealed a two-factor structure explaining 64% of the variance in the data. Regression analyses revealed each global property was predicted by at least one acoustic variable (R-squared = 0.33-0.87). These results provide evidence for the ability to perceive auditory scenes from a global perspective. Some of the acoustic measures predicted ratings of global scene perception, suggesting representations of auditory objects may be transformed through many stages of processing in the ventral auditory stream, similar to what has been proposed in the ventral visual stream. These findings and the open availability of our scene collection will make future studies on perception, attention, and memory for natural auditory scenes possible.

Keywords: auditory scene perception, acoustic analysis, natural scenes

Acknowledgements

Thank you to my advisor, Dr. Joel Snyder, for providing support and guidance throughout this project. Your patience and insight have made this an invaluable learning experience.

Thank you to my committee members, Dr. Colleen Parks, Dr. Erin Hannon, and Dr. Beiyu Lin, for help with designing this study and for providing feedback on this document. Your expertise served to better this project.

Thank you to all past and current members of the Auditory Cognitive Neuroscience Lab and the Auditory Cognition Development Lab for their support over the past four years. In particular, I would like to thank Dr. Nathan Higgins, Chastity Balagtas, Cassy Doolittle, Sarah Flood, Lan Nguyen, Lucy Pineda-Roman, Jared Leslie, Dr. Karli Nave, and Dr. Jessica Nave-Blodgett for their enthusiastic support and assistance with collecting and editing auditory scenes for our database.

Thank you to the Department of Defense National Defense Science and Engineering Graduate Fellowship program for supporting my research.

Thank you to Dr. Melissa Gregg at the University of Wisconsin, Parkside for introducing me to the world of auditory cognitive neuroscience and supporting my pursuit of attending graduate school.

Thank you to my family and friends – Katie McMullin, Maci DeOcampo, Kade McMullin, Gia DeOcampo, Rylan DeOcampo, Shannon Allegra, Kendra Bittner, and Kevin Mohawk – for their continuous support and laughs.

Thank you to Rodica Constantine for her unwavering support, keen insights, and never-ending laughs. I am forever grateful for our friendship.

Finally, I would like to thank my Mom and Dad for teaching me how to work hard and dream big.

Table of Contents

List of Tables

Chapter 1: Introduction

Every day, our auditory system undertakes the complex task of organizing various incoming sounds in a coherent manner, allowing us to not only decipher where sounds are coming from, but to also interpret *what* we are listening to. For example, when conversing with a friend at a noisy café, the auditory system maintains the exceptional ability to segregate the noisy background (e.g., music, espresso machines, other conversations) from your friend's voice and further group the sound components of their speech into an intelligible stream of words. The process of perceptually segregating and grouping numerous acoustic objects is known as 'Auditory Scene Analysis' (ASA; Bregman, 1990).

Historically, theories of both auditory and visual scene analysis have suggested that our perception of a scene relies on the identification and segregation of multiple objects within it, resembling a detail-oriented processing style (Bregman, 1990; Biederman, 1987). However, it is possible that a more global process may also occur when observers evaluate auditory scenes. In the visual modality, there is evidence for global properties that enable visual scenes to be rapidly recognized, even without recognition of individual objects comprising the scene (Greene & Oliva, 2006; Greene & Oliva, 2009a; 2009b; Ross & Oliva, 2010). The significance of the scene-centered approach proposed by Greene and Oliva (2009) is that the representation composed by the visual system exists at the level of the entire scene and not individual objects. Instead of building a visual representation using low-level geometric information, the visual system prefers to use high-level global properties that provide information about the scene's structure, function, and overall layout to guide perception. The global properties identified by Oliva and colleagues fall into three categories: structural properties (openness, expansion, mean depth), constancy properties (temperature, transience), and functional properties (concealment, navigability).

Greene and Oliva (2009b) asked participants to view a series of rapidly presented visual scenes ranging in duration from 10 to 200 msec and indicate whether the scenes were consistent with a basic-level category (e.g., identifying a scene as a mountain or waterfall) or a global category (e.g., identifying a scene as an open environment or hot place). The results indicated participants performed more accurately during the global categorization task relative to the basic-level task for the visual scenes of the shortest durations. Other studies have demonstrated the importance of global image features (Greene & Oliva, 2006; Oliva & Torralba, 2006; Ross & Oliva, 2010) and collectively suggest that these global properties could act as automatic heuristics used to analyze natural visual scenes.

Our primary research goal was to evaluate the contribution of high-level semantic knowledge and low-level acoustic information during auditory scene perception. As previously mentioned, Oliva and colleagues have demonstrated that when asked to categorize rapidly presented visual scenes, observers find global information within the scene to be more useful compared to the individual objects that constitute the scene. In the auditory domain, the role of high-level semantic information and low-level acoustic features has been explored in the change deafness literature. Characterized as a perceptual error, change deafness is the inability to detect changes in auditory scenes (Snyder et al., 2012). This error is useful in our study of ASA because it informs us of our auditory system's limitations. To study change deafness, participants are typically presented with a simultaneous array of sounds (e.g., dog barking, piano, phone ringing, bell). After a short interruption (usually white noise), the scene is presented again. Finally, participants must indicate whether the second scene was the same as or different than the first scene. In one such study, Gregg and Samuel (2009) presented participants with two types of *different* trials (i.e., the scene changed in some way from the first to second presentation). These trials exhibited either a within-category change (e.g., a small dog bark changing to a large dog bark) or a between-category change (e.g., a small dog bark changing to a bell chime). Results demonstrated that participants' inability to detect a within-category change was significantly worse than their ability to detect a between-category change. As the low-level acoustic features were controlled for, the authors hypothesized that semantic information is useful when constructing representations of auditory scenes. They further observed that listeners used both high-level semantic information and low-level acoustic information when constructing auditory representations of sounds, but the low-level acoustic information was not used to the same extent as semantic knowledge of sounds.

Additional research has also evaluated the influence of acoustic features of sounds on listeners' ability to identify and discriminate recognizable objects, both in isolation and when presented concurrently with complex auditory scenes. Leech, Gygi, Aydelott, and Dick (2009) addressed the possible existence of semantic knowledge-driven expectancies about auditory scenes. In their study, participants were presented with multiple target sounds embedded into an auditory scene. The target sounds were either congruent (e.g., the target being a goat and the background being a farm) or incongruent (e.g., the target being a goat and the background being a casino) with the auditory scene in which the sounds were embedded. An acoustic analysis of all target sounds and background auditory scenes was conducted to evaluate whether acoustic similarity or dissimilarity between the targets and backgrounds may have influenced target identification. Participants more accurately identified target sounds that were contextually incongruent with the background scenes and the acoustic variables that significantly influenced this effect were correlogram-based pitch measures and peak autocorrelation statistics. However, since acoustic similarity was not an exclusive predictor of target congruency or incongruency with the background scene, the findings from this study suggest that high-level semantic factors may significantly influence listeners' ability to detect and identify meaningful sounds within complex auditory scenes.

Change deafness tasks are highly useful in the study of ASA. However, change deafness is typically examined using a mixture of simultaneously presented sounds (Gregg & Samuel, 2008; 2009). This presents a fundamental limitation to studies of this type: the stimuli are artificial in nature, especially since some of the sound combinations may not typically occur in the real-world. Some work has been done using natural auditory scenes as stimuli, but clips of environmental sounds were superimposed onto the auditory scenes, making the scenes more artificial (Leech et al., 2009). An example of a study using more naturalistic sounds is by McDermott & Simoncelli (2011), which examined sound texture perception using a computational model of the human auditory system. Sound textures, which are the result of numerous similar acoustic events occurring in succession (e.g., rainstorm, galloping horses), were processed using an auditory model based on the tuning properties of neurons from the cochlea to the thalamus. To better understand how sound textures are represented in the brain, the authors then synthesized the sound textures based on the output of their model (i.e., the statistics of real-world sounds). They hypothesized that if the novel synthesized sound textures were statistically matched with those of the real-world sounds, then the brain should be able to achieve texture recognition due to the synthesized signal sounding like a version of the originally presented sound texture. In a series of behavioral experiments, they found that synthetic sound textures were recognizable to participants but eliminating some of the statistics in the model reduced performance. Additionally, the authors modified the model so that it was less representative of the mammalian auditory system, which resulted in reduced recognizability of the synthetic sound textures. Some of the synthesized sounds (e.g., wind chimes, tapping rhythm, and a person speaking English) were not recognizable, though. These findings suggest that sound texture perception arises from the recognition of simple statistics in early auditory representations, which are potentially computed in downstream neural populations. Ultimately, the results of this study are important to our understanding of how the auditory systems analyzes naturalistic sounds and could inform us of how the auditory system processes more complex stimuli, like naturalistic auditory scenes. By recording and using naturalistic auditory scene stimuli, we hope to further increase the range of ecologically valid stimuli and abilities under study in the field of auditory perception.

Our secondary research goal was to record and use real-world auditory scenes. One major limitation in the current body of literature is the consistent use of pure tones, noise bursts, or artificially contrived auditory scenes as stimuli to study ASA. While using such stimuli has revealed much about the fundamental mechanisms of ASA and auditory perceptual awareness, the findings resulting from this work has limited power in educating us about natural auditory scene processing. In the field of visual scene perception, the use of naturalistic stimuli is highly evident (Greene & Oliva, 2009a; 2009b; Ross & Oliva, 2010; Greene & Oliva, 2010; Harel et al. 2016; Hansen et al., 2018). This is perhaps the case because there are numerous large databases of natural visual scenes openly available for public use (Xiao, et al., 2010; Geisler & Perry, 2011). While there are some databases that include high quality clips of individual sound objects (e.g., a single dog bark; Gygi & Shafiro, 2010), no database of high quality, real-world auditory scenes currently exists to our knowledge. To address this issue, we recorded a large volume of audio/visual scenes, which will be made available on a public database for other researchers to use.

Purpose of Present Study

The present study aimed to evaluate the contributions of high-level global properties and low-level acoustic features to natural auditory scene perception. Participants listened to 200 auditory scenes and made a series of global property judgments on them. Additionally, we conducted an acoustic analysis on all 200 scenes with the goal of understanding how these features are related to global processing of auditory scenes. We predicted there would be a general consistency on all eight global property ratings of each auditory scene across participants, which was measured using intraclass correlations of each rating scale. We conducted two separate factor analyses on the average global property ratings and acoustic measures of each scene to determine the number of factors that characterize the variability found within scene judgments and within the array of acoustic features. An additional eight multiple linear regression analyses were also conducted to predict performance on the global property rating task based on the acoustic features of the scenes. We did not originally plan on conducting this analysis; however, we decided it was more helpful for directly testing the relationships between the average global property ratings and acoustic measures of each scene.

Chapter 2: Method

Auditory Scene Collection and Database

We collected and processed auditory scenes from various locations (e.g., parks, hiking trails, city streets, cafés, etc.) across the United States. Using a standardized recording procedure, we placed a Zoom Q8 camcorder (Zoom North America, Inc., Hauppauge, NY) on a tripod and made one-minute recordings at each location, noting various aspects of the scene, such as the date, time of day, cardinal direction, temperature (°F), sounds observed, and any additional notes about the recording. After each recording session, the field notes were digitized into a spreadsheet for ease of organization and file identification. Each recording was then listened to and all sounds identified in the field notes were confirmed as being present in the recording. We then edited each minute-long recording into a four-second-long version which best characterized the scene location and also included more than one sound object. Our collection of auditory scenes will be made available for other researchers to use via an online database.

Ethics Statement

All experimental procedures were approved by the University of Nevada, Las Vegas (UNLV) Institutional Review Board. All data collected from participants were anonymized.

Participants

We recruited 68 English-speaking adults (48 female) aged 18-35 (M_{age} = 21.19) with no known hearing, visual, or neurological deficits from the UNLV subject pool and across the United States. Participants from the UNLV subject pool were reimbursed for course credit and participants external to the university volunteered for no compensation. In total, 142 participants were excluded from this study if they did not speak English, did not complete the experiment, did not have normal hearing, had any type of severe neurological or psychiatric disorder (e.g., schizophrenia, bipolar disorder, stroke, traumatic brain injury), or failed the headphone check, compliance check, and/or attention check (see below for descriptions).

Stimuli

Stimuli consisted of 200 naturalistic auditory scenes originating from our database of acoustic scenes. Each scene was four seconds in length and matched for root-mean-square (RMS) amplitude. A linear on-ramp from zero amplitude was imposed on the first and last 10 msec of each sound clip to avoid introducing artifacts due to abrupt sound onsets and offsets. (Gregg & Samuel 2008; 2009, Gregg, Irsik, & Snyder, 2014, Gregg, Irsik, & Snyder, 2017).

Procedure

Participants were provided a link to complete the experiment online via Qualtrics (Qualtrics, Provo, UT; experiment links can be found on our project's Open Science Foundation Repository: https://osf.io/zj4xe/?view_only=). Informed consent was obtained online from each participant before they began the experiment. Participants were asked to complete the study on a desktop or laptop computer using headphones and while in a quiet environment. In total, the experiment took 60-120 minutes to complete. The experiment consisted of four sections: 1) a headphone check, 2) the global property rating task, 3) a compliance and attention check, and 4) a demographic questionnaire.

Headphone Check.

Because this was an online study of auditory perception, we tested each participant's sound quality by administering a headphone check. This test consists of 6 trials of a 3-AFC intensity discrimination task (Woods, Siegel, Traer, & McDermott, 2017). Participants were asked to indicate which tone had the lowest volume by selecting one of three button options labeled “Tone 1”, “Tone 2”, or “Tone 3.” Any participants who did not correctly answer five out of the six trials were excluded from the study.

Global Property Rating Task.

Each participant was asked to judge all 200 scenes on four different global properties on a Likert scale ranging from 1 (lowest extreme) to 7 (highest extreme; See Figure 1 for full description of each rating scale). Participants were allowed to listen to each scene as many times as they needed to make each of the four judgments. To avoid requiring participants to complete a two-day study, the eight category judgments were split amongst participants. The conditions in this experiment followed a Latin square design, with each global judgment type only appearing in each possible position once. Participants were randomly assigned to one of eight condition groups. Groups 1-4 made the following global property judgments on each scene: Open vs. Enclosed, Outdoor vs. Indoor, Natural vs. Human-Influenced, and Temperature. Groups 5-8 made judgments on each scene’s Season, Transience, Navigability, and Sparseness (See Table 1 for order of questions in each condition).

Table 1.

Order of Rating Scale Questions in Each Condition.

Group	Order of Rating Scale Questions			
1	Open vs. Enclosed	Outdoor vs. Indoor	Natural vs. HI	Temperature
2	Outdoor vs. Indoor	Open vs. Enclosed	Temperature	Natural vs. HI
3	Natural vs. HI	Temperature	Open vs. Enclosed	Outdoor vs. Indoor
4	Temperature	Natural vs. HI	Outdoor vs. Indoor	Open vs. Enclosed
5	Season	Transience	Navigability	Sparseness
6	Transience	Season	Sparseness	Navigability
7	Navigability	Sparseness	Season	Transience
8	Sparseness	Navigability	Transience	Season

Note. HI = Human-Influenced

Compliance and Attention Check.

We utilized a set of questions from Mehr, Singh, York, Glowacki, & Krasnow, (2018) to ensure participants were adequately attending to the experimental task. The following question was dispersed throughout the global property rating task: 1) "What color is the sky? Please answer this incorrectly, on purpose, by choosing RED instead of blue.", with the response options of "Green," "Red," "Blue," or "Yellow." The correct response option ("Red") was changed upon each presentation (e.g., the correct response was only presented in each answer slot once). Any participant who did not select this response option was excluded.

Upon completion of the rating task, participants were asked the following compliance questions:

1) "People are working on this task in many different places. Please tell us about the place you worked on this task. Please answer honestly." The response options for this question were: "I worked on this study in a very noisy place, I worked on this study in a somewhat noisy place, I worked on this study in a somewhat quiet place, or I worked on this study in a very quiet place." Any participant who answered with "I worked on this study in a very noisy place" or "I worked on this study in a somewhat noisy place" was excluded.
2) "Please tell us if you had difficulty loading the sounds. Please answer honestly." The response options for this question were "Yes" or "No." Any participant who responded with "No" was excluded.
3) "How carefully did you complete this experiment? Please answer honestly. The response options for this question were: "Not at all carefully," "Slightly carefully," "Moderately carefully," "Quite carefully," or "Very carefully." Any participant who answered with "Not at all carefully," "Slightly carefully," or "Moderately carefully" were excluded.

Demographic Questionnaire.

Lastly, participants completed a demographics questionnaire which asked about their health history and engagement with music. Additional questions were asked about participants' auditory environment (e.g., time spent in various environments). Data collected from this questionnaire has not been included in the analyses reported here.

Figure 1

Global Property Rating Scales.

Open vs. Enclosed						
How OPEN vs. CLOSED is this scene? Please rate this on a scale ranging from 1 (extremely open) to 7 (extremely closed).						
○	○	○	○	○	○	○
Extremely OPEN	Moderately OPEN	Slightly OPEN	Neither OPEN nor CLOSED	Slightly CLOSED	Moderately CLOSED	Extremely CLOSED
Outdoor vs. Indoor						
How OUTDOOR vs. INDOOR is this scene? Please rate this on a scale ranging from 1 (extremely outdoor) to 7 (extremely indoor).						
○	○	○	○	○	○	○
Extremely OUTDOOR	Moderately OUTDOOR	Slightly OUTDOOR	Neither OUTDOOR nor INDOOR	Slightly INDOOR	Moderately INDOOR	Extremely INDOOR
Natural vs. Human-Influenced						
How NATURAL vs. HUMAN-INFLUENCED is this scene? Please rate this on a scale ranging from 1 (extremely natural) to 7 (extremely human-influenced).						
○	○	○	○	○	○	○
Extremely NATURAL	Moderately NATURAL	Slightly NATURAL	Neither NATURAL nor HUMAN-INFLUENCED	Slightly HUMAN-INFLUENCED	Moderately HUMAN-INFLUENCED	Extremely HUMAN-INFLUENCED
Temperature						
How WARM vs. COLD is this scene? Please rate this on a scale ranging from 1 (extremely warm) to 7 (extremely cold).						
○	○	○	○	○	○	○
Extremely WARM	Moderately WARM	Slightly WARM	Neither WARM nor COLD	Slightly COLD	Moderately COLD	Extremely COLD
Season						
What season does this scene sound like?						
○	○	○	○	○	○	○
Winter	Between Winter and Spring	Spring	Between Spring and Summer	Summer	Between Summer and Fall	Fall
Transience						
How much change is happening in this scene? Please rate this on a scale ranging from 1 (no change) to 7 (extreme amount of change).						
○	○	○	○	○	○	○
No Change	Very Little Change	Slight Amount of Change	Moderate Amount of Change	Medium-High Amount of Change	High Amount of Change	Extreme Amount of Change
Navigability						
How difficult or easy would it be to navigate through this scene? Please rate this on a scale ranging from 1 (extremely difficult) to 7 (extremely easy).						
○	○	○	○	○	○	○
Extremely Difficult	Moderately Difficult	Slightly Difficult	Neither Difficult or Easy	Slightly Easy	Moderately Easy	Extremely Easy
Sparseness						
How much sound is in this scene? Please rate this on a scale ranging from 1 (very little sound) to 7 (high amount of sound).						
○	○	○	○	○	○	○
No Sound	Very Little Sound	Slight Amount of Sound	Moderate Amount of Sound	Medium-High Amount of Sound	High Amount of Sound	Extreme Amount of Sound

Acoustic Analysis

To quantitatively gauge how low-level information may be utilized to understand auditory scenes, an extensive acoustic analysis was conducted on all 200 auditory scenes. The acoustic analyses chosen for this study have been used in various prior studies (Houtgast & Steeneken, 1985; Ballas, 1993; Slaney, 1995; Gygi et al., 2007; Leech et al., 2009) and were executed in MATLAB (The MathWorks, Inc., Natick, Massachusetts). A description of each acoustic measure is listed below.

Envelope-Based Intensity and Rhythm Measures

(1) Long-term RMS/Pause corrected RMS, which indicates the amount of silence present within each auditory scene; (2) number of peaks, where a peak is defined as a point in the vector that has a greater amplitude than the previous point by at least 80% of the range of amplitudes present in the vector; (3) number of bursts, showing an increase in amplitude of at least 4 dB lasting at least 20 msec (Ballas, 1993); (4) total duration; and (5) burst duration/total duration, a measure of how "rough" the envelope is.

Autocorrelation Pitch Statistics

(1) Number of peaks; (2) maximum peak; and (3) standard deviation (SD) of the peaks. In this autocorrelation function, the peaks express periodicities in the waveform. The distribution of periodicities across various frequencies as well as the strength of a periodicity are evaluated in this measure.

Correlogram-Based Pitch Measures

(1) mean pitch; (2) median pitch; (3) *SD* pitch; (4) maximum pitch; (5) mean pitch salience; and (6) maximum pitch salience. This measure evaluates pitch and pitch salience by autocorrelating in sliding 16 msec time windows.

Moments of the Spectrum

(1) mean; (2) *SD*; (3) skew; and (4) kurtosis. This measures the distribution of energy related to the overall timbre of the scene.

RMS Energy in Octave-Wide Frequency Bands

Gygi et al., (2007) used frequency bands ranging from 63-16,000 Hz. This measures the distribution of energy separately for different frequencies.

Spectral Shift in Time Measures

(1) Centroid mean; (2) centroid *SD*; (3) mean; (4) *SD*; and (5) maximum centroid velocity. The measures of the centroid mean and *SD* are established on sequential 50-msec time windows throughout the entirety of the waveform, while the measure of spectral centroid velocity is determined by calculating the overall change in the spectral centroid across sliding 50-msec time widows.

Modulation Spectrum Statistics

Proposed by Houtgast and Steeneken (1985), the modulation spectrum displays intermittent temporal variations in the envelope of a scene. This measure "divides the signal into frequency bands approximately one critical band wide, extracts the envelope in each band, filters the envelope with low-frequency bandpass filters (upper f_o ranging from 1-32 Hz), and determines the power at that frequency." The result is a plot of the depth of modulation-by-modulation frequency. The statistics measured here will be the height and frequency of the maximum point in the modulation spectrum, as well as the number, mean, and variance of bursts in the modulation spectrum (using the burst algorithm described above; Gygi et al., 2007, p. 846).

Spectral Flux Statistics

Spectral flux evaluates how much change occurs in the spectrum at various frequency bands. This measure can potentially show salient changes in energy, which can be deduced as moments in time where a change in energy may capture the observer's attention, further influencing their perception of the scene.

Chapter 3: Results

All the auditory stimulus files used in this study as well as the raw and analyzed data can be found on our project's Open Science Foundation Repository (https://osf.io/zj4xe/?view_only=).

Inter-Rater Reliability

To evaluate inter-rater reliability of ratings made by participants on each of the eight global property scales, intra-class correlation (ICC) coefficients and their 95% confidence intervals were calculated using SPSS statistical software version 28 (SPSS Inc., Chicago, IL). We used a two-way random effects model based on average ratings to assess consistency across participants. The ICCs for all global property scales were statistically significant (all p-values < .001) and ranged from good to excellent (0.758 - 0.980; Koo & Li, 2016).

Table 2

Inter-rater Reliability as Measured by Intra-class Correlations (ICCs).

		95% CI			F Test with True Value 0		
	ICC	Lower Bound	Upper Bound	F value	*df*1	*df*2	p value
Sparseness	0.968	0.962	0.974	31.428	199	6965	< .001
Transience	0.944	0.933	0.955	17.919	199	6965	< .001
Season	0.758	0.707	0.804	4.135	199	6965	< .001
Navigability	0.787	0.742	0.827	4.667	199	6965	< .001
Openness	0.925	0.909	0.939	13.288	199	6169	< .001
Outdoor vs. Indoor	0.977	0.972	0.981	43.402	199	6169	< .001
Natural vs. Human-Influenced	0.980	0.975	0.984	49.211	199	6169	< .001
Temperature	0.801	0.760	0.839	5.032	199	6169	< .001

Results of ICC(2,k). Two-way random effects model, consistency definition, average measures. *df* = degrees of freedom.

Correlations between Global Properties

Pearson correlations between average global property ratings of each scale are reported in Table 3. Overall, there are many moderate, strong, and very strong correlations between global property rating scales, which justifies their use in the exploratory factor analysis to determine the underlying factor structure of auditory scene perception.

Table 3

Summary of Means, Standard Deviations, and Correlations Among Average Global Property Ratings

Variable	*M*	*SD*	1	2	3	4	5	6	7	8
1. Natural vs. Human-Influenced	4.55	1.61	—							
2. Openness	3.05	0.95	.61***	—						
3. Outdoor vs. Indoor	3.15	1.55	.63***	.94***	—					
4. Temperature	3.81	0.47	.41***	.44***	.35***	—				
5. Navigability	4.45	0.54	-.25***	-.17*	-.12	-.35***	—			
6. Transience	3.62	0.73	.52***	.14*	.30***	-.03	.24***	—		
7. Season	4.17	0.56	.45***	.07	.17*	.03	.24***	.73***	—	
8. Sparseness	4.12	0.82	.35***	.05	.22**	-.07	.32***	.87***	.73***	—

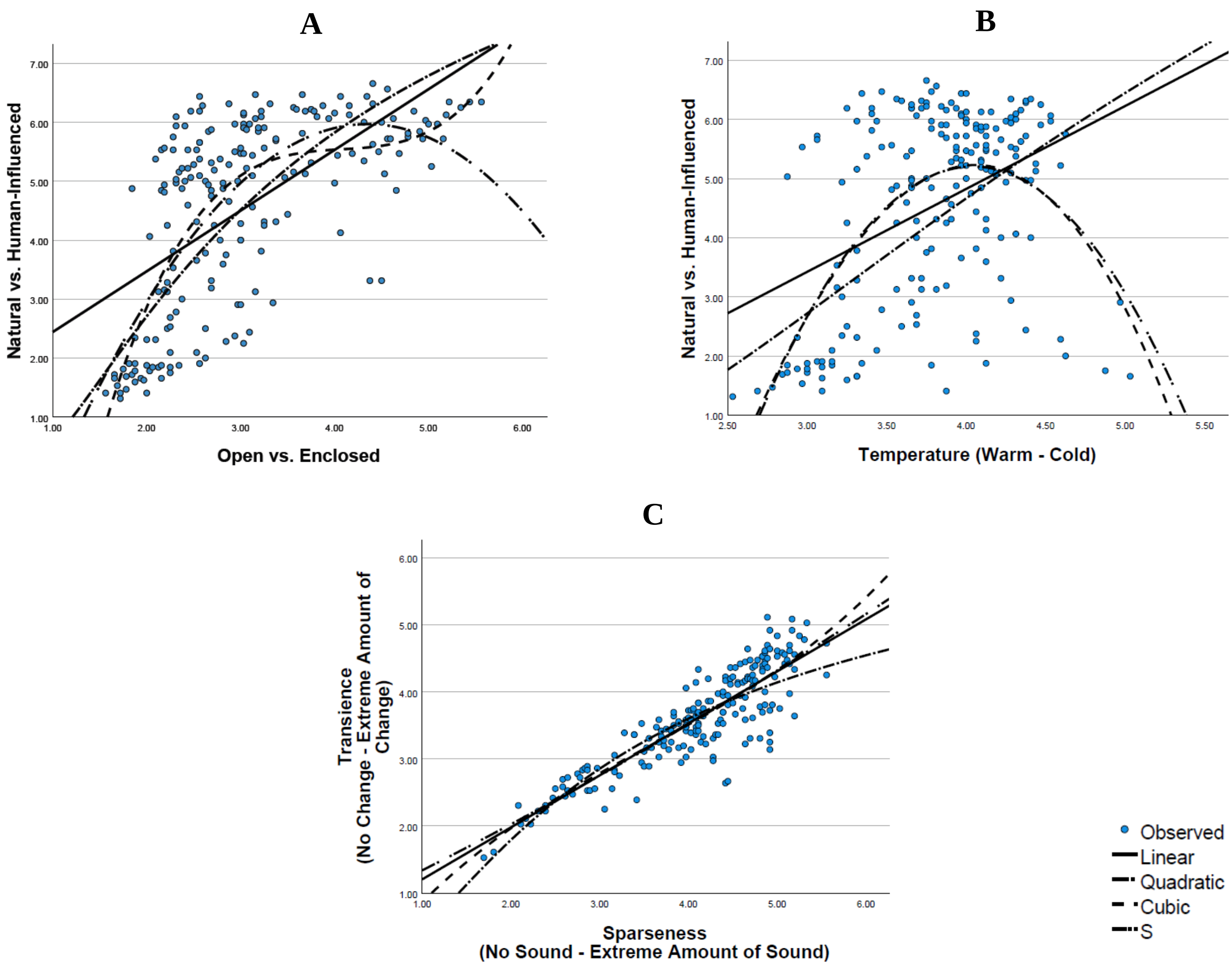
Note. Correlations between average global property ratings for all auditory scenes (n = 200). * $p < .05$, ** $p < .01$, *** $p < .001$

Curve Estimation of Global Property Ratings

We next created scatterplots to evaluate the relationships between each of the global property rating variables. Upon visual inspection of the scatterplots, both linear and non-linear (quadratic, cubic, or sigmoid) relationships were apparent between pairs of variables. Therefore, an additional curve estimation analysis was conducted on each apparently non-linear scatterplot to better characterize these relationships. The results of this analysis showed much of the data having a combination of significant linear, quadratic, cubic, and sigmoid relations, which indicates that many of the relationships between the global property rating scales are non-monotonic and had much variation. An additional measure of local effect size, Cohen's f^2, was also calculated for each type of curve estimation, results showed that some relationships had larger effect sizes for one or two types of curves. Figure 2A depicts a partly cubic relationship between the Natural vs. Human-Influenced and Open vs. Enclosed rating scales ($p < .001$, Cohen's $f^2 = 1.02$), Figure 2B depicts a partly quadratic relationship between the Natural vs. Human-Influenced and Temperature scales ($p < .001$, Cohen's $f^2 = 0.51$), and Figure 2C depicts a linear and sigmoid shaped relationship between the Transience and Sparseness rating scales (linear line: $p < .001$, Cohen's $f^2 = 3.18$; sigmoid curve: $p < .001$, Cohen's $f^2 = 3.48$).

Figure 2

Curve Estimations of Global Property Rating Scales.



Note. (A) The significant cubic relationship between the ratings of all auditory scenes on the Natural vs. Human-Influenced and Open vs. Enclosed scales. (B) The significant quadratic relationship between the Natural vs. Human-Influenced and Temperature scales. (C) The significant linear/sigmoidal relationship between the Transience and Sparseness rating scales.

Exploratory Factor Analyses

To evaluate the dimensionality of scene perception, we submitted the average global property ratings of each scale and all 35 acoustic measures to two separate exploratory factor analyses (EFA) using JASP statistical software version 16.1 (JASP Team, 2022).

Exploratory Factor Analysis: Global Properties

The average global property ratings of each scale (Naturalness, Openness, Sparseness, Navigability, Temperature, Outdoor vs. Indoor, Season, Transience) were entered into an EFA, with maximum likelihood factor extraction and Oblimin (oblique) rotation. The Kaiser-Meyer-Olkin test revealed sufficient sampling adequacy for the EFA, KMO = 0.69. Bartlett's test of sphericity indicated the correlation structure of the variables was adequate for the EFA as well, χ^2 (28) = 1253.84, $p < .001$. Upon visual inspection of the scree plot as well as a parallel analysis (see Figure 3), a two-factor solution was revealed and accounts for 64.0 % of the variance in the data. Table 4 displays the variables and factor loadings, with loadings less than |.40| excluded for clarity.

Factor 1. After rotation, Factor 1 consisted of four variables that accounted for 33% of the variance in the model. The global property variables which loaded onto this factor were Transience (0.94), Sparseness (0.93), Season (0.78). Although Navigability (0.37) also loaded onto this factor, its interpretation should be made with caution as its loading is below the |.40| threshold.

Factor 2. Factor 2 consisted of four variables that accounted for 31% of the variance in model. The global property variables which loaded onto this factor were Openness (1.01), Outdoor vs. Indoor (0.93), Natural vs. Human-Influenced (0.58), and Temperature (0.46).

Figure 3

Scree plot of eigenvalues revealing a 2-factor model.

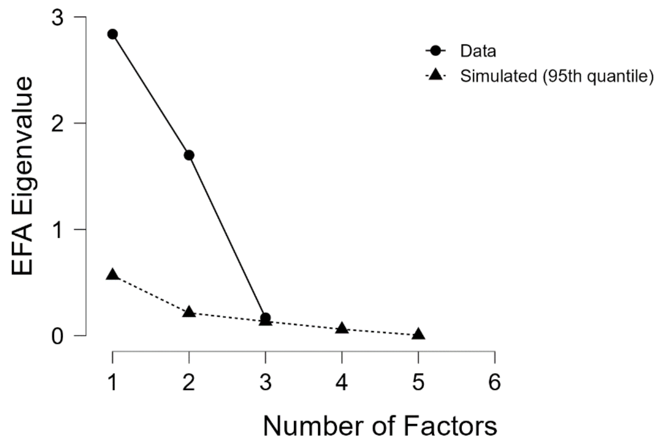


Table 4

Factor Loadings of Global Property Scales.

Factor Loadings	Factor 1	Factor 2
Transience	0.94	
Sparseness	0.93	
Season	0.78	
Navigability	0.37	
Openness		1.01
Outdoor vs. Indoor		0.93
Natural vs. Human-Influenced		0.58
Temperature		0.46

Note. Extraction method: maximum likelihood; Rotation method: Oblimin (oblique).

Exploratory Factor Analysis: Acoustic Variables

All 35 acoustic measures (envelope-based intensity and rhythm measures, autocorrelation pitch statistics, correlogram-based pitch measures, moments of the spectrum, RMS energy in octave-wide frequency bands, spectral shift in time measures, modulation spectrum statistics) were submitted to a separate factor analysis using maximum likelihood factor extraction and

Oblimin (oblique) rotation. Table 5 displays the variables and factor loadings for the rotated factors for the final model, with loadings less than |.40| excluded for clarity.

The Kaiser-Meyer-Olkin test revealed sufficient sampling adequacy for the final EFA analysis, KMO = 0.71. Bartlett's test of sphericity indicated the correlation structure of the variables was adequate for EFA as well, χ^2 (595) = 7490.33, $p < .001$. Upon visual inspection of the scree plot as well as a parallel analysis (see Figure 4), a seven-factor solution was revealed and accounts for 57% of the variance in the data.

Factor 1. After rotation, Factor 1 consisted of four variables that accounted for 10% of the variance in the model. One of the RMS energy measures of octave-wide frequency bands centered at 8000 Hz (0.89), two Moments of the Spectrum measures, the centroid (0.87) and standard deviation (0.81), and the mean pitch (0.61) all loaded onto this factor.

Factor 2. Factor 2 consisted of four variables which accounted for 10% of the variance in the model. Three Moments of the Spectrum measures, the standard deviation (1.00), maximum (0.90), and mean (0.86), as well as one modulation statistics measure, the maximum peak (0.49) loaded onto this factor.

Factor 3. Factor 3 consisted of three variables which accounted for 8% of the variance in the model. All of these variables loaded negatively onto the factor, and they include two Moments of the Spectrum measures, the skew (-0.91) and kurtosis (-0.86), and the mean Spectral Flux (-0.43).

Factor 4. Factor 4 consisted of two variables which accounted for 8% of the variance in the model. Both variables were measures of the autocorrelation, the mean peak (1.02) and standard deviation of peaks (0.98).

Factor 5. Factor 5 consisted of seven variables which accounted for 8% of the variance in the model. Two variables also loaded onto other factors; the mean pitch (0.41) also loaded onto Factor 1, and the spectral flux mean (0.45) also loaded onto Factor 3. In addition, the maximum peak of the autocorrelation (0.62), maximum pitch salience (0.62), mean pitch salience (0.58), and the maximum peak of spectral flux (0.48) loaded onto this factor as well. One measure of RMS energy in octave-wide frequency bands centered at 250 Hz loaded negatively on this factor (-0.43).

Factor 6. Factor 6 consisted of two variables which accounted for 7% of the variance in the model. Both variables were measures of RMS amplitude: the pause-corrected RMS (0.97) and overall RMS (0.96).

Factor 7. Factor 7 consisted of two variables which accounted for 6% of the variance in the model. Both variables were measures of the envelope: burst duration/total duration (0.90) and number of bursts (0.86).

Figure 4

Scree plot of eigenvalues revealing a 7-factor model.

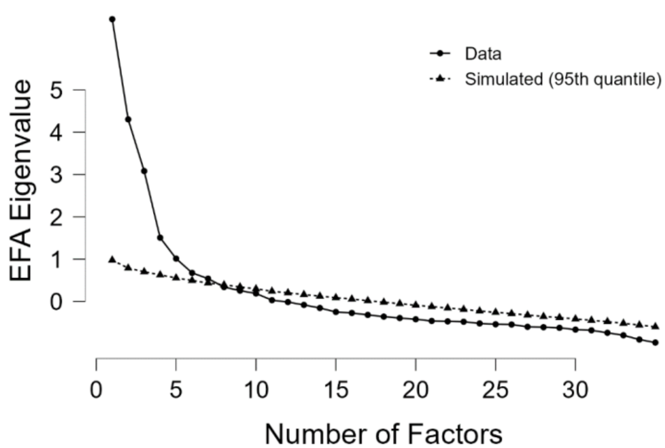


Table 5

Factor Loadings of Acoustic Variables.

Factor Loadings

	Factor						
	1	2	3	4	5	6	7
RMS in band f_c = 8000 Hz	0.89						
Moments of the Spectrum (Centroid)	0.87						
Moments of the Spectrum (SD)	0.81						
Mean Pitch	0.61				0.41		
Spectral Velocity (SD)		1.00					
Spectral Velocity (Maximum)		0.90					
Spectral Velocity (Mean)		0.86					
Modulation Statistics (Max Peak)		0.49					
Moments of the Spectrum (Skew)			-0.91				
Moments of the Spectrum (Kurtosis)			-0.86				
Spectral Flux (Mean)			-0.43		0.45		
Mean Peak in Autocorrelation				1.02			
SD of Peaks in Autocorrelation				0.98			
Maximum Peak in Autocorrelation					0.62		
Maximum Pitch Salience					0.62		
Mean Pitch Salience					0.58		
Spectral Flux (Maximum Peak)					0.48		
RMS in band f_c = 250 Hz					-0.43		
Pause-Corrected RMS Amplitude						0.97	
Overall RMS Amplitude						0.96	
Total Number of Bursts							0.90
Moments of the Spectrum (# of Peaks)							0.86

Note. Extraction method: maximum likelihood; Rotation method: Geomin (oblique); Loadings less than |.40| are not displayed.

Multiple Linear Regression Analyses

Next, eight multiple linear regression analyses were calculated to predict average performance on each global property rating scale based on the acoustic measures. Overall, each global property rating scale was significantly predicted by at least one acoustic variable (see Table 5).

Natural vs. Human-Influenced Regression

The first regression was calculated to predict performance on the Natural vs. Human-Influenced rating scale based on all 35 acoustic variables and was statistically significant, $R^2 = 0.57$, $R^2_{adj} = 0.47$, $F(35, 162) = 6.09$, $p < .001$. The significant acoustic predictors were pause-corrected/overall RMS amplitude ($\beta = 0.16$, $p < .05$), mean pitch ($\beta = -0.44$, $p < .001$), standard deviation of pitch ($\beta = -0.26$, $p < .05$), RMS energy in octave-wide frequency bands centered at 500 Hz ($\beta = 0.22$, $p < .05$), the spectral flux standard deviation ($\beta = -0.25$, $p < .05$), and the range of peaks in the autocorrelation ($\beta = -0.14$, $p = .05$).

Open vs. Closed Regression

The regression predicting performance on the Openness rating scale based on all acoustic variables was statistically significant, $R^2 = 0.57$, $R^2_{adj} = 0.48$, $F(35, 162) = 6.21$, $p < .001$. The significant acoustic predictors were the Moments of the Spectrum centroid ($\beta = 0.39$, $p < .05$), pause-corrected RMS amplitude ($\beta = 5.96$, $p < .05$), overall RMS amplitude ($\beta = -6.05$, $p < .05$), mean pitch ($\beta = -0.50$, $p < .001$), standard deviation of pitch ($\beta = -0.20$, $p < .05$), spectral flux standard deviation ($\beta = -0.30$, $p < .05$), range of peaks in the autocorrelation ($\beta = -0.23$, $p < .05$), RMS energy in octave-wide frequency bands centered at 250 Hz ($\beta = 0.18$, $p < .05$), and 1000 Hz ($\beta = -0.18$, $p < .05$).

Outdoor vs. Indoor Regression

The regression predicting performance on the Openness rating scale based on all acoustic variables was statistically significant, $R^2 = 0.58$, $R^2_{adj} = 0.49$, $F(35, 162) = 6.33$, $p < .001$. The significant acoustic predictors were the Moments of the Spectrum centroid ($\beta = 0.35$, $p < .05$), pause-corrected RMS amplitude ($\beta = 4.89$, $p < .05$), overall RMS amplitude ($\beta = -4.94$, $p < .05$), mean pitch ($\beta = -0.57$, $p < .001$) and standard deviation of pitch ($\beta = -0.26$, $p < .05$), mean pitch salience ($\beta = 0.32$, $p < .05$), spectral flux standard deviation ($\beta = -0.37$, $p < .001$), maximum peak in spectral flux ($\beta = 0.20$, $p = .05$), maximum peak in the autocorrelation ($\beta = -0.23$, $p < .05$), the range of peaks in the autocorrelation ($\beta = -0.17$, $p < .05$), RMS energy in octave-wide frequency bands centered at 1000 Hz ($\beta = -0.14$, $p = .05$) and 2000 Hz ($\beta = 0.13$, $p = .05$).

Temperature Regression

A regression was calculated to predict performance on the Temperature rating scale based on all acoustic variables and was statistically significant, $R^2 = 0.33$, $R^2_{adj} = 0.19$, $F(35, 162) = 2.32$, $p < .001$. The significant predictors were the Moments of the Spectrum standard deviation ($\beta = 0.48$, $p < .05$) and number of peaks ($\beta = -0.34$, $p < .001$), pause-corrected/overall RMS amplitude ($\beta = 3.68$, $p < .05$), spectral velocity mean ($\beta = -0.60$, $p < .05$), and spectral velocity standard deviation ($\beta = 0.86$, $p < .05$).

Navigability Regression

The regression predicting performance on the Navigability rating scale based on all acoustic variables was statistically significant, $R^2 = 0.41$, $R^2_{adj} = 0.28$, $F(35, 162) = 3.18$, $p < .001$. The only significant predictor was pause-corrected/overall RMS amplitude ($\beta = -0.17$, $p < .05$).

Season Regression

A regression was calculated to predict performance on the Season rating scale based on all acoustic variables and was statistically significant, $R^2 = 0.56$, $R^2_{adj} = 0.47$, $F(35, 162) = 5.99$, $p < .001$. The significant predictors were the Moments of the Spectrum skew ($\beta = -0.80$, $p < .05$) and kurtosis ($\beta = 0.46$, $p < .05$), as well as RMS energy in octave-wide frequency bands centered at 1000 Hz ($\beta = 0.15$, $p < .05$).

Transience Regression

A regression was calculated to predict performance on the Transience rating scale based on all acoustic variables and was statistically significant, $R^2 = 0.78$, $R^2_{adj} = 0.73$, $F(35, 162) = 16.28$, $p < .001$. The significant predictors were the Moments of the Spectrum skew ($\beta = -0.80$, $p < .001$), kurtosis ($\beta = 0.43$, $p < .05$), and number of peaks ($\beta = 0.12$, $p < .05$), mean pitch ($\beta = -0.22$, $p < .05$), mean pitch salience ($\beta = 0.23$, $p < .05$), as well as RMS energy in octave-wide frequency bands centered at 500 Hz ($\beta = 0.19$, $p < .001$) and 1000 Hz ($\beta = 0.20$, $p < .001$).

Sparseness Regression

A regression was calculated to predict performance on the Sparseness rating scale based on all acoustic variables and was statistically significant, $R^2 = 0.87$, $R^2_{adj} = 0.84$, $F(35, 162) = 31.51$, $p < .001$. The significant predictors were the Moments of the Spectrum skew ($\beta = -0.97$, $p < .001$), kurtosis ($\beta = 0.51$, $p < .001$), and number of peaks ($\beta = 0.09$, $p < .05$), mean pitch ($\beta = -0.15$, $p < .05$), maximum pitch ($\beta = 0.13$, $p < .05$), mean pitch salience ($\beta = 0.27$, $p < .001$), spectral flux mean ($\beta = -0.21$, $p < .05$), spectral flux standard deviation ($\beta = -0.11$, $p < .05$), as well as RMS energy in octave-wide frequency bands centered at 500 Hz ($\beta = 0.10$, $p < .001$) and 1000 Hz ($\beta = 0.12$, $p < .001$).

Table 6 *Significant Acoustic Predictors for Each Global Property Rating Scale.*

Variable	β	R^2
Natural vs. Human-Influenced		0.57
Pause-Corrected/Overall RMS Amplitude	0.16	
Pitch (Mean)	-0.44	
Pitch (SD)	-0.26	
Spectral Flux (SD)	-0.25	
Autocorrelation (Range of Peaks)	-0.14	
RMS in band f_c = 500 Hz	0.22	
Openness		0.57
Moments of the Spectrum (Centroid)	0.39	
Pause-Corrected RMS Amplitude	5.96	
Overall RMS Amplitude	-6.05	
Pitch (Mean)	-0.50	
Pitch (SD)	-0.20	
Spectral Flux (SD)	-0.30	
Autocorrelation (Range of Peaks)	-0.23	
RMS in band f_c = 250 Hz	0.18	
RMS in band f_c = 1000 Hz	-0.18	
Outdoor vs. Indoor		0.58
Moments of the Spectrum (Centroid)	0.35	
Pause-Corrected RMS Amplitude	4.89	
Overall RMS Amplitude	-4.94	
Pitch (Mean)	-0.57	
Pitch (SD)	-0.26	
Pitch Salience (Mean)	0.32	
Spectral Flux (SD)	-0.37	
Spectral Flux (Max Peak)	0.2	
Autocorrelation (Max Peak)	-0.23	
Autocorrelation (Range of Peaks)	-0.17	
RMS in band f_c = 1000 Hz	-0.14	
RMS in band f_c = 2000 Hz	0.14	

Variable	β	R^2
Temperature		0.33
Moments of the Spectrum (SD)	0.48	
Moments of the Spectrum (# of Peaks)	-0.34	
Pause-Corrected/Overall RMS Amplitude	3.68	
Spectral Velocity (Mean)	-0.6	
Spectral Velocity (SD)	0.86	
Navigability		0.41
Pause-Corrected/Overall RMS Amplitude	-0.17	
Season		0.56
Moments of the Spectrum (Skew)	-0.8	
Moments of the Spectrum (Kurtosis)	0.46	
Transience		0.78
Moments of the Spectrum (Skew)	-0.8	
Moments of the Spectrum (Kurtosis)	0.43	
Moments of the Spectrum (# of Peaks)	0.12	
Pitch (Mean)	-0.22	
Pitch Salience (Mean)	0.23	
RMS in band f_c = 500 Hz	0.19	
RMS in band f_c = 1000 Hz	0.2	
Sparseness		0.87
Moments of the Spectrum (Skew)	-0.97	
Moments of the Spectrum (Kurtosis)	0.51	
Moments of the Spectrum (# of Peaks)	0.09	
Pitch (Mean)	-0.15	
Pitch (Maximum)	0.13	
Pitch Salience (Mean)	0.27	
Spectral Flux (Mean)	-0.21	
Spectral Flux (SD)	-0.11	
RMS in band f_c = 500 Hz	0.1	
RMS in band f_c = 1000 Hz	0.12	

Chapter 4: Discussion

Here, we investigated the contributions of high-level global information and low-level acoustic features to auditory scene perception. Participants listened to complex, real-world auditory scenes and made judgments on a series of global properties (Sparseness, Transience, Season, Navigability, Openness, Outdoor vs. Indoor, Natural vs. Human-Influenced, Temperature). We found high between-participant consistency on ratings of all eight global property rating scales (indicated by the significant intra-class correlations). This particular result provides preliminary evidence for the ability to perceive auditory scenes from a global perspective, which is consistent with findings in the visual domain (Greene & Oliva, 2009a). A variety of acoustic measures were useful in predicting each of the global property ratings, though some of the acoustic-global relationships were non-linear.

These results are consistent with the hypothesis that global scene properties serve as high-level scene dimensions to inform a scene's layout, function, and constancy, allowing observers to rapidly understand the scene without needing to identify individual objects that are present (Greene & Oliva, 2009a; 2009b, Ross & Oliva, 2010, Greene & Oliva, 2010). Using both behavioral and computational methods, these studies demonstrated that observers can more quickly and accurately categorize visual scenes into a global category (e.g., an open environment) than a basic level category (e.g., a waterfall; Greene & Oliva, 2009a), utilize global information to perform basic-level categorization tasks (Greene & Oliva, 2009b), and adapt to global properties of visual scenes (Greene & Oliva, 2010). Additional electrophysiological studies have indicated the P2 event related potential (ERP) as a neural marker for global scene properties (Harel et al., 2016) and have revealed that global scene information is extracted in early ERP components (P1, N1, and P2; Hansen et al., 2018). The mounting evidence for the use of global scene properties in the visual domain provides a promising foundation for future behavioral, electrophysiological, neuroimaging, and computational studies in the auditory domain. Although our study only measured eight global properties, many more may be uncovered and investigated by future research to provide a well-rounded understanding of how people use objects to interpret complex auditory scenes. For example, future studies can evaluate how well people can identify the setting of a scene compared to the objects within it. Additionally, it would be useful to measure how quickly and in what order global scene judgments are made (e.g., is openness perceived prior to temperature?) and whether observers can adapt to these properties.

Dimensionality of Auditory Scene Perception

The results of our study indicate a high amount of dimensionality reduction along the auditory pathway when we listen to auditory scenes. Dimensionality reduction has been demonstrated in the perception of environmental sounds (Gygi, Kidd, & Watson, 2007), timbre (Grey, 1977), musical tonality (Shepard, 1982; Krumhansl, 1979; Toiviainen, Kaipainen, & Louhivouri, 1995), and rhythm (Jacoby & McDermott, 2017; Desain & Honing, 2003), suggesting this is a common feature of auditory processing. The multiple linear regressions calculated to predict performance on the global property rating scales based on the acoustic properties of each scene revealed each rating scale was significantly predicted by at least one acoustic variable. This finding, along with the difference in the number of reduced factors in the exploratory factor analyses (2 global property factors vs. 7 acoustic factors), suggest global variables may be processed at a higher level of the auditory pathway where acoustic features have been abstracted out of or have nonlinear relationships with global variables. The transformation of low-level acoustic information into high-level global representations of auditory scenes could occur similarly to processing along the ventral visual stream, where low-level information about visual objects (e.g., an object's geometric shape, position in space, etc.) culminates into high level representations of visual objects which allow for object recognition (DiCarlo, Zoccolan, & Rust, 2012).

Investigating responses to auditory scenes along the auditory pathway will be essential to our understanding of how the auditory system integrates low-level acoustic features of auditory scenes into high-level global representations of scenes. The auditory system functions hierarchically, showing tuning specificity for simple stimuli and acoustic features, such as pure tones, pitch (Patterson et al., 2002; Bendor & Wang, 2005; Norman-Haignere et al., 2013), frequency (Humphries et al., 2010; Da Costa et al., 2011), spatial cues (Higgins et al., 2017; Rauschecker & Tian, 2000; Stecker et al., 2005), and spectral and temporal modulations (Chi et al., 2005; Schönwiesner & Zatorre, 2009; Barton et al., 2012; Santoro et al., 2014) in primary auditory areas as well as tuning specificity for more complex stimuli, such as noise bursts (Kaas & Hackett, 2000), vocalizations (Rauschecker & Tian, 2000; Belin et al., 2000; Petkov et al., 2008), speech (Scott et al., 2000; Mesgarani et al., 2014; Overath et al., 2015; Norman-Haignere, Kanwisher, & McDermott, 2015), song (Norman-Haignere et al., 2022), and music (Boebinger, Norman-Haignere, McDermott, & Kanwisher, 2021) in non-primary auditory areas. Taken together, these findings indicate the auditory system processes information with increased complexity along the pathway, which parallels findings in the visual system (Tootell et al., 1988; 1998; Hubel and Wiesel, 1962; De Valois & De Valois, 1988; Movshon et al., 1978; Carandini et al., 1999; Cumming & DeAngelis, 2001; Gegenfurtner & Kiper, 2003; Horwitz & Hass, 2012; DiCarlo et al., 2012).

Neural Pathways for Scene Processing

Since the existence of global properties is supported by behavioral, computational (Greene & Oliva, 2006; Greene & Oliva, 2009a, 2009b; Ross & Oliva, 2010, Greene & Oliva, 2010), and neural (Harel et al., 2016; Hansen et al., 2018) studies in the visual domain as well as the results of our behavioral study, important questions regarding the neural pathways allowing for global processing of auditory scenes are prompted. Similar to the organization of the visual cortex (Mishkin, Ungerleider, & Macko, 1983; Goodale & Milner, 1992; Milner & Goodale, 2006), it has been proposed that the auditory system processes information using two parallel processing streams: a dorsal stream and a ventral stream (Rauschecker, 1998; Rauschecker & Tian, 2000, Alain et al., 2001; Griffiths, 2008; Lomber & Malhotra, 2008). The dorsal stream affords us the ability to localize auditory stimuli in space (Rauschecker & Scott, 2009; Rauschecker, 2012) and map sounds onto motor-based representations involved in speech production (Hickok & Poeppel, 2004), while the ventral stream allows us to identify and extract the content and meaning of auditory stimuli (including speech). Although the dorsal stream plays a critical role in processing spatial information, this discussion will focus on the ventral stream as its function is more relevant to identification and processing of non-spatial information in auditory scenes.

The ventral stream originates in the core auditory fields A1 and R, continues to the anterolateral and middle lateral belt regions, and terminates in the ventrolateral prefrontal cortex (vlPFC; Kaas & Hackett, 2000; Rauschecker & Tian, 2000). As previously mentioned, neurons in the core prefer lower-level sound features such as frequency and intensity, while neurons in the anterolateral belt respond to vocalizations, frequency-modulated sweeps, and band-passed noise (Chang et al., 2010; Rauschecker et al., 1995; Rauschecker & Tian, 2000, 2004; Tian & Rauschecker, 2004; Tian et al., 2001). The processing of sound stimuli in the vlPFC reflects post-sensory processing such as auditory attention, working memory, the meaning of sounds, and integration of multisensory information (Cohen et al., 2009; Plakke & Romanski, 2014; Poremba et al., 2003; Lee et al., 2009; Ng et al., 2014; Romanski et al., 2005; Russ et al., 2008a, 2008b). In the ventral visual stream, there is evidence of both object-selective and scene-selective regions (Epstein & Baker, 2019). Epstein and Kanwisher (1998) identified the parahippocampal place area (PPA), a region of the cortex which responds more strongly to pictures of scenes (e.g., houses) than objects (e.g., bodies, faces) during both active perception and mental imagery of visual scenes. More recent studies have highlighted the role of PPA in recognition of non-visual information as well, such as descriptions of famous places (Aziz-Zadeh et al., 2008) and audio descriptions of spatial information (Hausler, Eickhoff, & Hanke, 2022). A second visual scene-specific region exists in the retrosplenial complex (RSC) called the medial place area (MPA; Epstein & Kanwisher, 1998; O'Craven & Kanwisher, 2000), and a third in the occipital place area (OPA; Nakamura et al., 2000; Dilks et al., 2013). A recent account suggests the PPA is important for scene categorization whereas the OPA and RSC support visually guided navigation and map-based navigation, respectfully (Dilks, Kamps, & Persichetti, 2021). An additional fMRI study identified a series of distributed cortical networks which show tuning specific to various scene categories (e.g., navigation, social interaction, human activity, motion-energy, texture, non-human animals, civilization, natural environment), demonstrating the complexity of scene processing in the human cerebral cortex (Çelik et al., 2021).

Many object-specific areas have been identified as well; some areas respond most to faces (fusiform and occipital face areas; Kanwisher, McDermott, & Chun, 1997; Haxby, Hoffman, & Gobbini, 2000), shapes (posterior fusiform and lateral occipital complex (LOC); Malach et al., 1995; Grill-Spector & Malach, 2004), or bodies (fusiform and extrastriate body area; Peelen & Downing, 2005; Downing et al., 2001). A replication of Dilks et al. (2013)

provided further evidence of a double dissociation in scene and object processing in the OPA and LOC. Transcranial magnetic stimulation (TMS) delivered to OPA impaired the recognition of scenes while TMS delivered to LOC impaired recognition of objects (Wischnewski & Peelan, 2021). Scene and object processing has also been explored using intracranial electroencephalography (iEEG), which offers high anatomical and temporal resolution. Vlcek et al. (2020) collected iEEG data as epileptic patients viewed images containing both objects and scenes. Their results support the roles of the PPA and LOC in scene and object processing, respectively, as well as scene-selective areas MPA and OPA. This scene network was shown to extend to regions involved in processing scene novelty (anterior temporal lobe regions, including the hippocampus and parahippocampal gyrus). Additional object-selective areas were identified, including areas selective for tool use (intraparietal sulcus, supramarginal gyrus and middle temporal cortex; Vidal et al., 2010) and object recognition (inferior frontal gyrus and perirhinal cortex; Nakamura et al., 2000; Clarke & Tyler, 2014; Bar et al., 2001).

While these studies in the visual domain have identified brain regions related to visual object and scene processing, most studies investigating the ventral auditory stream have used simple tones, noise bursts, isolated speech, music (both instrumental and with voice) or individual environmental sounds. Future neuroimaging studies or invasive neurophysiological studies will be necessary to evaluate how complex auditory scenes are processed along auditory-specific pathways. An additional research avenue could investigate the potential relationship between the ventral visual stream and auditory scene processing. Since the amount of cortex devoted to processing visual information is far greater than the amount devoted to auditory processing, it could be possible that visual areas may aid in the perception and representations of auditory scenes. Additionally, computational modeling studies could increase our understanding of the neural and computational processes contributing to auditory scene processing. Most models of auditory scene analysis focus on the segregation of two auditory stimuli (e.g., tones, noise bursts, speech, foreground/background) into perceptual streams (Krishnan. Elhilali, & Shamma, 2014; Elhilali & Shamma, 2008; Ma, 2011), but these models are limited and do not explain how auditory objects and scenes are identified or understood. Although one model of the auditory system assessed the recognition and understanding of synthesized sound textures (i.e., temporally homogenous sounds such as a rainstorm or a choir of crickets; McDermott & Simoncelli, 2011), future studies are necessary to evaluate how auditory objects and scenes are processed from early to late stages in the auditory system.

In summary, our results provide preliminary evidence for the ability to perceive auditory scenes from a global perspective. Additionally, our results suggest a high degree of dimensionality reduction along the auditory pathway wherein global properties of scenes are processed at a high level and the acoustic features of scenes are processed at a low level. Examining the role of global properties in our perception of auditory scenes is essential to gain a finer understanding of the underlying processes which construct our representations of the auditory environment.

References

Alain, C., Arnott, S. R., Hevenor, S., Graham, S., & Grady, C. L. (2001). “What” and “where” in the human auditory system. *Proceedings of the national academy of sciences*, 98(21), 12301-12306.

Aziz-Zadeh, L., Fiebach, C. J., Naranayan, S., Feldman, J., Dodge, E., & Ivry, R. B. (2008). Modulation of the FFA and PPA by language related to faces and places. *Social Neuroscience*, 3(3-4), 229-238.

Ballas, J. A. (1993). Common factors in the identification of an assortment of brief everyday sounds. *Journal of experimental psychology: Human perception & performance*, *19*(2), 250-267. doi: 10.1037/0096-1523.19.2.250

Bar, M., Tootell, R. B., Schacter, D. L., Greve, D. N., Fischl, B., Mendola, J. D., Rosen, B. R., & Dale, A. M. (2001). Cortical mechanisms specific to explicit visual object recognition. *Neuron*, 29(2), 529-535.

Barton, B., Venezia, J. H., Saberi, K., Hickok, G., & Brewer, A. A. (2012). Orthogonal acoustic dimensions define auditory field maps in human cortex. *Proceedings of the National Academy of Sciences*, 109(50), 20738-20743.

Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature*, 403(6767), 309-312.

Bendor, D., & Wang, X. (2005). The neuronal representation of pitch in primate auditory cortex. *Nature*, 436(7054), 1161-1165.

Boebinger, D., Norman-Haignere, S. V., McDermott, J. H., & Kanwisher, N. (2021). Music-selective neural populations arise without musical training. *Journal of Neurophysiology*, 125(6), 2237-2263.

Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound*. Cambridge, MA: MIT Press.

Carandini, M., Heeger, D. J., & Anthony Movshon, J. (1999). Linearity and gain control in V1 simple cells. In *Models of cortical circuits* (pp. 401-443). Springer, Boston, MA.

Çelik, E., Keles, U., Kiremitçi, İ., Gallant, J. L., & Cukur, T. (2021). Cortical networks of dynamic scene category representation in the human brain. *Cortex*, 143, 127-147.

Chang, E. F., Rieger, J. W., Johnson, K., Berger, M. S., Barbaro, N. M., & Knight, R. T. (2010). Categorical speech representation in human superior temporal gyrus. *Nature Neuroscience*, 13(11), 1428-1432.

Chi, T., Ru, P., & Shamma, S. A. (2005). Multiresolution spectrotemporal analysis of complex sounds. *The Journal of the Acoustical Society of America*, 118(2), 887-906.

Clarke, A., & Tyler, L. K. (2014). Object-specific semantic coding in human perirhinal cortex. *Journal of Neuroscience*, 34(14), 4766-4775.

Cohen, M. A., Horowitz, T. S., & Wolfe, J. M. (2009). Auditory recognition memory is inferior to visual recognition memory. *Proceedings of the National Academy of Sciences*, 106(14), 6008-6010.

Cumming, B. G., & DeAngelis, G. C. (2001). The physiology of stereopsis. *Annual Review of Neuroscience*, 24(1), 203-238.

Da Costa, S., van der Zwaag, W., Marques, J. P., Frackowiak, R. S., Clarke, S., & Saenz, M. (2011). Human primary auditory cortex follows the shape of Heschl's gyrus. *Journal of Neuroscience*, 31(40), 14067-14075.

De Valois, R. L., & De Valois, K. K. (1980). Spatial vision. *Annual review of psychology*, 31(1), 309-341.

Desain, P., & Honing, H. (2003). The formation of rhythmic categories and metric priming. *Perception*, 32(3), 341-365.

DiCarlo, J. J., Zoccolan, D., & Rust, N. C. (2012). How does the brain solve visual object recognition?. *Neuron*, 73(3), 415-434.

Dilks, D. D., Julian, J. B., Paunov, A. M., & Kanwisher, N. (2013). The occipital place area is causally and selectively involved in scene perception. *Journal of Neuroscience*, 33(4), 1331-1336.

Dilks, D. D., Kamps, F. S., & Persichetti, A. S. (2021). Three cortical scene systems and their development. *Trends in Cognitive Sciences.*

Downing, P. E., Jiang, Y., Shuman, M., & Kanwisher, N. (2001). A cortical area selective for visual processing of the human body. *Science*, 293(5539), 2470-2473.

Elhilali, M., & Shamma, S. A. (2008). A cocktail party with a cortical twist: how cortical mechanisms contribute to sound segregation. *The Journal of the Acoustical Society of America*, 124(6), 3751-3771.

Epstein, R. A., & Baker, C. I. (2019). Scene perception in the human brain. *Annual Review of Vision Science*, 5, 373.

Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature*, 392(6676), 598-601.

Gegenfurtner, K. R., & Kiper, D. C. (2003). Color vision. *Annual Review of Neuroscience,* 26(1), 181-206.

Geisler, W. S., & Perry, J.S. (2011). Statistics for optimal point prediction in natural images. *Journal of Vision. 11*(12). doi: 10.1167/11.12.14

Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences*, 15(1), 20-25.

Greene, M. R., & Oliva, A. (2009a). The briefest of glances: the time courses of natural scene understanding. *Psychological Science*, *20*(4), 464-472. doi:10.1111/j.1467-9280.2009.02316.x

Greene, M. R., & Oliva, A. (2009b). Recognition of natural scenes from global properties: seeing the forest without representing the trees. *Cognitive Psychology*, *58*(2), 137-176. doi:10.1016/j.cogpsych.2008.06.001

Greene, M. R., & Olivia, A. (2006). Natural scene categorization from conjunction of ecological global properties. 291-296.

Gregg, M. K., & Samuel, A. G. (2008). Change deafness and the organizational properties of sounds. *Journal of Experimental Psychology: Human Perception and Performance, 34*(4), 974-991. doi:10.1037/0096-1523.34.4.974

Gregg, M. K., & Samuel, A. G. (2009). The importance of semantics in auditory representations. *Attention, Perception, and Psychophysics, 71*(3), 607-619. doi:10.3758/APP.71.3.607

Gregg, M. K., Irsik, V. C., & Snyder, J. S. (2014). Change deafness and object encoding with recognizable and unrecognizable sounds. *Neuropsychologia, 61,* 19-30.

Gregg, M. K., Irsik, V. C., & Snyder, J. S. (2017). Effects of capacity limits, memory loss, and sound type in change deafness. *Attention, Perception, & Psychophysics, 79,* 2564 - 2575.

Grey, J. M. (1977). Multidimensional perceptual scaling of musical timbres. *The Journal of the Acoustical Society of America*, 61(5), 1270-1277.

Griffiths, T. D. (2008). Sensory systems: auditory action streams?. *Current Biology*, 18(9), R387-R388.

Grill-Spector, K., & Malach, R. (2004). The human visual cortex. *Annual Review of Neuroscience,* 27, 649-677.

Gygi, B., & Shafiro, V. (2010). Development of the database for environmental sound research and application (DESRA): design, functionality, and retrieval considerations. *EURASIP Journal on Audio, Speech, and Music Processing,* 1-12.

Gygi, B., Kidd, G. R., & Watson, C. S. (2007). Similarity and categorization of environmental sounds. *Perception and Psychophysics*, *69*(6), 839-855.

Hansen, N. E., Noesen, B. T., Nador, J. D., & Harel, A. (2018). The influence of behavioral relevance on the processing of global scene properties: an ERP study. *Neuropsychologia*, 114, 168-180.

Harel, A., Groen, I. I., Kravitz, D. J., Deouell, L. Y., & Baker, C. I. (2016). The temporal dynamics of scene processing: A multifaceted EEG investigation. *Eneuro*, 3(5).

Häusler, C. O., Eickhoff, S. B., & Hanke, M. (2022). Processing of visual and non-visual naturalistic spatial information in the" parahippocampal place area". *Scientific data*, 9(1), 1-15.

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences*, 4(6), 223-233.

Hickok, G., & Poeppel, D. (2004). Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition*, 92(1-2), 67-99.

Higgins, N. C., McLaughlin, S. A., Rinne, T., & Stecker, G. C. (2017). Evidence for cue-independent spatial representation in the human auditory cortex during active listening. Proceedings of the National Academy of Sciences, 114(36), E7602-E7611.

Horwitz, G. D., & Hass, C. A. (2012). Nonlinear analysis of macaque V1 color tuning reveals cardinal directions for cortical color processing. *Nature Neuroscience,* 15(6), 913-919.

Houtgast, T., & Steeneken, H. J. M. (1985). A review of the MTF concept in room acoustitcs and its use for estimating speech intelligibility in auditoria. *The Journal of the Acoustical Society of America*, *77*(3), 1069-1077. doi: 10.1121/1.392224

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of Physiology*, 160(1), 106.

Humphries, C., Liebenthal, E., & Binder, J. R. (2010). Tonotopic organization of human auditory cortex. *Neuroimage*, 50(3), 1202-1211.

image features. *Journal of Vision, 10*(1):2, 1-25. doi: 10.1167/10.1.2

Jacoby, N., & McDermott, J. H. (2017). Integer ratio priors on musical rhythm revealed cross-culturally by iterated reproduction. *Current Biology*, 27(3), 359-370.

Kaas, J. H., & Hackett, T. A. (2000). Subdivisions of auditory cortex and processing streams in primates. *Proceedings of the National Academy of Sciences,* 97(22), 11793-11799.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience*, 17(11), 4302-4311.

Krishnan, L., Elhilali, M., & Shamma, S. (2014). Segregating complex sound sources through temporal coherence. *PLoS Computational Biology*, 10(12), e1003985.

Krumhansl, C. L. (1979). The psychological representation of musical pitch in a tonal context. *Cognitive Psychology*, 11(3), 346-374.

Lee, J. H., Russ, B. E., Orr, L. E., & Cohen, Y. (2009). Prefrontal activity predicts monkeys' decisions during an auditory category task. *Frontiers in Integrative Neuroscience*, 16.

Leech, R., Gygi, B., Aydelott, J., Dick, F. (2009). Informational factors in identifying environmental sounds in natural auditory scenes. *Acoustical Society of America, 126*(6), 3147-3155. doi: 10.1121/1.3238160

Lomber, S. G., & Malhotra, S. (2008). Double dissociation of 'what' and 'where' processing in auditory cortex. *Nature neuroscience*, 11(5), 609–616. https://doi.org/10.1038/nn.2108

Ma, L. (2011). Auditory streaming: Behavior, physiology, and modeling. Doctoral Dissertation, Bioengineering Program. College Park, MD: University of Maryland.

Malach, R., Reppas, J. B., Benson, R. R., Kwong, K. K., Jiang, H., Kennedy, W. A., Ledden, P. J., Brady, T. J., Rosen, B. R., & Tootell, R. B. (1995). Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proceedings of the National Academy of Sciences*, 92(18), 8135-8139.

MATLAB. (2021). *version 9.10.0.1851785 (R2010a)*. Natick, Massachusetts: The MathWorks Inc.

McDermott, J. H., & Simoncelli, E. P. (2011). Sound texture perception via statistics of the auditory periphery: evidence from sound synthesis. *Neuron, 71*(5), 926-940. doi:10.1016/j.neuron.2011.06.032

Mesgarani, N., Cheung, C., Johnson, K., & Chang, E. F. (2014). Phonetic feature encoding in human superior temporal gyrus. *Science*, 343(6174), 1006-1010.

Milner, D., & Goodale, M. (2006). The visual brain in action (Vol. 27). *OUP Oxford.*

Mishkin, M., Ungerleider, L. G., & Macko, K. A. (1983). Object vision and spatial vision: two cortical pathways. *Trends in Neurosciences*, 6, 414-417.

Movshon, J. A., Thompson, I. D., & Tolhurst, D. J. (1978). Receptive field organization of complex cells in the cat's striate cortex. *The Journal of Physiology*, 283(1), 79-99.

Nakamura, K., Kawashima, R., Sato, N., Nakamura, A., Sugiura, M., Kato, T., Hatano, K., Fukuda, H., Schormann, T., & Zilles, K. (2000). Functional delineation of the human occipito-temporal areas related to face and scene processing: a PET study. *Brain,* 123(9), 1903-1912.

Ng, C. W., Plakke, B., & Poremba, A. (2014). Neural correlates of auditory recognition memory in the primate dorsal temporal pole. Journal of neurophysiology, 111(3), 455-469.

Norman-Haignere, S. V., Feather, J., Boebinger, D., Brunner, P., Ritaccio, A., McDermott, J. H., Schalk, G., & Kanwisher, N. (2022). A neural population selective for song in human auditory cortex. *Current Biology*, 32(7), 1470-1484.

Norman-Haignere, S., Kanwisher, N. G., & McDermott, J. H. (2015). Distinct cortical pathways for music and speech revealed by hypothesis-free voxel decomposition. *Neuron*, 88(6), 1281-1296.

Norman-Haignere, S., Kanwisher, N., & McDermott, J. H. (2013). Cortical pitch regions in humans respond primarily to resolved harmonics and are located in specific tonotopic regions of anterior auditory cortex. *Journal of Neuroscience*, 33(50), 19451-19469.

O'Craven, K. M., & Kanwisher, N. (2000). Mental imagery of faces and places activates corresponding stimulus-specific brain regions. *Journal of Cognitive Neuroscience*, 12(6), 1013-1023.

Oliva, A., & Torralba, A. (2006). Building the gist of a scene: the role of global image features in recognition. *Progress in brain research: Visual perception, 155,* 23-36. doi: 10.1016/S0079-6123(06)55002-2

Overath, T., McDermott, J. H., Zarate, J. M., & Poeppel, D. (2015). The cortical analysis of speech-specific temporal structure revealed by responses to sound quilts. *Nature Neuroscience,* 18(6), 903-911.

Patterson, R. D., Uppenkamp, S., Johnsrude, I. S., & Griffiths, T. D. (2002). The processing of temporal pitch and melody information in auditory cortex. *Neuron*, 36(4), 767-776.

Peelen, M. V., & Downing, P. E. (2005). Selectivity for the human body in the fusiform gyrus. *Journal of Neurophysiology*, 93(1), 603-608.

Petkov, C. I., Kayser, C., Steudel, T., Whittingstall, K., Augath, M., & Logothetis, N. K. (2008). A voice region in the monkey brain. *Nature Neuroscience*, 11(3), 367-374.

Petkov, C. I., O'Connor K., N., Benmoshe, G., Baynes, K., & Sutter, M. L. (2005). Auditory perceptual grouping and attention in dyslexia. *Cognitive Brain Research*, *24*(2), 343-354.

Plakke, B., & Romanski, L. M. (2014). Auditory connections and functions of prefrontal cortex. *Frontiers in Neuroscience*, 8, 199.

Poremba, A., Saunders, R. C., Crane, A. M., Cook, M., Sokoloff, L., & Mishkin, M. (2003). Functional mapping of the primate auditory system. *Science*, 299(5606), 568-572.

Rauschecker, J. P. (1998). Parallel processing in the auditory cortex of primates. *Audiology and Neurotology*, 3(2-3), 86-103.

Rauschecker, J. P. (2012). Ventral and dorsal streams in the evolution of speech and language. *Frontiers in Evolutionary Neuroscience*, 4, 7.

Rauschecker, J. P., & Tian, B. (2000). Mechanisms and streams for processing of "what" and "where" in auditory cortex. *Proceedings of the National Academy of Sciences*, 97(22), 11800-11806.

Rauschecker, J. P., & Tian, B. (2004). Processing of band-passed noise in the lateral auditory belt cortex of the rhesus monkey. *Journal of Neurophysiology*, 91(6), 2578-2589.

Rauschecker, J. P., Tian, B., & Hauser, M. (1995). Processing of complex sounds in the macaque nonprimary auditory cortex. *Science*, 268(5207), 111-114.

Romanski, L. M., Averbeck, B. B., & Diltz, M. (2005). Neural representation of vocalizations in the primate ventrolateral prefrontal cortex. *Journal of Neurophysiology*, 93(2), 734-747.

Ross, M. G., & Oliva, A. (2010). Estimating perception of scene layout properties from global image features. Journal of vision, 10(1), 2-2.

Russ, B. E., Ackelson, A. L., Baker, A. E., and Cohen, Y. E. (2008a). Coding of auditory-stimulus identity in the auditory non-spatial processing stream. *Journal of Neurophysiology,* 99, 87–95.

Russ, B. E., Orr, L. E., and Cohen, Y. E. (2008b). Prefrontal neurons predict choices during an auditory same-different task. *Current Biology,* 18, 1483–1488.

Santoro, R., Moerel, M., De Martino, F., Goebel, R., Ugurbil, K., Yacoub, E., & Formisano, E. (2014). Encoding of natural sounds at multiple spectral and temporal resolutions in the human auditory cortex. *PLoS Computational Biology*, 10(1), e1003412.

Schönwiesner, M., & Zatorre, R. J. (2009). Spectro-temporal modulation transfer function of single voxels in the human auditory cortex measured with high-resolution fMRI. *Proceedings of the National Academy of Sciences*, 106(34), 14611-14616.

Scott, S. K., Blank, C. C., Rosen, S., & Wise, R. J. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain*, 123(12), 2400-2406.

Shepard, R. N. (1982). Geometrical approximations to the structure of musical pitch. *Psychological Review*, 89(4), 305.

Slaney, M. (1995). *Auditory toolbox: A MATLAB toolbox for auditory modeling work* (Apple Tech. Rep. No. 45), Cupertino, CA: Apple Computer.

Snyder, J. S., Gregg, M. K., Weintraub, D. M., & Alain, C. (2012). Attention, awareness, and the perception of auditory scenes. *Frontiers in Psychology, 3*, 15. doi:10.3389/fpsyg.2012.00015

Stecker, G. C., Harrington, I. A., & Middlebrooks, J. C. (2005). Location coding by opponent neural populations in the auditory cortex. *PLoS Biology*, 3(3), e78.

SUN Database: Large-scale Scene Recognition from Abbey to Zoo. J. Xiao, J. Hays, K. Ehinger, A. Oliva, and A. Torralba. *IEEE Conference on Computer Vision and Pattern Recognition*, 2010.

Tian, B., & Rauschecker, J. P. (2004). Processing of frequency-modulated sounds in the lateral auditory belt cortex of the rhesus monkey. *Journal of Neurophysiology*, 92(5), 2993-3013.

Tian, B., Reser, D., Durham, A., Kustov, A., & Rauschecker, J. P. (2001). Functional specialization in rhesus monkey auditory cortex. *Science*, 292(5515), 290-293.

Toiviainen, P., Kaipainen, M., & Louhivuori, J. (1995). Musical timbre: Similarity ratings correlate with computational feature space distances. *Journal of New Music Research*, 24(3), 282-298.

Tootell, R. B., Hadjikhani, N. K., Vanduffel, W., Liu, A. K., Mendola, J. D., Sereno, M. I., & Dale, A. M. (1998). Functional analysis of primary visual cortex (V1) in humans. *Proceedings of the National Academy of Sciences*, 95(3), 811-817.

Tootell, R. B., Switkes, E., Silverman, M. S., & Hamilton, S. L. (1988). Functional anatomy of macaque striate cortex. II. Retinotopic organization. *Journal of Neuroscience*, 8(5), 1531-1568.

Vidal, J. R., Ossandón, T., Jerbi, K., Dalal, S. S., Minotti, L., Ryvlin, P., et al. (2010). Category-specific visual responses: an intracranial study comparing γ, β, α, and ERP response selectivity. *Frontiers in Human Neuroscience*. 4:195. doi: 10.3389/fnhum.2010.00195

Vlcek, K., Fajnerova, I., Nekovarova, T., Hejtmanek, L., Janca, R., Jezdik, P., Kalina, A., Tomasek, M., Krsek, P., Hammer, J., & Marusic, P. (2020). Mapping the scene and object processing networks by intracranial EEG. *Frontiers in Human Neuroscience*, 411.

Wischnewski, M., & Peelen, M. V. (2021). Causal evidence for a double dissociation between object-and scene-selective regions of visual cortex: a preregistered TMS replication study. *Journal of Neuroscience*, 41(4), 751-756.

Curriculum Vitae

MARGARET MCMULLIN
Department of Psychology
University of Nevada, Las Vegas
4505 Maryland Parkway
Las Vegas, Nevada 89154
mcmullinmags@gmail.com

Education

2018-present **_Ph.D. in Psychological and Brain Sciences_**, *Cognitive Emphasis*
University of Nevada, Las Vegas
Advisor: Joel Snyder, Ph.D.

2018-2022 **_M.A. in Psychology,_** *Cognitive Emphasis*
University of Nevada, Las Vegas
Advisor: Joel Snyder, Ph.D.
Thesis: *Dimensionality of Natural Auditory Scene Perception: A factor analysis study*

2017-2018 Completed 18 credits towards:
M.S. in Psychological Research, *Cognitive and Affective Emphasis*
University of Wisconsin, Oshkosh
Advisor: James Koch, Ph.D.

2013-2017 **_B.S. in Psychology_**, *Magna Cum Laude*
University of Wisconsin, Parkside
Certificates: Neuroscience, Human Measurement and Research, Mental Health Skills
Minor in Geography
University of Wisconsin, Parkside
Advisor: Melissa Gregg, Ph.D.

Research Interests

Cognitive Neuroscience, Auditory and visual perception and memory

Awards

2021-2022 **_Summer Session Scholarship_** (University of Nevada, Las Vegas)
2020-2023 **_National Defense Science and Engineering Graduate (NDSEG) Fellow_**
United States Department of Defense
Office of Naval Research
2020-2021 **_Patricia Sastaunik Scholarship_** (University of Nevada, Las Vegas)
2020 **_Summer Session Scholarship_** (University of Nevada, Las Vegas)
2019-2020 **_Access Grant_** (University of Nevada, Las Vegas)
2018-2019 **_Access Grant_** (University of Nevada, Las Vegas)

2017 ***College of Natural and Health Sciences Travel Award***
University of Wisconsin – Parkside
2017 ***Department of Psychology Travel Award*** (University of Wisconsin, Parkside)
2017 ***Academic Achievement Award*** (University of Wisconsin, Parkside)

Manuscripts in Preparation

McMullin, M. A., Higgins, N. C., Gygi, B., Kumar, R., Elhilali, M., Snyder, J. S. (in prep). *Perception of global properties, objects, and settings in natural auditory scenes*.

Parks, C. M., Werner, L., **McMullin, M. A.,** & Snyder, J.S. (in prep). *Auditory and visual object recognition: equating modalities with repetition.*

Invited Talk(s)

McMullin, M. A. (2022, November). *Natural Auditory Scene Perception.* Oral Presentation presented at the University of Nevada, Reno Cognitive and Brain Sciences Department's Early Career Seminar Series, Reno, Nevada.

Conference and Academic Presentations

Lee, S., Werner, L., Mohawk, K. D., **McMullin, M. A.,** Snyder, J. S., & Parks, C. M. (2022, November). *Differing effects of divided attention on visual and auditory recognition.* Poster session presented at the annual meeting of the Psychonomic Society. Boston, MA.

McMullin, M. A., Higgins, N. C., Gygi, B., & Snyder, J. S. (2022, July). *Dimensionality of natural auditory scene perception: A factor analysis study*. Poster session presented at the Department of Defense National Science and Engineering Graduate Fellowship Conference. Boston, MA. **Also presented at:**

- Poster session presented at the International Conference on Auditory Cortex, Magdeburg, Germany. (2022, September).
- Poster session presented at the annual meeting of the Psychonomic Society. Boston, MA. (2022, November).
- Poster session presented at the Auditory Perception, Cognition, and Action Meeting. Boston, MA. (2022, November).
- Poster session at the 2023 MidWinter Meeting of the Association for Research in Otolaryngology. Orlando, FL. (2023, February)

Werner, L., **McMullin, M. A.,** Snyder, J. S., & Parks, C. (2021, November). *Recognition memory mechanisms and attention-dependent encoding may account for the auditory disadvantage for remembering objects*. Poster session presented at the annual meeting of the Psychonomic Society.

McMullin, M. A. (2020, January). *Dimensionality of natural auditory scene perception: a factor analysis study*. Oral Presentation presented at the UNLV Psychology Department's Proseminar Course. Las Vegas, NV.

McMullin, M. A. (2019, April). *Natural auditory scene analysis*. Oral Presentation presented at the UNLV Psychology Department's Proseminar Course. Las Vegas, NV.

McMullin, M. A., Bharvani, S., & Gregg, M. K. (2018, April). *A comparison of auditory and visual gist perception*. Poster session presented at the Midwestern Psychological Association Annual Meeting, Chicago, IL. **Also presented at:**

- Poster session presented at the annual meeting of the Psychonomic Society, Vancouver, B.C., Canada. (2017, November).
- Poster session presented at the University of Wisconsin, Parkside Student Showcase, Kenosha, WI. (2017, April).
- Poster session presented at the UW-System Symposium for Undergraduate Research & Creative Activity, Stevens Point, WI. (2017, April).
- Poster session presented at the Midwestern Psychological Association Annual Meeting, Chicago, IL. (2017, April).
- Poster session presented at the National Conference for Undergraduate Research, Memphis, TN. (2017, April).

Research Experience

Acquired Skills: EEG recording and analysis (BESA), Presentation by Neurobehavioral Systems, MATLAB by MathWorks, SuperLab by Cedrus, Qualtrics, R, SPSS

2018-2020 ***Graduate Research Assistant & Lab Manager***
Auditory Cognitive Neuroscience Lab
University of Nevada, Las Vegas
Advisor: Joel Snyder, Ph.D.
Projects:
Dimensionality of auditory scene perception: A factor analysis study
Auditory scene perception and object identification
Exploring the neural correlates of illusory and actual music perception
Neural Correlates of Auditory and Visual Long-term Memory (Collaboration with Dr. Colleen Parks)

2017-2018 ***Graduate Research Assistant***
Human Electrophysiology Lab
University of Wisconsin, Oshkosh
Advisor: James Koch, Ph.D.
Projects:
Effects of brain training programs on memory (Assistant)

2016-present ***Undergraduate Research Assistant, Current Collaborator***
Perception, Cognition, and Neuroscience Lab
University of Wisconsin, Parkside
Advisor: Melissa Gregg, Ph.D.
Projects:
A comparison of auditory and visual gist perception
Neural correlates of illusory and actual music perception
Echolocation and auditory memory (Assistant)

Professional Experience

2018-2020 ***Graduate Assistant***
Department of Psychology
University of Nevada, Las Vegas

2017-2018 ***Graduate Assistant***
College of Business
University of Wisconsin, Oshkosh

2017 ***Intern***
Neuropsychological Consultants, Inc.
Kenosha, Wisconsin

Teaching Experience

Fall 2021- Fall 2022
Guest Lecturer
Psychology 330: Infant and
Child Development
Topics: Brain Development, Genetics
Instructor: Rodica Constantine, M.A.
University of Nevada, Las Vegas

Spring 2019 & 2020
Teaching Assistant
Instructor: Dr. Joel Snyder
Psychology 425: Cognitive Neuroscience
University of Nevada, Las Vegas

Spring 2017
Teaching Assistant
Instructor: Dr. Melissa Gregg
Psychological Statistics
University of Wisconsin, Parkside

Summer 2019
Co-Instructor
Psychology 101: General Psychology
University of Nevada, Las Vegas

Fall 2018 & 2019
Teaching Assistant
Instructor: Dr. Joel Snyder
Psychology 305: Foundations of Perception
University of Nevada, Las Vegas

Service

2018-present ***UNLV Outreach Undergraduate Mentoring Program (OUMP) Mentor***
2019-2020 ***UNLV Experimental Student Committee Member***

Professional Affiliations

Association for Research in Otolaryngology (Student Member)
Association for Psychological Science (Student Member)
Psychonomic Society (Student Member)

Workshops

June 2019 ***Methods for Analyzing Sound and Modeling Auditory Plasticity (MASMAP)***

The State University of New York at Buffalo
Organizers: Drs. Eduardo Mercado III, Micheal Dent, and Peter Pfordresher
Goals of Course: Utilize sound processing software to characterize physical features of acoustic signals and learn to create basic computational models simulating auditory perception.

Professional References

Dr. Joel Snyder
Associate Professor of Psychology
Director of the Neuroscience Ph.D. Program
Department of Psychology
University of Nevada, Las Vegas
Phone: (702) 895-4692
Email: joel.snyder@unlv.edu

Dr. Melissa Gregg
Assistant Professor of Cognitive Psychology and Department Chair
Department of Psychology
University of Wisconsin, Parkside
Phone: (262) 595-2142
Email: greggm@uwp.edu

Dr. Colleen Parks
Associate Professor
Department of Psychology
University of Nevada, Las Vegas
Email: colleen.parks@unlv.edu