Distribution Agreement

Jonathan B. Lin March 30, 2012

Are there two kinds of auditory imagery?

by

Jonathan B. Lin

Krish Sathian, MD, PhD
Adviser

Neuroscience & Behavioral Biology Program

Krish Sathian, MD, PhD

Adviser

Simon Lacey, PhD

Committee Member

Paul R. Lennard, PhD

Committee Member

Kristin F. Wendland, PhD

Committee Member

2012

Are there two kinds of auditory imagery?

by

Jonathan B. Lin

Krish Sathian, MD, PhD

Adviser

An abstract of
a thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Sciences with Honors

Neuroscience & Behavioral Biology Program

2012

Abstract

Are there two kinds of auditory imagery?

By Jonathan B. Lin

Past research into cognitive styles has revealed that there are two types of visual and haptic imagery: object and spatial imagery. Object imagers tend to generate vivid, detailed mental representations, while spatial imagers tend to generate more schematic mental representations. However, it is unknown whether there are two different kinds of auditory imagery mapping onto this distinction. Although vision and audition are very different sensory systems, there are some similarities. For example, auditory imagery and perception rely on some (but not all) of the same neural processing resources, much like visual imagery and perception do. Moreover, visual and auditory imagery have been described in similar domains of vividness and clarity, respectively. We utilized two auditory imagery tasks that required participants to memorize and recognize unfamiliar stimuli; task 1 focused on the auditory surface property of timbre, while task 2 focused on temporal transformations as an auditory structural manipulation. The results revealed no differences between object and spatial imagers on either task. However, task 1 identified a novel inter-individual difference independent of this object-spatial distinction: one group of participants integrated timbre into their auditory images, while the other group did not. Also, consistent with past studies, musicians were superior at both tasks compared to non-musicians. Further research is required to clarify the exact nature of the inter-individual differences identified in the present study and how well they correspond to findings in other modalities.

Are there two kinds of auditory imagery?

by

Jonathan B. Lin

Krish Sathian, MD, PhD

Adviser

A thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Sciences with Honors

Neuroscience & Behavioral Biology Program

2012

Acknowledgements

I would like to thank Dr. Sathian and Dr. Lacey for their mentorship over the past three and a half years. I would also like to thank my Honors thesis committee members for their valuable insight and guidance, and the countless people who volunteered their time to participate in this study. This work was funded by an independent research grant from the Emory College of Arts and Sciences SIRE program for undergraduate research. Support to Dr. Sathian from the National Eye Institute, the National Science Foundation, and the Veterans Administration is also gratefully acknowledged.

Table of Contents

Are there two kinds of auditory imagery?

Introduction

Mental imagery “resembles perceptual experience, but occurs in the absence of the appropriate external stimuli” (Thomas, 2011). Visual imagery, one type of mental imagery, is often described as *seeing with the mind’s eye*. However, mental imagery is not identical in everyone; different people tend to create different kinds of mental representations depending on their individual predispositions (Kosslyn et al., 1984). The first attempt at identifying such inter-individual differences yielded the Verbalizer-Visualizer Questionnaire (Richardson, 1977), placing individuals along a continuum with verbalizers on one end and visualizers on the other. Verbalizers were thought to have low imagery abilities, whereas visualizers were thought to have high imagery abilities. However, subsequent studies suggested that it may not be appropriate to conceptualize these two styles as being on opposite ends of a continuum; rather, they may be parallel processes that are independent from one another (Edwards & Wilkins, 1981).

Recent behavioral research investigating cognitive styles has expanded on this knowledge, revealing that in addition to the ‘verbalizer’ style, there are two main types of visualizers: object and spatial imagers (Kozhevnikov et al., 2005). When asked to imagine a previously-seen item, object imagers tend to generate mental representations that are pictorial, vivid, and detailed; whereas spatial imagers tend to generate mental representations that are more schematic. Moreover, object imagers tend to integrate surface property information, e.g. color and texture, into their mental representations, while spatial imagers do not. An easy way to think of this distinction is as the difference between a photograph, in the case of object imagers, and a blueprint or diagram, in the case of spatial imagers. Furthermore, spatial imagers tend to be superior at performing complex transformations compared to their object imager counterparts. In order to quantify individual imagery preferences, Kozhevnikov and colleagues devised the Object-Spatial Imagery Questionnaire (OSIQ; Blajenkova et al., 2006), which scores individuals along the dual continua of object and spatial imagery. Subsequently, this questionnaire was expanded to incorporate the verbalizer dimension as well, resulting in the Object-Spatial Imagery and Verbal Questionnaire (OSIVQ; Blazhenkova & Kozhevnikov, 2009).

More recently, work in our laboratory has not only corroborated the existence of object and spatial dimensions in visual imagery but also demonstrated that these dimensions extend to haptically-derived representations (Lacey et al., 2011). For these studies, participants performed two tasks – a shape recognition task and a texture recognition task – which were aimed at distinguishing between object and spatial imagers on the basis of whether the participants integrated the surface property of texture into their mental representations. In the shape recognition task, participants were instructed to memorize the shapes of two objects (Shape 1, Shape 2), which happened to have different texture schemes (Figure 1A, left pair). At test, they were presented with the two original objects (Figure 1A, left pair), plus two additional objects that retained the same shapes but exchanged their texture schemes (Figure 1A, right pair). The task was to decide whether the test objects had the same shape as Shape 1 or 2. Participants completed this task in both visual and haptic within-modal conditions. The results demonstrated that, as expected, object imagers were significantly impaired at identifying the shapes with texture changes because they tend to integrate the surface property of texture into their mental representations and therefore misidentified objects when there was a change in texture. This misidentification occurred in both the visual and haptic conditions. Conversely, spatial imagers were unaffected by texture changes in both vision and touch.

The second task, i.e. the texture recognition task, was similar except that participants were asked to memorize the texture schemes of two objects (Texture 1 & 2; Figure 1B, left pair), which happened to have different shapes. At test, participants were presented with the two original objects (Figure 1B, left pair), plus two additional objects that retained the same texture schemes but exchanged their shapes (Figure 1B, right pair). The task now was to decide whether the test objects had the same texture schemes as Texture 1 or 2 and, as before, participants completed this task in both visual and haptic within-modal conditions. As expected, the spatial imagers, but not the object imagers, were impaired by the shape changes because they are typically not accustomed to recalling surface properties and were confused by the change in shape. This behavioral effect was clear in both vision and touch. Cumulatively, this experimental evidence corroborates the existence of these cognitive styles in both vision and touch.

Although vision and touch are very different sensory systems, the fact that object and spatial dimensions extend into visual and haptic imagery makes sense because vision and touch, which are both used to assess many properties of objects, share a number of commonalities. For example, previous work in our own lab has demonstrated that both vision and touch exhibit view-dependence, meaning that changing the orientation of an object impairs performance on a recognition task (Lacey et al., 2009a). Similarly, both visual and haptic object recognition involve robust activations in the lateral occipital complex, which was previously thought to be vision-specific (Amedi et al., 2001; Stilla & Sathian, 2008). All of these past studies demonstrate that visual and haptic processing share many features. Likewise, visual and auditory processing also share several common features. Similar to visual imagery and perception, auditory imagery and perception share some (but not all) neural processing resources (Vlek et al., 2011; Wu et al., 2011). Such sharing of resources between perception and imagery has also been observed in the somatosensory (Yoo et al., 2003) and olfactory (Djordjevic et al., 2005) systems.

Even more interestingly, Daselaar and colleagues showed, through functional magnetic resonance imaging (fMRI), that auditory and visual imagery share some “modality-independent” neural substrates (Daselaar et al., 2010). In these studies, participants saw a cue word and were instructed to do one of four things: 1) imagine a sound associated with the word, i.e. auditory imagery; 2) imagine a picture associated with that word, i.e. visual imagery; 3) listen to a sound associated with that word, i.e. auditory perception; or 4) observe an image associated with that word, i.e. visual perception. A modality-independent neural network including the superior prefrontal cortex, posterior cingulate cortex, lateral parietal cortex, and medial prefrontal cortex was shown to be activated in all four of these conditions using conjunction analyses (Daselaar et al., 2010). Finally, both visual and auditory imagery have been measured in similar domains of vividness (Marks, 1973) and clarity (Willander & Baraldi, 2010), respectively. All in all, it appears that despite the major differences between vision and audition, there are many significant similarities.

Given these similarities between vision and audition, we hypothesized that there may be different kinds of auditory imagery, which may be analogous to the object and spatial dimensions that have been observed in visual and haptic imagery (Lacey et al., 2011). Although this topic of research has been unexplored so far to our knowledge, there has been a significant body of work focused on inter-individual differences in auditory perception (see review by Schneider & Wengenroth, 2009), which might form the basis of differences in auditory imagery preferences. Previous studies have found that regardless of musical ability, individuals can be divided into three groups based on how they tend to perceive auditory stimuli: holistic, spectral, and intermediate listeners (Schneider et al., 2005a). (It is important to note that holistic listeners have also been referred to as synthetic or fundamental listeners and spectral listeners as analytical listeners (Schneider & Wengenroth, 2009); although the terms differ, they describe the same continuum of listening preferences). Most sounds are made up of a complex spectral pattern, i.e. of multiple frequencies; heard all together, these spectral patterns confer characteristics such as timbre. Holistic listeners tend to hear the whole spectral pattern holistically, integrating pitch and timbre. In contrast, spectral listeners tend to break up a complex sound into its spectral components, processing each individual frequency separately. Because these individuals break up the complex pattern of frequencies, they do not process the sound all together and therefore focus less on elements such as timbre. Intermediate listeners fall between these extremes. In these studies, the authors also found that holistic listeners had increased gray matter volume, as measured by structural MRI, and greater neural activity, as measured by a positive magnetoencephalography (MEG) waveform 50 ms after stimulus onset (P50m), in the left lateral Heschl's gyrus (HG), whereas spectral listeners had increased gray matter volume and P50m activity in the right lateral HG (Schneider et al., 2005a). It is logical that differences in HG underlie these two types of auditory perception given that HG – also known as the transverse temporal gyrus – is the primary auditory cortex, the first neural substrate involved in auditory processing.

Moreover, another study in the same laboratory examined the relationship between spectral/holistic listening tendencies and instrument preferences in musicians (Schneider et al., 2005b). The authors reported that musicians who played percussive or higher-pitched instruments, e.g. percussionists, trumpeters, guitarists, exhibited a strong tendency toward holistic listening, while musicians who played lower-pitched instruments, e.g. bassoonists, bassists, organists, had a bias toward spectral listening (Schneider et al., 2005b). However, the authors conceded that the preferences for a particular type of listening in certain instrumentalists, e.g. pianists, violinists, flutists, were more difficult to predict (Schneider et al., 2005b). The authors speculated that the type of listening in which an individual tends to engage may have an impact on what instrument they choose to play based on its timbre, tone, and/or size (Schneider et al., 2005b), though it is also possible that extended musical experience of a particular type can induce brain plasticity and thereby confer such tendencies.

Because perception and imagery are closely related, i.e. imagery generally draws on previously-observed percepts, these inter-individual differences in auditory perception may potentially translate into inter-individual differences in auditory imagery. In this case, because holistic listeners tend to hear the entire spectral pattern and therefore integrate the surface property of timbre, they may be the auditory analog of object imagers. On the other hand, spectral listeners may be the equivalent of spatial imagers because they tend not to encode this surface property but, instead, are more analytically oriented. If this is the case, we recognize that we may have to come up with new designations to describe these two imagery types.

In the present study, we utilized two tasks, based on the visual and haptic texture/shape tasks described previously (Lacey et al., 2011), to test whether there are two kinds of auditory imagery and whether the differences map onto the object-spatial distinction. In the first behavioral task, we focused on the auditory surface property of timbre – the sound quality that distinguishes different instruments. We expected to observe two types of performance, in line with the two types of auditory perception; specifically, we hypothesized that the auditory 'object' imagers would be impaired by changes in timbre, as they tend to integrate surface properties into their mental representations. In the second behavioral task, we focused on the auditory structural property of temporal order, treating time as the auditory analog of space. We expected auditory 'spatial' imagers to be unaffected by the temporal transformations because they are superior at complex transformations. Past studies have confirmed that mental representations of sound retain temporal properties; for example, participants performing an auditory imagery task with familiar songs have a longer reaction time when they have to compare two parts of the song that are farther apart in time (Halpern, 1988).

The immediate aim of this study was to determine whether commonalities observed in visual and haptic imagery extend to auditory imagery, offering the possibility of describing an organizing principle of imagery that applies to all three modalities. We also considered the potential effect of musical expertise, as past studies have revealed behavioral differences between musicians and non-musicians. Most relevant to this study, Aleman and colleagues found that musicians had higher levels of performance in both musical and non-musical auditory imagery tests (Aleman et al., 2000), suggesting some kind of domain-specific expertise. We hypothesized that our auditory imagery tasks, requiring participants to memorize and recognize unfamiliar auditory stimuli, would differentiate between two groups of individuals analogous to object and spatial imagers. We also hypothesized that, consistent with past evidence, musical expertise would improve performance, independent of these inter-individual differences.

Methods

Participants

Twenty-four non-musicians (12 male, 12 female; mean age 21 years, 9 months; standard deviation (sd) 4 years, 3 months) and twenty-four musicians (10 male, 14 female; mean age 20 years, 2 months; sd 2 years, 7 months) took part and were compensated for their time. The musicians and non-musicians did not differ significantly by age: $t_{38.023} = 1.564$, $p = .126$ (corrected for unequal variances). Participants were recruited via advertisements posted on the Emory campus and intranet; all gave informed written consent. All procedures conformed to the Declaration of Helsinki and were approved by the Emory University Institutional Review Board.

Non-musicians were defined as individuals who met three specific criteria. These participants 1) did not play a musical instrument (either currently or any time after age 10), 2) could not read music, and 3) had not taken any formal music lessons/classes for more than six months (because primary/secondary education in the United States sometimes includes some compulsory general music courses). Conversely, musicians were defined as individuals who met the opposite three criteria; musicians 1) currently played a musical instrument in a regular and organized setting, 2) were able to read music, and 3) had taken formal music lessons for at least three years. Although some neuroimaging studies have identified a critical period ending at the age of seven years during which musicians must begin their musical training to bring on musical training-induced neuroplastic changes (Schlaug et al., 2009), we elected not to follow these guidelines, as we considered them to be overly stringent for our purposes. In fact, our criteria were similar to those used in other studies demonstrating significant differences between musicians and non-musicians (e.g. Aleman et al., 2000; Paraskevopoulos et al., 2012). In the former study, the "musically-trained" group consisted of students from Utrecht University who 1) actively played a musical instrument at the date of testing and 2) had received more than two years of formal musical training (Aleman et al., 2000); in the latter, musicians were all students of the Music Conservatory in Münster (Paraskevopoulos et al., 2012). The vast majority of our musician participants were undergraduate music majors/minors enrolled at Emory College and/or members of the Emory University music ensembles, e.g. Emory Symphony Orchestra, Emory

Wind Ensemble, etc. The musician participants collectively played a large variety of instruments, which is important because inter-individual differences in auditory perception in musicians have been shown to be associated with the instrument they play (Schneider et al., 2005b). Details about the non-musician and musician participants are listed in Tables 1 and 2, respectively.

Questionnaires

Immediately preceding both auditory imagery tasks, all participants completed the Object-Spatial Imagery and Verbal Questionnaire (OSIVQ; Blazhenkova & Kozhevnikov, 2009). An OSIVQ difference score (OSIVQd) was calculated for each participant by subtracting the spatial subscore from the object subscore, as described previously (Lacey et al. 2011). A positive score indicates a preference for object imagery, while a negative score indicates a preference for spatial imagery. Moreover, participants completed both the Vividness of Visual Imagery Questionnaire (VVIQ; Marks, 1973) and an analogous questionnaire for auditory imagery, the Clarity of Auditory Imagery Scale (CAIS; Willander & Baraldi, 2010). At the end of the study, we wondered whether the observed inter-individual differences might be explained by inter-individual differences in the ability to control auditory images of voices, music, or environmental sounds. Accordingly, participants were recalled (n = 34; n = 14 unreachable) to complete the unpublished Bucknell Auditory Imagery Control Scale (BAICS), which has been described previously (Zatorre et al., 2010) and was kindly provided to us by the authors.

Auditory stimuli

Auditory stimuli were constructed in *Finale Notepad 2012* (Make Music, Inc., Eden Prairie, MN) and exported in mp3 format at 1411 kbps bit rate. All auditory stimuli used in auditory tasks 1 and 2 were constructed as a sequential combination of three fragments from a collection of fifty-four four-note fragments. These fragments were created first to maintain consistency between the two tasks and for ease of proper counterbalancing.

The notes making up each fragment were selected quasi-randomly with a random integer generator from 0 to 11, such that 0 represented C4 (middle C; ~261.6 Hz), 1 represented C#4 (one semitone above middle C; ~277.2 Hz), 2 represented D4 (two semitones above middle C; ~293.7 Hz), etc.; each note was used only once in each fragment. Nine instruments in the same relative pitch range were used to construct all fifty-four fragments, i.e. six fragments were played on each instrument. These nine instruments came from three broad categories: percussive (piano, marimba, and glockenspiel), plucked (guitar, dulcimer, and harp), and woodwind (flute, recorder, and clarinet). Examining the spectrographs for each of these instruments revealed that within each broad category, spectral patterns were qualitatively similar, whereas patterns characterizing instruments from different categories were distinct from one another. All of the auditory manipulations were performed with the open-source audio-editor *Audacity* (audacity.sourceforge.net/). Prior to our experiments, we normalized the energy content of all fifty-four auditory stimuli with the freeware *WaveGain frontend* (members.home.nl/w.speek/wavegain.htm).

We conducted pilot studies with separate participants from the main study (n = 9) to ensure that none of these fragments were recognizable as originating from a familiar melody. During these pilot experiments, participants were asked to rate each auditory fragment on a 5-point Likert scale in response to the following question: *"Do these four notes remind you of a tune or melody that you have heard before?"* Four familiar four-note fragments from the Westminster Quarters, i.e. the melody played by many clock bells to demarcate each quarter hour, were added as catch-trials. As expected, the catch-trials were indeed identified as familiar with a collective mean rating of 4.56. The majority of the auditory fragments were identified as not familiar (ranging from 1.22 to 3.00). Four fragments that were rated as somewhat familiar (means between 3.00 and 3.50) were replaced by new randomly-constructed fragments, which were then rated satisfactorily, i.e. < 3.00. The fifty-four four-note fragments were used to construct the auditory stimuli used in both task 1 and task 2.

Auditory imagery task 1 stimuli

For task 1, the stimuli were constructed by combining three different fragments, each from a different instrument category, i.e. one percussive fragment, one plucked fragment, and one woodwind fragment. Therefore, each stimulus had both a note order, i.e. the order of the twelve notes, and an instrument order, i.e. the order of the three instruments. These stimuli were placed into pairs (Sequence 1-original & 2-original), such that the two stimuli in each pair used a different set of three instruments (Figure 2A-Original; Figure 2B-Original). For each of these pairs, a corresponding pair of stimuli was constructed (Sequence 1-changed & 2-changed), and these stimuli retained the note orders but exchanged their instrument orders (Figure 2A-Changed; Figure 2B-Changed). In other words, Sequence 1-changed utilized the note order of Sequence 1-original but the instrument order of Sequence 2-original. Conversely, Sequence 2-changed utilized the note order of Sequence 2-original but the instrument order of Sequence 1-original. The spectrographs (showing 172 Hz to 7 kHz) for a representative group of four stimuli constructed in this way are presented in Figure 2. We constructed a total of nine groups of four sequences. The order of instruments was counterbalanced, and each of the fifty-four fragments was used once to create the original (unchanged) stimuli pairs.

Auditory imagery task 1 procedure

During the experimental task, participants were asked to memorize two sequences with different note orders and different instrument orders (note order #1/instrument order #1 & note order #2/instrument order #2), which they memorized as Sequence 1 and Sequence 2. The auditory stimuli were presented through stereo computer speakers at ~70 dB. During the experiment, participants could see a laptop screen, which displayed visual cues to guide them through the task. Participants heard each sequence twice; this decision was based on pilot studies that revealed floor effects when participants were able to hear the sequences only once. At test, participants were presented with the original sequences plus two additional sequences in which the note orders were the same but the instrument orders were exchanged (note order #1/instrument order #2 & note order #2/instrument order #1). The task was to decide whether the sequence was the same as Sequence 1 or Sequence 2 in a two-alternative forced choice (2AFC) paradigm. The specific instructions given to the participants are listed in Appendix A. Participants were instructed to indicate their responses by pressing one of two mouse buttons; these responses were recorded with the Presentation software (Neurobehavioral Systems, Inc., Albany, CA).

Performance on this task was calculated as the percentage of correct responses. Two scores were calculated, one for the accuracy in identifying the original sequences (*task 1 orig*), i.e. the ones that were the same as those memorized, and another for the accuracy in identifying the sequences with the exchanged instrument orders (*task 1 changed*). We expected that comparing these two different scores would allow us to identify whether individuals were affected by the instrument changes. Another score, the *task 1 percentage drop* was calculated by the following formula: (100 × [*task 1 changed – task 1 orig*]) / *task 1 orig*. This new score was intended to quantify the impairment on the task caused by the changes in timbre.

Auditory imagery task 2 stimuli

For task 2, the auditory stimuli also consisted of a sequential combination of three different fragments, but here, they were all played on the same instrument. We constructed a temporally transformed version of each of these stimuli, i.e. we reordered the three fragments within the stimulus. There are five possible temporal transformations, and they were used in roughly equal proportions. During the task, the two sequences of the same instrument were presented as pairs (Figure 3A; 3C), along with the corresponding temporally-transformed sequences (Figure 3B; 3D). Because we used nine instruments to construct the fragments, there were nine groups of four sequences, using each of the fifty-four fragments twice. A representative group of four stimuli constructed in this way is presented in Figure 3.

Auditory imagery task 2 procedure

During the experimental task, participants were asked to memorize the three fragments that made up each of the stimuli as Set 1 and Set 2. As in the previous task, the auditory stimuli were presented through stereo computer speakers at ~70 dB with visual cues displayed on the laptop screen. Participants heard each of these stimuli twice; this decision was based on pilot studies that revealed floor effects when participants were able to hear the sequences only once. At test, participants heard both the original, unchanged stimuli, plus the two corresponding temporally-transformed stimuli. They were asked to decide whether the stimulus they heard was made up of the fragments from Set 1 or Set 2 in a 2AFC paradigm. The specific instructions given to the participants are listed in Appendix B. As in task 1, participants were instructed to indicate their responses by pressing one of two mouse buttons, and these responses were recorded with the Presentation software (Neurobehavioral Systems, Inc., Albany, CA).

Performance on this task was calculated as the percentage of correct responses. Two scores were calculated, one for the accuracy in identifying the original stimuli (*task 2 orig*), i.e. the ones that were the same as those memorized, and another for the accuracy in identifying the temporally-transformed stimuli (*task 2 changed*). Another score, the *task 2 percentage drop* was calculated by the following formula: (100 × [*task 2 changed* – *task 2 orig*]) / *task 2 orig*. This score was intended to quantify the impairment on the task caused by the temporal transformations.

Results

Auditory imagery task 1

Participants were classified as object or spatial imagers by their OSIVQd score with a positive score indicating an object imager and a negative score indicating a spatial imager, as previously described (Lacey et al., 2011). Based on this method of classification, there were more object imagers (n = 35) than spatial imagers (n = 12). One participant with an OSIVQd = .00 was omitted from the subsequent analysis. We ran a mixed ANOVA (between-subjects factor: imagery preference, i.e. object versus spatial; within-subjects factor: stimulus type, i.e. original versus changed). Although performance was impaired when the instrument orders changed ($F_{1,45} = 4.807$, $p < .05$; Figure 4), there was no difference between the object and spatial imagers ($F_{1,45} = 2.513$, $p = .120$; Figure 4). There was also no interaction ($F_{1,45} = .598$, $p = .554$; Figure 4). Although unequal sample sizes are sometimes a concern, ANOVA is relatively robust against violations of this assumption as long as there is homogeneity of variance between the two groups (object and spatial imagers), as there was in this case (Levene's Test, $p = .332$ in *task 1 orig*; $p = .847$ in *task 1 changed*). (A non-significant p-value on Levene's test indicates that the variances of the two data sets are similar).

To explore whether musical experience affects performance on auditory imagery task 1, we performed a second mixed ANOVA (between-subjects factor: musical expertise, i.e. musicians ($n = 24$) versus non-musicians ($n = 24$); within-subjects factor: stimulus type, i.e. original versus changed). Consistent with our hypothesis, musicians were superior at auditory imagery task 1 compared to non-musicians ($F_{1,46} = 12.568$, $p = .001$; Figure 5). Moreover, participants performed better on the original sequences compared to the sequences with changed instrument orders ($F_{1,46} = 41.275$, $p < .001$; Figure 5), although there was no interaction between these two factors: $F_{1,46} = .247$, $p = .622$ (Figure 5). We interpreted these results cautiously, as we observed significant heterogeneity of variance in *task 1 orig* scores (Levene's test, $p = .008$) and near-significance in *task 1 changed* scores ($p = .054$). We could not correct this heterogeneity through either log or square-root transformations. Although ANOVA can be considered to be robust against heterogeneity of variance with equal samples sizes, we took a more conservative approach by using non-parametric tests. Non-parametric tests based on rank transformations, including those described below, generally do not require homogeneity of variance (Conover & Iman, 1981; Siegel, 1957). The Mann-Whitney U Test corroborated the main effect of musical expertise: $U = 658.5$, $z = -3.648$, $p < .001$; the Wilcoxon Signed-Rank Test corroborated the main effect of changing instrument orders: $Z = -4.980$, $p < .001$. Because the more conservative non-parametric tests yielded the same conclusions, we were comfortable reporting the parametric results also.

Auditory imagery task 2

The same analysis, as described for auditory imagery task 1, was performed for auditory imagery task 2. We performed a mixed ANOVA (between-subjects factor: imagery preference; within-subjects factor: stimulus type, i.e. original versus changed/transformed). The participant with an OSIVQd = .00 was omitted from this analysis. Object and spatial imagers had indistinguishable levels of performance on auditory task 2: $F_{1,45} = 3.170$, $p = .082$ (Figure 6). Moreover, participants performed similarly on both the original stimuli and those that were temporally transformed ($F_{1,45} = 1.568$, $p = .217$; Figure 6), with no interaction between the two factors: $F_{1,45} = .039$, $p = .845$ (Figure 6). As before, we were not concerned by the unequal sample sizes because there was homogeneity of variance (Levene's Test, $p = .783$ in *task 2 orig*; $p = .192$ in *task 2 changed*).

To explore whether musical expertise affects performance on auditory imagery task 2, we performed a mixed ANOVA (between-subjects factor: musical expertise; within-subjects factor: stimulus type). As expected, musicians performed significantly better than non-musicians ($F_{1,46} = 35.702$, $p < .001$; Figure 7). However, participants performed equally well when identifying the original stimuli and those that were temporally transformed ($F_{1,46} = 2.785$, $p = .102$; Figure 7). There was no interaction between these two factors: $F_{1,46} = .309$, $p = .581$ (Figure 7). There was no need to consider non-parametric alternatives because of equal sample sizes and homogeneity of variance (Levene's test, $p = .073$ for *task 2 orig*; $p = .799$ for *task 2 changed*).

Inter-individual differences in performance

Although there were no inter-individual differences in performance on either auditory imagery task 1 or 2 based on imagery preference as measured by OSIVQd scores, there did appear to be high variability in performance. Particularly, we noticed that some individuals seemed to be more affected by the changes in instrument order and temporal transformations compared to others. Accordingly, two new variables were calculated, *task 1 drop* and *task 2 drop*, to represent the percentage drop in performance for the changed stimuli versus the original stimuli in both tasks. To test this observed individual variability systematically, we used k-means clustering (designating k = 2) to determine whether there were two levels of performance.

For the variable *task 1 drop*, k-means clustering revealed two groups that were distinct and equal in size. Cluster 1 (n = 24) had a mean *task 1 drop* of -77.26%, while cluster 2 (n = 24) had a mean *task 1 drop* of -2.58%, corroborating the existence of two levels of performance (Figure 8). Cluster membership was independent of musical expertise and imagery preference as measured by OSIVQd scores: $X^2_{1,\, n = 48} = 1.333$, p = .248 and $X^2_{1,\, n = 47} = 1.570$, p = .210, respectively.

For the variable *task 2 drop*, k-means clustering revealed two groups, but the majority of participants fell into one instead of the other (n = 6 in cluster 1; n = 42 in cluster 2; Figure 9). Participants in cluster 2 included both musicians and non-musicians. However, the majority of participants falling into cluster 1 (mean task 2 drop = +46.0%) were non-musicians (n = 5 of 6, or 83.3%). Most participants in the smaller cluster 1 performed near or below chance levels of accuracy on the original, untransformed stimuli, suggesting that these positive percentage changes may have been caused by random chance rather than being meaningful.

Basis of inter-individual differences in task 1

Although these inter-individual differences in task 1 did not appear to map onto the object-spatial distinction previously observed in visual and haptic imagery (Lacey et al., 2011), we hypothesized that they could fall along other imagery-related scales. To this end, we performed correlation analyses on *task 1 drop* and scores on the VVIQ, the CAIS, and the BAICS. There was no significant correlation between VVIQ scores and *task 1 drop* ($r = .041$, $p = .781$). The correlation between CAIS scores and *task 1 drop* was weak but trended toward significance ($r = .263$, $p = .071$). Further analysis revealed that this correlation between CAIS scores and *task 1 drop* was moderate but significant in musicians ($r = .486$, $p = .016$), while not significant in non-musicians ($r = .064$, $p = .765$). This was not due to either differences in the mean CAIS scores of musicians versus non-musicians ($t_{46} = -.118$, $p = .906$) or differences in the variance of these scores (Levene's test; $p = .336$). Also of note, the CAIS scores were found to be significantly correlated with VVIQ scores ($r = -.320$, $p = .027$) in all participants. The VVIQ is scored such that lower scores indicate more vivid visual imagery; therefore, this negative correlation can be interpreted to mean that individuals with more vivid visual images tend to have clearer auditory images and vice versa. Scores on the BAICS were uncorrelated with *task 1 drop* ($r = .275$, $p = .116$, $n = 34$), even when the musicians and non-musician were analyzed separately ($r = .140$, $p = .649$, $n = 13$; $r = .369$, $p = .099$, $n = 21$; respectively).

Discussion

Although we expected to observe analogs of object and spatial imagery in our auditory imagery experiment, neither of our tasks revealed an effect of imagery preference on performance. In fact, object and spatial imagers performed almost identically when ignoring the differences between musicians and non-musicians. These data support two possible conclusions: either 1) auditory imagery does not possess equivalents of object and spatial imagery or 2) our auditory imagery tasks did not appropriately tap into the elements of an auditory image that would have revealed these object-spatial dimensions. However, auditory task 1 revealed a novel inter-individual difference, distinguishing between individuals who integrated timbre in their auditory images and those who did not. These differences may very well imply the presence of imagery preferences that correspond with the holistic and spectral modes of perception.

Object and spatial properties in auditory imagery

Because a number of similarities between visual and auditory imagery have been observed, we hypothesized that we could identify inter-individual differences in performance on two auditory imagery tasks, which would map onto the object-spatial distinction. This hypothesis was informed by previous studies reporting common object-spatial dimensions in vision and touch (Lacey et al., 2011). However, unlike visual and haptic images, auditory images are very different because they deal with characteristics in time rather than in physical space. In our past experiments linking vision and touch, we chose to use the surface property of texture to identify the object imagers and the structural property of shape to identify the spatial imagers. Both of these characteristics can be easily perceived and imagined both visually and haptically. Although we ultimately were able to identify an auditory analog of a surface property, we faced more difficulty identifying the auditory equivalent of shape and conceptualizing what an auditory spatial/structural manipulation would entail.

In designing the task to identify the auditory equivalent of object imagery, we chose to focus on the surface property of timbre, i.e. the spectral pattern of an auditory stimulus that yields its instrument quality. Differences in this spectral pattern, which generally involve keeping the same fundamental frequency while drastically altering the pattern of overtones, yield a completely different sound (e.g. comparing the sound of a violin versus that of a piano), although the pitch is generally perceived to be the same. Much like color and texture, timbre is a surface property that can be modulated while keeping all other properties the same. The choice of timbre as an auditory surface property for identifying inter-individual differences in imagery is also supported by past literature investigating inter-individual differences in auditory perception. The designations of holistic versus spectral listeners (Schneider et al., 2005a) could feasibly lead to two different forms of imagery preferences because holistic listeners *hear* timbres when they incorporate the entire spectral pattern of the sound, whereas spectral listeners hear individual spectral elements separately, thereby paying less attention to timbre. Previous studies have also shown that auditory input, e.g. sounds of crinkling and crumpling, can contribute information about an object's material composition, much like somatosensory texture information can (Arnott et al., 2008). This acoustically derived information likely arises from perceiving the sound's pattern of spectral frequencies, i.e. timbre, drawing a further parallel between somatosensory texture and auditory timbre as comparable surface properties.

However, the spatial analog in auditory imagery is much more difficult to pinpoint. Spatial imagers are much better than object imagers at performing complex transformations in the visual and haptic domains. Accordingly, it would be logical that if auditory 'spatial' imagers exist, they would be superior at performing the equivalent of a complex transformation in the auditory domain. In the present study, we chose to utilize temporal transformations, as described for auditory imagery task 2, to identify individuals who are superior at this type of task. This task was based on a past fMRI study, during which participants heard the first few notes of a familiar melody and then a string of notes; participants had to identify whether the string of notes was an exact or an inexact reversal of the fragment from the familiar melody (Zatorre et al., 2010). This melody reversal task was shown to associate with strong activations in the intraparietal sulcus (Zatorre et al., 2010), a region that has been implicated in spatial imagery in visuo-haptic representations (see review by Lacey et al., 2009b). However, we decided to use unfamiliar pitch sequences to maintain consistency between the two auditory imagery tasks and also used a slightly modified type of transformation, as we imagined that melody reversals would have been too difficult with unfamiliar pitch sequences.

Although our choices of potential candidates for surface and structural properties of auditory stimuli are supported by past literature, we concede that these dimensions do not map perfectly in the auditory domain and therefore complicate the conclusions that we can draw from our experimental data. Our data analyses revealed comparable performances in object and spatial imagers, as identified by their OSIVQd scores, on both auditory tasks. However, it remains unclear whether this is because there are no equivalents of object-spatial dimensions in auditory imagery or whether the tasks we designed do not tap into the appropriate elements of an auditory image.

Effect of musical expertise

In our studies, we did find a significant main effect of musician expertise on both auditory tasks, supporting the fact that musicians are broadly superior at auditory imagery tasks compared to non-musicians. This finding is consistent with our hypothesis and not entirely surprising, as past studies have reported similar behavioral results (e.g. Aleman et al., 2000). This effect is likely because musicians both have had extensive instrumental training and are exposed to music on a regular basis. Herholz and colleagues conducted a MEG study to elucidate the neural basis of musical imagery and the effect of musical experience (Herholz et al., 2008). Based on their findings, they cautiously argued that although musicians and non-musicians have similar neural networks for music imagery, they differ in their musical imagery abilities because musicians' intense training can induce neuroplastic changes, thereby fine-tuning the network (Herholz et al., 2008). Moreover, it is also possible that musicians may approach auditory imagery tasks in a different way due to their expertise. This possibility is supported by a previous study showing that musicians have superior pre-attentive auditory processing and can extract more information from musically-relevant stimuli compared to their non-musician peers (Koelsch et al., 1999). We concede that a few of our musician participants may have possessed absolute pitch – or, perfect pitch – which may have confounded our results because these individuals would have been able to perform this task as a simple recall task rather than as an imagery task. Nonetheless, absolute pitch is relatively rare; previous studies have estimated that approximately 7.3% of students in university music programs like that at Emory have absolute pitch (Gregersen et al., 1999). This estimate corresponds to one or two musicians with absolute pitch out of our group of n = 24.

Given that the inter-individual differences in auditory perception have been reported to align with particular instrument preferences (Schneider et al., 2005b), we examined whether there was a qualitative relationship between OSIVQd scores and instrument preferences. There was no obvious association (Table 2), suggesting that a preference for object versus spatial imagery may not be related to the specific instrument a musician plays.

Presence of inter-individual differences

Despite the fact that we did not find any inter-individual differences mapping onto the object-spatial distinction, this does not rule out the possibility that inter-individual differences in auditory imagery exist along other dimensions. Supporting this possibility, k-clustering for the auditory imagery task 1 percentage drop after timbre changes revealed two distinct and coherent clusters of the same size (n = 24 each); one of which clustered at -2.58%, and the other clustered at -77.26%. These results suggest that there are two types of individuals: individuals who are impaired by the timbre changes and those who are not. For further studies, we would need to identify a task on which the latter group is impaired. It is important to note that cluster membership was found to be independent of both musical expertise and object-spatial imagery preference. These findings suggest that this inter-individual difference exists in both musicians and non-musicians alike and reinforces the idea that these inter-individual differences are independent of the object-spatial imagery dimensions, respectively. Therefore, these labels of object versus spatial do not appear to be as relevant in the auditory domain. There may be more general labels for these inter-individual differences in imagery, e.g. depictive versus schematic, that may not be modality-specific, but precisely what they are remains to be determined.

Although we anticipated that our auditory imagery tasks would reveal two types of auditory imagery that would map onto the object-spatial distinction, we instead identified an equally interesting inter-individual difference along a different dimension. Our results demonstrate that some individuals integrate the property of timbre into their auditory images, whereas others do not, implying the existence of the inter-individual difference of whether timbre is maintained in the auditory image. If we define auditory object imagers as individuals who incorporate the surface property of timbre into their auditory images, this group clearly fits this description. However, because this preference cannot be identified by the OSIVQ, it appears that a preference for auditory object imagery does not necessarily correspond with a preference for visuo-haptic object imagery. These differences could very well be explained by the presence of inter-individual differences in auditory perception, i.e. holistic versus spectral listeners. Because holistic listeners tend to perceive timbre, they should logically integrate this surface property into their auditory images; conversely, because spectral listeners tend not to perceive timbre, they would not be expected to integrate this property into their auditory images. However, it remains to be tested whether this clear link exists.

For further exploration, we tested whether there was any correlation between any established imagery-related scales and *task 1 drop*. No such correlation was observed with the VVIQ. This is not surprising given that this questionnaire measures the vividness of visual imagery rather than an auditory-specific characteristic. Furthermore, no correlation was observed with the BAICS, implying that these inter-individual differences are independent of the control an individual has on his/her auditory images. Interestingly, there was a moderate positive correlation between the CAIS scores and *task 1 drop* in musicians. This finding is logical: individuals who can mentally form clearer auditory images can better distinguish between instrument changes and are thus less likely to be confused and make errors because of instrument changes. However, this correlation was absent in non-musicians, implying that expertise may indeed underlie auditory imagery task performance. In other words, consistent with previous experiments, musicians and non-musicians may be using similar neural networks (Herholz et al., 2008) to perform these tasks and fall along similar ranges of CAIS scores. Nonetheless, intense musical training grants musicians an advantage in these auditory tasks, perhaps because of the “fine-tuning” of these neural networks (Herholz et al., 2008). Training-induced neuroplastic changes have been reported in musicians, e.g. increased grey matter in motor, auditory, and visual-spatial regions (Gaser & Schlaug, 2003), but the functional role of these changes in the context of musical and auditory imagery has been relatively unexplored to date.

Interestingly, we observed a moderate, negative correlation between VVIQ and CAIS scores in all participants regardless of their musical expertise, suggesting that individuals with more vivid visual images tend to have clearer auditory images (low VVIQ scores denote more vivid visual imagery, whereas high CAIS scores denote clearer auditory imagery). This observation is surprising because Willander and Baraldi (2010) reported that the CAIS is designed to assess clarity, which they argue is distinct from vividness. Nonetheless, a similar correlation has also been observed between the Spanish version of the CAIS and the VVIQ-2 ($r = .42$, $p < .001$; Campos & Pérez-Fabello, 2011). This finding is consistent with our results, even though the sign of the r-value is opposite: the VVIQ-2 scoring system is reversed compared to the original VVIQ (high VVIQ-2 score denote more vivid visual imagery; Marks, 1995). The moderate correlations found in our study and that of Campos & Pérez-Fabello (2010) suggest that vividness of visual images and clarity of auditory images may be linked or associated in some way.

Finally, the existence of two types of auditory imagers, as revealed by auditory imagery task 1, begs the question of what characterizes these two groups. One group appears to be analogous to ‘object’ imagers, as these individuals seem to integrate the surface property of timbre into their auditory images. However, it is still uncertain what the best way to describe the opposite group would be. Temporal transformations do not appear to capture this difference, as neither musicians nor non-musicians, on average, seemed to be significantly impaired by these changes. Furthermore, k-clustering of auditory imagery task 2 percentage drop did not yield two groups that were clearly distinct; instead, this analysis yielded one group of n = 42 and another of the remaining n = 6, which were not particularly coherent upon further inspection. It is possible that the temporal transformation was not challenging enough to reveal two levels of performance since no participants seemed to be impaired. If this is the case, it would be pertinent to test a more drastic transformation, such as the reversals which were used in the past (Zatorre et al., 2010).

To this end, we piloted a reversal task with fragments from familiar melodies along with auditory imagery task 1, which was described herein. However, while Zatorre and colleagues reported an average accuracy above 80% (Zatorre et al., 2010), performance on our melody reversal task was near chance levels (n = 9; mean accuracy 54.2%; range 31.3% to 68.8%). Moreover, there was no correlation between these scores and *task 1 drop* (r = -.267, p = .488), suggesting that any potential inter-individual differences that could be identified by the melody reversal task would not fall along the inter-individual differences that we identified in the present study. Finally, there were no significant correlations between accuracy on the melody reversal task and scores on any of the imagery-related questionnaires (OSIVQ, VVIQ, CAIS, & BAICS). Cumulatively, we interpret this pilot data to mean that melody reversals may not tap into the auditory analog of a spatial transformation.

Further experiments

To more thoroughly answer the question motivating this research study, it may be valuable to test a number of other candidates for object and spatial equivalents in auditory imagery. Although we did not find inter-individual differences along the object-spatial dimensions in the present study, this finding does not rule out the possibility that they exist. We simply may not have identified the appropriate object and spatial analogs. Some other possibilities could include melodic contour and different types of temporal transformations, as object and spatial properties, respectively. Melodic contour may confer a particular mood to an auditory stimulus, much like an auditory *color*; moreover, more difficult types of temporal transformations may better distinguish participants, as our participants, on average, were unaffected by the transformations we used in this study. The tasks to test these candidates could be designed in a similar way, and the results could be analyzed to determine whether these tasks are tapping into appropriate object and spatial dimensions.

Moreover, it would be informative to examine the novel inter-individual differences that have been identified with auditory imagery task 1. These inter-individual differences may potentially arise from the two main types of perception, as described in previous studies. To test this, it would be relatively straightforward to test participants on auditory imagery task 1 from the present study and classify the same participants as holistic or spectral listeners based on their performance on a pitch test described previously (Schneider et al., 2005a; 2005b). However, these inter-individual differences could also be based on some other dimension that has not been described to date. Ultimately, further exploration of inter-individual differences in auditory imagery may lead to the conceptualization of a new organizing principle to describe auditory imagery, with potential applications in neurorehabilitation for individuals with sensory deficits.

Conclusion

In this study, we sought to identify whether there are two types of auditory imagery along the object-spatial distinction, which has been previously described in visual and haptic imagery. Interestingly, we did not observe any significant differences between the object and spatial imagers on either of the two auditory imagery tasks, implying either 1) that object/spatial imagery do not map neatly in auditory imagery or 2) that our tasks do not tap into the appropriate dimensions. Moreover, we observed a significant effect of musical expertise, such that musicians were superior at both tasks, which is consistent with past findings.

Nonetheless, despite these results, auditory imagery task 1 distinguished between two groups of individuals: those who were impaired on an auditory recognition task by timbre changes and those who were not. Although we conceived this experiment to identify two different kinds of auditory imagery that map onto the object-spatial distinction, we instead identified an equally novel kind of inter-individual difference along a different dimension. The origin of these inter-individual differences is yet to be determined. Cumulatively, the bulk of the experimental evidence suggests that there may indeed be inter-individual differences in auditory imagery; further studies are necessary to clarify the exact nature of these differences and how well they correspond to findings in other modalities.

References

Aleman, A., Nieuwenstein, M. R., Böcker, K. B. & de Haan, E. H. (2000). Music training and mental imagery ability. *Neuropsychologia*, 38(12): 1664-1668.

Amedi, A., Malach, R., Hendler, T., Peled, S. & Zohary, E. (2001). Visuo-haptic object-related activation in the ventral visual pathway. *Nat Neurosci*, 4(3): 324-330.

Arnott, S. R., Cant, J. S., Dutton, G. N. & Goodale, M. A. (2008). Crinkling and crumpling: An auditory fMRI study of material properties. *Neuroimage*, 43(2): 368-378.

Blajenkova, O., Kozhevnikov, M. & Motes, M. A. (2006). Object-spatial imagery: a new self-report imagery questionnaire. *Appl Cognit Psychol*, 20: 239-263.

Blazhenkova, O. & Kozhevnikov, M. (2009). The new object-spatial-verbal cognitive style model: Theory and measurement. *Appl Cognit Psychol*, 23(5): 638-663.

Campos, A. & Pérez-Fabello, M. J. (2011). Some psychometric properties of the Spanish version of the Clarity of Auditory Imagery Scale. *Psychol Rep*, 109(1): 139-146.

Conover, W. J. & Iman, R. L. (1981). Rank transformations as a bridge between parametric and nonparametric statistics. *The American Statistician*, 35(3): 124-129.

Daselaar, S. M., Porat, Y., Huijbers, W. & Pennartz, C. M. (2010). Modality-specific and modality-independent components of the human imagery system. *Neuroimage*, 52(2): 677-685.

Djordjevic, J., Zatorre, R. T., Petrides, M., Boyle, J. A. & Jones-Gotman, M. (2005). Functional neuroimaging of odor imagery. *Neuroimage*, 24(3): 791-801.

Edwards, J. E. & Wilkins, W. (1981). Verbalizer-Visualizer Questionnaire: Relationship with imagery and verbal–visual ability. *Journal of Mental Imagery*, 5(2): 137-142.

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

Gregersen, P. K., Kowalsky, E., Kohn, N. & Marvin, E. W. (1999). Absolute pitch: prevalence, ethnic variation, and estimation of the genetic component. *Am J Hum Genet*, 65(3): 911-913.

Halpern, A. R. (1988). Mental scanning in auditory imagery for songs. *J Exp Psychol Learn*, 14(3): 434-443.

Herholz, S. C., Lappe, C., Knief, A. & Pantev, C. (2008). Neural basis of music imagery and the effect of musical expertise. *Eur J Neurosci*, 28(11): 2352-2360.

Koelsch, S., Schröger, E. & Tervaniemi, M. (1999). Superior pre-attentive auditory processing in musicians. *Neuroreport*, 10(6): 1309-1313.

Kosslyn, S. M., Brunn, J., Cave, K. R. & Wallach, R. W. (1984). Individual differences in mental imagery ability: a computational analysis. *Cognition,* 18(1-3): 195-243.

Kozhevnikov, M., Hegarty, M. & Mayer, R. (2002). Revising the visualizer-verbalizer dimension: Evidence for two types of visualizers. *Cognition Instruct*, 20: 37-77.

Kozhevnikov, M., Kosslyn, S. M. & Shephard, J. (2005). Spatial versus object visualisers: a new characterisation of cognitive style. *Mem Cognition*, 33: 710-726.

Lacey, S., Lin, J. B. & Sathian, K. (2011). Object and spatial imagery dimensions in visuo-haptic representations. *Exp Brain Res*, 213(2-3): 267-273.

Lacey, S., Pappas, M., Kreps, A., Lee, K. & Sathian, K. (2009a). Perceptual learning of view-independence of visuo-haptic object representations. *Exp Brain Res*, 198(2-3): 329-337.

Lacey, S., Tal, N., Amedi, A. & Sathian, K. (2009b). A putative model of multisensory object representation. *Brain Topogr*, 21(3-4): 269-274.

Marks, D. F. (1973). Visual imagery differences in the recall of pictures. *Br J Psychol*, 64(1): 17-24.

Marks, D. F. (1995). New directions for mental imagery research. *Journal of Mental Imagery*, 19(3-4): 153-167.

Paraskevopoulos, E., Kuchenbuch, A., Herholz, S. C. & Pantev, C. (2012). Statistical learning effects in musicians and non-musicians: An MEG study. *Neuropsychologia*, 50(2): 341-349.

Richardson, A. (1977). Verbalizer-visualizer: A cognitive style dimension. *Journal of Mental Imagery*, 1(1): 109-125.

Schlaug, G., Forgeard, M., Zhu, L., Norton, A., Norton, A. & Winner, E. (2009). Training-induced neuroplasticity in young children. *Ann N Y Acad Sci*, 1169: 205-208.

Schneider, P., Sluming, V., Roberts, N., Bleeck, S. & Rupp, A. (2005b). Structural, functional, and perceptual differences in Heschl's gyrus and musical instrument preference. *Ann N Y Acad Sci*, 1060: 387-394.

Schneider, P., Sluming, V., Roberts, N., Scherg, M., Goebel, R., Specht, H. J., Dosch, H. G., Bleeck, S., Stippich, C. & Rupp, A. (2005a). Structural and functional asymmetry of lateral Heschl's gyrus reflects pitch perception preference. *Nat Neurosci*, 8: 1241-1247.

Schneider, P. & Wengenroth, M. (2009). The neural basis of individual holistic and spectral sound perception. *Contemporary Music Review*, 28(3): 315-328.

Siegel, S. (1957). Nonparametric statistics. *The American Statistician*, 11(3): 13-19.

Stilla, R. & Sathian, K. (2008). Selective visuo-haptic processing of shape and texture. *Hum Brain Mapp*, 29(10): 1123-1138.

Thomas, N. J. T. (2011). Mental Imagery. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2011 Edition). Retrieved March 4, 2012, from plato.stanford.edu/

Vlek, R. J., Schaefer, R. S., Gielen, C. C. A. M., Farquhar, J. D. R. & Desain, P. (2011). Shared mechanisms in perception and imagery of auditory accents. *Clin Neurophysiol*, 122: 1526-1532.

Willander, J. & Baraldi, S. (2010). Development of a new Clarity of Auditory Imagery Scale. *Behav Res Methods*, 42: 785-790.

Wu, J., Yu, Z., Mai, X., Wei, J. & Luo, Y. (2011). Pitch and loudness information encoded in auditory imagery as revealed by event-related potentials. *Psychophysiology*, 48: 415-419.

Yoo, S.-S., Freeman, D. K., McCarthy, J. J. & Jolesz, F. A. (2003). Neural substrates of tactile imagery: a functional MRI study. *Neuroreport*, 14(4): 581-585.

Zatorre, R. J., Halpern, A. R. & Bouffard, M. (2010). Mental reversal of imagined melodies: A role for the posterior parietal cortex. *J Cognitive Neurosci*, 22: 775-789.

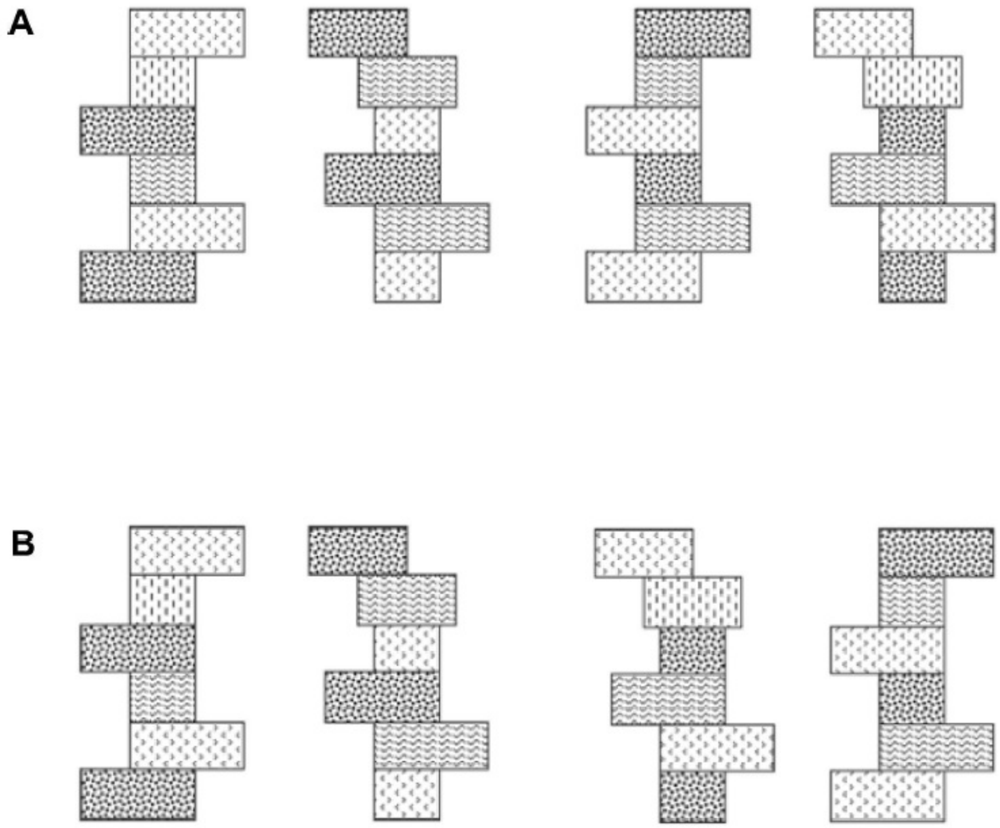

Figure 1 A) Schematic example of Shapes 1 and 2 with (left pair) original texture schemes and with (right pair) the texture schemes exchanged. (B) Example of Textures 1 and 2 with (left pair) original shapes and with (right pair) the shapes exchanged. Reproduced from Lacey et al., 2011.

Table 1 Demographic information of non-musician participants

Age	Gender	Instrument	OSIVQd (pref)
19	M	Non-musician	.067 (object)
22	F	Non-musician	.200 (object)
19	F	Non-musician	1.133 (object)
24	F	Non-musician	-2.400 (spatial)
18	F	Non-musician	1.800 (object)
26	F	Non-musician	.267 (object)
18	M	Non-musician	-.067 (spatial)
19	F	Non-musician	-.533 (spatial)
19	M	Non-musician	0.000 (N/A)
19	M	Non-musician	.267 (object)
29	F	Non-musician	1.800 (object)
28	M	Non-musician	.533 (object)
30	M	Non-musician	1.467 (object)
18	F	Non-musician	-1.200 (spatial)
27	M	Non-musician	-.333 (spatial)
19	M	Non-musician	-1.467 (spatial)
18	M	Non-musician	-1.600 (spatial)
18	M	Non-musician	.467 (object)
18	F	Non-musician	-.200 (spatial)
18	F	Non-musician	.533 (object)
26	M	Non-musician	.733 (object)
28	F	Non-musician	.733 (object)
21	F	Non-musician	1.000 (object)
21	M	Non-musician	.667 (object)

Table 2 Demographic information of musician participants

Age	Gender	Instrument	OSIVQd (pref)
30	F	Viola	1.933 (object)
19	F	Clarinet & piano	-.067 (spatial)
21	F	Oboe & English horn	1.400 (object)
18	M	Trumpet	-.400 (spatial)
20	F	Flute	.133 (object)
19	M	Piano & violin	1.133 (object)
21	M	Piano (jazz)	1.267 (object)
18	M	Violin	-.533 (spatial)
18	F	Trombone	1.267 (object)
21	M	Piano & violin	.667 (object)
18	F	Piano	1.267 (object)
18	F	French horn	.600 (object)
20	F	Cello	.733 (object)
19	F	Violin	-.133 (spatial)
23	M	Guitar	.533 (object)
20	F	Piano	2.267 (object)
20	F	Piano	1.133 (object)
22	M	Piano	1.133 (object)
21	F	Harp	2.133 (object)
18	F	Violin	2.000 (object)
19	F	Piano	1.200 (object)
18	M	Flute	.733 (object)
21	M	Trumpet & piano	.800 (object)
22	M	Flute & piccolo	1.600 (object)

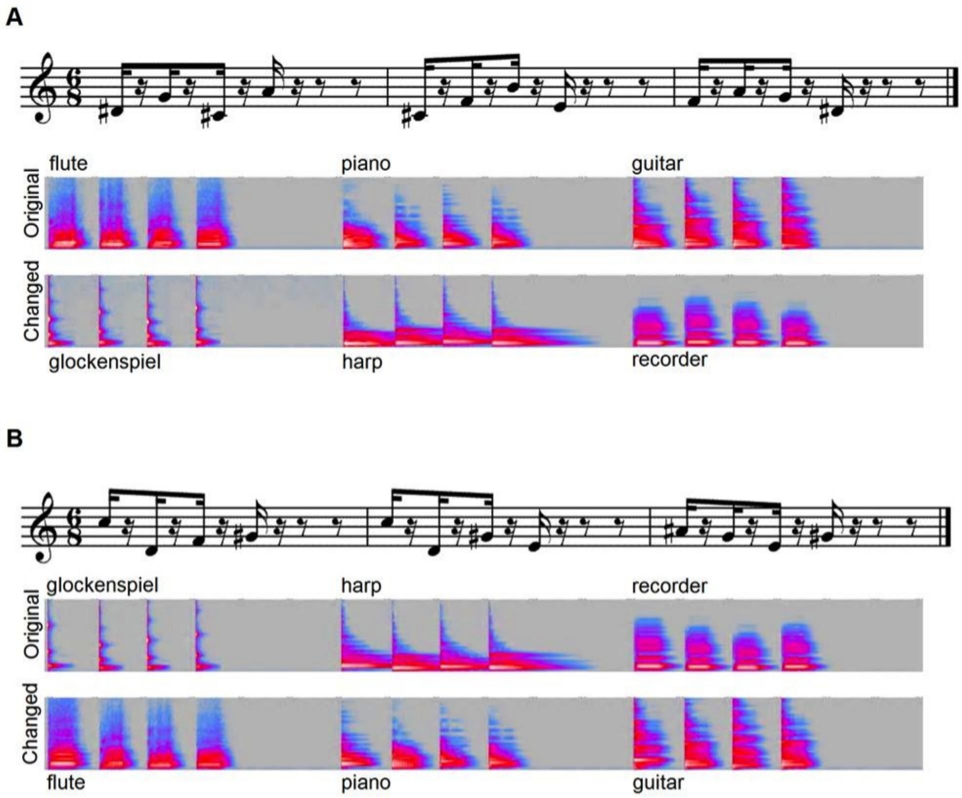

Figure 2 Spectrographic representations of sample Sequence 1 and 2 with original instruments (A-Original; B-Original) and changed instruments (A-Changed; B-Changed). The musical notation denotes the note order of the two spectrographs below it. Note that the instruments used in A-Original are the same as those used in B-Changed (flute, piano, guitar), while those used in B-Original are the same as those used in A-Changed (glockenspiel, harp, recorder).

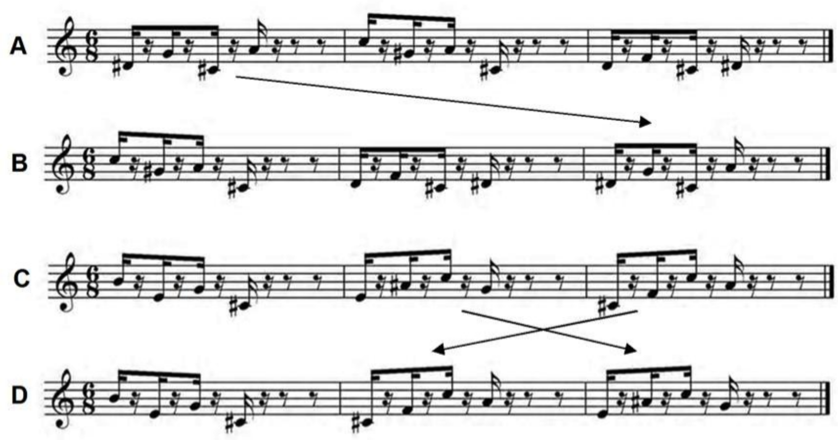

Figure 3 Schematic examples of sample Stimulus 1 (A) and Stimulus 2 (C) with their temporally-transformed counterparts (B and D, respectively). Note that the arrows indicate how the stimuli were temporally-transformed to produce the new stimuli and that the instrument is the same for all four of these stimuli.

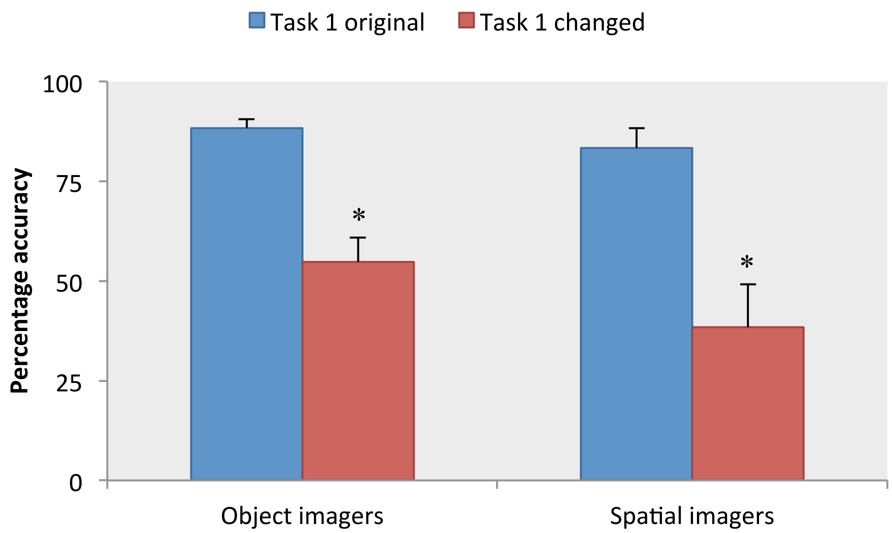

Figure 4 Although performance was impaired when the instrument orders were changed, there was no difference between object and spatial imagers. Error bars represent SEM.

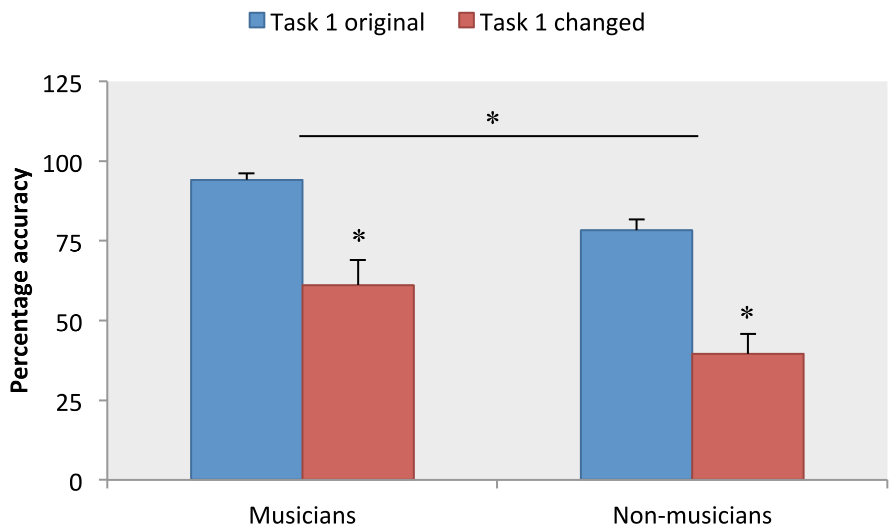

Figure 5 Musicians performed better than non-musicians, and both groups were less accurate when instrument orders changed. However, there was no interaction. Error bars represent SEM.

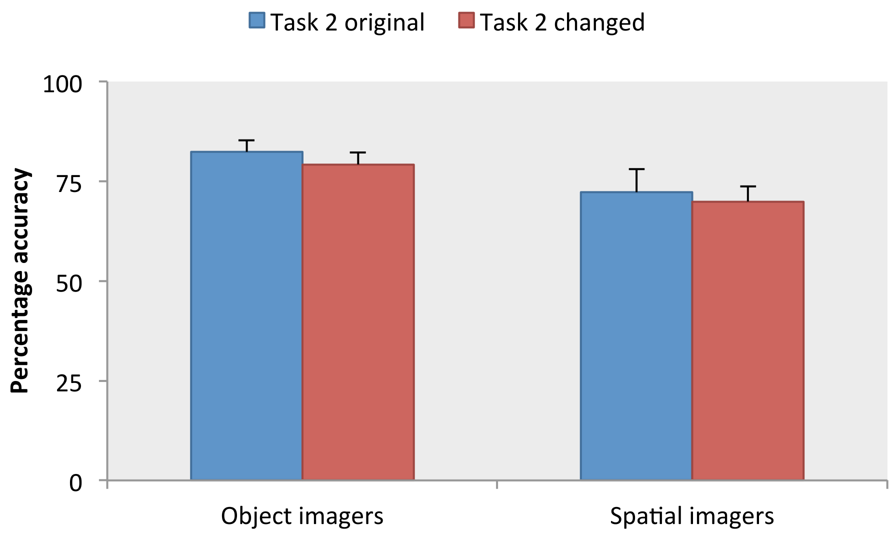

Figure 6 Object and spatial imagers, as classified by their OSIVQd scores, performed similarly across all conditions, and neither group was affected by the temporal transformations. There was also no interaction. Error bars represent SEM.

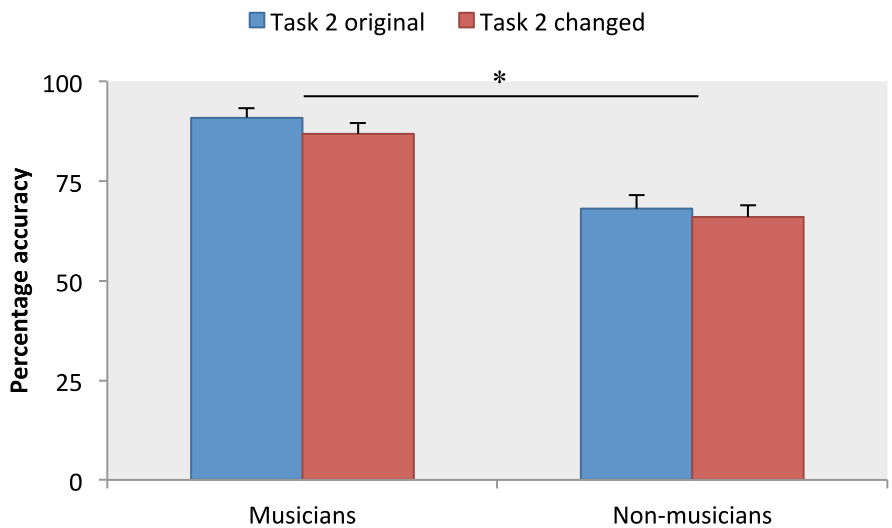

Figure 7 Musicians performed better than non-musicians, but neither group was affected by the temporal transformations. There was also no interaction. Error bars represent SEM.

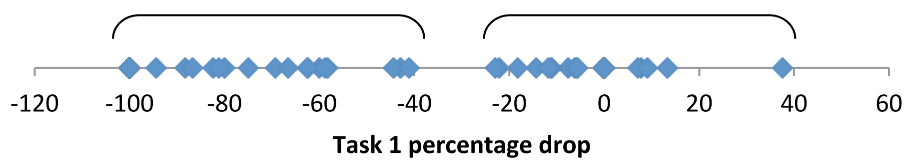

Figure 8 K-means clustering of *task 1 drop* delineated two coherent groups of equal size (n = 24), with averages of -77.26% and -2.58%. Each point represents an individual participant; brackets demarcate the clusters.

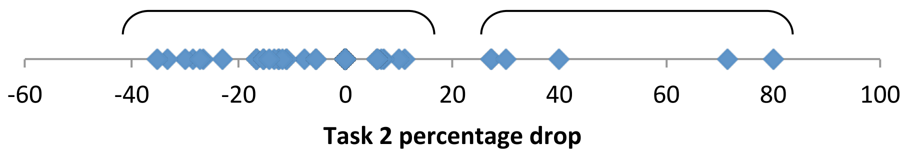

Figure 9 K-means clustering of task 2 drop revealed two groups that were very unequal in size (n = 42 & n = 6). Moreover, the smaller cluster with a mean task 2 drop of +46.0% included mainly participants who were near or below chance levels of accuracy on the original (untransformed) stimuli, suggesting that these positive changes may be caused by random chance. Each point represents an individual participant; brackets demarcate the clusters.

Appendix A: Task 1 instructions

You will be performing a task of memorizing auditory stimuli, each consisting of three four-note blocks. You will first see the cue “listen,” which means to get ready. You will then hear each of the auditory stimuli that you have to memorize as sequence 1 and sequence 2. The appropriate number will appear on the screen during this phase. You will then hear these sequences a second time. Next, you will see the cue “test,” which means to get ready to be tested. You will then hear four auditory stimuli (also of three four-note blocks), one at a time, and you will have to determine whether each is sequence 1 or 2. After each of these stimuli, you will see the cue “respond”; at this point, you will have five seconds to click to indicate your response. Please do not indicate your response until after the response cue appears. There will be a short pause after the trial, and this whole process will then repeat.

Appendix B: Task 2 instructions

You will be performing a task of memorizing auditory stimuli, each consisting of three four-note blocks. You will first see the cue "listen," which means to get ready. You will then hear the auditory stimuli that you have to memorize as set 1 and set 2. The appropriate number will appear on the screen during this phase. You will then hear these stimuli a second time. Next, you will see the cue "test," which means to get ready to be tested. You will then hear four auditory stimuli, one at a time, which will possess the same blocks as either set 1 or set 2. These blocks may or may not be presented in the same order. After each one, you will see the cue "respond"; at this point, you will have five seconds to click to indicate that it is either set 1 or 2. Please do not indicate your response until after the response cue appears. There will be a short pause after the trial, and this whole process will then repeat.