Spring 2018

Components of Auditory Imagery in Healthy Aging

Arley K. Schenker
aks019@bucknell.edu

I, Arley Schenker, do grant permission for my thesis to be copied.

COMPONENTS OF AUDITORY IMAGERY IN HEALTHY AGING

by

Arley K. Schenker

A Thesis

Presented to the Faculty of
Bucknell University
In Partial Fulfillment of the Requirements for the Degree of
Master of Science in Psychology

Approved:

Adviser: Andrea Halpern

Department Chairperson

4/2018
(Date: month and Year)

Acknowledgements

I would like to thank my advisor, Professor Andrea Halpern, for her extensive support and guidance in all aspects of this project. I would also like to thank my committee members, Professors Aaron Mitchel and Anna Baker, for their suggestions and different perspectives that allowed me to improve my research in a multitude of ways. Finally, I would like to thank Professors Eliza Congdon and Peter Judge for providing me with much needed guidance regarding the statistical design and analysis of my experiments.

For their help with older adult recruitment, I would like to thank the following individuals: Ruth Burnham and Annie Smith of the Bucknell Institute of Lifelong Learning (BILL), Andrea Newbury, Sue Jamison, Mick Smyer, and Kathy Schenker. This project could not have been completed without their help, and I very much appreciate their efforts in helping with my experiments.

Finally, I would like to thank fellow members of the Halpern Lab Sam Santomartino, Elizabeth Lampe, Ashlyn Kollar, and Brandon Mills for their help in recruitment, data collection, and analysis.

Table of Contents

List of Tables

List of Figures

Abstract

Mental imagery, a complex cognitive task that can be conceptualized into separable components, has been seldom studied in older adults. Auditory imagery is a particularly good modality to study throughout the lifespan, given that sounds can be both highly familiar and unfamiliar and that they inherently take place over a period of time. We tested for age differences in each of the four components of auditory imagery: generation, maintenance, inspection, and transformation. Furthermore, we investigated the degree to which certain cognitive measures that vary among individuals, such as musical background, self-reported auditory imagery, and working memory, predict performance differences. Across three tasks, we investigated whether or not there were age and individual differences in the four auditory imagery components. No age group differences were observed in generation and transformation, although musical background and self-reported auditory imagery were significant predictors of some results. Some age differences were found in a musical test of maintenance and inspection, and musical background and self-reported auditory imagery related to some aspects of the results. No age differences were found in a non-musical task of maintenance and inspection. Overall, this research contributes to the knowledge of age differences in cognitively complex tasks.

Introduction

Mental imagery, the ability to imagine events in the absence of perception, is remarkably useful in daily life. Individuals may use visual, auditory, or motor imagery in order to accomplish tasks, rehearse future scenarios, or reflect on past events. For example, visual imagery may be used to mentally reorganize a living room or conceptualize a new product design, auditory imagery can be used to remember the words of a popular song or the voice of a familiar individual, and motor imagery is useful in practicing the movements for physical rehabilitation or a sport. These types of mental imagery are necessary for making sense of perception, and it is a critical feature of daily life.

Although mental imagery is sometimes spontaneous, it is quite a complex process. Mental imagery relies on sensory processing (depending on which modality or modalities are involved), memory processes, including working and long-term memory, and executive functions, to name a few (McNorgan, 2012). Mental images can be quite vivid and automatic due to their use of sensory processes, which are themselves vivid. However, these images can also be quite effortful to produce, given that several processes must work simultaneously to create a realistic mental image and to process it in light of previous knowledge. This interplay of automaticity and effort makes mental imagery an intriguing study of cognitive processing in general, as these tasks can reveal how several cognitive processes interact in order to carry out a single task and how this interaction may change over time.

To account for the complexity of a single imagery task, mental imagery is generally studied with respect to its four component parts: *generation*, first producing a mental image, *maintenance,* keeping that image in working memory for several seconds, *inspection*, focusing on and answering questions about specific parts of the image, and *transformation*, manipulating the image in the mind (Dror & Kosslyn, 1994). Certain imagery tasks may focus on one of these components alone, but more often, mental imagery tasks are concerned with the overlap of several components. For example, image generation is necessary for virtually all mental imagery processes and is therefore straightforward to study in isolation; meanwhile, some imagery tasks make simultaneous use of image maintenance and inspection, such as mental scanning, wherein it is more difficult to isolate the mechanisms of maintenance and inspection individually (Kosslyn, Ball, & Reiser, 1978). Despite these challenges, these four components provide a useful framework for understanding the most important steps of mental imagery.

Aging and Mental Imagery

Given the complexity and usefulness of mental imagery, one might consider how performance of these component tasks might change in older adulthood. Aging is known to be associated with changes to several cross-domain cognitive processes, many of which are involved in mental imagery. For example, working memory, the act of maintaining and manipulating information in consciousness, has been robustly shown to decline in older adulthood (Old & Naveh-Benjamin, 2008; Salthouse, 1991; Salthouse, 1994). Speed of processing has been shown to decline in older adulthood as well, and one theory argues that this deficit may explain other deficits associated with aging, including that of working memory (Salthouse, 1996). Working memory and processing speed are important in mental imagery tasks. Images are maintained in consciousness using working memory, and these images must be processed quickly before they fade from working memory. Finally, older adults show deficits compared to younger adults in processing completely novel stimuli, as opposed to stimuli with which they are already familiar (Salthouse, 2012). In some imagery tasks, it is necessary to manipulate a mental image into a less familiar or completely novel representation, and this may be challenging for older adults. For these reasons, one might conclude that older adults perform mental imagery tasks more poorly than younger adults.

On the other hand, performance of other cross-domain cognitive processes appears to be maintained throughout the lifespan. For example, implicit memory retrieval, which includes procedural memory and unintentional recall, shows slight or no age differences between older and younger adults compared to explicit memory retrieval processes (Old & Naveh-Benjamin, 2008). Furthermore, semantic memory, including memory for knowledge and other general facts independent of a specific time or place, seems to be relatively spared from aging-related decline as compared to episodic memory (Old & Naveh-Benjamin, 2008). These types of memory are heavily recruited in mental imagery. Details such as the shape and color of an animal, the lyrics to a familiar song, or a body movement are learned through incidental encoding, and they eventually become so well learned that they are retrieved as semantic memories. Although working memory is needed to maintain these memories for any inspection or manipulation that may be required for an imagery task, the retrieval of these memories during generation should likely show no differences between younger and older adults. By studying mental imagery in its components, we can begin to discern how stability or loss in different cognitive processes throughout the lifespan relates to cognition in general.

Visual and Motor Imagery

Few studies have attempted to answer the basic question of whether there are differences between older and younger adults on performance of mental imagery tasks. Of these studies, the vast majority are concerned with visual imagery, and only some consider it in terms of the four components specified above. However, this last group of studies shows differing results about whether the abilities to perform these component tasks are preserved with age or not; in other words, some studies find no age differences in a component task, whereas others find differences in the same component (Briggs, Raz, & Marks, 1999; De Beni, Pazzaglia, & Gardini, 2007; Dror & Kosslyn, 1994; Kemps & Newson, 2005; Palermo, Piccardi, Nori, Giusberti, & Guariglia, 2016). One striking example of this involves image inspection. One study found no age differences between older and younger adults on an image inspection task (Dror & Kosslyn, 1994); however, in a second investigation using the same task, the same group of researchers found a decline in performance in older adulthood (Brown, Kosslyn, & Dror, 1998). None of the four components has been found to unilaterally show or not show age differences; it is thought that these inconsistencies may be related to variability in age ranges and specific component tasks between studies (Palermo et al., 2016).

Perhaps the most robust finding in this area is that error rate and reaction time in performing mental rotation, an image transformation task, increase with age (Adduri &

Marotta, 2009; Band & Kok, 2000; Berg, Hertzog, & Hunt, 1982; De Simone, Tomasino, Marusic, Eleopra, & Rumiati, 2013; Dror & Kosslyn, 1994; Hertzog & Rypma, 1991; Joanisse, Gagnon, Kreller, & Charbonneau, 2008; Kemps & Newson, 2005; Palermo, Iaria, & Guariglia, 2008; Saimpont, Malouin, Tousignant, & Jackson, 2013; Sharps & Gollin, 1987). This finding may be caused by age-related slowing in spatial information processing (Berg et al., 1982), qualitative differences in representation of visual mental rotations between older and younger adults (Dror, Schmitz-Williams, & Smith, 2005), or time pressure induced by experimenters, which is related to lower accuracy in older adults (Sharps & Gollin, 1987). Transformation tasks, in this way, seem to depend on those cross-domain cognitive processes that tend to decline with older age, which may be a reason that this finding has so much support.

There is also evidence of neural differences during visual imagery tasks between older and younger adults. One functional magnetic resonance imaging (fMRI) study showed that older adults showed lower selectivity in visual association areas during imagery of faces and moving stimuli (Kalkstein, Checksfield, Bollinger, & Gazzaley, 2011). Furthermore, an event-related potential (ERP) investigation regarding response monitoring in mental rotation found that older adults, compared to younger adults, displayed a larger number of slow errors in the mental rotation task and had lower amplitude ERP error potentials (Band & Kok, 2000). These results suggest that older adults detected only some of the errors, particularly when the task was more complex, and that this showed a rapid, as opposed to gradual, loss of efficiency in mental rotation.

On the whole, it is difficult to conclude from these studies how older adults compare to younger adults with respect to their performance on visual imagery tasks.

Some studies also show differences between older and younger adults in the performance of motor imagery tasks. Motor imagery expands our knowledge of mental imagery because, unlike visual images, imagined movements take place over a series of several seconds. One study showed that scores on two mental chronometry tests decrease with increasing age (Schott, 2012), suggesting that a discrepancy between perceived and imagined timing develops throughout the lifespan, particularly after the age of 70. Interestingly, these results were found to be mediated by scores on a working memory test. This result underscores the importance of working memory in mental imagery of all modalities.

Auditory Imagery

Although most mental imagery work has focused on visual imagery, auditory imagery is particularly important to consider with respect to aging because auditory information, by nature, is temporally dynamic. Auditory tasks such as detecting changes in tone durations and discriminating the temporal order of sounds are known to be challenging for older adults, likely because of age-related deficits in temporal structure processing (Rimmele, Sussman, & Poeppel, 2015). Auditory imagery, especially imagining songs that last several seconds, also requires extensive temporal processing (Halpern, 1988b). Furthermore, internal representations of auditory information have been shown to update in real-time, necessitating precise temporal perception and imagery (Halpern, 1988a). Working memory is also implicated in auditory imagery tasks, particularly in image generation (Halpern & Zatorre, 1999). As previously discussed, working memory tasks are often difficult for older adults. Such deficits suggest that auditory imagery performance in general would decline with aging.

However, this may not be true for all auditory stimuli. Auditory imagery of familiar melodies may be more stable across the lifespan than that of other auditory stimuli. Older adults can retrieve memories of familiar melodies, learned several decades previously, just as well as younger adults, although they are poorer in retrieving unfamiliar melodies (Halpern & Bartlett, 2010). Furthermore, implicit retrieval of familiar melodies is preserved in older adults, even while explicit retrieval is not (Gaudreau & Peretz, 1999). These are key skills in musical imagery tasks. Therefore, for older adults, auditory imagery of familiar melodies may be an exception to the general trend of difficult auditory tasks.

An understudied, but important, area of inquiry involves auditory imagery of environmental sounds, or non-musical, non-speech sounds that occur in the environment. Environmental sounds are encountered often in daily life and are thus closely linked to semantic memory (Lewis et al., 2004). Older adults have greater difficulty than younger adults in naming environmental sounds that are presented to them (Fabiani, Kazmerski, Cycowicz, & Friedman, 1996). Few studies have investigated auditory imagery of such sounds; however, one study showed that fMRI activation during auditory imagery of environmental sounds resembles that of auditory imagery of musical sounds, in that secondary auditory cortex, but not primary, is active (Bunzeck, Wuestenberg, Lutz, Heinze, & Jancke, 2005). Another study, focusing on animal sounds, compared ERPs when viewing a picture of an animal with those of viewing a picture of an animal while simultaneously imagining the sound typically made by that animal (Wu, Mai, Chan, Zheng, & Luo, 2006). Two components were more active in the imagery condition than the control condition; however, Hubbard (2010) notes that those authors did not show behavioral evidence that participants were generating auditory images, or that the ERP components reflected activity specific to auditory imagery. To determine whether auditory image generation actually occurs in these sorts of tasks, one could correlate accuracy on the tasks with a validated measure of self-reported auditory imagery; if this score is positively correlated with accuracy, then it is likely that participants are using auditory imagery to successfully complete the task.

Individual Differences Approach

Individual differences are an influential factor in cognitive performance, and mental imagery is no exception. This is particularly true given that mental imagery encompasses such a variety of cognitive skills. By measuring various individual differences that may contribute to auditory imagery performance regardless of age, we can begin to develop a full understanding of the factors that contribute to better or worse performance in auditory imagery tasks. The factors that we consider to be most relevant are musical background, self-reported auditory imagery vividness and control, and working memory.

Musical background has been shown to play a key role in both musical and non-musical auditory imagery. Musical background has been shown to be related to self-reported auditory imagery vividness (Pfordresher & Halpern, 2013). Additionally, a comparison of musicians' and non-musicians' performances on a musical imagery task, a non-musical imagery task, and a visual imagery control task revealed an advantage for the musicians in the auditory tasks, but not the visual task (Aleman, Nieuwenstein, Bocker, & de Haan, 2000). These results suggest that musical experience is related to better imagery of sounds overall. Musical background can also be relevant in auditory and cognitive performance in older adults. Older musicians, compared to older non-musicians, perform better in auditory frequency and duration discrimination tasks, as well as having higher visuospatial working memory (Grassi, Meneghetti, Toffalini, & Borella, 2017). Although the direction of causality in this relationship is unclear, i.e., it is unknown whether musical experience improves auditory imagery or vice versa (or both), these results suggest that the relationship between musical training and auditory imagery is an important one that persists throughout the lifespan and is thus important to consider in the current investigation.

Self-reported auditory imagery vividness and control are also important to consider. In visual imagery, self-reported imagery vividness and objective imagery performance have been found to be correlated; however, other studies have shown no correlation or even a negative correlation (Baddeley & Andrade, 2000). A prominent measure of these abilities for auditory imagery is the Bucknell Auditory Imagery Scale (BAIS). Interestingly, it has been shown that age does not correlate with BAIS vividness (Lima et al., 2015); however, a recent thesis showed that increasing age was associated with increasing vividness of imagery (van Hardeveld de Jong, 2016). The role of self-reported imagery vividness and control in objective imagery performance is unclear, particularly as adults age. However, by including this variable, we hoped to provide additional insight into the nature of this relationship.

Working memory, unsurprisingly, is also highly related to auditory imagery. Working memory is key in all components of auditory imagery, as the image must be held in working memory until it has been sufficiently processed for the task. Indeed, self-reported vividness of both visual and auditory images was found to correlate with working memory (Baddeley & Andrade, 2000). Although working memory spans have been shown to decline with age overall, aging does not cause working memory deficits; rather, working memory deficits in older age are mediated by processing efficiency and storage capacity (Salthouse, 1991). Furthermore, there are considerable age differences in working memory span, particularly depending on level of education (Gregoire & Van Der Linden, 2004). As such, working memory is an important variable to consider in this investigation.

Current Research Overview

In this series of studies, we compared the performance of younger and older adults in auditory imagery tasks across several imagery components. In addition to comparing performance in the imagery tasks, we also related performance within and across age groups in the various measures of individual differences. Both of these analyses allowed us to determine whether there are age differences in auditory imagery and, furthermore, whether there are individual differences in performance beyond any age-related effects. The first study investigated the components of generation and transformation in a pitch imagery task, and the second investigation used both musical and non-musical stimuli to investigate the components of maintenance and inspection. These investigations together thus examined each of the four imagery components, as well as both musical and non-musical stimuli. Through this research, we hoped to gain a comprehensive understanding of potential mechanisms and age-related effects in auditory imagery performance.

Experiment 1: Pitch Generation and Transformation

Introduction

Our first investigation sought to address these issues by focusing on the generation and transformation components of auditory imagery. We created a Generation task and a Transformation task for auditory imagery of the pitches of familiar melodies. When generating an image of a familiar tune, one must retrieve that tune from long-term memory before using it in working memory. Transformation, on the other hand, involves the steps involved in generation as well as a manipulation of the tune in working memory, so this task may be more taxing on the working memory system. It is also the case that, as well as being behaviorally separable, these two components are neurally separable. For instance, auditory image generation has been associated with activation in the supplementary motor area and several left frontal regions (Halpern & Zatorre, 1999), whereas posterior parietal cortex is active during tasks of auditory image transformation (Zatorre, Halpern, & Bouffard, 2009).

In each condition, participants were played the first phrase of a culturally familiar tune and were asked whether the last note of the phrase was correct or incorrect. In the Perception task, participants heard the entirety of the first phrase. In the Generation task, participants heard only the first few notes of the tune and the last note. This guided participants to generate an auditory image of the middle of the phrase. In the Transformation task, participants heard the first few notes of the same tune they had just encountered in the Generation trial, but transposed into a different key (i.e., the starting note was different than the one they had just heard).

In accordance with pilot data, we predicted that overall performance would be close to ceiling in the Perception task, fairly high in the Generation task, and lower, but significantly above chance, in the Transformation task. We predicted that older and younger adults would perform similarly on the Perception task. However, our predictions for the Generation task depended on whether the advantage of using familiar melodies or the disadvantage of relying on working memory would have a greater effect on the performance of the older adults. We predicted that an advantage of using familiar melodies would result in similar performance across age groups in the Generation task, as this should indicate that older and younger adults process familiar melodies in a similar way. On the other hand, the disadvantage of relying on working memory would result in worse performance in the older adults, as older adults as a group have lower working memory capacities than younger adults. Because the Transformation task relies so heavily on working memory, we predicted that in any case, this task would be more difficult for the older adults, particularly those with lower working memory spans.

Methods

Participants

In all studies, participants in the young adult group were recruited from enrollment in an introductory psychology course. They were able to choose from a list of psychology studies in which they could participate, and they received partial course credit for their participation. All participants were required to have normal or age-corrected hearing, verify by self-report that they did not have absolute pitch, and have been raised in the United States or Canada in order to ensure familiarity with the musical stimuli. Older adults were recruited via email and word of mouth within the Lewisburg area.

Twenty-eight undergraduates at Bucknell University constituted the young adult group in Experiment 1 (Table 1). Fifteen participants were female, and participants ranged in age from 18 to 21 years. Twenty-one older adults also participated in Experiment 1. Three additional older adults also participated; however, they were later excluded for being outside of the age range or having too many missing trials during the experiment. Eleven participants were female, and participants ranged in age from 60 to 93 years. They had significantly more years of education than did the younger adults.

Stimuli

Twelve familiar tunes were used for the three task conditions. These tunes were Christmas, folk, patriotic, or pop songs and had been previously rated to be highly familiar to adults raised in the United States and Canada (Halpern, 1984). Tunes had a tempo of 130 beats per minute on average (range: 100-150). Only the first phrase of each tune was used, and tunes were chosen such that their first phrase did not contain repeating words. This was done so that participants could end their imagery on a specific word.

Stimuli were played as a single melody line in a piano timbre with no lyrics. In the Generation and Transformation tasks, the tunes were accompanied by a metronome track. This was added to help participants maintain the correct tempo during imagery. As we were investigating pitch imagery rather than temporal imagery, we assumed this addition would not interfere with the difficulty of the pitch imagery. Tunes were created as MIDI files before being converted to WAV files. Tunes were presented in the same key in the Perception and Generation conditions, but were of course presented in a different key in the Transformation condition. Incorrect last notes in all three conditions had two versions. In one version, the correct last note was replaced with a note higher in pitch than the penultimate note, and in the other version, it was replaced with a note lower in pitch than the penultimate note. This was done to ensure that correct-incorrect determinations could not be performed based simply on whether the last note sounded too high or too low in every case. There was a 50% chance that the last note was correct on any given trial.

Background questionnaires

Musical background questionnaire. This questionnaire asked participants how many hours they spend daily listening to music, whether they had sung in an amateur or highly selective choir, whether they had had private instrument or voice lessons, whether they had played in an amateur band or orchestra, and whether they had semi-professional or professional musical experience. Participants responded either yes or no to each of these, and if yes, they were asked to provide details such as the number of years involved in each of these and what instrument they had played. Number of years of private music lessons was used as the final measure of musical background. Participants were not recruited on the basis of their musical experience; we sought participants at all levels of experience. Musical experience did not significantly differ between age groups (see Table 1 for all demographic information).

Bucknell Auditory Imagery Scale. This scale is a self-report measure of auditory imagery vividness and control (Halpern, 2015). Participants encountered items asking them to imagine environmental and musical sounds, and they were asked to rate the vividness, then control, of their imagery on a scale from one to seven. Older adults had significantly higher mean BAIS (average of the vividness and control subscales) scores (M = 5.49) than younger adults (M = 4.76).

Vocabulary test. Participants completed the second half of the vocabulary section of the Wechsler Adult Intelligence Scale – Revised (WAIS-R). Specifically, a researcher read the words to participants, who wrote down brief definitions on an answer sheet. These responses were then given 0, 1, or 2 points, depending on their accuracy. This task was used to measure general verbal intelligence. Additionally, it served as a means to compare the older and younger groups, as older adults generally have more comprehensive vocabularies than younger adults (Harada, Love, & Triebel, 2013). This represents the fact that the older adults who participated in this study are generally not cognitively deficient, as they can outperform younger adults in certain cognitive tasks. In the current study, older adults had significantly better vocabularies than younger adults, $t(47) = 6.32$, $p = .0001$.

Backwards digit span task. This task of working memory was conducted orally. A researcher read lists of numbers to participants, who were asked to repeat those lists in reverse order. Lists were of 3 to 10 digits in length, and two lists were given at each length. If participants reported the list incorrectly, they were given an additional chance to correctly repeat the list. Participants were not given explicit feedback on their accuracy. Digit span did not significantly differ between age groups.

Familiar tunes task

The main experimental task was programmed in Superlab 4.0 (Cedrus Corporation, San Pedro, CA) and was presented on an Apple Macbook laptop. Participants completed trials in one of two pseudorandomized orders, which ensured that a tune was not played twice in a row and that an equal number of correct and incorrect trials was presented.

Procedure

Participants were brought into the psychology lab and were asked to sign an informed consent form. They then completed the musical background questionnaire, BAIS, vocabulary test, and backwards digit span test, which were always given in that order. At the beginning of the Superlab procedure, participants heard each of the 15 tunes that would be used during the experiment (3 training, 12 experimental) and were asked whether they were familiar with each of the tunes. The majority of participants were familiar with all of the tunes, and no participants were unfamiliar with more than one tune. When participants were unfamiliar with a tune, they were instructed to guess on the trials involving that tune, which were then removed from analysis. The structure of the experiment is outlined in Figure 1 and is adapted from a previous study (Herholz, Coffey, Pantev, & Zatorre, 2016). There were two sections of the experiment: Perception and Imagery. These sections were constructed such that each trial of the Perception task consisted of one tune played once, whereas each trial of the Imagery task consisted of one tune played twice (once for a Generation trial, and once for a Transformation trial). Generation and Transformation versions of the same tune were paired in the Imagery condition in order to ensure that the transformation was indeed perceived as a transformation compared to the other tasks.

The Perception task was included to control for general pitch discrimination ability. At the beginning of this task, participants were instructed to report whether the last note of each tune was correct or incorrect, and they were discouraged from humming or singing the tunes. Following the presentation of the instructions, participants were given three training trials using tunes that were not present in the main experiment. During only these trials, participants were given feedback on their correctness. In each experimental trial, participants heard the first phrase of a familiar tune, whose title and lyrics were presented visually on the computer screen. The last note of the phrase was either the correct pitch or an incorrect pitch. Each tune was played twice, once with a correct last pitch and once with an incorrect last pitch, for a total of 24 trials.

At the beginning of the Imagery section, participants received further instruction describing the stimuli for the Imagery task. Specifically, participants were told that the middle of the tunes would be removed and that they would have to continue the tunes in their heads in order to perform the tasks. Participants were then presented with six practice trials with both Generation and Transformation pairs, receiving feedback on each trial. The Generation tune in the imagery section was always the first tune in the pair. Each tune used the same phrase as in the Perception task; however, in this task, participants heard only a 3-5 note cue of the beginning of the phrase. After this cue, participants were instructed to imagine the tune. The last note of the tune was presented at the expected time in the phrase, but its pitch was either correct or incorrect as varied in the Perception task. Participants made a correctness judgment at the end of each trial, as in the Perception task. Phrases were presented in the same key in both this task and the Perception task. Again, each tune was played twice, for a total of 24 Generation trials.

Following this, in the Transformation task, participants heard a version of the phrase consisting only of the cue and the last note; however, the phrase was transposed into a different key. Phrases were transposed up or down an interval of a perfect fourth (5 semitones). This distance resulted in a new key that was musically similar to the old key, to ease the transformation of the image. Tunes were counterbalanced to have equal numbers of higher and lower transpositions and correct and incorrect last notes, and there were 24 Transformation trials.

Analysis

Analysis of accuracy, measured in d', was completed with a 2 (age group: younger and older; between subjects) x 2 (imagery condition: Generation and Transformation; within subjects) mixed measures ANOVA. D prime, or d', is a signal detection theory measure that, by comparing the distributions of hits and false alarms, calculates one's ability to discriminate types of stimuli (Macmillan & Creelman, 2004). To analyze any effects of the individual difference measures, we performed two linear regression analyses for each imagery condition. The first analysis entered years of education and vocabulary score as potential predictors, and the second analysis added musical background, BAIS vividness, BAIS control, and backwards digit span. This allowed us to determine how much additional variance was accounted for by the variables of interest, musical background, self-reported auditory imagery, and working memory, while controlling for education and vocabulary. Because the BAIS subscales were highly correlated ($r = .742$, $p < 0.0001$), each of these analyses was performed once with the individual subscales as predictors, and a second time with their average as a predictor (mean BAIS). The model that accounted for the most variance was selected. All analyses were performed using SPSS 24.

Results

In these analyses, we sought to determine whether older and younger adults performed differently on either the Generation or Transformation tasks. Performance was higher in the Generation task, with a d' of 1.29, than the Transformation task, with a d' of .964 (Figure 2). D prime values around 0 represent performance at chance level, so the d' values observed here represent rather low, but reliably above chance, performance. Interestingly, older adults tended to have slightly higher performance in both the Generation and Transformation tasks; however, this effect did not appear to be significant.

These results were confirmed by a mixed measures ANOVA, which showed a main effect of imagery task, $F(1, 47) = 5.129$, $p = .028$. This main effect was such that participants in both age groups performed better in the Generation task than the Transformation task. Additionally, this served as a manipulation check, as we intended the Transformation task to require more mental work than the Generation task, and this was supported by the results. It is also worth noting that the Generation task was much more difficult than anticipated: we had predicted that performance on this task would be fairly high, which is associated with a d' of approximately 2, but the observed mean of

1.2 is well below that prediction. There was no main effect of age, $F(1, 47) = .559$, $p = .458$, and no interaction between age and imagery task, $F(1, 47) = 3.330$, $p = .074$. Interestingly, there was a trend of older adults having higher performance than the younger adults in the Transformation task, which is in the opposite direction than predicted; however, this effect was insignificant. Contrary to our hypotheses, these results suggest that older and younger adults equally found the imagery tasks difficult.

We included several individual differences measures in order to determine whether performance could be moderated by individual differences in relevant cognitive skills. In order to establish whether these measures were independent from one another, we analyzed bivariate correlations of all of the measures. Although for the most part, these measures were not related to one another, some were highly correlated, and this varied based on age group (Tables 2-4). Overall, number of years of education was highly correlated with vocabulary score, $r = .587$, $p < .0001$, which was entirely explained by the fact that older adults were both more educated and had higher vocabulary scores. Overall, number of years of education was also correlated with BAIS control, $r = .296$, $p = .039$; however, these factors were not correlated in either age group alone, which suggests that this correlation was driven by group differences. Additionally, BAIS vividness was negatively correlated with backwards digit span, $r = -.381$, $p = .007$.

After examining the relationships between the individual differences measures, we asked whether better scores in these measures were predictive of better task performance. A linear regression relating mean BAIS, musical background, and working memory to performance in the Generation task, controlling for education and vocabulary score, showed that the variables of interest accounted for 22.2% of the variance, which was significant, F(5, 48) = 3.112, *p* = .017. Of the variables of interest, only musical background was observed to be a significant predictor, *p* = .012, although mean BAIS was marginally significant, *p* = .071. Higher musical background scores were associated with higher performance, and this variable accounted for 14.1% of the overall variance. A similar regression for the Transformation task revealed that the variables of interest accounted for 23.9% of the variance, and this was significant, F(5, 48) = 4.451, *p* = .002. Both musical background and mean BAIS were significant predictors of the results, *p* = .005 and *p* = .045, respectively. Both variables were positively correlated with performance. Musical background accounted for 15.9% of the overall variance, and mean BAIS accounted for an additional 4.3% of the variance. Given the differences in BAIS scores between the two age groups, we ran another analysis that subsampled each group to equate BAIS scores, which found similar results.[1]

Discussion

In this experiment, we predicted performance on the Generation task would exceed that in the Transformation task, and indeed, that was the case. This supports our intent that the Transformation task be more cognitively demanding than the Generation

---

[1] Running the same statistical tests, we found no main effects of imagery task, F(1, 28) = 1.215, p = .28, or age group, F(1, 28) = .194, p = .663, and no interaction, F(1, 28) = 1.564, p = .078. The regression analyses in this subsample found that no individual differences variables predicted better performance in the Generation task; however, musical background continued to predict better performance in the Transformation task, F(1, 28) = 4.505, p = .043. This factor accounted for 13.9% of the total variance (adjusted: 10.8%). These results suggest that a factor above and beyond self-reported auditory imagery, musical background, contributed to the main effect of imagery task observed in the full sample.

task, as intended. However, performance on the Generation task was much lower than expected, suggesting that participants in both age groups found the task difficult.

Regarding our predictions for differences between age groups, we predicted that older and younger adults would perform similarly on the Generation task. The Generation task results were consistent with our predictions in that there were no age differences in this task. One explanation for this result is that older adults indeed show an advantage in processing familiar melodies that allows them to perform on the same level as younger adults. However, given that both participant groups performed worse than expected, an alternative explanation for this result is that this was a particularly difficult instance of auditory image generation.

We also predicted that older adults would find the Transformation task more difficult than would the younger adults, given that the Transformation task requires a higher working memory load. Contrary to our predictions, there were no age differences in the Transformation task, and even a slight trend in the opposite direction, wherein older adults performed better than younger adults in this task. This was certainly a curious result, as older adults tend to do more poorly than younger adults in working memory tasks. One potential explanation for this result may be that the sample of older participants was significantly more educated than the sample of younger participants, so the older participants may show a pattern of performance that is better than typical older adult samples. This explanation does not fully account for the results, however, because education was not a significant predictor in the regression ($p > .2$).

Another interesting result was that older and younger adults differed in their self-reported auditory imagery scores, as measured by the Bucknell Auditory Imagery Scale. Specifically, older adults reported substantially higher auditory imagery vividness and control than younger adults. This result is inconsistent with previous research, which has tended to find no age differences in this scale (Lima et al., 2015). As shown in the regression results, higher self-reported auditory imagery predicted higher performance in the Transformation task. This may be another reason why older adults performed at the same level as younger adults in that task.

Experiment 2A: Mental Scanning

Introduction

Our second investigation expanded upon the first in two ways. First, it tested how performance of auditory imagery maintenance and inspection change with age. Second, it examined these components in the context of both musical and non-musical stimuli. We were motivated to study non-musical stimuli in addition to familiar musical stimuli because older adults have been shown to have difficulty naming these stimuli, and study of these sounds, particularly in older adults, is somewhat lacking.

The musical task was a mental scanning task, replicating Halpern (1988a). This study, in turn, drew upon the ideas of Kosslyn et al. (1978), who conceptualized this task for the visual modality. Kosslyn and colleagues devised a task where participants were shown a map containing various objects, at varying distances from one another. Participants were then asked to scan through the map from one object to the other. In order to complete this task with success, participants must maintain knowledge of the locations of both objects while simultaneously inspecting their mental representations; in this way, this task tests the components of maintenance and inspection at the same time. It was found that scanning time increased linearly with the distance between the objects on the map, which indicated that visual mental representations possess similar characteristics to visual percepts; namely, that they can be processed and that they encode metric distances. Similarly, Halpern (1988a) performed a similar task involving lyrics in culturally familiar songs, showing that time to scan between the pitches associated with two lyrics in the song increased with the distance (in musical beats) between the lyrics.

The current study replicated the original 1988 study in both younger and older adults, whereas the original study sampled only younger adults.

Overall, we predicted that older adults would make more errors than younger adults. This is due to the fact that processing speed and working memory decrease significantly in older adulthood (Salthouse, 1996), and these mental representations must be processed quite quickly in order to be processed fully in working memory. Furthermore, we predicted that the rate of reaction time as lyric distance increases would differ between age groups. In particular, the rate would be greater (i.e., steeper) in older adults compared to younger adults. Again, this would likely be due to age-related deficits in the key processes of working memory and processing speed. With regard to individual differences measures, we predicted that higher self-reported auditory imagery vividness and control would positively correlate with mental scanning accuracy and negatively correlate with reaction time, for both older and younger adults. We predicted that this relationship would also hold for musical background.

Methods

Participants

Nineteen undergraduates at Bucknell University comprised the younger adult group in Experiment 2A. Sixteen participants were female, and participants ranged in age from 18 to 20 years (Table 2). The majority of participants completed both parts of Experiment 2 in the same session. Two younger participants completed only the serial order memory task because they were not familiar with enough songs in the mental scanning task. Eleven additional participants performed the mental scanning task but were excluded from the final analyses due to insufficient pitch discrimination ability (see Procedure). An additional nine participants performed the task but were excluded from the final analyses because their pattern of responses resulted in missing values in the reaction time analysis (see Analysis).

Fifteen older adults comprised the older adult group in Experiment 2A. Nine participants were female, and participants ranged in age from 60 to 83 years (Table 2). All participants completed both parts of Experiment 2. Six of these participants also participated in Experiment 1. Individual differences measures for these participants were taken from Experiment 1. Audiometry was used to screen hearing for all older adults. Frequencies of 250, 500, 1000, 2000, 4000, and 8000 Hz were tested in each ear on a GSI-17 audiometer using standard procedure. For some participants, a threshold at 8000 Hz could not be obtained; however, this should not affect their perception of the stimuli, as no stimuli contained fundamental frequencies of 8000 Hz. All audiograms from 250 to 4000 Hz resembled the median values of expected hearing threshold deviations reported in the International Organization for Standardization (ISO) statistical distribution of hearing thresholds (2017). Six additional participants performed the mental scanning task but were excluded from final analyses due to insufficient pitch discrimination ability (see Procedure). An additional two participants performed the task but were excluded from the final analyses because their pattern of responses resulted in missing values in the reaction time analysis (see Analysis).

Task materials

The mental scanning task was a replication of a previous mental scanning task (Halpern, 1988a). Nine familiar songs were used as stimuli for the mental scanning task. They had the characteristic that each lyric coincided with a beat, rather than being syncopated. Trials consisted of one of three *startpoints*, indicating the first, third, and fifth lyric of the song, and one of four *stepsizes*, indicating the temporal distance between the startpoint and the second lyric. Difficulty was expected to increase with increasing temporal distance. All startpoint/stepsize combinations were used except for those in which the two lyrics had the same pitch. The two lyrics always had different pitches, and participants were asked whether the pitch of the second lyric was higher or lower than the first. Trials in which one of the lyrics would semantically interfere with the task (e.g., "up" from "Somewhere Over the Rainbow" or "down" from "The Itsy Bitsy Spider") were removed. There were 5 trials used for training at the beginning of the task and 58 trials used after the training. One of the nine tunes was used exclusively in the training section. Of the 58 experimental trials, 6 were included as filler trials to ensure that each song was presented an approximately equal number of times and that the startpoints and stepsizes were approximately equally represented.

Background questionnaires

In both parts of Experiment 2, participants were assessed on the following skills: musical background, self-reported auditory imagery, vocabulary, and working memory. Musical background was assessed using the Musical Training subscale of the Goldsmiths Musical Sophistication Index (Gold-MSI) (Mullensiefen, Gingras, Musil, & Stewart, 2014). This replaced the musical background questionnaire from Experiment 1 due to its increased specificity and ease of scoring. Responses were scored according to the Subscales Scoring Template provided on the Goldsmiths website. Self-reported auditory imagery, vocabulary, and working memory were assessed as in Experiment 1. The six older participants who had also participated in Experiment 1 completed the Gold-MSI during the Experiment 2 session, and their Experiment 1 scores on the other background questionnaires were used in Experiment 2 analyses.

Procedure

Before participants completed either of the two imagery tasks, they first completed an informed consent form. Older participants then completed an audiogram. Following this, all participants completed the Gold-MSI Musical Training Subscale, BAIS, WAIS-R vocabulary test, and backwards digit span task. These measures were always administered in the listed order. Following this, participants used Superlab 4.5 to complete the mental scanning and the serial order memory tasks, the order of which was counterbalanced.

In order to ensure familiarity with the songs in the experiment, participants were provided with a lyrics sheet after the completion of the preliminary tasks. This lyrics sheet contained the lyrics of the first phrases of the nine songs that would be included in the experiment. Lyrics that would be queried in the experiment were denoted in all caps. Participants were told to familiarize themselves with the lyrics on the sheet and how they were denoted (for example, if multiple pitches occurred during one word, only the first part of the word was written in all caps). At this point, participants were asked to tell an experimenter if they were unfamiliar with any of the songs. Of the nineteen participants in the younger adult group, nine were unfamiliar with one song, two were unfamiliar with two songs, and one was unfamiliar with three songs. The rest of the younger adults, and all of the older adults, were familiar with all of the songs. If they were unfamiliar with four or more songs, they did not complete the mental scanning task. Participants could not refer back to the lyrics sheet during the task.

The mental scanning task was programmed in Superlab 4.5 (Cedrus Corporation, San Pedro, CA) and was presented on an Apple Macbook laptop. Clear Post-it notes were placed on top of the relevant keys such that the letter on the key was still visible, and they were labeled with their corresponding responses (i.e., “high” or “low”). Participants first completed a pitch discrimination screening task, in which they heard between 9 and 13 pairs of sine wave tones and were asked whether the second tone was higher or lower in pitch than the first tone. Responses were made using the M and N keys on the laptop keyboard. The pitch discrimination task allowed us to control for overall pitch discrimination ability independent of pitch discrimination during a mental imagery task. As a result, participants were excluded from final analyses if they did not perform the pitch discrimination task above chance. Given that chance was 50%, participants were excluded if they did not achieve at least 70% performance.

Following the screening task, participants were told that they would see two words on each trial and that their task was to imagine the pitches of the two words in their heads and report whether the second pitch was higher or lower than the first. They then completed five training trials wherein the distance between the two lyrics increased as training continued. The training trials were always given in the same order, and participants received feedback on their performance.

In each trial of the experimental task, participants saw the title of the song they would be asked to imagine, which was presented for 1000 ms (Figure 3). They were then shown two lyrics of the cued song in all caps in the center of the screen, as well as instructions on the bottom of the screen reminding participants of which keys to press to make their responses. The lyrics were shown until the participants made a keyboard response, for which they were given unlimited time. Participants were given no feedback after the training trials. After the response was made, there was a 1000 ms delay, after which the next trial began. Trials with different startpoints and stepsizes were intermixed, and trials were completed in a different random order for each participant. This task took approximately 10 minutes to complete.

Analysis

Participants were presented with 52 experimental trials, comprising all combinations of the three startpoints (1, 3, 5) and four stepsizes (2, 4, 6, 8). Although trials with a stepsize of four were presented to participants, they were removed from reaction time and error rate analyses due to an insufficient number of trials at that level due to these being naturalistic materials. Only one trial was presented with startpoint 3 and stepsize 4, whereas 3-6 trials were present in all other startpoint and stepsize combinations. As a result, the three stepsizes that were included in the final analyses were 2, 6, and 8, and 42 trials were included in the final analysis.

The main results were thus computed with two 3 (startpoint: 1, 3, 5; within subjects) x 3 (stepsize: 2, 6, 8; within subjects) x 2 (age: younger, older; between subjects) mixed measures ANOVAs, one for median reaction time and one for mean error rate. As is standard for reaction time analyses, only correct trials were included. Due to significant skew to the right in the distributions of untransformed correct reaction times for both age groups (see Figures 4.1 and 4.2), we report a log transformation of reaction time (Schott, 2012). There were 11 participants, referred to above, who were missing trials in at least one combination of startpoint and stepsize due to their pattern of responses. These participants were excluded from the final analyses for this reason. After running the ANOVA, we then ran an ANCOVA, entering musical background, BAIS vividness, BAIS control, and working memory as covariates. Because the BAIS subscales were highly correlated (r = .806, $p < 0.0001$), this analysis was performed once with the individual subscales as predictors, and a second time with their average as a predictor (mean BAIS). We report the analysis using mean BAIS given its increased power. Although it would have been preferable to analyze the effects of these variables with linear regression, the mixed design of the experiment made it unfit for analysis with regression. We thus decided to use an ANCOVA to answer this question.

A secondary dependent variable in the reaction time analysis was slope, which allowed us to compare differences between age groups in the rate of reaction time change across startpoint and stepsize. Because the three stepsizes that were included were not at linear intervals, a singular slope of reaction times as stepsize increased could not be calculated. Therefore, we separately analyzed the slopes between stepsizes two and six and between stepsizes six and eight, as well as analyzing the average of these slopes. We analyzed any effects of age group on slope using a 3 (startpoint: 1, 3, 5; within subjects) x 2 (slope position: 2 to 6, 6 to 8; within subjects) x 2 (age: younger, older; between subjects) mixed measures ANOVA. We then related individual differences measures with the reaction time slopes using a mixed measures ANCOVA, with BAIS vividness, BAIS control, working memory, and musical background included as covariates.

Results

First, we analyzed the overall difficulty of the task. In all conditions, error rates were rather high, ranging from 15.8% to 31.9% across startpoint and stepsize conditions (Table 6). They tended to increase as startpoint and stepsize increased. Indeed, a mixed measures ANOVA on error rate revealed a main effect of startpoint, $F(2, 64) = 5.422$, $p = .007$, wherein error rate significantly increased between startpoints 3 and 5 ($p = .008$, with Bonferroni correction). There was also a marginally significant effect of stepsize, $F(2, 64) = 2.536$, $p = .087$, wherein error rate increased as stepsize increased. These main effects show that the startpoint and stepsize manipulations were working as intended, as the task should have been more difficult when the startpoint was not at the beginning or as the distance between the lyrics increased. There was no main effect of age group, $p = .288$. Finally, there was a marginally significant three-way interaction between startpoint, stepsize, and age group, $F(4, 128) = 2.000$, $p = .098$. A mixed measures ANCOVA with BAIS vividness, BAIS control, musical background, and working memory included as covariates did not reveal any significant main effects or interactions (all *p*'s greater than .12). This suggests that individual differences in self-reported auditory imagery, musical background, and working memory accounted for the main effects of startpoint and stepsize observed in the ANOVA.

The main measurement of interest in the mental scanning task was log reaction time, as reaction time reflects the effort of participants to process the stimuli in real time. We expected reaction times to increase as startpoint and stepsize increased, for similar reasoning as in the error rates analysis. A mixed measures ANOVA on median reaction time indeed revealed a significant main effect of startpoint, $F(1.230, 39.356) = 5.605$, $p = .006$, wherein reaction time significantly increased between startpoints 1 and 5 ($p = .010$). Additionally, there was a main effect of stepsize, $F(1.698, 54.327) = 22.379$, $p < 0.0001$, where reaction time increased between all combinations of stepsizes (all *p*'s less than .002). Finally, there was a main effect of age group, $F(1, 32) = 7.855$, $p = .009$, wherein older adults were slower overall than younger adults.

We then examined interactions between age group and startpoint or stepsize. These interactions would indicate that the age groups were differentially affected by the difficulty manipulations, as predicted. However, neither the startpoint by age group ($p = .398$) nor the stepsize by age group ($p = .169$) interactions were significant. There was a significant interaction between startpoint and stepsize, $F(4, 128) = 3.638$, $p = .008$, wherein as startpoint increased, reaction time increased more steeply for stepsizes 6 and 8 than for stepsize 2 (Figure 5). The three way interaction between age group, startpoint, and stepsize was not significant ($p = .529$).

To investigate the role of the individual differences measures in reaction time performance, we ran a mixed measures ANCOVA with BAIS vividness, BAIS control, musical background, and working memory included as covariates. In this analysis, the previously significant main effects of startpoint and stepsize, and their interaction, were no longer significant (all *p*'s greater than .3). This suggests that the individual differences variables accounted for all effects of the experimental variables; in other words, when these variables were controlled for, we did not observe any effects of the experimental manipulations alone. The age group by startpoint and age group by stepsize interactions remained insignificant, $p = .530$ and $p = .176$ respectively. The three way interaction between age group, startpoint, and stepsize was also not significant, $p = .867$. These data therefore suggest, interestingly, that controlling for these individual differences variables did not reveal any relationships between age and task performance.

Our final reaction time analysis dealt with slopes, or rates of change between startpoints and stepsizes. First, all slopes were positive, indicating that the task took longer to complete in higher stepsizes, as intended. A mixed measures ANOVA revealed a main effect of startpoint, $F(2, 64) = 5.758$, $p = .005$. This main effect indicates that slopes (across stepsizes) associated with startpoint 3 were higher than those for the other two startpoints. There was no main effect of slope, $p = .350$, showing that slopes between stepsizes 2 and 6 did not differ from those between 6 and 8, across startpoints. There was also no main effect of age group, $p = .315$. This result suggests that even though there was a main effect of age group in the first-order reaction time analysis, the rates of change in reaction time did not change did not differ across age groups. Additionally, there was a significant startpoint by slope interaction, $F(2, 64) = 4.364$, $p = .017$, but no three way interaction with age group, $p = .416$. The interaction of startpoint and slope is driven by a much higher slope from stepsize 6 to stepsize 8 in startpoint 3 than any of the other slopes.

As in Experiment 1, we analyzed bivariate correlations of the individual differences measures, which varied based on age group (Tables 7-9). As in Experiment 1, education and vocabulary score were highly correlated due to group differences in both of these variables. Furthermore, in younger adults, Gold-MSI was positively correlated with backwards digit span, whereas in older adults, BAIS control was positively correlated with backwards digit span. Overall, the majority of these measures were not correlated with one another, which suggests that they tested different aspects of cognition.

Discussion

In this task, participants were presented with two lyrics of well-known songs and were asked to compare the pitches of those lyrics. They presumably accomplished this by generating an auditory image of the song using auditory imagery, inspecting the auditory image to find the two target pitches, and maintaining the first pitch while imagining the second pitch. We predicted that older adults would make more errors and have a slower rate of response as startpoint, the starting point of the first lyric, and stepsize, the distance between the lyrics, increased. We found no differences in error rate across age groups, although both groups showed higher errors as startpoint and stepsize increased. This shows that our startpoint and stepsize manipulations did indeed require more cognitive effort when they were intended to do so.

Regarding reaction time, older adults were slower overall than younger adults. This result is not surprising, as many studies show slowing down in older adults. Furthermore, reaction time significantly increased with both startpoint and stepsize for both age groups. This represents that our difficulty manipulations worked as intended and that participants were performing the task in an expected way. Beyond this, we expected to find one or more interactions between age group and the difficulty manipulations. Whether or not covariates were included, the three way interaction of age group, startpoint, and stepsize was not significant, nor were interactions between age group and startpoint or age group and stepsize. One reason for this may be that the age group by startpoint interaction was underpowered. A post-hoc power analysis using G*Power 3.1 revealed that the study designed achieved 63% power (Faul, Erdfelder, Lang, & Buchner, 2007). The age group by stepsize interaction, however, was sufficiently powered. Moreover, both analyses were likely underpowered due to a low number of trials in each cell and the fact that this number of trials varied significantly across participants.

Future studies should seek to recruit a greater number of older adults, which will help to clarify the role of age differences in auditory imagery tasks. More pertinent to this task, however, future implementations of this task should be sure to choose songs that cover all the conditions being tested in the study. This will ideally result in a sufficient number of trials for analysis, regardless of whether participants are incorrect on some of those trials.

Experiment 2B: Serial Order Memory Task

Introduction

The non-musical task was a serial order memory task for transformed environmental sounds. We sought to investigate the functioning of the auditory system independent of semantic memory, as participants may use their semantic knowledge to answer questions rather than generating images (Hubbard, 2010). Older adults tend to complete new, unfamiliar tasks by relating them to more familiar ones (Park & Reuter-Lorenz, 2009), and in so doing, they use a different strategy than the one being tested. Because environmental sounds are so tied to semantic memory, we reversed the audio files for our experimental stimuli and ensured through pilot testing that they were not reliably nameable. This allowed the sounds to be similarly acoustically complex; however, the reversed versions of the sounds did not evoke any semantic representation. Serial order memory tasks investigate participants' ability to remember the order of a list of items, usually numbers or letters (Lewandowsky & Murdock, 1989). This generally involves playing a list of items, then giving a retention interval in which participants are encouraged to mentally rehearse the list, then playing a second list with the same items, whose order is either the same or different from the first list. In order to successfully complete the task, participants must be able to maintain the first list while also inspecting where any differences may occur between the first and second lists. Similarly to the mental scanning task, this task thus analyses the imagery components of maintenance and inspection. To our knowledge, such a task has not been carried out with non-speech sounds. Critically, we investigated whether performance on this task was correlated with self-reported auditory imagery. If so, then it is likely that auditory imagery was necessary to successfully complete the task. If not, interpretation is more difficult; it may be that participants created names for sounds and remembered the names rather than the sounds themselves. To that end, we also asked participants to report their strategy on the task, which would allow us to see whether participants were consciously using imagery or were explicitly choosing a more semantic strategy.

We predicted that in both age groups, performance on this task would decrease as the length of the list increased (from four to six items). We also predicted that this difference would be larger in the older group, given age-related deficits in working memory and novel stimuli processing. Concerning individual differences, we thought it highly likely that working memory would correlate with performance, as working memory is critical for successful performance of serial order memory tasks. However, this correlation would not show that participants used auditory imagery to complete the task. To show this, performance would have to correlate with self-reported auditory imagery, and we predicted that this would indeed occur.

Methods

Participants

Thirty-one younger adults participated in Experiment 2B, recruited as discussed above. Twenty-four participants were female, and participants ranged in age from 18 to 20 years. Twenty-three older adults also participated in Experiment 2B, and participants ranged in age from 60 to 83 years. Other demographic information for both groups can be found in Table 10.

Materials

Stimuli in the serial order memory task were adapted from the Environmental Sounds Library collected by Marcell, Borella, Greene, Kerr, and Rogers (2000). This library consists of 120 common environmental sounds and was previously available in a public repository. All sounds were reversed using Audacity, and those sounds whose reversed version sounded significantly different from the forward version (and were thus non-nameable) were kept. Additional sounds were taken from a free sounds website and from a podcast episode about sounds used in technology.

The sounds were evaluated in multiple stages of pilot testing to ensure that no sounds were named consistently or correctly. In the first stage, five young adult pilot participants were played 38 reversed sounds and six forward sounds that were not reversed elsewhere in the experiment. Each sound was 2 s in length and was normalized using Audacity. After each sound, participants were asked whether they recognized the sound, and, if so, whether they could name the sound and to give confidence ratings for that name. Forward trials were included to allow pilot participants to accurately calibrate their confidence ratings. Sounds that had been given the same name by three of the five participants, even if the name was incorrect, were excluded. Six sounds were excluded using these criteria. In the second stage, two different young adult pilot participants were played the 31 reversed sounds that had not been excluded from the first stage and five forward sounds. Sounds from the first stage were shortened to 1 s in length. Again, pilot participants were asked to name the sounds if possible and to provide confidence ratings. No sounds were excluded in this stage, so 31 sounds were used in the final experiment.

Sounds were then combined into lists consisting of four or six sounds, representing easy and hard difficulty conditions, respectively. A Python script was used to generate 40 text lists of sounds for each difficulty condition that would be combined into audio files for each trial (two lists per trial), as well as eight lists to be used in training trials at the beginning of the experiment. The sounds in the second list in each trial were always the same sounds as in the first list but were shuffled 50% of the time; in other words, in half of the trials, the sounds in the second list were in a different order from the first list (Figures 6.1 and 6.2). To avoid reliance on primacy and recency short-term memory effects, the second list in each trial always had the same first and last sounds as the first list, even when the middle sounds were in a different order. Consequently, to ensure that participants did not learn to ignore the first and last sounds when making their decisions, five filler trials were generated for each difficulty condition wherein all sounds could be shuffled into a different order in the second list, and the first sounds of the first and second lists were always different. These filler trials were not included in the final analysis. This resulted in 50 total trials (25 trials per difficulty condition), where 40 were included in the final analysis.

Each list was then created as a separate WAV file in Audacity. Individual sounds were separated by 500 ms of silence, with no silence at the beginning or end of the files. In the third stage of piloting, two different young adult pilot participants were played each of these files and were asked to verbally report whether any sounds in each file 1) were indistinguishable from any other sounds in the file or 2) sounded like a different, nameable sound when played adjacent to another sound. No trials were excluded in this stage.

Procedure

The serial order memory task was programmed in Superlab 4.5 (Cedrus Corporation, San Pedro, CA) and was presented on an Apple Macbook laptop. Clear Post-it notes were placed on top of the relevant keys such that the letter on the key was still visible, and they were labeled with their corresponding responses (i.e., “same” or “different”). Participants were instructed that they would hear a list of sounds, followed by a period of silence. They were told to imagine the sounds they had just heard during this period of silence. Participants were not given specific instructions for the best way to imagine the sounds; this was purposefully left open-ended. Then, they were told that hear the same sounds again, but they may be in a different order. They were then asked to make one of two keyboard responses depending on whether the second list was in the same order or a different order compared to the first list. Responses were made using the Z and X keys on the laptop keyboard. Response keys were designated to be adjacent but arbitrary, and participants were told to respond with their dominant hand to avoid any timing effects related to the use of both hands to respond. Participants were instructed not to sing, hum, or tap, and they were told to respond as quickly as possible while still being accurate. They then completed four training trials in which they received feedback about their accuracy. The training trials occurred in a different random order for each participant. The sounds in these trials were the six least recognized sounds that were excluded during piloting.

Following these trials, participants heard the 50 experimental and filler trials. At the beginning of each trial, participants saw a screen displaying the words “Playing sounds…” for 1000ms. This signaled to the participant that a new trial was about to begin. Then, they heard the first list while the words “List 1” appeared on the screen. Following this, participants were shown a screen instructing them to imagine the sounds they had just heard, and this screen was shown for 10s. Then, participants saw a screen displaying the words “Playing sounds… List 2” while hearing the second list. Participants were then shown a reminder of which buttons were associated with each response, and this screen was shown until participants made a keyboard response. Trials were blocked by condition, with the easy condition always first. Trials were randomized within each condition in a different order for each participant. This task took approximately 30 minutes to complete. Afterwards, participants were asked to write down their strategy or strategies for remembering the sounds in the task they had just completed. Because the instruction to imagine the sounds during the retention interval was open-ended, we anticipated that participants used one or multiple different strategies, so we found it useful to measure them. These strategy self-reports were qualitatively coded by two independent coders using a procedure adapted from Charmaz (2008). These coders then worked together to create a codebook, which they worked together to apply to all of the self-reports. A third independent coder then applied the codebook to the self-reports. Reliability between the ratings of the third coder and the first two coders was 82%. The third coder and one of the first two coders then worked together to refine the codebook and resolve any discrepancies, which resulted in a finalized list of codes. This list was used in the analyses to determine whether participants who reported using a certain strategy had higher performance in the serial order memory task.

Analysis

The main results were computed with a 2 (age: younger, older; between subjects) x 2 (difficulty condition: easy, hard; within subjects) mixed measures ANOVA, with difficulty condition as a within-subjects variable. To assess any effects of the individual differences variables on performance, BAIS vividness, BAIS control, working memory, and musical background were included as potential predictors in linear regression analyses. Because the BAIS subscales were highly correlated with one another ($r = .791$, $p < 0.0001$), one analysis was performed with the individual subscales as predictors, and a second was performed with their average as a predictor (mean BAIS). The analysis accounting for the highest proportion of variance was chosen.

Results

Overall, d' values for both age groups and difficulty conditions were lower than expected (Figure 7). In the easy condition, the average d' was 1.3, whereas in the hard condition, the average d' was 1.1. These values suggest that participants found the task difficult, but still performed reliably above chance. There was also a trend of higher performance, in both conditions, in the younger adults compared to the older adults. However, a mixed measures ANOVA on the effect of difficulty condition showed no effect of condition, $F(1, 52) = 1.335$, $p = .253$, no effect of age group, $F(1, 52) = 2.171$, $p = .147$, and no interaction between condition and age group, $F(1, 52) = .849$, $p = .361$ (Figure 7). In a linear regression analysis on easy d' including musical background, BAIS vividness, BAIS control, and backwards digit span as covariates, controlling for education and vocabulary, the variables of interest accounted for 16.1% of the variance, and this was significant, $F(6, 53) = 2.350$, $p = .046$. However, no individual predictors were significant (all *p*'s greater than .11). In a similar analysis on hard d', the variables of interest accounted for 11.1% of the variance; however, this model was not significant, $p = .442$. This suggests that the variables of interest were more predictive of performance in the easy condition than in the hard condition, i.e., that other factors beyond the variables measured here contributed to variance in performance in the hard condition.

Again, we analyzed bivariate correlations between the individual differences variables, and for the third time, they differed between age groups (Tables 11-13). Across both age groups, education was positively correlated with vocabulary score, but within groups, this was only true for the older adults. In younger adults, there were marginally significant positive relationships between musical background and backwards digit span and between education and vocabulary score. In older adults, there were many more significant relationships. Most interestingly, musical background was positively related to both BAIS vividness and BAIS control scores, which was not observed in the younger adults. Overall, most of the variables were not correlated with each other, again suggesting that they were independent, for the most part.

The six strategies identified through the coding process, in decreasing order of use, were association, visualization, finding pattern, identifying pitch, pick and choose, and repetition (Table 14). Association refers to word association; that is, participants reported associating sounds with words, and remembering the placement of the words in the list. Visualization indicates that the participants created a visual image of one or more sounds and used their image to remember their order in the list. Finding pattern indicates that participants isolated a subset of sounds from the first list, for example, the third and fourth sounds in the list, to compare to the second list. Identifying pitch refers to an auditory imagery strategy of rehearsing the actual sounds in their mind's ear, or listening for particular acoustic characteristics (i.e., placement of high or low pitched sounds in the list. Pick and choose is similar to a combination of finding pattern and identifying pitch; participants focused on the placement of particularly salient sounds between the first and second without making explicit mention to any acoustic characteristics of the sounds. Finally, repetition indicates that participants tried to rehearse the entire sequence of sounds during the retention interval without using a particular subset or pitch strategy.

Frequencies of strategy use differed between age groups (Table 14). Comparisons of note are that older adults reported using a finding pattern strategy much more than younger adults (47.8% of older adults compared to 12.9% of younger adults). To a lesser extent, older adults also reported using an identifying pitch strategy more often than younger adults (13.0% of older adults compared to 6.5% of younger adults). Repetition, the only strategy that involved the entire list, was exclusively used by younger adults, although it was not used by very many young adults.

Analysis of performance differences based on age group and strategy usage is skewed by the fact that there are unequal numbers of participants who used each strategy. Nevertheless, this analysis reveals interesting differences in strategy effectiveness between difficulty conditions and age groups. Participants who used an association strategy had higher performance than those who did not (Table 15). This was also the case for participants who used a visualization strategy. Finding pattern, the strategy most used by older adults, was associated with lower performance in those who reported using it. Additionally, the strategy most related to auditory imagery, identifying pitch, was associated with lower performance in those who reported using it; however, again, it is important to note that few participants reported using this strategy, so these results may be spurious.

Discussion

In this experiment, we tested whether age or other individual differences measures would play a role in serial order memory for non-musical auditory stimuli, specifically, transformed environmental sounds. Although older adults have been found to do more poorly in tasks involving novel stimuli, particularly in a task that requires considerable working memory, the current study did not find any relationship between age and performance. The only variable we tested that played a role in the results was working memory, such that higher backwards digit span scores were associated with higher performance. Given that the task almost exclusively tested working memory, this result was expected. We predicted that self-reported auditory imagery would also be associated with better performance; however, this prediction was not supported by the results.

Because there was a trend towards lower performance in the hard condition and in the older group, we ran a power analysis using G*Power to determine whether our study design was sufficiently powered to detect a significant effect. The study design achieved an observed power of 48% for the main effect of difficulty condition, given an observed partial eta squared of .16. Additionally, there was an observed power of 33% for the interaction between age group and difficulty condition, given an observed partial eta squared of .13. Both of these levels of observed power are substantially lower than the standard of 80%, indicating that this design did not have enough power to find a significant effect. To achieve this level of power, we would have needed to recruit 2-3 times more participants, which was unfeasible for this particular research project. In the future, it would be wise to design and allocate resources for a similar study such that it would be sufficiently powered to find any effects that may exist.

Most participants used either an association or a visualization strategy. Older adults reported using the strategies of finding pattern and pick and choose more often than younger adults. Both of these strategies involve choosing salient subsets of the lists to focus on, rather than using a more holistic strategy. To that end, the most holistic strategy, repetition, was exclusively used by younger adults. This suggests that older adults may have reduced their working memory loads by using these sorts of strategies, thus enabling them to perform as well as younger adults using a holistic strategy. However, this would imply that use of such compensatory strategies would be associated with better performance, which was not the case (Table 15). Thus, it may be the case that older adults may have simply had different strategy preferences than younger adults, which may be due to their higher levels of education or generational effects.

General Discussion

In this investigation, we tested whether differences between younger and older adults would be found in the performance of auditory imagery generation, maintenance, inspection, and transformation tasks. We included both musical and non-musical tasks, as well as a range of measures of individual differences. Although we expected to find some age differences due to the high working memory load in the imagery tasks, overall, we found fewer than expected age differences in auditory imagery performance in the component measures. Specifically, there were no age differences in generation and transformation, some qualitative age differences in maintenance and inspection of musical stimuli, and no age differences in performance of maintenance and inspection of non-musical stimuli, although there were some differences in the strategies used in the non-musical task.

A major issue in the mental imagery and aging literature is that results are not consistent across imagery components. In other words, one study may find age differences in a test of an imagery component, whereas another may find no age differences in a similar task of the same component (e.g., Brown et al., 1998; Dror & Kosslyn, 1994). Regarding the components of generation and transformation, our pitch imagery task from Experiment 1 can be compared to the task on which it was based, which only tested younger adults (Herholz et al., 2016). Our pitch imagery task was found to be as difficult as that used by Herholz et al. (2016), who reported low, but significantly above chance, performance. This result verifies that this task is difficult across samples of individuals. Furthermore, older adults and younger adults performed approximately equally. It may be the case that because this task involved perceptual processing to some extent, by playing out loud the first and last parts of each phrase, older adults may not have shown deficits relative to younger adults because these aspects of processing are spared in older adulthood.

The tests of auditory image maintenance and inspection, the mental scanning and serial order memory tasks, can be compared both to each other and to other tasks in the literature. Both Kosslyn et al. (1978) and Halpern (1988a) studied mental scanning in younger adults, Kosslyn et al. in the visual modality and Halpern in the auditory modality. Both of these studies found that reaction time reliably increased as distance between the stimuli increased, which we replicated here. Our findings also expand on those of these studies by showing that when musical background, self-reported auditory imagery, and working memory were controlled for, this effect went away, suggesting that these variables may be related to differences in processing during this task. Furthermore, our study found that the reaction time increase with distance was heightened in older adults in certain circumstances. Our other task of maintenance and inspection, the serial order memory task using transformed environmental sounds, does not have an equivalent in the literature. Nevertheless, serial order memory tasks using other stimuli have shown both decreases in performance with increasing list lengths and decreases in performance in older adulthood, neither of which was supported by our data (Lewandowsky & Murdock, 1989; Naveh-Benjamin, Cowan, Kilb, & Chen, 2007). This could partially be explained by a floor effect, as both younger and older adults may have found processing of these stimuli to be quite difficult, which would have eliminated a possible advantage in younger adults. It is unclear as to why our data did not support the finding that performance decreases as list length increases.

We made several predictions regarding performance in tasks with different stimuli. Namely, both the pitch imagery task and the mental scanning task made use of songs that participants in both age groups found highly familiar. Meanwhile, the serial order task involved the use of entirely novel sounds, transformed from familiar environmental sounds. We predicted that auditory imagery of familiar stimuli should be easier than that of novel stimuli, particularly for older adults: indeed, older adults have been shown to have difficulties processing novel stimuli (Salthouse, 2012), and they have been shown to have advantages in retrieving memories of familiar melodies compared to unfamiliar ones (Halpern & Bartlett, 2010). Thus, we predicted that that performance in the serial order memory task would be lower than that of the other tasks. However, the data did not support this hypothesis: d' values were approximately equal in the pitch imagery and serial order memory tasks, even when comparing older and younger adults, and error rates in the mental scanning task were comparable to the d' values observed in the other tasks. This strongly suggests that participants of all ages found the tasks equally difficult, and that there was no effect of stimuli despite the rather different stimuli and task demands. We offer a few possible explanations for this result. First, the novel stimuli in the serial order task were designed to have no connection to semantic memory. In other comparisons of novel and familiar stimuli, novel stimuli are things like unfamiliar melodies, which have much more structure than the transformed environmental sounds presented here. Because of this difference, it may be difficult to group the novel stimuli in this experiment with novel stimuli in other experiments. Second, the previously reported disadvantages in processing novel stimuli may have assumed that the tasks with familiar and novel stimuli were equally difficult, and the tasks presented here may not have been. Participants may have learned throughout life to adopt certain strategies for remembering the order of a list, which is a somewhat common task in everyday life. Conversely, rehearsing specific sections of familiar songs in the mind's ear is not a common task, and it is possible that these inherent differences in the task demands led to fewer differences in performance than we had predicted.

As relationships between the individual differences variables were measured in three different samples of younger and older adults, it is useful to compare these relationships. In all samples, education and vocabulary were highly correlated, due to the fact that older adults were both more educated and had higher vocabulary scores in all cases. Furthermore, the two subscales of the BAIS, the measure of self-reported auditory imagery, were highly correlated with one another in all three samples, with *r* values ranging from .7 to .8. Although this is a high correlation, it is not perfect, which suggests that vividness and control predict different aspects of auditory imagery. Indeed, performance in some tasks has been found to be predicted by only one of these subscales (Colley, Keller, & Halpern, 2017), which suggests that these subscales had the ability to differentially predict results in the current investigation as well. The other correlations between these variables were not as reliably found across samples, which suggests that they may be less predictable. Namely, musical background was only sometimes found to be correlated with either of the BAIS subscales, although this result has been found in prior studies (Pfordresher & Halpern, 2013).

In a similar vein, the three different participant samples varied as to whether the individual differences measures differed across age groups. Specifically, across all studies, older adults were much more educated than younger adults, and they had lower musical background in two of the experiments than did the younger adults. These have important implications for the results, as musical background and education can both affect participants' approach to a given task; however, these do not represent cognitive age differences, but rather, they show that the younger and older samples were not equally matched in these measures. Additionally, older adults had higher vocabulary scores than did younger adults across all studies. This is a robust result (e.g., Harada et al., 2013) and is usually taken to indicate higher semantic memory and crystallized intelligence in older adults. The subscales of the BAIS were found to be higher in older adults in one sample, but there were no differences in the other two samples. There has been previous evidence suggesting both of these outcomes (Lima et al., 2015; van Hardeveld de Jong, 2016), so an experiment with a large sample size must be done to determine the true extent of age differences in the BAIS. Were we to do this, we might find that auditory imagery is heavily linked with perceptual processes that are spared in older age, and thus there would be no age differences in this task; on the other hand, in a larger sample, the extensive temporal processing and working memory requirements could negatively affect older adults' ability to maintain an auditory image in the mind's ear long enough to evaluate its vividness or control (Baddeley & Andrade, 2000).

The final individual difference variable included here, with perhaps the most surprising results, was working memory, which was not found to vary between age groups in any of the three samples, although typically older adults have lower spans (Baddeley & Andrade, 2000; Salthouse, 1991). One reason for this may be our older adults' additional years of education. This hypothesis would imply, however, that education would be positively correlated with working memory scores, and this was not supported in any of the three samples. Another possibility is that some other factor associated with high education, which was not measured directly here, may also be responsible for increased working memory spans. Both the Transformation task, the higher startpoints and stepsizes in the mental scanning task, and both difficulty conditions of the serial order memory task were designed to be dependent on working memory. The mental scanning task did find age differences in performance, in that older adults were slower in particular conditions; however, the other two tasks did not yield such differences.

Limitations & Future Directions

Perhaps the largest limitation of these studies was that the older adult group was not fully representative of older adults in general. As of 2015, only 11 percent of Americans over the age of 65 had attained advanced degrees (U.S. Census Bureau, 2016). However, across all three older adult samples in the current study, the average years of education attained was 17.8, where 16 years of education represents having attained a bachelor's degree. This indicates that half of the participants had earned advanced degrees. Many of the older residents of the area are retired professors or otherwise well-educated, and our biggest source of recruitment was from a group of older adults who actively participated in lifelong learning classes; of the 44 older adults who participated in all three studies, 12 had more than 20 years of education, indicating having attained a PhD or multiple master's degrees. Thus, it is reasonable that the older adult sample was skewed in this manner. This educational difference may have caused other abnormal results, such as the lack of age differences between working memory scores observed in all three samples. Moreover, it may be difficult to compare performance between these advanced older adults and Bucknell undergraduates, given that many Bucknell undergraduates do not go on to complete an advanced degree. Future studies should attempt to sample more broadly in order to have more generalizable results. For example, recruitment strategies could involve posting flyers in locations frequented by older adults, such as the movie theater or a church, rather than limiting recruitment to university groups.

An additional limitation is that the design of the serial order memory study was significantly underpowered. As previously mentioned, two to three times as many participants were needed to achieve sufficient power, which was unfeasible in this situation given the length of time of a master's thesis and the limited scope of our recruiting efforts. This limitation could be improved through the same broadening of recruitment strategies described above, as those strategies would hopefully result in more participants.

Furthermore, the cross-sectional design of these studies does not allow for the investigation of cohort effects. Although the familiar songs in the pitch imagery and mental scanning tasks were chosen to be familiar to both younger and older adults, the older adults were more familiar with the songs; several younger adults reported not knowing one or more songs. Participants were asked only whether or not they were familiar with the songs; had they been asked to rate how familiar they are with each song, the older adults may have reported being more familiar than the younger adults due to additional decades of exposure. More generally, because the study was not longitudinal, it is difficult to control for the effects of upbringing or other generational differences when investigating the cognitive effects of aging across generations. Ideally, future investigations of these questions will include longitudinal studies, which will allow for the isolation of any effects of aging from the effects of generation or changing culture. Such studies might select familiar songs based on which songs were popular during childhood for each age range of participants, which might better ensure that participants are equally familiar with the stimuli.

Dependent measures may also show age differences differentially. Most of the tasks presented here measured performance with accuracy, either error rate or d'. The mental scanning task was the only task to measure both reaction time and error rate. This decision was made for several reasons. First, the mental scanning task was the only task where reaction time indicated real-time processing and could be compared across task conditions. Second, reaction time can be easily over-interpreted if there is not a theoretical motivation for its measurement, and the mental scanning task was the only one in which reaction time was theoretically relevant to processing of the task. However, for many tasks in the cognitive aging literature, measuring a combination of accuracy and reaction time often provides the most informative picture of age differences, even when the task is not directly testing reaction time (Vaportzis, Georgiou-Karistianis, & Stout, 2013). Older adults often sacrifice speed for accuracy, which is not surprising given that cognitive processing and neural transmission tend to slow down in older age (Salthouse, 1996). Given that accuracy measures in the current investigation did not show significant age differences, it is possible that tasks more dependent on reaction time may have revealed a speed-accuracy tradeoff. Therefore, future investigations of auditory imagery and aging should be sure to measure both reaction time and accuracy, if it is appropriate to do so, in order to determine whether a speed-accuracy tradeoff is occurring.

A final limitation for all three studies is that the majority of the methods used were new. The generation/transformation and serial order memory tasks, although adapted from prior tasks, were completely new. Furthermore, the BAIS and the mental scanning task have only been used a handful of times in other studies, so it is not clear whether or not we should have expected to find age differences. This was, however, a necessary limitation, as many of the questions we asked in this research have not been answered in other studies. Future studies should be sure to continue using these or similar methods in order to determine their reliability and ability to measure age and individual differences in auditory imagery.

Future investigations of auditory imagery in healthy aging should use a variety of tasks. The musical imagery tasks included in Experiments 1 and 2A focused solely on pitch imagery; however, there are other aspects of auditory imagery that could be investigated, including tempo, timbre, and loudness. It is possible that these facets of imagery could be differentially affected by healthy aging and other individual differences factors, so these questions would be interesting to consider in the future. Conversely, multiple studies should be sure to use the same tasks in different samples so that researchers can have consistent methods. A limitation of the mental imagery and aging literature is that studies continually make small adaptations to previous tasks, such as using the same structure but different stimuli, or vice versa. Although this is also a strength in that the tasks become more precise over time, this adaptation makes results difficult to compare across studies. Therefore, future investigations should be structured such that they can both replicate an existing task and then make any adaptations that they see fit, in order to increase generalizability and sensibly interpret their results.

Environmental sounds should also be further studied in the future. The current study tested the effects of non-nameable sounds that were derived from environmental sounds. These were designed to be processed without the use of semantic memory, in the hope that the general processing of the auditory system could be investigated. However, the question of whether there would be age differences in an auditory imagery task of non-transformed environmental sounds remains open. Experiment 2B tested imagery of transformed environmental sounds in a working memory-heavy task, and there may be other designs that can test imagery of environmental sounds in a less taxing way while maintaining a realistic context for imagining those sounds. Nevertheless, because these stimuli are present and informative in navigating the environment, and because these sounds may vary in vividness according to their role in predicting events in the environment, it is important to investigate potential age differences in these types of stimuli.

In conclusion, we found fewer age differences than expected in various tests of auditory imagery. The pitch imagery and serial order memory tasks showed a trend towards age differences but did not find significant effects, suggesting that a more powered design may be better able to confirm those trends one way or another. Furthermore, the individual differences variables of musical background, self-reported auditory imagery, and working memory sometimes predicted better performance, but this was not true for all included studies. Future studies should be sure to measure these and other relevant cognitive variables, which may explain performance in imagery tasks.

References

Adduri, C. A., & Marotta, J. J. (2009). Mental rotation of faces in healthy aging and Alzheimer's disease. *PLoS ONE*, *4*(7).

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

Baddeley, A. D., & Andrade, J. (2000). Working memory and the vividness of imagery. *Journal of Experimental Psychology: General*, *129*(1), 126–145.

Band, G. P., & Kok, A. (2000). Age effects on response monitoring in a mental-rotation task. *Biological Psychology*, *51*(2–3), 201–21.

Berg, C., Hertzog, C., & Hunt, E. (1982). Age differences in the speed of mental rotation. *Developmental Psychology*, *18*(1), 95–107.

Briggs, S. D., Raz, N., & Marks, W. (1999). Age-related deficits in generation and manipulation of mental images: I. The role of sensorimotor speed and working memory. *Psychology and Aging*, *14*(3), 427–435.

Brown, H. D., Kosslyn, S. M., & Dror, I. E. (1998). Aging and Scanning of Imagined and Perceived Visual Images. *Experimental Aging Research*, *24*, 181–194.

Bunzeck, N., Wuestenberg, T., Lutz, K., Heinze, H., & Jancke, L. (2005). Scanning silence: Mental imagery of complex sounds. *NeuroImage*, *26*, 1119–1127.

Charmaz, K. (2008). Grounded Theory as an Emergent Method. In *Handbook of Emergent Methods* (pp. 155–170).

Colley, I. D., Keller, P. E., & Halpern, A. R. (2017). Working Memory and Auditory Imagery Predict Sensorimotor Synchronization with Expressively Timed Music.

The Quarterly Journal of Experimental Psychology, (August), 1–49.

De Beni, R., Pazzaglia, F., & Gardini, S. (2007). The generation and maintenance of visual mental images: Evidence from image type and aging. *Brain and Cognition*, *63*(3), 271–278.

De Simone, L., Tomasino, B., Marusic, N., Eleopra, R., & Rumiati, R. I. (2013). The effects of healthy aging on mental imagery as revealed by egocentric and allocentric mental spatial transformations. *Acta Psychologica*, *143*(1), 146–156.

Dror, I. E., & Kosslyn, S. M. (1994). Mental imagery and aging. *Psychology and Aging*, *9*(1), 90–102.

Dror, I. E., Schmitz-Williams, I. C., & Smith, W. (2005). Older adults use mental representations that reduce cognitive load: Mental rotation utilizes holistic representations and processing. *Experimental Aging Research*, *31*(4), 409–420.

Educational attainment in the United States: 2015. (2016). *United States Census Bureau*.

Fabiani, M., Kazmerski, V. A., Cycowicz, Y. M., & Friedman, D. (1996). Naming norms for brief environmental sounds: Effects of age and dementia. *Psychophysiology*, *33*, 462–475.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191.

Gaudreau, D., & Peretz, I. (1999). Implicit and explicit memory for music in old and young adults. *Brain and Cognition*, *40*(1), 126–129.

Grassi, M., Meneghetti, C., Toffalini, E., & Borella, E. (2017). Auditory and cognitive performance in elderly musicians and nonmusicians. *PLoS ONE*, *12*(11), e0187881.

Gregoire, J., & Van Der Linden, M. (2004). Effect of age on forward and backward span tasks. *Journal of the International Neuropsychological Society*, *10*(4), 475–481.

Halpern, A. R. (1984). Organization in memory for familiar songs. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*(3), 496–512.

Halpern, A. R. (1988a). Mental scanning in auditory imagery for songs. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(3), 434–443.

Halpern, A. R. (1988b). Perceived and imagined tempos of familiar songs. *Music Perception*, *6*(2), 193–202.

Halpern, A. R. (2015). Differences in auditory imagery self-report predict neural and behavioral outcomes. *Psychomusicology: Music, Mind, and Brain*, *25*(January 2015), 37–47.

Halpern, A. R., & Bartlett, J. (2010). Memory for melodies. In M. R. et al Jones (Ed.), *Music Perception* (Vol. 36, pp. 233–258).

Halpern, A. R., & Zatorre, R. J. (1999). When that tune runs through your head: A PET investigation of auditory imagery for familiar melodies. *Cerebral Cortex*, *9*(7), 697–704.

Harada, C. N., Love, M. C. N., & Triebel, K. (2013). Normal cognitive aging. *Clin Geriatr Med.*, *29*(4), 737–752.

Herholz, S. C., Coffey, E. B. J., Pantev, C., & Zatorre, R. J. (2016). Dissociation of neural networks for predisposition and for training-related plasticity in auditory-motor learning. *Cerebral Cortex*, *26*(7), 3125–3134.

Hertzog, C., & Rypma, B. (1991). Age differences in components of mental rotation task performance. *Bulletin of the Psychonomic Society*, *29*(3), 209–212.

Hubbard, T. L. (2010). Auditory imagery: empirical findings. *Psychological Bulletin*, *136*(2), 302.

International Organization for Standardization (2017). Acoustics - Statistical distribution of hearing thresholds related to age and gender (ISO 7029).

Joanisse, M., Gagnon, S., Kreller, J., & Charbonneau, M. C. (2008). Age-related differences in viewer-rotation tasks: is mental manipulation the key factor? *J Gerontol B Psychol Sci Soc Sci*, *63B*(3), 193–200. Retrieved from

Kalkstein, J., Checksfield, K., Bollinger, J., & Gazzaley, A. (2011). Diminished top-down control underlies a visual imagery deficit in normal aging. *Journal of Neuroscience*, *31*(44), 15768–15774.

Kemps, E., & Newson, R. (2005). Patterns and predictors of adult age differences in mental imagery. *Aging, Neuropsychology, and Cognition*, *12*(1), 99–128.

Kosslyn, S. M., Ball, T. M., & Reiser, B. J. (1978). Visual images preserve metric spatial information: evidence from studies of image scanning. *Journal of Experimental Psychology. Human Perception and Performance*, *4*(1), 47–60.

Lewandowsky, S., & Murdock, B. B. (1989). Memory for serial order. *Psychological Review*, *96*(1), 25–57.

Lewis, J. W., Wightman, F. L., Brefczynski, J. A., Phinney, R. E., Binder, J. R., & DeYoe, E. A. (2004). Human brain regions involved in recognizing environmental sounds. *Cerebral Cortex*, *14*(9), 1008–1021.

Lima, C. F., Lavan, N., Evans, S., Agnew, Z., Halpern, A. R., Shanmugalingam, P., … Scott, S. K. (2015). Feel the noise: Relating individual differences in auditory imagery to the structure and function of sensorimotor systems. *Cerebral Cortex*, *25*(11), 4638–4650.

Macmillan, N. A., & Creelman, C. D. (2004). *Detection Theory: A User's Guide*. Psychology Press.

Marcell, M. M., Borella, D., Greene, M., Kerr, E., & Rogers, S. (2000). Confrontation naming of environmental sounds. *Journal of Clinical and Experimental Neuropsychology*, *22*(6), 830–864.

McNorgan, C. (2012). A meta-analytic review of multisensory imagery identifies the neural correlates of modality-specific and modality-general imagery. *Frontiers in Human Neuroscience*, *6*.

Müllensiefen, D., Gingras, B., Musil, J., & Stewart, L. (2014). The musicality of non-musicians: An index for assessing musical sophistication in the general population. *PLoS ONE*, *9*(2).

Naveh-Benjamin, M., Cowan, N., Kilb, A., & Chen, Z. (2007). Age-related differences in immediate serial recall: Dissociating chunk formation and capacity. *Memory & Cognition*, *35*(4), 724–737.

Old, S. R., & Naveh-Benjamin, M. (2008). Age-related changes in memory: Experimental approaches. In S. M. Hofer & D. F. Alwin (Eds.), *Handbook of Cognitive Aging: Interdisciplinary Perspectives* (pp. 151–167).

Palermo, L., Iaria, G., & Guariglia, C. (2008). Mental imagery skills and topographical orientation in humans: A correlation study. *Behavioural Brain Research*, *192*, 248–253.

Palermo, L., Piccardi, L., Nori, R., Giusberti, F., & Guariglia, C. (2016). The impact of ageing and gender on visual mental imagery processes: A study of performance on tasks from the Complete Visual Mental Imagery Battery (CVMIB). *Journal of Clinical and Experimental Neuropsychology*, *38*, 752–763.

Park, D. C., & Reuter-Lorenz, P. (2009). The Adaptive Brain: Aging and Neurocognitive Scaffolding. *Annual Review of Psychology*, *60*(1), 173–196.

Pfordresher, P. Q., & Halpern, A. R. (2013). Auditory imagery and the poor-pitch singer. *Psychonomic Bulletin and Review*, *20*(4), 747–753.

Rimmele, J. M., Sussman, E., & Poeppel, D. (2015). The role of temporal structure in the investigation of sensory memory, auditory scene analysis, and speech perception: A healthy-aging perspective. *International Journal of Psychophysiology*, *95*(2), 175–183.

Saimpont, A., Malouin, F., Tousignant, B., & Jackson, P. L. (2013). Motor imagery and aging. *Journal of Motor Behavior*, *45*(1), 21–8.

Salthouse, T. A. (1991). Decomposing adult age differences in working memory. *Developmental Psychology*, *27*(5), 763.

Salthouse, T. A. (1994). The aging of working memory. *Neuropsychology*, *8*, 535–543.

Salthouse, T. A. (1996). The processing speed theory of adult age differences in cognition. *Psychological Review*, *103*, 403–428.

Salthouse, T. A. (2012). Consequences of age-related cognitive declines. *Annual Review of Psychology*, *63*, 201–26.

Schott, N. (2012). Age-related differences in motor imagery: Working memory as a mediator. *Experimental Aging Research*, *38*, 559–583.

Sharps, M. J., & Gollin, E. S. (1987). Speed and accuracy of mental image rotation in young and elderly adults. *Journal of Gerontology*, *42*(3), 342–344.

van Hardeveld de Jong, F. (2016). *Aging and imagery: Does aging affect imagery ability?* Leiden University.

Vaportzis, E., Georgiou-Karistianis, N., & Stout, J. C. (2013). Dual task performance in normal aging: A comparison of choice reaction time tasks. *PLoS ONE*, *8*(3), e60265.

Wu, J., Mai, X., Chan, C. C. H., Zheng, Y., & Luo, Y. (2006). Event-related potentials during mental imagery of animal sounds. *Psychophysiology*, *43*(6), 592–597.

Zatorre, R. J., Halpern, A. R., & Bouffard, M. (2009). Mental reversal of imagined melodies: A role for the posterior parietal cortex. *Journal of Cognitive Neuroscience*, *22*(4), 775–89.

Tables and Figures

Table 1

Demographic Information for Experiment 1.

Characteristic	Younger adults (N = 28) M (SD)	Older adults (N = 21) M (SD)	*p*-value
Age	18.75 (0.75)	72.29 (8.31)	< 0.0001**
Education	12.93 (0.81)	17.48 (2.50)	< 0.0001**
Musical Background	3.11 (4.10)	1.82 (4.94)	.324
BAIS Vividness	4.71 (0.59)	5.35 (0.92)	.005**
BAIS Control	4.80 (0.69)	5.64 (0.69)	<0.0001**
WAIS-R Vocabulary	18.50 (4.82)	27.52 (5.11)	<0.0001**
Backwards Digit Span	4.50 (0.75)	3.95 (1.28)	.067

Note: ** denotes $p < 0.01$.

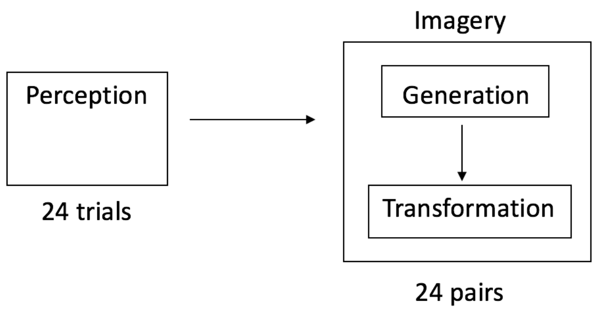

Figure 1. Diagram of Experiment 1 Procedure and Example Trial.

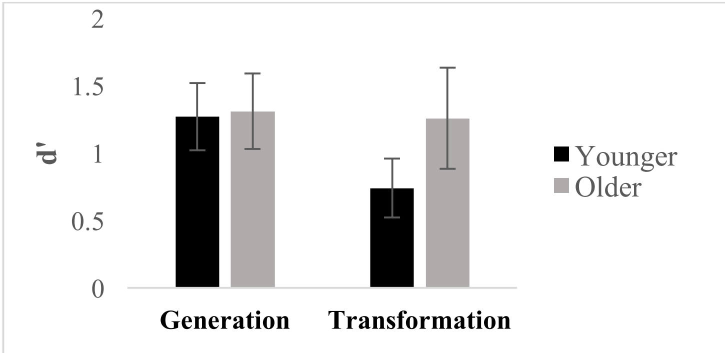

Figure 2. Accuracy results for Experiment 1. Error bars represent standard error of the mean.

Table 2

Correlations of Individual Differences Measures in Experiment 1.

	Education	Vocab	Musical Background	BAIS-V	BAIS-C
Vocab	.587**				
Musical Background	-.142	.026			
BAIS-V	.118	.214	.200		
BAIS-C	.296*	.279^{+}	.088	.742**	
WM	-.130	-.033	.026	-.381**	-.191

N = 49. Table denotes r values between individual differences measures and experimental performance. Vocab refers to WAIS-R vocabulary score. BAIS-V refers to the vividness subscale of the Bucknell Auditory Imagery Scale. BAIS-C refers to the control subscale. WM refers to backwards digit span. $^{+}$ denotes $p < 0.10$, * denotes $p < 0.05$; ** denotes $p < 0.01$.

Table 3

Correlations of Individual Differences Measures and Younger Adult Performance in Experiment 1.

	Education	Vocab	Musical Background	BAIS-V	BAIS-C
Vocab	.132				
Musical Background	-.208	.413*			
BAIS-V	-.274	.342⁺	.309		
BAIS-C	-.111	.198	.280	.778**	
WM	-.183	.144	.163	-.098	-.033

N = 28. Table denotes r values between individual differences measures and experimental performance. Vocab refers to WAIS-R vocabulary score. BAIS-V refers to the vividness subscale of the Bucknell Auditory Imagery Scale. BAIS-C refers to the control subscale. WM refers to backwards digit span. $^{+}$ denotes $p < 0.10$, * denotes $p < 0.05$; ** denotes $p < 0.01$.

Table 4

Correlations of Individual Differences Measures and Older Adult Performance in Experiment 1.

	Education	Vocab	Musical Background	BAIS-V	BAIS-C
Vocab	.112				
Musical Background	.009	-.087			
BAIS-V	-.392^{+}	-.425^{+}	.266		
BAIS-C	-.342	-.535*	.098	.634**	
WM	.225	.263	-.263	-.422^{+}	-.093

N = 21. Table denotes r values between individual differences measures and experimental performance. Vocab refers to WAIS-R vocabulary score. BAIS-V refers to the vividness subscale of the Bucknell Auditory Imagery Scale. BAIS-C refers to the control subscale. WM refers to backwards digit span. $^{+}$ denotes $p < 0.10$, * denotes $p < 0.05$; ** denotes $p < 0.01$.

Table 5

Demographic Information for Experiment 2A.

Characteristic	Younger adults (N = 19) M (SD)	Older adults (N = 15) M (SD)	*p*-value
Age	18.53 (0.61)	70.20 (5.75)	< 0.0001**
Education	12.40 (0.61)	18.20 (3.36)	< 0.0001**
Gold-MSI: Training	23.89 (8.66)	16.33 (7.68)	.012*
BAIS Vividness	4.61 (0.80)	4.91 (1.06)	.354
BAIS Control	4.90 (0.62)	5.34 (0.95)	.113
WAIS-R Vocabulary	20.74 (5.83)	26.73 (6.25)	.007*
Backwards Digit Span	4.74 (1.41)	4.33 (0.98)	.352

Note: * denotes $p < 0.05$. ** denotes $p < 0.01$.

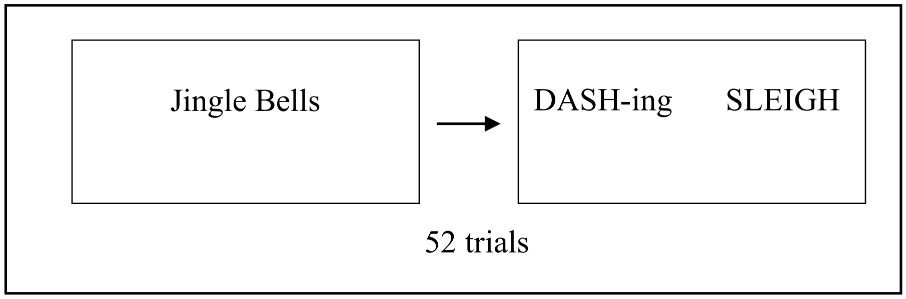

Figure 3. Diagram of Experiment 2A Procedure and Example Trial. Titles were shown for 1000ms, and the lyrics were shown until participant gave a keyboard response. The word “DASHING” has two distinct pitches, so all caps denotes the pitch on which participants should focus (“DASH”). Key press instructions were provided on every trial.

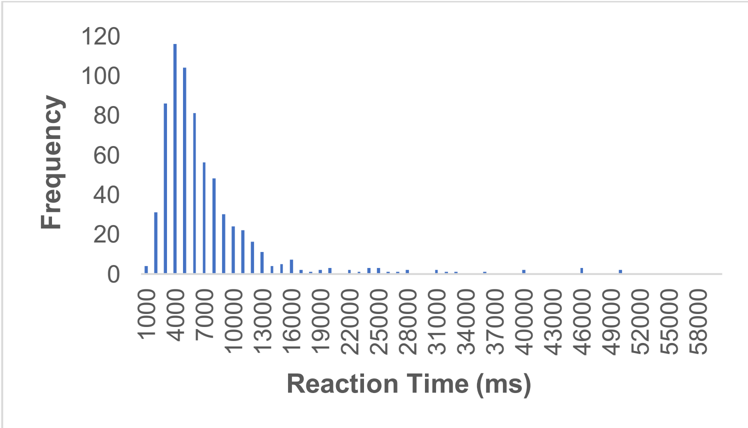

Figure 4.1. Distribution of untransformed reaction times for younger adults in Experiment 2A. Only correct trials were included.

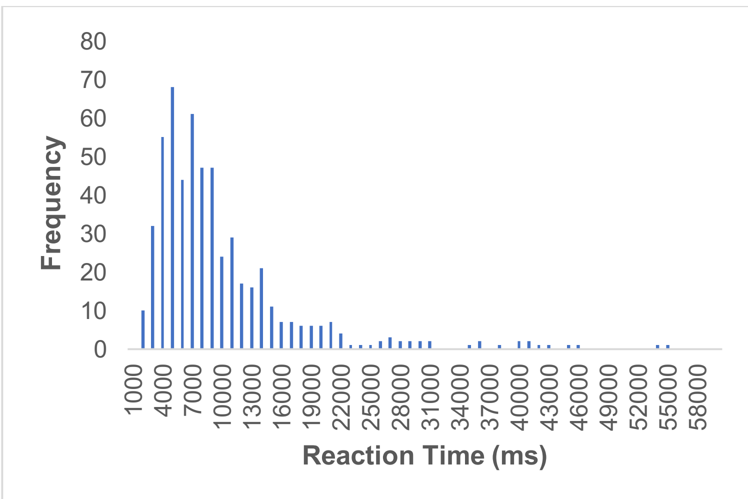

Figure 4.2. Distribution of untransformed reaction times for older adults in Experiment 2A. Only correct trials were included.

Table 6

Error Rates by Startpoint, Stepsize, and Age Group in Experiment 2A.

Age Group	Startpoint	Stepsize	Mean	Standard Error
Younger	1	2	0.228	0.048
		6	0.184	0.06
		8	0.36	0.064
	3	2	0.116	0.046
		6	0.218	0.06
		8	0.259	0.049
	5	2	0.279	0.046
		6	0.291	0.054
		8	0.272	0.058
Older	1	2	0.2	0.054
		6	0.417	0.067
		8	0.267	0.072
	3	2	0.213	0.052
		6	0.28	0.068
		8	0.3	0.055
	5	2	0.36	0.051
		6	0.32	0.061
		8	0.317	0.065

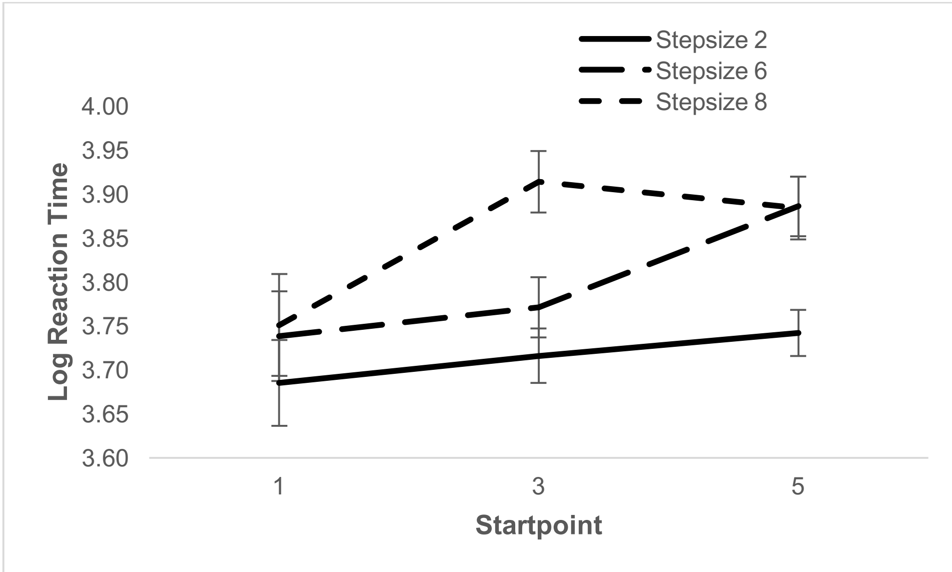

Figure 5. Startpoint by stepsize interaction for reaction time in Experiment 2A. Data is collapsed across age groups, as the three way interaction was not significant. Error bars represent standard error of the mean.

Table 7

Correlations of Individual Differences Measures in Experiment 2A.

	Education	Vocab	Gold-MSI	BAIS-V	BAIS-C
Vocab	.584**				
Gold-MSI	-.319$^+$	-.316$^+$			
BAIS-V	-.110	-.104	.031		
BAIS-C	.025	.073	.122	.806**	
WM	-.074	.084	.294$^+$	-.006	-.045

N = 34. Table denotes r values between individual differences measures and experimental performance. Vocab refers to WAIS-R vocabulary score. Gold-MSI refers to the musical training subscale of the Goldsmiths Musical Sophistication Index. BAIS-V refers to the vividness subscale of the Bucknell Auditory Imagery Scale. BAIS-C refers to the control subscale. WM refers to backwards digit span. $^+$ denotes $p < 0.10$, * denotes $p < 0.05$; ** denotes $p < 0.01$.

Table 8

Correlations of Individual Differences Measures in Younger Adults in Experiment 2A.

	Education	Vocab	Gold-MSI	BAIS-V	BAIS-C
Vocab	.225				
Gold-MSI	-.425^{+}	-.302			
BAIS-V	.016	.050	-.179		
BAIS-C	.075	.158	-.044	.803**	
WM	.030	.093	.531*	.114	.132

N = 19. Table denotes r values between individual differences measures and experimental performance. Vocab refers to WAIS-R vocabulary score. Gold-MSI refers to the musical training subscale of the Goldsmiths Musical Sophistication Index. BAIS-V refers to the vividness subscale of the Bucknell Auditory Imagery Scale. BAIS-C refers to the control subscale. WM refers to backwards digit span. $^{+}$ denotes $p < 0.10$, * denotes $p < 0.05$; ** denotes $p < 0.01$.

Table 9

Correlations of Individual Differences Measures in Older Adults in Experiment 2A.

	Education	Vocab	Gold-MSI	BAIS-V	BAIS-C
Vocab	.563*				
Gold-MSI	.169	.050			
BAIS-V	-.542*	-.434	.438		
BAIS-C	-.437	-.236	.598*	.806**	
WM	.174	.343	-.330	-.107	-.138

N = 15. Table denotes r values between individual differences measures and experimental performance. Vocab refers to WAIS-R vocabulary score. Gold-MSI refers to the musical training subscale of the Goldsmiths Musical Sophistication Index. BAIS-V refers to the vividness subscale of the Bucknell Auditory Imagery Scale. BAIS-C refers to the control subscale. WM refers to backwards digit span. $^{+}$ denotes $p < 0.10$, * denotes $p < 0.05$; ** denotes $p < 0.01$.

Table 10

Demographic Information for Experiment 2B.

Characteristic	Younger adults (N = 31) M (SD)	Older adults (N = 23) M (SD)	*p*-value
Age	18.55 (0.57)	68.78 (6.47)	< 0.0001**
Education	12.29 (0.48)	18.13 (3.18)	< 0.0001**
Gold-MSI: Training	22.42 (7.62)	15.78 (8.47)	.0039**
BAIS Vividness	4.50 (0.88)	4.91 (1.03)	.1215
BAIS Control	4.82 (0.62)	5.18 (0.92)	.0918
WAIS-R Vocabulary	20.06 (5.41)	25.57 (7.17)	.0022**
Backwards Digit Span	4.77 (1.43)	4.13 (1.14)	.0829

Note: ** denotes $p < 0.01$.

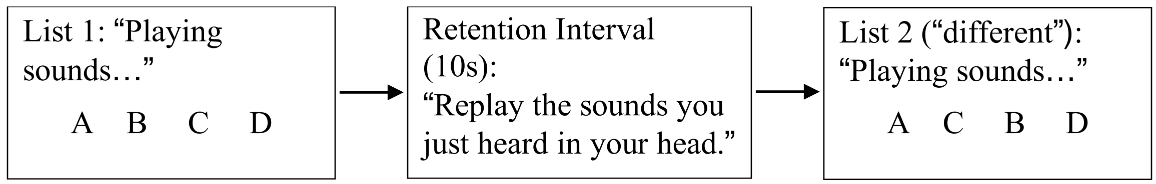

Figure 6.1. Diagram of Experiment 2B "different" trial in the easy condition. A, B, C, and D represent four unique sounds, each 1000ms in length. They are presented with 500ms of silence separating them. In List 2, A and D remain in the same position in both lists.

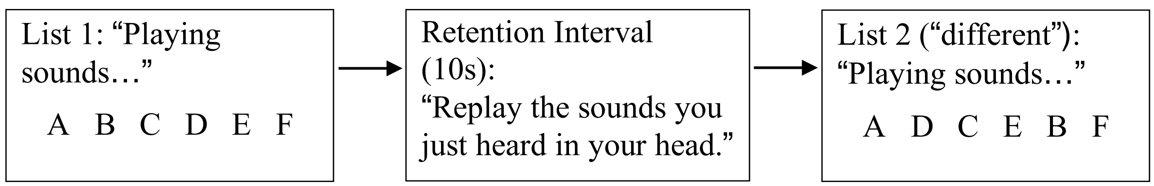

Figure 6.2. Diagram of Experiment 2B "different" trial in the hard condition. A, B, C, D, E, and F represent six unique sounds, each 1000ms in length. They are presented with 500ms of silence separating them. In List 2, A and F remain in the same position in both lists.

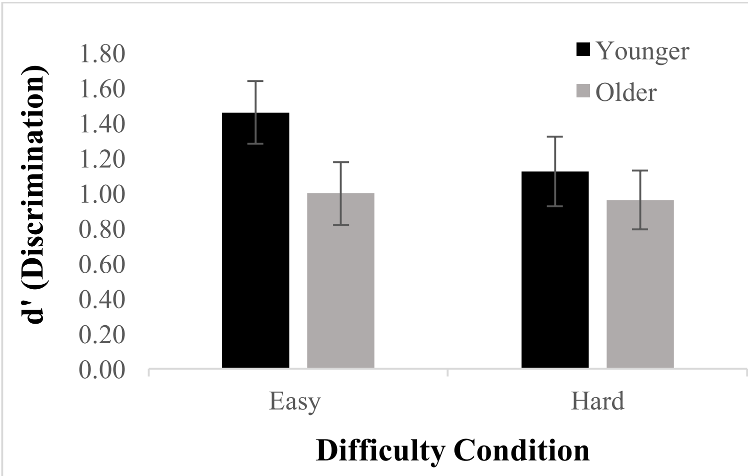

Figure 7. Performance in Experiment 2B. This experiment tested the serial order memory task. Lists in the easy condition consisted of four sounds, and lists in the hard condition consisted of six sounds. Error bars represent standard error of the mean. No comparisons were significant.

Table 11

Correlations of Individual Differences Measures in Experiment 2B.

	Education	Vocab	Gold-MSI	BAIS-V	BAIS-C
Vocab	.606**				
Gold-MSI	-.247^{+}	-.203			
BAIS-V	.013	-.127	.128		
BAIS-C	.041	.002	.237^{+}	.791**	
WM	-.130	.067	.182	-.246^{+}	-.195

N = 54. Table denotes r values between individual differences measures and experimental performance. Vocab refers to WAIS-R vocabulary score. Gold-MSI refers to the musical training subscale of the Goldsmiths Musical Sophistication Index. BAIS-V refers to the vividness subscale of the Bucknell Auditory Imagery Scale. BAIS-C refers to the control subscale. WM refers to backwards digit span. $^{+}$ denotes p < 0.10, * denotes p < 0.05; ** denotes p < 0.01.

Table 12

Correlations of Individual Differences Measures in Younger Adults in Experiment 2B.

	Education	Vocab	Gold-MSI	BAIS-V	BAIS-C
Vocab	.353^{+}				
Gold-MSI	-.003	-.165			
BAIS-V	.130	-.092	.003		
BAIS-C	.233	.005	.127	.746**	
WM	-.047	.045	.333^{+}	-.113	-.003

N = 31. Table denotes r values between individual differences measures and experimental performance. Vocab refers to WAIS-R vocabulary score. Gold-MSI refers to the musical training subscale of the Goldsmiths Musical Sophistication Index. BAIS-V refers to the vividness subscale of the Bucknell Auditory Imagery Scale. BAIS-C refers to the control subscale. WM refers to backwards digit span. $^{+}$ denotes $p < 0.10$, * denotes $p < 0.05$; ** denotes $p < 0.01$.

Table 13

Correlations of Individual Differences Measures in Older Adults in Experiment 2B.

	Education	Vocab	Gold-MSI	BAIS-V	BAIS-C
Vocab	.646**				
Gold-MSI	.185	.046			
BAIS-V	-.425*	-.368^{+}	.471*		
BAIS-C	-.376^{+}	-.183	.569**	.818**	
WM	.221	.380^{+}	-.256	-.352	-.329

N = 23. Table denotes r values between individual differences measures and experimental performance. Vocab refers to WAIS-R vocabulary score. Gold-MSI refers to the musical training subscale of the Goldsmiths Musical Sophistication Index. BAIS-V refers to the vividness subscale of the Bucknell Auditory Imagery Scale. BAIS-C refers to the control subscale. WM refers to backwards digit span. $^{+}$ denotes $p < 0.10$, * denotes $p < 0.05$; ** denotes $p < 0.01$.

Table 14

Self-Reported Strategies in Experiment 2B.

Strategy	Younger adults (N = 31)	Older adults (N = 23)	Both Groups
Association	35.5	39.1	37.0
Visualization	29.0	30.4	29.6
Finding Pattern	12.9	47.8	27.8
Identifying Pitch	6.5	13.0	9.3
Pick and Choose	9.7	4.3	7.4
Repetition	9.7	0	5.6

Percentages of each age group, and total percentages, of participants who reported using each strategy. Participants were permitted to report multiple strategies, thus, percentages do not add to 100.

Table 15

Difference Scores in Strategy Usage in Experiment 2B.

	Visual.	Assoc.	Pick and Choose	Finding Pattern	Identifying Pitch	Repetition
Easy						
Younger	0.67	0.37	-0.07	-0.64	-0.97	-1.23
Older	0.32	0.41	N/A	-0.44	-0.27	N/A
Hard						
Younger	-0.12	-0.23	0.30	-0.52	0.08	-1.03
Older	0.00	0.20	N/A	-0.06	-0.76	N/A

Scores represent differences in average d' between those who reported using each strategy and those who did not. Strategies ordered by effectiveness for younger adults in the Easy condition. Visual. represents a visualization strategy, and Assoc. represents an association strategy. Values of N/A are due to the lack of older adults who reported using those strategies: one older adult reported using Pick and Choose, and none reported using Repetition; thus, difference scores could not be calculated for those cells.