

PRECISION AND CAPACITY DEVELOPMENT
IN AUDITORY WORKING MEMORY

A Thesis
presented to
the Faculty of the Graduate School
at the University of Missouri-Columbia

In Partial Fulfillment
of the Requirements for the Degree
Master of Arts

by
KATHERINE MICHELLE CLARK
Dr. Todd Schachtman and Dr. Nelson Cowan, Thesis Supervisors

DECEMBER 2014

The undersigned, appointed by the dean of the Graduate School, have examined the thesis entitled

PRECISION AND CAPACITY DEVELOPMENT
IN AUDITORY WORKING MEMORY

presented by Katherine Michelle Clark,

a candidate for the degree of Master of Arts

and hereby certify that, in their opinion, it is worthy of acceptance.

Professor Todd Schachtman

Professor Nelson Cowan

Professor Jeff Rouder

Professor Mary Fagan

ACKNOWLEDGEMENTS

I would like to thank my advisors, Todd Schachtman and Nelson Cowan, for their support of this project. Without their guidance, challenges, and dedication, none of this would have been possible. I would also like to thank Kyle Hardman, without whom the modelling of the data would have taken significantly longer. His patience and guidance were instrumental in the formation of this thesis. I would like to thank Scott Saults for his programming expertise and enthusiasm for this project. Additionally, I would like to thank Bret Glass, who was in charge of the majority of scheduling for both child and adult participants; he was integral to the running of this study and I am extremely grateful. Thanks also go to my thesis committee members, Jeff Rouder and Mary Fagan, for agreeing to be a part of this process. And finally I would like to thank Garrett Hinrichs, Evie Vergauwe, and Timothy Ricker for their contributions and support throughout the development and implementation of this study. Thank you all, so very much. Without you this project would never have become a reality.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	ii
LIST OF FIGURES	v
LIST OF TABLES	vi
ABSTRACT	vii
Chapter	
1. INTRODUCTION	1
Memory for Items in an Array	
Adult Working Memory Capacity and Precision	
Developmental Capacity and Precision in Visual Working Memory	
Auditory Working Memory Capacity and Precision	
Development of Auditory Working Memory	
The Present Study	
2. METHODS	16
Subjects	
Procedure	
3. OUR MODEL	21
Guessing Strategies	
Modelling	
Our Model vs. the Zhang and Luck Model	
Mathematical Description of Our Model	
4. RESULTS	27
Tone Task	

Raven's Progressive Matrices	
Counting Span	
Correlations Between Tasks and Regression Analyses	
5. DISCUSSION	36
REFERENCES	40
APPENDIX	
1. SOUND LEVEL READINGS	44

LIST OF FIGURES

Figure 1: Visual depiction of the tone task	18
Figure 2: Visual depiction of the response slider	19
Figure 3: Illustration of the multinomial process tree model	22
Figure 4: Scatterplot and line graphs of our model	24
Figure 5: Scatterplot matrix of all raw data	29
Figure 6: Examples of three individual subjects' response patterns	30
Figure 7: Model predictions scatterplot	31
Figure 8: Graphic depiction of means for each parameter across age groups	32

LIST OF TABLES

Table 1: Absolute response differences for each group at each set size	28
Table 2: Means and standard deviations for each parameter across age groups	32
Table 3: Raw and partial correlation matrix	34

PRECISION AND CAPACITY DEVELOPMENT IN AUDITORY WORKING MEMORY

Katherine Michelle Clark

Nelson Cowan and Todd Schachtman,
Thesis Co-Advisors

ABSTRACT

This research has explored the development of auditory working memory capacity and precision in children and adults. Very little is known about the capacity limits of auditory working memory, and even less is known about its precision. Previous research has focused primarily on visual working memory – its development, capacity, and precision; however no studies thus far have examined all three of these processes in a single experiment in either the visual or auditory modality. Our goal was to learn more about the limits and precision of auditory working memory and how it develops across one’s lifespan.

Introduction

The proposed research was designed to examine the childhood development of auditory working memory in a manner that allowed separate indices of two key components: capacity and precision of the representation. After a brief overview of working memory, several lines of research are explained that together provide the background for our inquiry: (1) memory for items in an array, which led to the first quantitative estimates of working memory capacity; (2) adult research that allowed a separation of estimates of capacity and precision in working memory; (3) the extension of this work on capacity and precision to child development; (4) the extension of the capacity and precision estimates to auditory materials; and (5) developmental auditory working memory. Then it is explained how these lines of research come together to inspire the proposed research on the development of auditory working memory capacity and precision.

Before we can focus on the topic of auditory working memory, it is important to look at the background of working memory in general. Working memory can best be described as the amount of information actively held in one's mind at any given point in time, as opposed to the vast amount of information saved in long-term memory. It is crucial for tasks such as reading, writing, and carrying on conversations. In 1956 George Miller first described its capacity limits in his classic paper "The magical number seven, plus or minus two." The first part of the introduction of that article sets the framework for the area of cognitive psychology that is now dedicated to the study of working memory, and many researchers have devoted their careers to studying this vital resource. He showed that normal adults can typically repeat a list of about 7 meaningful units, or chunks.

Of primary concern for the present project are the measures of working memory precision and capacity that have been developed over the years. Baddeley and Hitch (1974) created a classic model of working memory, describing it as having a limited “central executive” moderator that coordinates the input and output from either the visuospatial sketchpad, which is responsible for processing incoming visual or spatial information, or the phonological loop, which processes incoming auditory information. In 2000, Baddeley amended this model to include an episodic buffer, which is proposed to link together all incoming visual, spatial, and auditory information. Though there has been much support for this model of working memory, many other researchers have developed their own theories of those processes.

Cowan (2001) amended Miller’s original theory of capacity limits when he showed that working memory is much more likely limited to 3-4 distinct, meaningful units or chunks of information rather than Miller’s (1956) previously proposed range of 5-9 chunks. The smaller estimate only occurred for situations in which it was not possible to use the Baddeley and Hitch phonological loop type of mechanism to rehearse the units, which can increase the amount that could be retained in working memory. Examples of stimuli that are difficult to rehearse include briefly-presented, simultaneous spatial arrays of simple objects, and lists of nonverbal sounds such as tones.

Much of the research conducted on working memory has focused on capacity, precision, or development, and nearly all of this work has been in the visual domain. Some studies have looked at capacity and precision together in adults, and some have looked at either the development of capacity or the development of precision in children. Only a handful of studies have examined auditory working capacity, and none have looked at its development. This thesis

will examine the capacity and precision of auditory working memory from a developmental perspective, opening a new door in the area of working memory research.

Memory for Items in an Array

Working memory has been studied in one way or another for many years, though its capacity wasn't defined until Miller's seminal article on "The magical number seven, plus or minus two." Since then, hundreds of studies have been conducted examining its capacity, duration, and limitations in a variety of ways. One of the most important developments in the field of working memory research was that of array tasks. Up until 1960, the bulk of working memory research involved sequential presentation of visual information. George Sperling (1960) pioneered the use of visual arrays for the study of working memory. He determined that for full reports – where subjects are required to remember everything seen in a presented array – the number of items cannot exceed the limits of working memory, or performance will suffer. If partial reports are used instead, however, performance improves greatly – unless a long delay between presentation and report time occurs. He determined that simultaneous presentation was just as effective a measure as sequential presentation, and was amongst the first to suggest that the reason for enhanced performance with immediate recall was related to sensory memory, which appears to lack the restrictions of working memory aside from time; sensory memory in his study only lasted for several hundred milliseconds.

In 1974, Philips expanded the distinction between sensory and visual working memory. He defined sensory memory as extremely high-capacity but extremely time-limited, and conducted a series of experiments utilizing unfamiliar stimuli – matrices of varying sizes filled in with different numbers of dots. He found that within several hundred milliseconds, sensory

memory seemed to be tied to spatial position as performance declined when the dots changed location, and that it appeared to be almost completely lost when a mask was used. Working memory, on the other hand, outlasted sensory memory despite its limited capacity, and it was unaffected by both changes in serial position and by masking. This work was corroborated by Pashler, who examined performance as a function of display duration in 1988. He used familiar rather than unfamiliar stimuli to see whether or not performance on change-detection tasks would be affected by different interval lengths and masking. His work showed similar results to Philips' 1974 study; performance was best at the very shortest retention intervals only without the presence of a mask, which again suggests the presence of sensory storage. He also developed a formula to estimate the number of items in working memory (which we will discuss later) and found that several items could be held at once.

These array studies, while groundbreaking, neglected to look at whether or not memory for different types of features would result in differences in storage capacity. Luck and Vogel (1997) did just that in a series of experiments that tested memory across differing stimulus durations as well as memory for different stimulus features such as color and orientation. In their task, an array of items was followed by another array identical to the first, or with one item changed in some way from the original array. The task was to discern whether or not a change had occurred. In another variation of the procedure, the first array was followed by a second array in which a single item was demarcated, with the possible change occurring to just that one item (limiting the decision to one item). In all of these procedures, they found that capacity remained constant across manipulations. One conclusion was that there is a constant capacity in the number of objects (presumably chunks) in working memory. A further implication was that working memory does not contain separate systems for different characteristics of objects;

instead it seemed to integrate features to create a full picture of the to-be-remembered items. They suggested that capacity limits in working memory may exist to some extent because of this binding together of object features.

With simple visual objects, many studies have verified the finding of Pashler (1988) and Luck and Vogel (1997) suggesting that individuals can remember about 3-5 items in an array (Cowan, 2001). Some recent studies have conflicted with Luck and Vogel's finding that one can retain complex objects with all of their features as easily as one can retain simple objects. This research has shown that to some extent, there is competition between features for storage in working memory (Cowan, Blume, & Saults, 2013; Oberauer & Eichenberger, 2013, Hardman & Cowan, in press). This issue is of limited relevance for the present work, however, in that we will examine memory for simple tones varying only in frequency.

A remaining issue to be considered is that there is still some uncertainty about whether sensory memory ends after several hundred milliseconds, or whether there is a longer phase of more processed sensory memory for several seconds, as has been proposed (Cowan, 1984, 1988). In addition, for example, an analogue to Sperling's (1960) procedure produces an estimate of sensory memory lasting several seconds (Darwin, Turvey, & Crowder, 1972). This is important because an observed working memory capacity estimate could be contaminated by sensory memory.

Adult Working Memory Capacity and Precision

One limitation of the studies of working memory capacity is that the observed capacity may depend on not only the number of items or chunks stored in working memory, but also the precision of the representations. Alvarez and Cavanagh (2004) found that the observed capacity

is smaller using complex objects (e.g., Chinese characters) than it is using simple objects (e.g., colored squares). Awh, Barton, and Vogel (2007) showed that the capacity was the same for simple and complex objects, but that the complex objects often were retained with too low a precision to be used to detect a within-category change. For example, in an array that included a mixture of Chinese characters and cubes of different orientations, an item might be retained in working memory with enough precision to allow detection of a change from a Chinese character to a cube, but not enough precision to allow detection of a change from one Chinese character to a different such character. From these studies it became clear that it would be helpful to have some measure of the precision of representations in working memory.

Measures of precision were revolutionized by Zhang and Luck (2008), who developed a method of examining working memory precision for simple, continuous changes in visual items: changes from one color to another on the color wheel, or changes in the orientation of a form in space. Instead of a change-detection task, an array was followed by a cue indicating which item to recall, and that item was to be recalled by the subject selecting a location on a response wheel (a complete color wheel or a complete 360-degree spatial orientation wheel). We used a variation of this procedure in our study, taking their visual method and translating it to the auditory domain with a non-musical representation of frequency, which is necessarily monotonic rather than circular in its representation.

On every trial in their experiment, an array of several colored squares was to be remembered. Then, one square was singled out for recall, which was to be accomplished by selecting the location on the color wheel that best matched the stimulus square within the array. The basic assumption of their model is that the probability of reporting any given color in response to a stimulus is based on a mixture of two subsets of trials. In one subset, the subject

has the correct color in working memory to some degree of precision (not perfectly). On these trials, the selected color will most often match the correct response, but responses deviating from the correct response will also frequently occur. In another subset of trials, the subject is assumed not to have the correct item in memory, in which case it is assumed that the subject will guess randomly, resulting in a flat response distribution. What is actually observed is a mixture of these two bases of responding. The mathematical combination of these components of responding yields two parameters: P_m , the probability that an item is in memory, and SD, the standard deviation of the distribution of responses on trials in which the item is in working memory. This SD can be viewed as the imprecision of the representation.

There is a controversy in the field regarding the correct model of working memory. The model discussed above is a version of the slot model in which only some of the items in the array are entered into the available slots in working memory. Even Zhang and Luck (2008) had to modify this model by assuming that if the number of working memory slots available exceeded the number of items in the array, the extra capacity would be used to improve the precision of one or more of the representations (see also Anderson, Vogel, & Awh, 2011). Certain other theorists have disagreed with details of the findings and with the theoretical model, and have instead argued that working memory is a fluid resource that can be distributed among any number of items in an array (Bays & Husain, 2009; van den Berg, Awh, & Ma, 2014).

The two theories of visual working memory storage and precision to which we have just referred have been termed the flexible-resource and limited-item theories (Zhang and Luck, 2008). Flexible-resource theories view working memory capacity as fluid and unlimited. In these models, working memory is essentially a vast pool of resource that can be distributed amongst multiple to-be-remembered items. When there are fewer items to remember, the quality

of the memory representations are much more precise; as the number of to-be-remembered items increases, their representations worsen. This theory essentially says that working memory resources can be distributed indefinitely, but that precision declines as the number of items held in memory increases. In contrast to this, limited-item theories view working memory capacity as fixed, with a fixed number of ‘slots’ that can hold items to be remembered.

The debate between theories is ongoing, and has gotten very complex (e.g., van den Berg et al., 2014). The present work may not be able to resolve it, but for reasons we will explain, we favor the limited-item theory, in which the number of slots is limited. In our work we plan on adopting that theory in order to obtain estimates of the number and precision of working memory representations and their change with age. Therefore, our stance will be to ask conditional questions: Does the limited-item theory provide a good account of memory for a tone series and, if so, which parameters of the model change with age? Even if the model turns out to need modification, our study will provide basic new data that will be of use for this new line of inquiry and will address the question of how working memory develops in childhood.

Further support for the limited-items theory. To further demonstrate item limitations, Zhang and Luck (2011) replicated their original 2008 study but varied the amount of precision needed for each task by changing the number of distinct colors that a subject was allowed to select from. Subjects were presented with an array of colored squares and then asked to replicate a specific color from the probe array in either a high-precision task (with 180 different color choices arranged on a wheel) or a low-precision task (with only a small set of distinct spokes arranged on a similar wheel). They found an identical ‘k’ between the two conditions, suggesting that precision is not reduced between conditions – a result which supports the limited-item theory: that regardless of the number of items to be remembered, precision remains the

same. If the resources theory were correct, precision would increase at the cost of the number of items held in working memory. Even when adding motivational factors such as payments, the experimenters still found that precision remained stable across the number of items to be remembered, supporting their original model. This paper, and its predecessor (Zhang and Luck, 2008), provide strong evidence to support the theory that capacity is limited by number of items, not by an unlimited pool of resources.

Continuing along these lines, a study conducted by Anderson et al. (2011) expanded on Zhang and Luck's findings, showing that items in working memory appeared to plateau once a specific item limit was reached. The researchers focused on whether or not WM capacity was limited by number of items or if it could be distributed with decreasing precision amongst multiple items, as posited by flexible-resource models. This study introduced electrophysiological as well as behavioral methods to the current body of research, and found that when item limits were exceeded, working memory resolution appeared to stabilize. When the set number of slots was reached, resources were unable to be allocated to hold additional items in working memory. The brain imaging evidence suggested that the superior intraparietal sulcus (IPS) and the lateral occipital complex (LOC) were able to support and maintain visual details about specific items, while the inferior IPS was able to determine the amount of those items that could be held in memory at any given point in time.

The results of these studies, combined with the results of the two studies previously reviewed by Zhang and Luck, in our view, strongly support the hypothesis that visual working memory precision is determined by the number of items held within a set number of slots; it is not able to be distributed amongst an infinite number of items as the discrete-resource models postulate. There is also reaction time evidence to support this view (Donkin, Nosofsky, Gold, &

Shiffrin, 2013), showing a mixture of two separate pools of trials with reaction time distributions suggesting working memory knowledge and guessing, respectively. It is our hope to translate the results of this visual model to the auditory domain.

Developmental Capacity and Precision in Visual Working Memory

It is well-known that working memory capacity develops with age before reaching its final plateau of around 3-4 chunks of information (Cowan, 2001; Vogel et al., 2001). Children appear to reach this level of visual working memory maturity around the age of eleven (Rigges et al., 2006). However, there is some controversy as to how this development occurs; is it a result of actual capacity growth, or is it more related to the development of better rehearsal strategies. In a traditional change-detection paradigm developed by Luck and Vogel (1997) to study this phenomenon, subjects are presented with arrays of colored squares to remember. On half of the trials, no change is present; on the other half, the color of one of the squares in the array changes. Subjects are asked to indicate whether or not the target array is different than the probe array, typically by pressing a button to indicate “Same” or “Different.” By using this paradigm they found that adults performed best with arrays containing 1-2 squares, slightly worse with arrays containing 3-4 squares, and exhibited a definite decline in performance on arrays containing more than four squares, which supports the previously mentioned work.

Cowan et al. (2005) expanded this work to children by examining capacity differences between 3rd grade children, 5th grade children, and adults. They found that the younger children seemed to only be able to remember up to two items, suggesting that capacity was set at that limit, while the older children were able to remember up to four items. Rigges et al. (2006) continued this line of research by further modifying Luck and Vogel’s change-detection

paradigm to better examine the capacity development of working memory in children. They tested 5, 7, and 10-year olds to see how their performance differed. What they found was that the older children were able to more accurately perform the change-detection tasks at the larger array sizes, which again supports the work previously discussed. Their results seem to point towards an actual increase in capacity limit rather than to a development of rehearsal strategies over time, but it is possible that as children age they both gain greater capacity limits and begin to utilize better rehearsal strategies. Overall, the current work in the field suggests that visual working memory task performance improves as one ages.

Thus far, the majority of developmental visual working memory research has primarily focused on capacity. However, emerging studies have begun to look at the development of visual working memory precision. As mentioned previously, there are two major competing theories on visual working memory precision – again, the limited item theories and the flexible resource theories. In a recent study, the task that Zhang and Luck (2008) used on adults to test visual working memory precision was modified to test children (Burnett Heyes, Zokaei, van der Staaij, Bays, & Husain, 2012). Burnett Heyes et al. used a sample of boys ranging in age from 7-13 years old in order to gain an accurate picture of precision changes over time. What they found was that precision did indeed significantly increase with age. Because it did not plateau amongst the older children, they suggest that precision may continue to develop as an individual ages. Burnett Heyes et al. adopted a resource theory and cannot speak to the question of whether capacity also changes with age, inasmuch as there is no capacity limit in the resource theory. The results of this study on precision development, as well as the results of those conducted on capacity development strongly suggest that visual working memory is dynamic and develops

across an individual's lifespan. This thesis sought to examine if this work in the visual domain can be translated to the auditory domain.

Auditory Working Memory Capacity and Precision

Thus far, the bulk of the research on working memory precision and capacity has been limited to the visual domain. In 2012 Li, Cowan, and Saults attempted to expand this work to the auditory modality. Little work has been done in this area, although the general consensus has been that auditory working memory is more limited in capacity and that it is less precise.

In a set of three experiments with adults, Li et al. utilized a change-detection paradigm in which subjects were presented with sets of 2, 3, 4, 5, or 6 tones to be remembered. Multiple manipulations were used across the three experiments, including the use of articulatory suppression. They found that working memory did not reach a stable plateau across conditions as in visual working memory tasks. More capable subjects appeared to be able to remember a maximum of three tones, which is significantly less than performance on visual working memory trials, where the mean number of remembered items for all subjects is around three. One reason for these results might be related to the need for highly dissimilar tones. Their results suggested that in order to accurately remember nonverbal auditory stimuli, subjects require both categorical and detailed information; when timbre was added to the tone-memory tasks, subjects were able to perform better (although not as well as they do for visual stimuli). Although this was a groundbreaking article in the field of auditory working memory, it only examined capacity limits; it did not look at precision or development.

Supplementing this research, Kumar et al. (2012) examined auditory working memory precision not by using a change-detection task but by using a pitch-matching paradigm in which

subjects were required to recreate a previously presented tone by continuously adjusting its pitch. These authors claimed that they did not find support for the limited-items model but rather for the flexible -resource model of working memory. It is not clear why their results would not be compatible also with the limited-items model, but at any rate, there is a precedent indicating that tone memory can be successfully studied with the stimulus reproduction task of Luck and Vogel (2008). Because the authors did not fit Luck and Vogel's capacity model to their data, it is not possible to determine if Kumar et al.'s findings correspond to developmental changes in capacity or precision.

Development of Auditory Working Memory

Very few studies of auditory working memory have examined its development, instead focusing on either its capacity or its precision. Keller and Cowan (1994) investigated the development of auditory working memory capacity. They used tonal rather than verbal stimuli, making their work relevant to the present study, and asked subjects in three different age groups (6-7, 10-12, and adults) to indicate changes in a traditional change-detection method of saying whether or not probed items were the same or different as what had been presented earlier. They found a difference in precision when the inter-tone interval was held constant, and they also found that the persistence of memory for tonal information changed with age, with adults maintaining memories significantly longer than children.

The Present Study

The present study attempted to combine the previously discussed work on auditory working memory capacity and precision from a developmental perspective. As of now, no studies have looked at both capacity and precision in the auditory domain for either children or

adults, and none have looked at capacity or precision development separately or in the same experiment. This work further examined the flexible versus fixed resource models of working memory to see which model best applies to the auditory domain, and assessed the capacity limits of auditory working memory.

By utilizing varying set sizes, the study examined how performance varied across different numbers of to-be-remembered items. The subjects heard sets of 1, 2, 3, or 4 tones to remember, which allowed an estimate of how many tones they could hold in working memory, as well as the precision with which they recalled those tones. It was expected that they would be better able to remember tones in the 1- and 2- tone conditions than in the 3- and 4- tone conditions, with performance declining as the number of stimuli increased. This declining performance was expected to indicate the capacity limits of auditory working memory; it was predicted that capacity would be more limited in the auditory domain than the visual with expected capacity limits of around 2-3 items.

Because the present study used a tone-reproduction task instead of a change-detection task, we were able to measure the precision of auditory working memory. Subjects were asked to reproduce one of the tones that they heard during the presentation time by adjusting a slider until they created a tone that matched their memory of the probed tone. This procedure allowed us to compare the tone generated by the subjects to the actual frequency of the probed tone in order to evaluate the precision of the representation they held in working memory during the task.

The methods utilized in this study allowed for an examination of both the capacity and precision of auditory working memory. Additionally, we investigated different age groups of

children as well as adults, which will allowed us to not only to assess precision and capacity together but also to study its development.

Methods

Subjects

One hundred and twenty-five subjects were recruited for this study using both the University of Missouri's online announcement system advertising campus opportunities and the pool of subjects that had participated in previous studies for the lab. Subjects were compensated \$15/hour for their participation. These subjects were divided into four groups: Group One, was composed of 31 first and second graders (16 female, mean age = 7.47 years, SD = 0.68); Group Two, was composed of 32 third and fourth graders (15 female, mean age = 9.07 years, SD = 0.75); Group Three, was composed of 30 individuals in grades five, six, and seven (19 female, mean age = 11.3 years, SD = 0.95); and Group Four, was composed of 32 adults (25 female, mean age = 39.74 years, SD = 6.12) (primarily the parents of the children who were participating in the study). Five Group One subjects – two male and three female – were dropped from the final analyses due to computer error. Two male Group Two subjects were dropped from the final analyses due to a failure to complete all required tasks; an additional two female subjects were dropped from Group Four due to experimenter error. A total of 115 subjects were included in the final analyses – 25 from Group One, 29 from Group Two, 30 from Group Three, and 31 from Group Four.

Procedure

Recruitment and screening. Before participating in this study, subjects went through a brief screening process during which they were asked a standard set of questions about things such as first language, vision issues, hearing issues, and learning issues – as well as a set of questions about any previous musical training that they may have had. There is quite a bit of

evidence suggesting that individuals with musical training perform better on auditory working memory tasks than those without it (Franklin et al., 2008; Benassi-Werke et al., 2012; Williamson et al., 2010), which allowed us to compare the individuals with musical training to those without.

Fluid intelligence measure. Subjects were first asked to complete the Raven's Progressive Matrices task, which is a nonverbal test suitable for individuals from the age of five and up. It is a standardized test of intelligence and allowed us to examine whether or not intelligence is related to auditory working memory capacity and precision. The test contains five sets of twelve-item "puzzles" that progressively increase in difficulty as the subject proceeds through them. Tests are scored on a scale of 1-60, and the scores are interpreted based on the subject's age. Upon completion of the Raven task, subjects then were asked to move into a sound-attenuated booth for the duration of the study.

Tone memory task. The remainder of the study involved computer-based programs created in E-prime and using auditory stimuli generated using Praat. Eighty sine-wave tones were generated, beginning with a low tone frequency of 150 Hz and with each successive tone 1.045 times the frequency of the previous tone. The tones were adjusted to have an amplitude of 71+/- 5 dB when measured with a sound level meter (see Sound Level Readings chart, attached). Each

tone lasted 1000 ms, with 200 ms on- and offset ramps. Subjects heard all stimuli using Audiotechnica M50 headphones.

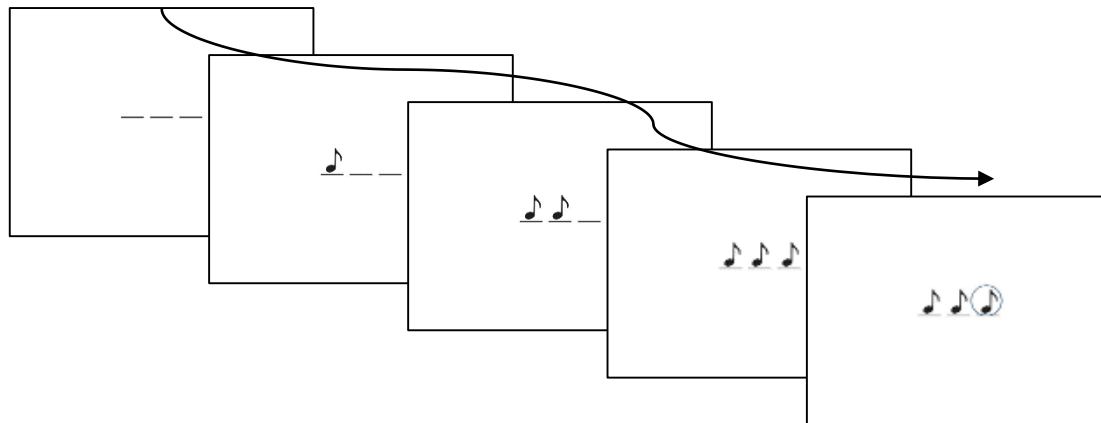


Figure 1. Illustration of the visual part of the stimulus display during tone presentation. The final panel shows the encircled item that is to be reproduced by the subject; it can be at any serial position.

Figure 1 illustrates the presentation of the tonal stimuli. Subjects were shown one, two, three, or four blank lines for 1000 ms; the number of lines shown corresponded with the number of tones played in the sequence to be remembered. A musical eighth note filled in each blank slot as the tone was played. Each tone lasted for 600 ms; on trials with more than one tone, there was a 400 ms pause between each tone. After the final tone played there was a 1000 ms retention interval, after which a blue circle surrounded the music note representing the probe tone for 1000 ms, indicating the serial position required for the reproduction task. That is, subjects were instructed to reproduce the circled tone from memory. The tones for each serial position in a sequence were randomly selected from the eighty possible tones without replacement on that trial, and the sequence to be remembered on a trial never contained duplicate tones.

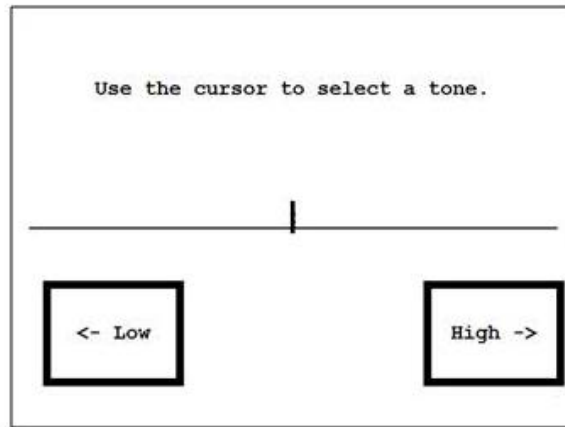


Figure 2. Illustration of the response display on a trial.

Subjects then performed a pitch-matching technique by dragging the slider (pictured above in Figure 2) to recreate the probed tone. There were 80 tone choices on the slider. Subjects had five attempts or could press enter if satisfied that the tone that they produced matched the one that they were attempting to remember. Subjects did not receive feedback on these trials. The experiment was randomized with ten practice trials (one of each possible trial condition) and forty-eight experimental trials so that each set size was presented 12 times, with an equal number of tests at each serial position. This task was performed twice per subject for a total number of ninety-six trials per person. The second run-through of the tone task occurred after the subjects completed the Counting Span Task, described below.

Counting Span Task. Upon completion of the first run of the main task, subjects were then instructed to complete a classic measure of working memory capacity: counting span. In the counting span task, subjects were asked to orally count the number of dark blue circles presented on a screen filled with an array of dark and light blue circles and squares. Subjects began with two screens to count and continued up to eight screens based on performance with 2-6 circles per screen to be remembered. This supplementary task aids in the identification of individuals with high vs low working memory spans in order to see if their performance is influenced by their

span scores. Complex span tasks have been used since the 1980's to effectively measure general fluid intelligence as well as working memory span (Daneman & Carpenter, 1980; Case et al., 1982; Cowan et al., 2005; Broadway & Engle, 2010).

Our Model

In order to allow us to extract parameters of performance in each age group, we adapted the Zhang and Luck model to fit our study. Because their model relied on a circular distribution of stimuli, and ours used a flat distribution, their model does not fit our study exactly. We developed a new set of parameters that included two guessing distributions: central guessing (as in Zhang and Luck) and uniform guessing; capacity (K), precision, and attention. The following section describes the guessing strategies and then discusses our model in further detail, in both a theoretical and a mathematical manner.

Guessing Strategies

There are two types of guessing strategies that we hypothesize were used by the subjects that participated in this study: central and uniform. Central guessing might be considered the ideal strategy when people have no information whatsoever about the tone that they are seeking to identify, as it minimizes the mean distance between their guess and the actual stimulus, given that all stimuli are equally likely in this procedure. In central guessing, responses are distributed in a band around the middle of the response scale.

The second guessing strategy is uniform guessing, where a subject seems to guess at random – they are just as likely to guess in the center as they are to guess anywhere else, creating a relatively uniform distribution of answers throughout their data. Uniform guessing may be considered an example of probability matching, a strategy that does not minimize the mean difference but does reproduce the distribution of stimuli in the response distribution (Vulkan, 2000).

Modelling

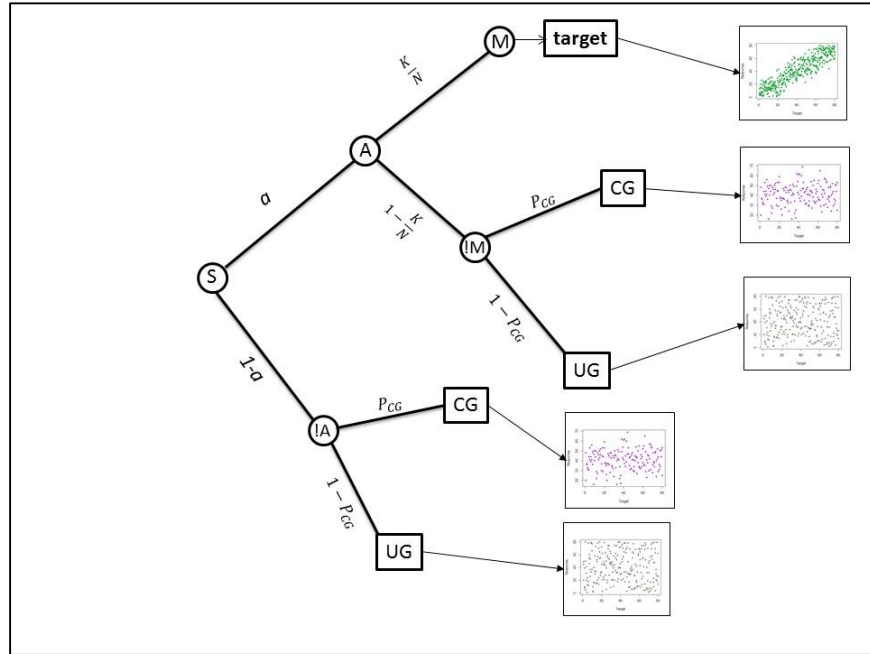


Figure 3. Illustration of the multinomial process tree depicting our theoretical model. Each outcome is followed by a panel depicting the pattern of results based on that outcome. See the text for a further explanation.

Our model is a multinomial process model, in which binary paths occur in a branching structure as shown in Figure 3. The first binary decision in this tree, extending from node ‘S,’ represents whether or not a subject is paying attention on the trial, as depicted by the parameter a . If a subject is not paying attention, he or she will follow along the branch represented by $1-a$, which represents that the subject is not, in fact, attending to the presentation, and as a result must guess. This attention parameter is a necessary part of the model to account for any factor that can prevent the entry of the studied items into working memory regardless of the set size.

The subjects who are attending on a given trial proceed to node “A”, where the tree branches depending on whether the target tone is in memory. It is in memory with probability $\frac{K}{N}$, where K is the number of items in working memory (with a maximum of 4) and N is the number of tones in the sequence. If so, the tone is correctly recalled with a certain imprecision, so that

the responses fall on a truncated normal distribution with a mean located at the probed tone's assigned number and an SD of σ .

If the subject was paying attention but failed to encode the target stimuli in memory, they follow the line to node “!M” (represented by $1 - \frac{K}{N}$, which divides into the two types of guessing strategies – Central and Uniform). The probability that a subject will use one of these strategies is represented by P_{CG} for Central Guessing and $1 - P_{CG}$ for Uniform Guessing.

The subjects who were not attending to the presented stimuli follow the line to node “!A”, which occurs with probability $1 - a$. From there, they split in the same manner as those from node “!M”, and proceed to utilize either a central or uniform guessing strategy with probabilities P_{CG} versus $1 - P_{CG}$.

Our Model vs. the Zhang and Luck Model

The Zhang and Luck model, depicted in Figure 1, relies on a circular distribution of possible responses. Because our model used auditory stimuli, the range of responses is represented by a flat line. One consequence of this difference in the stimulus representation is that Zhang and Luck's model only needs uniform guessing, whereas our model includes two guessing strategies – central and uniform. (With a circular representation, there is of course no central point.) Figure 5 depicts our model in a manner similar to the figure created by Zhang and Luck.

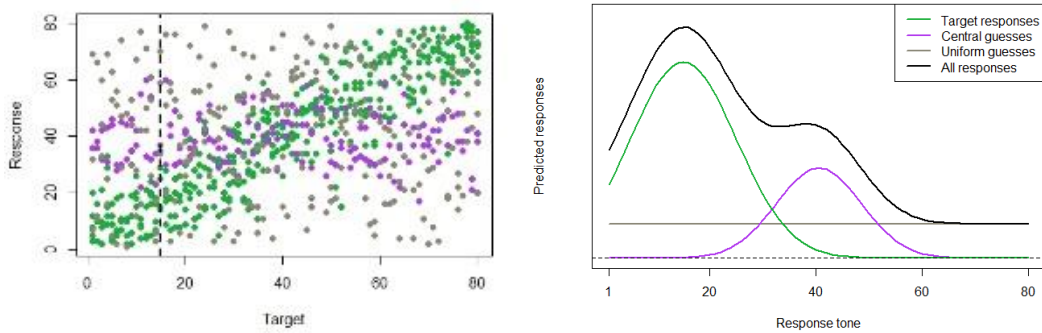


Figure 4. Scatterplot (left) and vertical slice (right) of our model using hypothetical data. The vertical slice comes from the points to the left of the dashed line in the scatterplot. In both panels, green represents the hypothetical target response distribution based on information in working memory, purple represents the hypothetical central guessing distribution, and grey represents the hypothetical uniform guessing distribution. The black line in the vertical slice shows the sum of all responses.

In the left-hand panel of Figure 4, the scatterplot shows the possible response distributions, with green representing responses based on target knowledge, purple representing responses coming from the central guessing strategy, and grey representing responses coming from the uniform guessing strategy. The right half of the figure is a slice of that scatterplot for all target frequencies represented to the left of the dashed vertical line in the scatterplot. The representation in the right-hand panel is rotated, with the response tone represented by the vertical axis in the left-hand panel but by the horizontal axis in the right-hand panel. In the right-hand panel, on the vertical axis is the number of data points at any one response tone frequency. The uniform guessing strategy, as in the Zhang and Luck model, is represented by the grey horizontal line indicating equal guesses at each tone frequency. The central guessing strategy, shown in purple, reflects a normal distribution centered at the middle two tones on the spectrum (40-41). The target distribution, shown in green, is represented by a curve in our model reflecting a normal curve centered on the correct target frequency and truncated at the ends of the response spectrum. (It is this truncation process that would not be necessary if the tones formed a circular pattern like visual stimuli, but this is impossible with tones.) Finally, the black irregular curve

reflects the summation of responses from all three strategies and is the form that the data presumably take. The modeling effort therefore serves to disentangle the contributions of three strategies to performance.

Mathematical Description of Our Model

The k th response at the j th set size for the i th participant will be denoted y_{ijk} and the studied tone to which they were responding will be denoted x_{ijk} . Additionally, we will denote the variability of the WM representation of remembered tones for the i th participant and the j th set size as σ_{ij} . The variability of the use of a central guessing strategy will be denoted δ_i . Given these definitions, we can define the likelihood functions for the three different types of responding as follows

$$L_T = \varphi|_1^{80}(y_{ijk}|x_{ijk}, \sigma_{ij})$$

$$L_{CG} = \varphi|_1^{80}(y_{ijk}|\bar{y}, \delta_i)$$

$$L_{UG} = Unif|_1^{80}(y_{ijk})$$

where, L_T is the likelihood that a response was from the target distribution, L_{CG} is the likelihood that the response was from the central guessing distribution, and L_{UG} is the likelihood that the response was from the uniform guessing distribution. Furthermore, $\varphi|_1^{80}(y|\mu, \sigma)$ is the probability density function of a truncated normal distribution for some quantile y given a mean μ and standard deviation σ , with the distribution truncated below at 1 and above at 80 (1 and 80 being the lowest and highest tones, respectively). For the target distribution, μ is equal to the studied tone, x_{ijk} , and for the central guessing distribution, μ is equal to the mean of the possible response values, $\bar{y} = 40.5$. The standard deviations of the target and central guessing distributions

are σ_{ij} and δ_i , respectively. Finally, $Unif|_1^{80}(y_i)$ is the probability density function of a uniform distribution with minimum value 1 and maximum value 80.

Given the definitions of the likelihoods of the two kinds of guessing, we can say that

$$L_G = \rho_i * L_{CG} + (1 - \rho_i) * L_{UG}$$

where L_G is the likelihood for a response given that that the response is one of the two kinds of guessing and ρ_i is the probability that, if a participant is guessing, they make a central guess.

Using the preceding definitions and the process tree given in Figure 4, we can write the full likelihood function for a single response as

$$L(y_{ijk}|x_{ijk}, \theta) = a_i * \left[\min\left(\frac{K_i}{N_{ijk}}, 1\right) * L_T + \left(1 - \min\left(\frac{K_i}{N_{ijk}}, 1\right)\right) * L_G \right] + (1 - a_i) * L_G \quad (1)$$

where a_i is the probability that the participant was paying attention on the trial, K_i is the number of tones in WM, N_{ijk} is the set size for the trial (i.e. $N_{ijk} = j$), and θ represents all of the parameters of the model. It can be shown with basic algebra that Equation 1 reduces to

$$L(y_{ijk}|x_{ijk}, \theta) = PM * L_T + (1 - PM) * L_G$$

where PM is the probability that the target tone will be in WM and is given by $PM = a_i * \min(K_i/N_{ijk}, 1)$. The subscripts of the parameters indicate the fact that each participant has a single a , K , δ , and ρ that apply to all set sizes and that each participant has one σ per set size.

Results

The results of the tone task procedure were examined in several ways. First, we present the results in an atheoretical form, as the mean absolute deviation of the response from the stimulus value, averaged across stimuli for each set size. Next we present our final selection of the model based on the data and results of the model in terms of parameter values that can change across age groups. Finally, we present the results of the Raven's Progressive Matrices and Counting Span tasks and their relation to parameters of tone task performance.

Tone Task

Absolute deviation scores. Table 1 shows the mean and standard deviation of the absolute discrepancy between the stimulus tone number and the response tone number for each set size and age group, averaged across stimulus number. For example, if the stimulus was 43, the score of 2 would be assigned to responses of either 41 or 45. Performance improves with age group (indicated by decreases in mean discrepancy across age groups) and it declines with set size (indicated by increases in discrepancy at higher set sizes). Note, though, that changes in either capacity or precision could account for these age group and set size effects, which will therefore be considered further only in terms of the theoretical modelling parameters.

Table 1

Mean absolute discrepancy for each group at each set size

Age Group	Absolute Difference Set Size 1	Standard Deviation, Set Size 1	Absolute Difference, Set Size 2	Standard Deviation, Set Size 2	Absolute Difference, Set Size 3	Standard Deviation, Set Size 3	Absolute Difference, Set Size 4	Standard Deviation, Set Size 4
One (n=25)	12.20	7.76	15.58	7.81	17.78	6.71	19.15	8.13
Two (n=29)	6.23	3.00	10.36	4.56	12.58	3.82	14.39	4.04
Three (n=30)	4.25	1.92	7.10	2.66	9.95	3.99	10.84	4.94
Four (n=32)	4.14	2.08	5.21	2.16	7.24	2.86	10.00	3.27

Final model selection. We tested whether we could simplify our model by removing one of the types of guessing from the model. If we could remove a type of guessing and still fit the data adequately well, our model would be more parsimonious as a result. The test was done using a nested model comparison in which the full model, with both types of guessing, was compared to two reduced models, each of which had only one type of guessing in the model. Relative to the full model with both types of guessing, the reduced model with only central guessing was rejected, $\Delta(116) = 285$, $p < .001$, as was the reduced model with only uniform guessing, $\Delta(116) = 1287$, $p < .001$ (see Huelsenbeck and Crandall, 1997, for more information on nested model comparison).

Full dataset. The raw data form the scatterplot matrix shown in Figure 6, which allows us to see the accuracy of each group as it changes across set size. Each point is a single trial's result. For each of 16 square panels in the figure, the X axis reflects the stimulus tone number and the Y axis reflects the response tone number. The diagonal line formed in each square panel of the matrix thus represents the most accurate responses possible. The blurriness as the responses deviate from the main diagonal represents imprecise responses or guesses as subjects'

accuracy decreases. There is a systematic degradation in responding as set sizes increase (from the left-hand column of panels progressing to the right) and a systematic improvement in responding for any one set size as age group increases (from the top row of panels progressing to the bottom). In the extreme, for example, the differences between Group One and Group Four are very clear; the youngest children perform fairly well with only 1 tone (top left panel of the figure) but by the time they reach the highest set size (top right panel), they are barely able to respond accurately. Groups Three and Four (third and fourth rows of panels) maintain a very strong pattern of accuracy even at the highest set size (right column of panels).

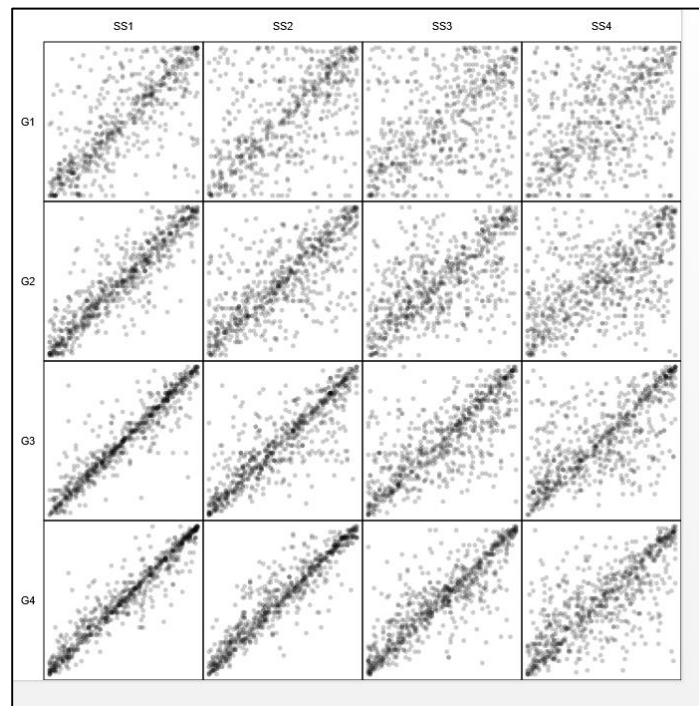


Figure 5. Scatterplot matrix of all data shown by set size (rows) and age group (columns). The individual scatterplots represent the target tones (X-axes) vs the subjects' final responses (Y-axes) within each block

To illustrate the potential influences on responding, Figure 6 depicts raw data from three individual subjects with extremely different patterns of responding. The first scatterplot shows the data from a subject that used a uniform, or random guessing strategy. The points represent

the subjects' last response made on a trial (x axis) vs the target tone (y axis). When a subject uses a random guessing strategy, their data shows no discernable pattern, as one can see (left-hand panel). The central plot shows the responses of someone with extremely high accuracy, who likely did not need to guess often, because he or she was able to successfully hold the target tone in working memory. Finally, the right-hand scatterplot represents a subject who often was correct but often used a central guessing strategy; a guessing band is formed directly across the middle of their raw data, showing the tendency to guess towards the middle. Among other subjects, one can see other mixtures of these extreme patterns.

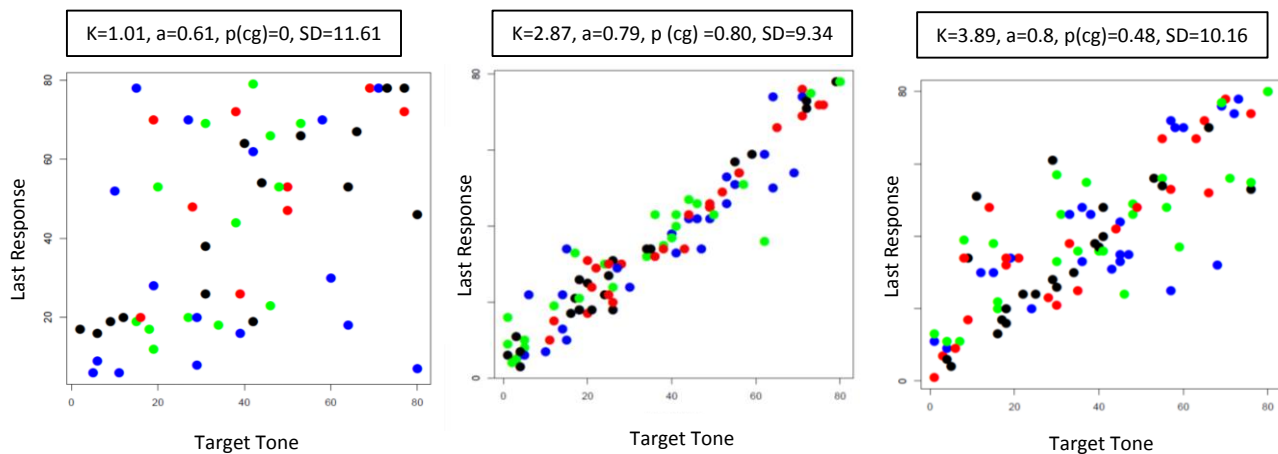


Figure 6. Individual subject data showing three different response strategies. The first plot shows uniform guessing, the second shows accurate responding, and the third shows central guessing. Graphs show target tone vs last response, with the colors representing each set size (black for SS1, red for SS2, green for SS3, and blue for SS4). The numbers at the top of each plot show parameter estimates for that individual subject.

Suitability of the model. Figure 7 shows the model predictions using randomly generated data. This figure was designed by sampling the data from each subjects' target and guessing distributions, as given by their parameter estimates, so that the number of data points would match those in Figure 5. Individual data points in Figure 7 are not identical to the data points in the actual data, given that the model includes random noise, but the overall patterns are strikingly similar. Our model thus does indeed seem to be an excellent fit to the data. We do not

provide fit statistics because our aim is not to compare different models (although some variants were initially ruled out as explained above); we wish to use this sufficiently good model to compare parameters in an effort to determine the nature of developmental changes in performance.

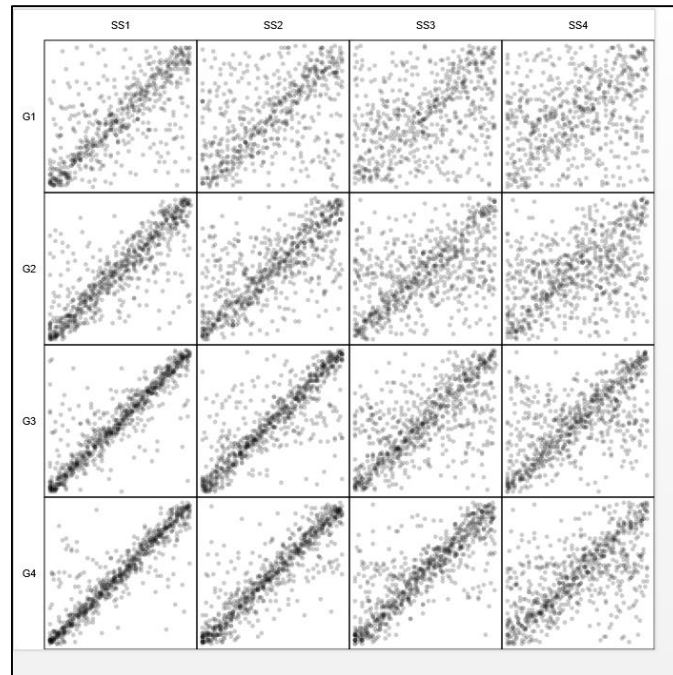


Figure 7. Predicted results of our model. Scatterplot matrix of predicted data shown by set size (rows) and age group (columns). The individual scatterplots represent the target tones (X-axes) vs predicted final responses (Y-axes). The pattern is strikingly similar to the results shown in Figure 6.

Age Group appeared to be a clear predictor of working memory capacity, imprecision (target variability), and the probability of a subject utilizing a central guessing strategy. Table 2 (below) shows the means and standard deviations of each age group for each parameter. Figure 8 shows these same results depicted graphically.

Table 2

Means and standard deviations of each parameter for each age group

Age Group	Mean Capacity (K)	Standard Deviation	Mean Target Variability	Standard Deviation	Mean Attention	Standard Deviation	Mean p(central guessing)	Standard Deviation
One (n=25)	2.13	1.22	12.03	5.27	0.82	0.24	0.48	0.46
Two (n=29)	2.18	0.97	9.15	3.30	0.90	0.11	0.78	0.38
Three (n=30)	3.03	0.95	6.51	3.33	0.87	0.10	0.95	0.17
Four (n=32)	3.07	0.73	6.03	2.36	0.91	0.08	0.77	0.34

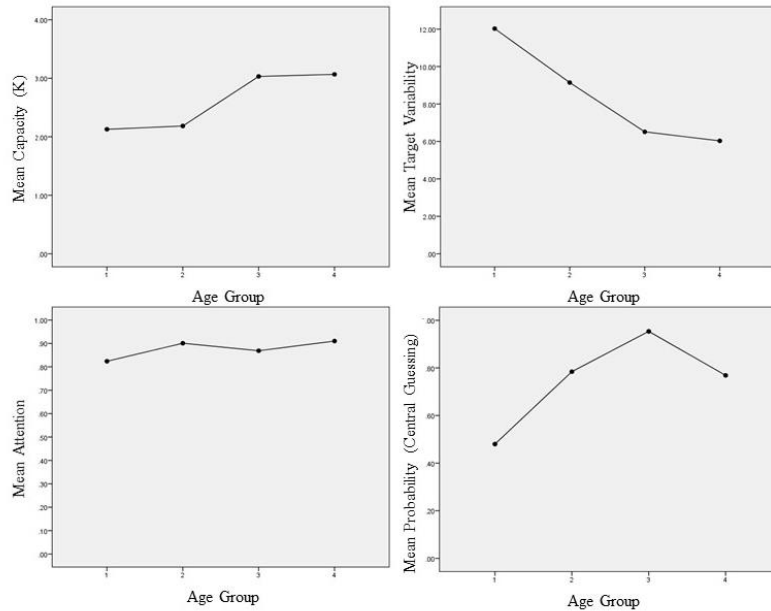


Figure 8. Graphs of Mean Capacity (K), Imprecision (target variability), Attention, and Probability (Central Guessing) for each Age Group (x axis).

A one-way MANOVA was calculated to examine the effects of age group on the different parameters of the model. The results of this MANOVA showed a significant multivariate effect for Age Group, Wilke’s $\lambda = 0.524$, $F(12, 286.03) = 6.60$, $p < .001$, partial eta squared = 0.194. The observed statistical power to test this effect was equal to 1, which suggests that it was indeed a very strong main effect. Upon conclusion that the multivariate main effect

was highly significant, the univariate main effects were subsequently examined. Significant univariate main effects were found for three of the four tested parameters. These were K ($F(3, 111) = 8.09, p < .001, \text{partial eta squared} = 0.179, \text{power} = 0.99$), imprecision ($F(3, 111) = 15.78, p < .001, \text{partial eta squared} = 0.30, \text{power} = 1.0$), and the probability of utilizing a central guessing strategy ($F(3, 111) = 8.57, p < .001, \text{partial eta squared} = 0.19, \text{power} = 0.93$). The attention parameter was not found to be significant.

Raven Progressive Matrices

The Raven scores were strongly related to age group ($F(3, 111) = 42.44, p < .001$). The mean Raven scores were 26.84 (SD=10.35) for Group One, 33.83 (SD=8.13) for Group Two, 41 (SD=7.8) for Group Three, and 49.15 (SD=4.29) for Group Four.

Counting Span

The scores for the counting span task were also strongly related to age group ($F(3, 111) = 15.42, p < .001$). The mean scores for the counting span task were as follows: 1.76 (SD=1.2) for Group One, 2.72 (SD=1.39) for Group Two, 3.87 (SD=1.80) for Group Three, and 4.29 (SD=1.62) for Group Four.

Correlations Between Tasks and Regression Analyses

Because Age Group was the strongest predictor of performance (see the top half of Table 3), Group was partialled out in order to examine any other correlation amongst the variables (shown in the bottom half of Table 3). As one can see in the table, when controlling for Age Group there are other factors that still correlate with capacity and precision. The Raven score was correlated with capacity (.19, $p < .05$) and musical training was correlated with precision (-

.29, $p < .001$). Precision was also correlated with subjects' Raven scores ($-.34$, $p < .001$) and capacity ($-.19$, $p < .05$). Attention was correlated with subjects' Raven scores (0.24 , $p < .05$) and capacity (-0.19 , $p < .05$). Another correlation was found between capacity and counting span (0.22 , $p < .05$).

Table 3. Raw Correlations (top right) and Partial Correlations controlling for Age Group (bottom left)

Control Variables		Raven	Musical Training	Counting Span	Attention	Capacity	p(CG)	Average Variability
Raven	Correlation	1.00	.34	.44	.40	.28	.19	-.58
	Significance (2-tailed)		.00	.00	.00	.00	.05	.00
Musical Training	Correlation	.07	1.00	.39	.20	.02	.18	-.44
	Significance (2-tailed)	.46		.00	.03	.83	.05	.00
Counting Span	Correlation	.08	.22	1.00	.38	.19	-.01	-.40
	Significance (2-tailed)	.38	.02		.00	.05	.90	.00
Attention	Correlation	.24	-.05	.11	1.00	-.11	.18	-.35
	Significance (2-tailed)	.01	.57	.22		.25	.05	.00
Capacity	Correlation	.19	.05	.22	-.19	1.00	.06	-.03
	Significance (2-tailed)	.04	.62	.02	.04		.52	.79
Probability of Central Guessing	Correlation	-.03	.08	-.20	.01	.08	1.00	-.30
	Significance (2-tailed)	.78	.41	.03	.88	.39		.00
Average Target Variability	Correlation	-.34	-.29	-.16	.07	-.19	-.18	1.00
	Significance (2-tailed)	.00	.00	.09	.43	.04	.05	

Although a negative correlation between K and attention was present after controlling for age group, this relationship was likely spurious. It is indicative of uncertainty in the model as it applies to individuals; there was not enough data per subject to produce entirely stable results. To some extent, K and attention trade off with one another. Nevertheless, the rest of our findings were strong enough to make statements about the different age groups and the relationships amongst the variables within each group.

After examining the correlations and partial correlations, it was determined that the use of regression analyses would be useful to help us better understand our key parameters – capacity (K) and target variability. The regressions were performed to determine how much variance in

each parameter could be explained by the factors that it correlates with, and also to examine which of those factors significantly contributed to the parameters aside from age group.

With K as the dependent variable, a stepwise regression was run to examine its relationship to both Raven score and Counting Span score. Both factors were found to be significant contributors to K's variance ($F(2,112) = 15.08, p < .001, R^2_{\text{change}} = 0.212$). However, subjects' Raven scores were slightly more related to the variance of K ($F(1,113) = 21.77, p < .001, R^2_{\text{change}} = 0.16$) than their Counting Span scores ($F(1,113) = 18.86, p < .001, R^2_{\text{change}} = 0.14$).

A second stepwise regression was conducted with Target Variability as the dependent variable, examining the extent to which Raven Score, capacity, and musical training could explain its variance. All three factors were found to contribute significantly to the variance of this parameter ($F(3,111) = 26.45, p < .001, R^2_{\text{change}} = 0.42$). Considered pairwise, Raven scores and musical training were the strongest contributors and could account for most of the shared variance, ($F(2,112) = 38.06, p < .001, R^2_{\text{change}} = 0.41$). On the individual level, each of these factors was found to be significant. Raven was the strongest contributor to the variance of Target Variability ($F(1,113) = 57.67, p < .001, R^2_{\text{change}} = 0.34$). Musical training was the second strongest ($F(1,113) = 27.48, p < .001, R^2_{\text{change}} = 0.20$). Capacity, while still highly significant, was the least strong of the three factors ($F(1,113) = 15.90, p < .001, R^2_{\text{change}} = 0.12$).

Discussion

This study was designed to examine precision and capacity development in auditory working memory by utilizing a reproduction task, different set sizes, and different age groups. The reproduction task was implemented rather than a change-detection task in order to allow us to estimate the precision with which subjects could remember the tonal stimuli. Differing set sizes – 1, 2, 3, or 4 tones to remember – were used so that capacity estimates could be calculated. The four age groups included children in grades 1 and 2 (Group One), children in grades 3 and 4 (Group Two), children in grades 5-7 (Group Three), and adults (Group Four).

The analyses showed that both target variability and working memory capacity changed significantly with age group. This finding supported our initial hypothesis – that with age, subjects were able to recall more tones with better precision. The changes in variability were most pronounced from Group One to Group Two, and from Group Two to Group Three. However, Group Four was still less variable than Group Three. This suggests that the development of precision greatly slows once individuals reach a certain age – around 10-12 years old. The capacity changes followed a somewhat different pattern - although Group One performed the worst and Group Four the best, performance by Groups One and Two were extremely close, followed by a sharp increase in performance between Groups Two and Three, and then a leveling out again between Groups Three and Four. The difference in the patterns of precision and capacity development are very interesting; the increase in capacity between Groups Two and Three could be due to some developmental milestone that the older children have just reached.

The probability of a subject using a central guessing strategy was also found to change significantly with age. The probability of a subject utilizing a central guessing strategy increased with age for the three groups of children, which makes sense as it is the ideal guessing strategy to minimize the absolute discrepancy between the presented tone number and the response tone number is to guess around a middle value; it stands to reason that the older an individual is, the more likely he or she is to use a more optimal strategy. The deviation in this pattern is that Group Three subjects were more likely to use central guessing than Group Four; however, this may be explained by the Group Four subjects' higher capacity and variability estimates – perhaps they were simply less likely to need to rely on guessing than their younger counterparts and thus had little need for central guessing.

Attention was not found to change significantly with age group, and cannot be used to explain the development in capacity (e.g. Cowan et al., 2010, Cowan et al., 2011). In the case of attention, the youngest subjects were again the worst, but the sharp increase in attention found between Groups One and Two slowed significantly between Groups Two and Three, and between Group Three and Group Four. This may help to explain why the Group Three subjects were so close to those in Group Four in capacity, but fell further short in estimates of their precision.

Our estimates for working memory capacity (K) for tones were much higher than the estimates produced by previous work (e.g. Li et al., 2013). However, we used a different method; Li and colleagues used a change-detection paradigm, while we used a reproduction task. The estimates that we found were comparable to previous working memory studies on the development of capacity in other domains. Because of the nature of Li et al.'s study, they had no means of separating K and variability, which may have caused high variability representations to

lead to incorrect responses. Another possibility is that the probe tone used in Li et al.'s study may have interfered with their subjects' memory representations of the to-be-remembered tones. In sum, this evidence suggests that tone memory relies on some of the same mechanisms as memory in other modalities, which would support an amodal model of working memory (Chein et al., 2011; Cowan et al., 2011, Cowan et al., in press; Li et al., 2014, Majerus et al., 2014).

Two other factors were found to be significantly related to performance on the tone task – musical training and Raven's score. At the beginning of the study, data on musical training was recorded for each subject. We separated the subjects by level of musical training – they were coded as 1 if they had less than one year of musical experience outside of any school requirements, 2 if they had 1-2 years of training, 3 if they had 3-4 years of training, and 4 if they had 5 or more years of musical training. It was found that the level of musical training that subjects had correlated significantly with their target variability – that is, the more training they had, the more precise they were in their responses. This supports previous work that suggests that musical training leads to enhanced auditory working memory skills (Franklin et al., 2008; Benassi-Werke et al., 2012; Williamson et al., 2010). It is equally notable, though, that musical training did not influence capacity. This suggests that capacity may be more biologically determined and precision or variability more open to environmental influences.

While musical training was specific to precision, capability as measured by Raven scores was not – they correlated both with capacity and variability. Those with a higher Raven score were found to have less variability, and they were also found to have a greater working memory capacity. This again supports previous work that suggests that intelligence is related to enhanced working memory performance (Cowan et al., 2005; Cowan et al., 2006).

The results of this study suggest that working memory precision and capacity improve with age. We also found increases in the probability of utilizing the ideal guessing strategy for this task. In the future, the results of this study could be used to further examine the relationship between musical training and auditory working memory.

References

- Alvarez, G.A., & Cavanagh, P. (2004). The capacity of visual short term memory is set both by visual information load and by number of objects. *Psychological Science, 15*, 106-111.
- Anderson, D. E., Vogel, E. K., & Awh, E. (2011). Precision in visual working memory reaches a stable plateau when individual item limits are exceeded. *The Journal of Neuroscience, 31*, 1128-1138.
- Awh, E., Barton, B., & Vogel, E.K. (2007). Visual working memory represents a fixed number of items regardless of complexity. *Psychological Science, 18*, 622-628.
- Bays, P.M., & Husain, M. (2008). Dynamic shifts of limited working memory resources in human vision. *Science, 321*, 851-854.
- Broadway, J.M., & Engle, R.W. (2010). Validating running memory span: Measurement of working memory capacity and links with fluid intelligence. *Behavior Research Methods, 42*, 563-570.
- Burnett Heyes, S., Zokaei, N., van der Staaij, I., Bays, P. M., & Husain, M. (2012). Development of visual working memory precision in childhood. *Developmental Science, 15*(4), 528-539.
- Case, R., Kurland, M., & Daneman, M. (1979). Operational efficiency and the growth of M-space. *Paper presented at the annual convention of the Society for Research in Child Development, San Francisco.*
- Chein, J. M., Moore, A. B., & Conway, A. R. A. (2011). Domain-general mechanisms of complex working memory span. *NeuroImage, 54*(1), 550-559.
- Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review & user's guide. *Psychonomic Bulletin & Review, 12*, 769-786.
- Cowan, N. (1984). On short and long auditory stores. *Psychological Bulletin, 96*, 341-370.
- Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information processing system. *Psychological Bulletin, 104*, 163-191.
- Cowan, N., AuBuchon, A.M., Gilchrist, A.L., Ricker, T.J., & Saults, J.S. (2011). Age differences in visual working memory capacity: Not based on encoding limitations. *Developmental Science, 14*, 1066-1074.
- Cowan, N., Blume, C.L., & Saults, J.S. (2013). Attention to attributes and objects in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*, 731-747.
- Cowan, N., Elliott, E.M., Saults, J.S., Morey, C.C., Mattox, S., Hismjatullina, A., & Conway, A.R.A. (2005). On the capacity of attention: Its estimation and its role in working memory and cognitive aptitudes. *Cognitive Psychology, 51*, 42-100.

- Cowan, N., Fristoe, N.M., Elliott, E.M., Brunner, R.P., & Saults, J.S. (2006). Scope of attention, control of attention, and intelligence in children and adults. *Memory & Cognition*, *34*(8), 1754-1768.
- Cowan, N., Keller, T., Hulme, C., Roodenrys, S., McDougall, S., & Rack, J. (1994). Verbal memory span in children: Speech timing clues to the mechanisms underlying age and word length effects. *Journal of Memory and Language*, *33*, 234-250.
- Cowan, N., Li, D., Moffitt, A., Becker, T.M., Martin, E.A., Saults, J.S., Christ, S.E. (2011). A Neural Region of Abstract Working Memory. *Journal of Cognitive Neuroscience*, *23*(10), 2852–286.3
- Cowan, N., Morey, C.C., AuBuchon, A.M., Zwilling, C.E., & Gilchrist, A.L. (2010). Seven-year-olds allocate attention like adults unless working memory is overloaded. *Developmental Science*, *13*, 120-133.
- Cowan, N., Saults, J.S., & Blume, C.L. (in press). Central and peripheral components of working memory storage. *Journal of Experimental Psychology: General*.
- Daneman, M., & Carpenter, P.A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning & Verbal Behavior*, *19*, 450-466.
- Darwin, C.J., Turvey, M.T., & Crowder, R.G. (1972). An auditory analogue of the Sperling partial report procedure: Evidence for brief auditory storage. *Cognitive Psychology*, *3*, 255-267.
- Donkin, C., Nosofsky, R.M., Gold, J.M., & Shiffrin, R.M. (2013). Discrete slot models of visual working-memory response times. *Psychological Review*, *4*, 873-902.
- Gorgoraptis, N., Catalao, R. G., Bays, P. M., & Husain, M. (2011). Dynamic updating of working memory resources for visual objects. *The Journal of Neuroscience*, *31*, 8502-8511.
- Hardman, K. O., & Cowan, N. (in press). Remembering Complex Objects in Visual Working Memory: Do Capacity Limits Restrict Objects or Features? *Journal of Experimental Psychology: Learning, Memory, and Cognition*.
- Huelsenbeck, J. P., & Crandall, K. A. (1997). Phylogeny estimation and hypothesis testing using maximum likelihood. *Annual Review of Ecology and Systematics*, *28*, 437–466.
- Luck, S.J., & Vogel, E.K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*, 279-281.
- Keller, T.A., & Cowan, N. (1994). Developmental increase in the duration of memory for tone pitch. *Developmental Psychology*, *30*, 855-863.
- Koelsch, S., Schulze, K., Sammler, D., Fritz, T., Müller, K., & Gruber, O. (2009). Functional architecture of verbal and tonal working memory: An fMRI study. *Human Brain Mapping*, *30*(3), 859-873.
- Kraus, N., Strait, D.L., & Parbery-Clark, A. (2012). Cognitive factors shape brain networks for auditory skills: spotlight on auditory working memory. *Annals of the New York Academy of Sciences*, *77*, 100-107.

- Kumar, S., Joseph, S., Pearson, S., Teki, S., Fox, Z.V., Griffiths, T.D., & Husain, M. (2012). Resource allocation and prioritization in auditory working memory. *Cognitive Neuroscience, First*, 1-9.
- Li, D., Christ, S.E., Cowan, N. (2014). Domain-general and domain-specific functional networks in working memory. *Neuroimage, 102*, 646-656.
- Li, D., Cowan, N., & Saults, J.S. (2013). Estimating working memory capacity for lists of nonverbal sounds. *Attention, Perception, & Psychophysics, 75*, 145-160.
- Majerus, S., Cowan, N., Péters, F., Van Calster, L., Phillips, C., Schrouff, J. (2014). Cross-Modal Decoding of Neural Patterns Associated with Working Memory: Evidence for Attention-Based Accounts of Working Memory. *Cerebral Cortex, 10*, 1-14.
- Miller, G.A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*, 81-97.
- Oberauer, K., & Eichenberger, S. (2013). Visual working memory declines when more features must be remembered for each object. *Memory & Cognition, 41*, 1212-1227.
- Pashler, H. (1988). Familiarity and visual change detection. *Perception & Psychophysics, 44*, 369-378.
- Phillips, W.A. (1974). On the distinction between sensory storage and short-term visual memory. *Perception & Psychophysics, 16*, 283-290.
- Riggs, K.J., McTaggart, J., Simpson, A., & Freeman, R.P.J. (2006). Changes in the capacity of visual working memory in 5 to 10 year olds. *Journal of Experimental Child Psychology, 95*, 18-26.
- Saults, J.S., & Cowan, N. (2007). A central capacity limit to the simultaneous storage of visual and auditory arrays in working memory. *Journal of Experimental Psychology: General, 136(4)*, 663-684.
- Shipstead, Z., Redick, T.S., Hicks, K.L., & Engle, R.W. (2012). The scope and control of attention as separate aspects of working memory. *Memory, 20*, 608-628.
- Skoe, E., & Kraus, N. (2012). A little goes a long way: How the adult brain is shaped by musical training in childhood. *The Journal of Neuroscience, 32(34)*, 11507-11510.
- Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs, 74* (Whole No. 498.)
- Van den Berg, R., Awh, E., & Ma, .W.J. (2014). Factorial comparison of working memory models. *Psychological Review, 121*, 124-149.
- Vulkan, N. (2000). An economist's perspective on probability matching. *Journal of Economic Surveys, 14*, 101-118.
- Zhang, W., & Luck, S.J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature, 453*, 23-35.

Zhang, W., & Luck, S.J. (2011). The number and quality of representations in working memory. *Psychological Science*, 22, 1434–1441.

Appendix A
SOUND LEVEL READINGS

Tones	Sound Level Readings, 70 dB range, A weighting
1	73
2	73
3	74
4	74
5	75
6	75
7	75
8	75
9	76
10	76
11	76
12	75
13	75
14	75
15	75
16	75
17	75
18	74
19	74
20	74
21	73
22	72
23	71
24	71
25	71
26	70
27	69
28	69
29	68
30	68
31	68
32	69
33	69
34	69
35	69
36	69
37	69
38	69

39	69
40	69
41	69
42	68
43	69
44	68
45	68
46	68
47	67
48	67
49	67
50	67
51	67
52	66
53	66
54	66
55	66
56	66
57	66
58	66
59	67
60	67
61	68
62	68
63	68
64	69
65	68
66	68
67	68
68	67
69	67
70	67
71	67
72	67
73	67
74	67
75	66
76	66
77	67
78	67
79	67
80	66