ESTIMATING WORKING MEMORY CAPACITY FOR LISTS OF NONVERBAL SOUNDS

A Thesis

presented to

the Faculty of the Graduate School

at the University of Missouri

In Partial Fulfillment

of the Requirements for the Degree

Master of Arts

---

by

Dawei Li

Dr. Nelson Cowan, Thesis Supervisor

May 2011

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

ESTIMATING WORKING MEMORY CAPACITY

FOR LISTS OF NONVERBAL SOUNDS

presented by Dawei Li,

a candidate for the degree of master of arts,

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

Professor Nelson Cowan

Professor Shawn Christ

Professor Judith Goodman

ACKNOWLEDGEMENTS

I would like to thank my advisor Dr. Nelson Cowan, who has been supporting me on my research ever since the start of my graduate study. I would also like to thank Dr. Scott Saults for his valuable help to create the various auditory stimuli, and to add the detailed information to the method section. Finally, I would like to thank Dr. Shawn Christ and Dr. Judith Goodman for their comments on my draft.

TABLE OF CONTENTS

LIST OF FIGURES

Figure Page

ESTIMATING WORKING MEMORY CAPACITY FOR LISTS OF NONVERBAL SOUNDS

Dawei Li

Dr. Nelson Cowan, Thesis Supervisor

ABSTRACT

Working memory (WM) capacity limit has been extensively studied in the domains of visual and verbal stimuli. The previous studies have suggested a constant WM capacity of typically about 3 or 4 items, based on the number of items in working memory reaching a plateau after several items as the set size increases. We designed a series of experiments to investigate nonverbal auditory WM capacity. Experiment 1 and 2 used simple tones and revealed the capacity limit of up to 2 tones following a 6-s retention interval. In Experiment 3 we added timbre information to the simple tones, and the capacity estimate improved to about 2.5 sounds, still somewhat lower than found previously for items in known categories. This study supports a critical role of categorical information for high WM performance.

Working memory (WM) refers to the cognitive process that involves maintenance and manipulation of a limited amount of information for a short period, usually a few seconds (Baddeley & Hitch, 1974; Cowan, 1995). WM is fundamental to a lot of higher-order cognitive functions, such as decision making, language processing, and planning. For decades, researchers have been curious about the limit of WM capacity – how much information can be stored in WM. Although WM has been investigated in great detail using visual and verbal stimuli, WM for tones has received far less attention and will be examined here. We will consider factors that affect performance in other domains in order to assess the capacity of a key attention-related component of WM for nonverbal sounds, uncorrupted by various mnemonic strategies.

WM in Other Domains

WM in other domains provides an important context in which to formulate the manner to examine WM for tones. Based on various empirical and experimental evidence, Miller (1956) proposed that people could keep in what we call WM lists of approximately seven items, plus or minus two. (Previous studies, including Miller's, have used a variety of terms including immediate memory and short-term memory, but we will ignore nuances possibly distinguishing these terms from WM.) Miller's work elicited many subsequent studies on humans' WM capacity. Some studies, however, indicated that Miller might have overestimated WM capacity (Sperling, 1960; Luck & Vogel, 1997; Cowan, 2001). In one experiment in Sperling's seminal work, the participants were briefly presented an array of 12 characters, and were instructed to write down the characters that they could remember after the array had disappeared. The results showed that only about 4 characters could be written down, meaning that WM capacity might be more restricted than that estimated by Miller.

Chunking. One reason for the higher estimate obtained by Miller (1956) is that, as he himself pointed out, people sometimes can group several items from a list into a larger meaningful unit or chunk, and remember the chunk instead of the individual items. In a straightforward illustration, although people usually cannot remember 9 random letters, such as *NGJLXISFH*, they can easily remember *IRSFBICIA*, if they are able to chunk these letters into 3 US government agencies – *IRS, FBI*, and *CIA*. In Sperling's experiments, due to the rapid and concurrent presentation of items, one can assume it was difficult to apply chunking, leading to a smaller estimate of WM capacity. Similar results have been obtained with the recognition of nonverbal items (e.g., Luck & Vogel, 1997). Later we will consider how musical knowledge may allow chunking for musical sounds, which we will discourage by selecting stimuli judiciously.

Rehearsal. Another factor affecting WM capacity estimates, especially verbal WM, is the strategy of rehearsal, or covertly repeating verbal items or labels in WM in order to refresh the representations of the items in WM. Some studies have shown that phonologically similar words, such as cat, bat, and mat, were more difficult to remember than phonologically dissimilar words, a phonological similarity effect (Conrad & Hull, 1964), and that people could memorize fewer words with longer length than words with shorter length, a word length effect (Baddeley, Thomson, & Buchanan, 1975). These effects at least partly reflected the use of rehearsal in verbal WM, which is more error-prone when the words are phonologically similar and takes longer when the words are longer. When participants are required to repeat a simple word, such as "the", while remembering the items (an *articulatory suppression* task), both the phonological similarity effect and the word length effect are greatly diminished or disappear entirely (for a review see Baddeley, 1986). Because humming might be used to rehearse nonverbal sounds (e.g., Hickock, Buchsbaum, Humphries, & Muftuler, 2003), we use suppression to prevent that possibility (cf. Schendel & Palmer, 2007).

Sensory memory. Another factor that can enhance the estimate of WM is memory for the physical properties of the stimuli, i.e., exactly how they look or sound. Sperling (1960) showed that exposure to the character array led to sensory memory of most of the array items for a short period (under 1 s), available for recall if a partial report cue was provided so that only one row of up to 4 items had to be recalled on a particular trial. Similar indications of a rich but short-lived sensory memory for a complex array are obtained in the auditory modality, with auditory sensory memory lasting several seconds (e.g., Darwin, Turvey, & Crowder, 1972). In some experiments, items to be recalled are followed by an interfering item in the same modality in order to overwrite sensory memory, making it necessary that the concepts rather than sensations be recalled (e.g., Saults & Cowan, 2007). We adopt that strategy here for nonverbal sounds.

Core WM capacity. Cowan (2001) suggested that the smaller limit of 3 to 5 items in WM is obtained under conditions in which it is not possible to use chunking, rehearsal, or sensory memory to enhance performance. Under those conditions, memory may be based on the number of conceptual items that can be held in the focus of attention. A key example upon which the present work is based is the recognition memory for colored squares, examined by Luck and Vogel (1997). In one experiment they instructed the participants to memorize a briefly-presented array of a few colored squares for several seconds, followed by the presentation of a second, probe array in which one square may have changed color; in another experiment yielding similar results, one item in the second array was marked to indicate which square might have changed. The task was to decide whether the new square had the same color as the previous square in that location. By this change-detection paradigm they estimated the participants' visual WM capacity at about 4 items. The brief presentation of the first array made the items difficult to chunk, and a secondary memory load of two digits further discouraged rehearsal. Sensory memory presumably could not be used to great advantage either, inasmuch as the probe array would have overwritten the critical sensory information before a judgment could be made. Given these restrictions, it is suggested that the results are indicative of a core WM capacity (Cowan, 2001).

This core WM capacity has been observed also when participants are taught pairs of words, in which case they can recall about 3 chunks from a list in the presence of articulatory suppression, no matter whether the chunks in the list were singletons or learned pairs (Chen & Cowan, 2009). Given the considerable evidence for a small core capacity for information from stimuli that can be labeled, we wished to examine WM for tonal stimuli that cannot easily be labeled.

Cowan (2001) presented a measure that can be used to estimate the number of items held in WM. This measure applies to the experimental situation in which the test probe display clearly indicates which item changed if any of them did (Rouder, Morey, Morey, & Cowan, in press). It assumes that the array includes N items and that k items fit in WM. Then when $N \geq k$, the proportion of correct detections of a change, or *hits*, can be estimated as $hits = k/N + (1 - k/N)g$, where g is the rate of guessing that there has been a change, in the absence of WM information. Guessing takes place only if the tested item was not in WM, so when there is no change, *false alarms= (1-k/N)g*. Combining these equations yields the estimate *k=(hits-false alarms)N.* We apply this formula to recognition memory for lists of tones.

WM for Tones

In contrast to the extensively investigated domains of visual and verbal WM, few studies have investigated the capacity limit of nonverbal auditory items in WM that are uncontaminated in that they both contain little verbal or verbalizable information, and are difficult to visualize. It is possible that such auditory items could be more difficult to remember due to their pure acoustic nature, which can be memorized only through their sound properties, instead of phonological, visual, or semantic properties.

Some early studies on absolute judgment of tones yielded an upper limit of 5 different pitches. In these studies the participants were instructed to identify an individual tone, selected from a few pre-generated tones, each assigned to a response number (Pollack, 1952; Pollack, 1953). Although it revealed a constant capacity limit for tones, the method of absolute judgment is not a direct index of auditory WM because of high task demands: the participant must retain all of the pre-generated tones along with their assigned response numbers while perceiving the test tone.

A series of studies on music sequence production, in which the participants, usually musicians, learned to perform several musical pieces, have shown that the musicians' pitch-ordering errors usually arose from sequences that have a range of 3 to 4 tones. Pitch-ordering error refers to the musical sequence that is reproduced in the wrong order, and reflects the activity level of the musical tones in the memory. The result indicated that the “wrong” tone still had higher availability in the memory when it was up to 3 to 4 pitches ahead of the current sequence location. This appears to indicate a constant capacity constraint in WM for musical sequences (Drake & Palmer, 2000; Palmer, 2005; Palmer & Pfordresher, 2003).

The use of melodies involves two levels of structure that could assist performance, similar to the chunking processes that we have discussed (Davies, 1979). Musical sequences are composed of scales of discrete pitch relationships, or intervals. Western music is based on the 12-tone chromatic scale of equal temperment which repeats every octave, when the frequency is doubled. Further, most melodies in Western music use only seven-interval subsets of the chromatic scale called diatonic scales (Burns & Ward, 1982). People can use familiarity with these scales to encode the melodic contour of a musical sequence, grouping or chunking intervals to achieve better memory of musical sequences compared to random tone sequences (Dewar, Cuddy & Mewhort, 1977; Idson & Massaro, 1976).

Capacity for lists of tones. There has been little research in which the number of tones in a sequence has been varied in order to assess the effect of that manipulation on the ability to detect a change in one tone. Watson, Foyle, and Kidd (1990) varied the number of component tones widely. They chose tones in a manner that eliminated conventional musical cues, dividing the frequency range 300-3kHz into *N* tones based on logarithmically equal intervals, where *N* was the list length, and shuffling the order of the resulting tones. Clearly, the number of tones made a very large difference for performance, though no estimate of the WM capacity for tones could be obtained from their procedure. Note that, using this method, the number of tones in the list is confounded with the frequency difference between adjacent tones.

Kidd and Watson (1992) found that what was important was not the number of tones per se but the proportion of the tone list taken up by the target tone. In their procedure, however, participants were held responsible for only one tone per series, the one in the middle of the pattern (or in one experiment, two tones flanking the middle tone and changing together), which would not place a load on WM commensurate with the list length.

In the closest precursor to the present study that we could find, Prosser (1995) chose 14 tones that were selected to avoid a musical scale and presented lists of 2, 4, or 6 randomly-selected tones per trial. The list was followed by a tone probe to be judged present or absent from the sequence. To evaluate the results, we apply the formula of Cowan (2001) to the means shown in Prosser's Figure 1. Doing so using data for a short (1-s) retention interval, for lists of 2, 4, and 6 tones *k*=1.5, 2.2, and 2.9 tones in WM, respectively. These estimates are roughly consistent with past evidence on non-tonal stimuli, or are slightly lower. The shift across list lengths is found also for visual arrays and may occur because certain individuals have a capacity higher than *N*, resulting in ceiling effects that limit the estimates for the smaller set sizes.

Capacity, attention, and time. Cowan (2001) suggested that a limited number of items comprising the core contents of WM is held in the focus of attention. The primary function of holding information that way would be to make the item representations resistant to interference or decay. In that regard, it is useful to examine the items in WM after a several-second retention interval, so that features and items susceptible to decay already would have decayed, and what remains is the items held firmly in mind. Cowan et al. (in press) presented a combination of colored squares and spoken letters followed by a mask and then an 8-s retention interval, and after that period still observed a capacity of 2.9 to 3.6 items.

Capacity might be lower, however, for tonal stimuli that do not correspond to known musical categories. Prosser (1995) included a 7-s retention interval and, for lists of 2, 4, and 6 tones, we estimate from his Figure 1 that *k*=1.5, 1.7, and 1.7 items, respectively.

The present study. We wished to explore further these rough estimates of tones in WM, derived from the findings of Prosser (1995) at a long retention interval, more systematically in order to understand WM capacity limits. We adapted the change-detection procedure by presenting sequences of tones, followed by a probe tone or probe tone list to be recognized as the same as the original list or changed (see Figure 1). To identify the WM capacity limit for individual tones without any available musical structure, we used lists of tones randomly selected from a nonmusical scale of 12 pitches that differ from notes of the chromatic scale and span several octaves. We used a retention interval of 6 s (following a list-final masking stimulus), which is long enough that any residual sensory memory that somehow survived the mask should already have decayed before the probe (see Darwin et al., 1972), leaving behind information that is protected from decay. We included only individuals without special music training, defined as participation in a band or orchestra or music instruction at a college level.

Several features distinguish our study from the past work of Prosser (1995) or any other study to our knowledge. First, as one step to eliminate sensory memory information, as mentioned we presented a masking sound after each list. There is a long history of auditory backward masking of recognition using interstimulus target-mask intervals of a fraction of a second (e.g., Massaro, 1975) but our purpose here was not to prevent recognition. Rather, similar to Saults and Cowan (2007), we waited long enough for recognition of all tones in the list to be completed and then presented a mask, in order to force participants to rely on the recognized abstract information in WM rather than a sensory memory trace, which otherwise might have persisted for several seconds (Cowan, 1984; Darwin et al., 1972).

Second, unlike most prior studies, in some conditions we suppressed articulation in case participants were able to vocalize tones covertly and rely on that process as subvocal rehearsal. Third, to equate the amount of inter-tone interference in memory, we included conditions in which the number of tones stayed the same across different memory loads, which was accomplished by presenting 6 tones and requiring memorization starting at a variable point in the middle of the list (Figure 1).

Fourth, and finally, we provided visual cues to indicate which serial position in the tone series was being probed. We did this because it is required for the *k* measure of items in working memory, which is based on the assumption that the participant has to search only the memory of one item. This measure of items in working memory has been psychometrically validated much more fully than any other measure; the data conform to a receiver operating characteristic function expected according to the model (Rouder et al., 2008, in press).

All of these precautions, taken together, should allow us to examine WM capacity for abstract information about tones without any pre-learned categories for the tones.

Experiment 1

Method

Participants. Twenty-seven undergraduate students (12 male, 15 female) participated in the experiment to fulfill introductory psychology course requirements.

Apparatus and stimuli. The stimuli were presented with E-Prime (Schneider, Eschman, & Zuccolotto, 2002) on 17-inch color monitors in soundproof booths. Twelve simple tones (sine waves) were generated by Praat software (Boersma & Weenink, 2009), with a lowest frequency of 200 Hz and a highest frequency of 3900 Hz. There was a 31% frequency difference between each two adjacent tones. Each tone had a duration of 500 ms, and included 25-ms linear onset and offset ramps.

We wanted the pitches of our 12 tones to be as far apart as possible, so they would be easy to discriminate, but still within a range with similar difference limens for frequency change, which increases sharply beyond 4000 Hz (Sek & Moore, 1995). We also wanted them to differ from familiar musical notes. Thus, our lowest tone was about 35 cents above the G below middle C (G3) while our highest tone was about 23 cents below B7, the second highest note on an 88-key piano (100 cents = I semitone). A 31% difference between tones avoids familiar musical intervals and harmonic relationships between tones. Adjacent semitones in music differ by about 5.9% (precisely $2^{1/12}$) in twelve-tone equal temperament, the common tuning system for Western music (Burns & Ward, 1982). Although our stimuli spanned about 4 octaves, no tone in our set had a simple harmonic relationship with another tone. For example, the second harmonic of 200 Hz is 800 Hz, but the closest frequency to that in our set was 771.6 Hz. Avoiding octaves minimizes the tendency to confuse two tones with different pitch height but equal chroma, based on octave generalization (Shepard, 1982).

Six circles were presented in the center of the screen on a gray background, as shown in Figure 1. The participants were seated approximately 50 cm from the screen. The sounds were presented through two speakers (left and right) in front of the participants, and fell between 60 and 70 dB as measured by a sound level meter.

Procedure. On each trial, participants had to try to remember 2, 3, 4, 5, or 6 tones and then perform a recognition task. At the beginning of each trial, a “+”appeared on the center of the screen for 1000 ms, which indicated the onset of a trial and provided a fixation point for the participant. Next, six circles were presented in the center of the screen as shown in Figure 1. Six auditory tones, selected from the twelve simple tones that we created before the experiment, were sequentially presented through the loudspeakers, each lasting for 500 ms with a 250-ms silent interval between tones. A printed character (*, &, $, @, #, %, or ->) accompanied each tone, and the characters were presented sequentially, with each character in one of the circles, always starting from the circle at the top. The character disappeared as soon as its corresponding tone ended. The participants were instructed to start remembering tones starting with the one accompanied by a forward arrow (->) and continuing until the end of the series. They were also instructed to ignore the characters except for the forward arrow (->).The position of the forward arrow (->) was manipulated such that the memory load was set to include five levels: 2, 3, 4, 5, and 6 tones. The other characters were randomly arranged, and there was no constant association between particular characters and particular tones.

Two additional types of trials were included in the experiment, and were the same as the other conditions except that the participants heard only 2 or 4 tones and saw 2 or 4 characters, respectively, during the encoding phase. The characters were presented sequentially each in one circle, starting from the circle on the top, and the first character was always a forward arrow (->). We included these additional conditions to estimate to what extent the different stimulus presentation methods would affect the participants' performance. In the following text we will denote these trials as "presentation method 2" (PM2), and the other trials as "presentation method 1" (PM1).

A masking tone, which was produced by simultaneous combination of the twelve different possible stimulus tones, was presented for 500 ms after the last one of the six tones, in the same temporal rhythm as these tones, to eliminate sensory memory. After a 6000-ms retention interval, a probe tone was presented, accompanied by a "?" symbol in one of the circles corresponding to a tone that was to be remembered. The participants were to decide whether the probe tone corresponding to the "?" location was the same as the one at that location during encoding, or was different. If the tone was different, it did not match any of the tones in the presented series, and the participants were made aware of that. In half of the trials, the correct answer would be "same", and in the other half of the trials, the correct answer would be '"different". The participants were instructed to press "s" for "same" and "d" for "different", and they had unlimited time to respond. Feedback that lasted for 500 ms was provided after the participant made a response. A blank period with a dot in the center of the screen lasted for 1000 ms before the next trial started.

The trials were allocated into 10 blocks. Each block contained 4 trials for each condition, adding up to 28 trials per block. Each trial lasted for 16 seconds, and the experiment lasted for 1.5 hours.

In half of the blocks, the participants were instructed to whisper “the” twice a second during the encoding and maintenance phases (“whisper” sessions); in the other half of the sessions, they were instructed to tap the right index finger on the table twice a second during these phases (“tap” sessions). The “whisper” and “tap” sessions were arranged in a counterbalanced manner within the participant (always whisper-tap-tap-whisper-whisper-tap-tap-whisper-whisper-tap).

Before the experiment, the participant was trained to whisper “the” and tap his or her finger, each for 1 minute. During the practice, there was a beep every second to help the participants keep the pace. The participants also performed two practice memory blocks, each consisting of 7 trials (1 trial per condition). The first practice session was a “whisper” block, and the second practice was a “tap” block.

Results and Discussion

A two-way repeated measure ANOVA of PM1 response accuracy with the set size of tones to be remembered (2, 3, 4, 5, or 6) and articulation condition (“whisper” and “tap”) as within-participant factors revealed significant main effects of set size, F(4, 104) = 16.57, $p < 0.01$, and articulation, F(1, 26) = 4.88, $p < 0.05$. The interaction between set size and articulation was not significant, F(4, 104) = 1.27, $p = 0.29$ (see Figure 2, top left). The main effect of articulation suggested that repeating a simple word could interrupt rehearsal of tones, replicating the results of previous work (Schendel & Palmer, 2007).

For each set size, we calculated the participants’ WM capacity using Cowan’s *k* formula as noted above. The results are shown in Figure 2 (bottom left). The highest capacity estimate (mean ± 95% within-subject confidence interval) was 2.01 ± 0.42 tones at Set Size 6, lower than those estimated in simple visual or verbal WM tasks (Luck & Vogel, 1997; Chen & Cowan, 2008), while similar to the capacity limit found in the previous studies on memory for tone sequences (Prosser, 1995).

To investigate the influence of the different stimulus presentation methods, we used only the accuracy rate data of set size 2 and 4 but both presentation methods, and conducted a two-way repeated measure ANOVA with set size (2 and 4) and presentation method (PM1 and PM2) as within-participant factors. The results revealed significant main effects of set size, F(1, 26) = 30.77, p < 0.01, and presentation method, F(1,26) = 6.85, p<0.05. The interaction was not significant, F(1, 26) = 0.95, p = 0.34. The main effect of presentation method suggested that people performed slightly better when they memorized all the stimuli that they heard, instead of starting to remember from a specific stimulus. However, the *k* value estimates for PM2 were still low, with 1.17 ± 0.34 at Set Size 2 and 1.56 ± 0.63 at Set Size 4, compared with 0.99 ± 0.35 at Set Size 2 and 1.40 ± 0.66 at Set Size 4 for PM1. The results indicate that changing the presentation method would not induce much improvement in terms of *k* value estimates.

Experiment 2

The estimated auditory WM capacity in Experiment 1 was much lower than the measured visual or verbal WM capacity in previous studies, even in the presence of a long retention interval (Cowan et al., in press).

An early study on short term memory for tone lists found better accuracy rates when the context tones were also presented together with the probe tone, compared with the single-tone probe (Dewar et al., 1979). The authors suggested that higher-order information, such as relational or pattern information, aided in the WM performance. In the second experiment, to examine this, we used the full-list probe to test this hypothesis.

Method

Participants. Twenty-four undergraduate students (7 male and 17 female) participated in the experiment to fulfill the introductory psychology course requirements.

Apparatus and Stimuli. The apparatus and stimuli were the same as those we used in Experiment 1.

Procedure. The procedure was similar to that of Experiment 1, except for one difference during the test phase. Instead of presenting only one test tone and a question mark, in the present experiment a list of tones was presented during the test. The number of the tones during the test was the same as the number of tones that the participants were supposed to remember. The same number of characters were also presented sequentially each in one of the circles, starting from the circle marked with the -> during the stimuli presentation. One of the characters was a question mark ("?"). The participants were instructed to decide whether the test tone corresponding to the "?" was the same as the tone at the same specific location during stimulus presentation, or whether it was different from any of the tones that they remembered.

Results and Discussion

For the trials of PM1, the same two-way repeated measure ANOVA was conducted and revealed significant main effects of set size, F(4, 92) = 18.24, $p < 0.01$, and articulation condition, F(1, 23) = 12.74, $p < 0.01$. The interaction between set size and articulation was not significant, F(4, 92) = 1.59, p = 0.18 (Figure 2, top middle). We again calculated the *k* value for each set size. The highest *k* value was 1.58 ± 0.41 tones at set size 6, even lower than the highest *k* value in Experiment 1 (Figure 2, bottom middle).

We combined the data from Experiment 1 and 2 and conducted a three-way ANOVA with accuracy rates as the dependent variable, the two experiment groups as between-participant factor, and set size and articulation condition as within-participant factors. The results revealed significant main effects of set size, F(4, 196) = 33.52, $p < 0.01$, and articulation, F(1, 49) = 16.23, $p<0.01$. However, the main effect of experiment did not reach significance, F(1,49) = 1.39, p = 0.24. The interaction effects were not significant either.

A similar analysis was conducted to investigate the presentation method effect as we did in Experiment 1. We again found significant main effects of set size, F(1, 23) = 29.28, $p < 0.01$, and presentation method, F(1, 23) = 7.51, $p < 0.05$. The interaction effect was not significant, F(1, 23)=0.005, p = 0.94. The *k* value estimates were 1.18 ± 0.26 (set size 2) and 1.53 ± 0.44 (set size 4) for PM2, compared with 1.04 ± 0.28 and 1.27 ± 0.52 for PM1. The minor improvement again suggested that the different presentation methods were not the reason for the low performance in the tone WM tasks.

It was obvious that presenting the tone list instead of a single tone did not improve people’s performance in the simple-tone WM task. The discrepancy between our experiments and the Dewar et al (1979) study was probably due to the more familiar musical stimuli that they used. They found that recognition memory was more accurate under full-context conditions than under no-context conditions even for sequences of random tones. However, their random-tone sequences were always selected from 12 tones of a chromatic scale, compared to musical sequences selected from 7 notes in the same octave of a major scale. Note that even the random (atonal) sequences used by Dewar et al. (1979) included a majority of intervals from a major scale, so that sequences still might be encoded as melodies with some ‘wrong’ notes. Certainly the tones of a chromatic scale include far more musical and familiar intervals than our tones, with frequencies that span nearly four octaves and notes without any consistent relationship between height and chroma (Shepard, 1984). For example, the first five of our stimuli are closest to the musical notes G3, C4, F4, A4, D5, with intervals that differ, on average, by 40 cents from any musical intervals. In that light, it is not surprising that context made so little difference to memory for our thoroughly nonmusical stimuli. Based on these results, the single tone probe should be appropriate for this study, so we continued with the single tone probe in Experiment 3.

Experiment 3

The results in Experiments 1 and 2 were discrepant from the previously observed higher capacity to maintain simple visual and verbal items in WM (e.g., Cowan, 2001). However, some previous studies also revealed relatively low capacity estimates, for example when people were instructed to memorize certain types of stimuli. When the items were complex, fewer items could be memorized. Complex item sets have included, for example, irregular shapes, faces, and novel characters (Alvarez & Cavanagh, 2004; Jha & McCarthy, 2000). However, our tones were not complex and were quite dissimilar from one another in frequency.

Some researchers have raised the possibility that categorical information is critical for high WM performance. Using visual items that were easy to distinguish but lacked categorical information, Olsson and Poom (2005) revealed visual WM capacity as low as only one item. After adding categorical information to the stimulus set, including discrete colors and shapes, the estimated visual WM capacity increased to slightly below three. The authors concluded that categorical information stored in long-term memory was crucial to visual WM performance.

In the domain of absolute judgment for tones, Pollack found improved performance when the "dimensionality" of the tones increased (Pollack, 1953). When the tones had only one dimension (frequency), the participants could identify about 5 tones. When the tones included two dimensions (frequency and sound level), the same participants could identify more than 8 tones. This result suggests that the same outcome as revealed by Olsson and Poom's work might also apply to WM for tones.

Considering the above evidence, we hypothesized that the low performance in

Experiments 1 and 2 could be due to the lack of categorical information in our stimuli set. The tones could only be distinguished by their different frequencies, in contrast to such stimuli as auditory letters and colored squares, which are defined by phonological or discrete color categories besides pure acoustic or visual information. Therefore, in Experiment 3, we included both frequency and quality differences in our stimuli, so that the new stimulus items would have more categorical information than the one used in Experiments 1 and 2.

Method

Participants. Twenty-four undergraduate students (6 male, 18 female) participated in the experiment to fulfill the introductory psychology course requirements.

Apparatus and stimuli. The apparatus in Experiment 3 was the same as that used in Experiments 1 and 2. The only difference is that the stimuli we used in Experiment 3 had different qualities (timbres), as well as having different frequencies. (These stimuli can be heard on the first author's web site, http://psychology.missouri.edu/dlmgf.) We selected twelve sounds generated with GarageBand (Apple Inc., Cupertino, California), a program in the Macintosh Operating System, each played by a distinct instrument (*Trumpet Section, Smooth Clav, Classic Rock Organ, Negril Bass, Tenor Sax, Space Harpsichord, Grand Piano, Live Pop Horns, Aurora Bell, Pop Flute, Hollywood Strings*, and *Clean Electric Guitar*). Then we varied the fundamental frequencies of these sound files to be the same as the frequencies that we used in Experiments 1 and 2, from lowest (200 Hz) to highest (3900 Hz) in the order shown.

Procedure. The procedure was the same as Experiment 1 except that we used the multidimensional sounds instead of pure tones. Additionally, just after the experiment, the participants answered a questionnaire to specify how many sounds they memorized by labeling them as objects such as instruments, instead of by purely acoustic properties. They also rated from 1 to 5 the extent to which they relied on the labels to remember the sounds, 1 being *mostly acoustic* and 5 being *mostly labeled*.

Results and Discussion

We conducted a similar two-way repeated measure ANOVA on PM1 accuracy with set size (2 to 6) and articulation condition ("whisper" and "tap") as within-participant factors. The results (Figure 2, top right) revealed significant main effects of set size, F(4,92) = 13.92, p < 0.01, and articulation condition, F(1,23) = 12.74, p < 0.01. The interaction was not significant, F(4,92) = 1.59, p = 0.27.

We calculated the *k* values for each set size, and found the highest *k* value of 2.48 ± 0.26 at set size 5 (Figure 2, bottom right). This capacity estimate is higher than the ones we calculated in Experiments 1 and 2. This result in auditory WM conceptually replicates the work by Olsson and Poom (2005) with visual stimuli when discrete color and shape information was included. Also, in contrast to the monotonic pattern found in Experiments 1 and 2, the *k* value curve in this study peaked at Set Size 5 and then leveled off, indicating that a capacity limit had been reached.

Again we conducted a two-way ANOVA with set size (2 and 4) and presentation method (PM1 and PM2) as within-participant factors. The results revealed significant set size effect, F(1, 23) = 27.34, p < 0.01, but the main effect of presentation method was not significant, F(1, 23) = 0.73, p = 0.40. The *k* value estimates were 1.25 ± 0.11 (set size 2) and 1.90 ± 0.24 (set size 4) for PM2, compared with 1.23 ± 0.12 and 1.75 ± 0.24 for PM1. Different presentation methods had no effect on WM performance in this task.

Although we wish to conclude that the capacity limit observed in this experiment is the limit in number of categorical acoustic items that can be held in WM, an alternative explanation might be that the participants labeled the sounds with certain instruments, and memorized the sounds by their labels instead of their acoustic properties. This possibility can be examined, however, using the questionnaires that participants completed after the main procedure. They rated the number of sounds they were able to label, as well as the extent to which they relied on the labels, from 1 to 5. The mean number of sound labeled (± SEM) was 4.92±0.63. We also calculated the weighted number of sounds for each individual as the number labeled multiplied by the rated reliance on labels divided by the maximum possible rating. For example, an individual who indicated that 3 sounds were labeled and that the reliance on those labels was 4 out of a possible 5 would receive a weighted score of 3x(4/5)=2.40. The average weighted score was 3.57±0.60, small compared with the maximum possible weighted number (12). Additionally, we also examined the correlation between the participants' overall accuracies and their ratings. No significant correlation was found between recognition accuracy with the number of sounds labeled, r(23)=-.28, or the weighted number, r(23)=-.33.

One participant performed near chance. Without that participant, the maximum capacity was 2.54 (for 5 sounds) and the correlations were close to zero (r=-.11 and -.10, respectively).

General Discussion

Many previous studies have revealed constant memory capacity of 3 or 4 items or chunks, when people were instructed to remember lists of simple items, such as auditory letters and visual colored squares (e.g., see Cowan, 2001; Rouder et al., 2008). The most important evidence is that the *k* value, which represents the number of items being kept in WM, increases with memory load, peaks at between 3 and 4 items in WM (or at about 3 items after a long retention interval), and then levels off. Such pattern strongly indicates the presence of a constant WM capacity. However, few researches have studied the capacity limit in the domain of nonverbal auditory items.

In the above experiments, we investigated the core auditory WM capacity limit by using different sets of auditory stimuli. Experiments 1 and 2 revealed capacity estimates of 2 items or fewer, when simple tones were used. In Experiment 3, we changed the simple tones to sounds with different qualities and frequencies, and found an improved capacity limit of about 2.5 items. We also observed a rise-and-plateau pattern of the *k* value estimates in Experiment 3, similar to the previous studies on verbal and visual WM.

The low capacity estimates in auditory WM for simple tones as revealed in Experiments 1 and 2 are consistent with the results in some previous studies on memory for tone sequences (Prosser, 1995). Camos and Tillmann (2008) presented to participants a list of rapid auditory tones differing in frequency, and instructed the participants to evaluate the number of the tones that they heard. A big discrepancy in terms of response time was found between the list of 2 and 3 tones, suggesting that the participants were able to keep up to 2 simple tones in the focus of attention.

The comparison between Experiment 3 and the previous two experiments indicated the critical role of categorical information in WM. Categorical information refers to the knowledge stored in long-term memory that situates the input stimuli into discrete classes, such as color, shape, phonological, or semantic category. When timbre was included in the stimulus set to allow categories to be formed, the estimated capacity improved to 2.5 items. Our results were in accord with Olsson and Poom's (2005) work, which indicated a capacity limit of slightly below 3 when discrete color and shape information were added into their visual stimulus set.

Both capacity estimates (that of Experiment 3 and of Olsson & Poom, 2005) were still less than the ordinary capacity limit of 3 or 4 (Cowan, 2001) or about 3 after a long delay (Cowan et al., in press). A possible reason for the slight discrepancy is that the timbre information in our study was not entirely discrete, such that people could still confuse one timbre with another. For Olsson and Poom's study, the participants needed to memorize the conjunction of shape and color, while the short stimulus presentation time might have prevented them from chunking the shape and color together into an object, leading to a lower capacity estimate.

Some previous research with relatively low WM capacities might be explained in terms of the absence of categorical information. The stimulus sets in these studies included novel characters, complex shapes, faces, etc (Alvarez & Cavanagh, 2004; Jha & McCarthy, 2000). It is difficult to form distinctive representations of many such stimuli in long-term memory, which was probably the reason for the low WM capacity in those studies. We suspect that the stimulus sets that lead to a constant WM capacity estimate of 3 or 4 items share a common feature, specifically that these stimuli have clear categorical information.

It is not the case that capacity can continue to grow by making stimuli more and more dissimilar from one another. Anderson, Vogel, and Awh (2011) used a procedure in which the precision of the recollection of an item's orientation could be examined and they found that with set sizes larger than 3 items, there were no further increases in the number of items recalled and no further loss in the precision of each item recalled. Therefore, for maximal WM storage, the stimuli must be dissimilar enough to allow clear categorization, but not necessarily any more dissimilar than that.

In this article we discussed three studies on core auditory WM capacity. Although people were able to retain up to 2 simple tones, their performance improved when timbres were added into the stimulus set. The average capacity was nevertheless slightly lower than the 3 to 4 items typically found for categorical stimuli such as known characters or colors after an extended retention interval. Further research is needed to measure core auditory WM capacity with different sets of stimuli, as well as different degrees of involvement of categorical information in the stimulus sets.

References

Alvarez, G.A., & Cavanagh, P. (2004). The capacity of visual short-term memory is set both by information load and by number of objects. *Psychological Science*, 15, 106-111.

Baddeley, A.D. (1986). *Working memory*. Oxford, England: Clarendon Press.

Baddeley, A. & Hitch, G. J. (1974). Working memory. In: *Recent advances in learning and motivation, vol. 8,* ed. G. Bower. Academic Press.

Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short term memory. *Journal of Verbal Learning and Verbal Behavior* **14**: 575–589.

Boersma, P., & Weenink, D. (2009). Praat: doing phonetics by computer (Version 5.2.02) [Computer program]. Retrieved Sept, 2010, from http://www.praat.org/

Burns, E. M. (1999). Intervals, scales, and tuning. In D. Deutsch, D. Deutsch (Eds.) , *The psychology of music (2nd ed.)* (pp. 215-264). San Diego, CA US: Academic Press.

Camos, V., & Tillmann, B. (2008). Discontinuity in the enumeration of sequentially presented auditory and visual stimuli. *Cognition*, 107, 1135-1143.

Chen, Z., & Cowan, N. (2009). Core verbal working-memory capacity: the limit in words retained without covert articulation. *The Quarterly Journal of Experimental Psychology*, *62*, 1420-1429.

Conrad, R., Hull, A.J. (1964). Information, acoustic confusion and memory span. *British Journal of Psychology*, 55, 429-432.

Cowan, N. (1984). On short and long auditory stores. *Psychological Bulletin*, *96*, 341-370.

Cowan, N. (2001). The magical number four in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences,24,* 87–114.

Cowan, N., Johnson, T.D., & Saults, J.S. (2005). Capacity limits in list item recognition: evidence from proactive interference. *Memory*, 13, 293-299.

Cowan, N., Li, D., Moffitt, A., Becker, T.M., Martin, E.A., Saults, J.S., & Christ, S.E. (in press). A neural region of abstract working memory. *Journal of Cognitive Neuroscience*.

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.

Davies, J. (1979). Memory for melodies and tonal sequences: A theoretical note. *British Journal of Psychology*, 70, 205-210.

Dewar, K.M., Cuddy, L.L., & Mewhort, D.J.K. (1977). Recognition memory for single tones with and without context. *Journal of Experimental Psychology: Human Learning and Memory*, 3, 60-67.

Drake, C., & Palmer, C. (2000). Skill acquisition in music performance: relations between planning and temporal control. *Cognition*, 74, 1-32.

Hollands, J.G., & Jarmasz, A. (2010). Revisiting confidence intervals for repeated measures designs. *Psychonomic Bulletin & Review*, 17, 135-138.

Idson, W. L., & Massaro, D. W. (1976). Cross-octave masking of single tones and musical sequences: The effects of structure on auditory recognition. *Perception & Psychophysics*, *19*, 155-175.

Jha, A.P., & McCarthy, G. (2000). The influence of memory load upon delay-interval activity in a working-memory task: an event-related functional MRI study.

Kidd, G.R., & Watson, C.S. (1992). The "proportion-of-the-total-duration rule" for the discrimination of auditory patterns. *Journal of the Acoustical Society of America, 92*, 3109-3118.

Lin, P.H., & Luck, S.J. (2009). The influence of similarity on visual working memory representations. *Visual Cognition*, 17, 356-372.

Luck, S. J., & Vogel, E. K. (1997, November 20). The capacity of visual working memory for features and conjunctions. *Nature, 390*, 279–281.

Massaro, D.W. (1975). Backward recognition masking. Journal of the Acoustical Society of America, 58, 1059-1065.

Olsson, H., & Poom, L. (2005). Visual memory needs categories. *Proceedings of the National Academy of Sciences*, 102, 8776-8780.

Palmer, C. (2005). Sequence memory in music performance. *Current Directions in Psychological Science, 14,* 247-250.

Palmer, C., & Pfordresher, P.Q. (2003). Incremental planning in sequence production. *Psychological Review*, 110, 683-712.

Pashler, H. (1988). Familiarity and visual change detection. *Perception & Psychophysics*, *44*, 369-378.

Pollack I. (1952). The information of elementary auditory displays. *The Journal of the Acoustical Society of America*, 24, 745-749.

Pollack I. (1953). The information of elementary auditory displays. II. *The Journal of the Acoustical Society of America*, 25, 765-769.

Prosser, S. (1995). Aspects of short-term auditory memory as revealed by a recognition task on multi-tone sequences. *Scandinavian Audiology*, 24, 247-253.

Rouder, J.N., Morey, R.D., Cowan, N., Zwilling, C.E., Morey, C.C., & Pratte, M.S. (2008). An assessment of fixed-capacity models of visual working memory. *Proceedings of the National Academy of Sciences (PNAS), 105*, 5975–5979.

Rouder, J.N., Morey, R.D., Morey, C.C., & Cowan, N. (in press). How to measure working-memory capacity in the change-detection paradigm. *Psychonomic Bulletin & Review*.

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*, 663-684.

Schendel, Z.A., & Palmer, C. (2007). Suppression effects on musical and verbal memory. *Memory & Cognition*, 35, 640-650.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime reference guide.* Pittsburgh, PA: Psychology Software Tools.

Sek, A., & Moore, B. J. (1995). Frequency discrimination as a function of frequency, measured in several ways. *Journal of the Acoustical Society of America*, 97(4), 2479-2486.

Shepard, R. N. (1982). Geometrical approximations to the structure of musical pitch. *Psychological Review*, 89(4), 305-333.

Todd, J.J., & Marois, R. (2004). Capacity limit of visual short-term memory in human posterior parietal cortex. *Nature, 428*, 751-754.

Watson, C.S., Foyle, D.C., & Kidd, G.R. (1990). Limits of auditory pattern discrimination for patterns with various durations and numbers of components. *Journal of the Acoustical Society of America, 88*, 2631-2638.

Figure 1. Procedure of Experiment 1. See text for details.

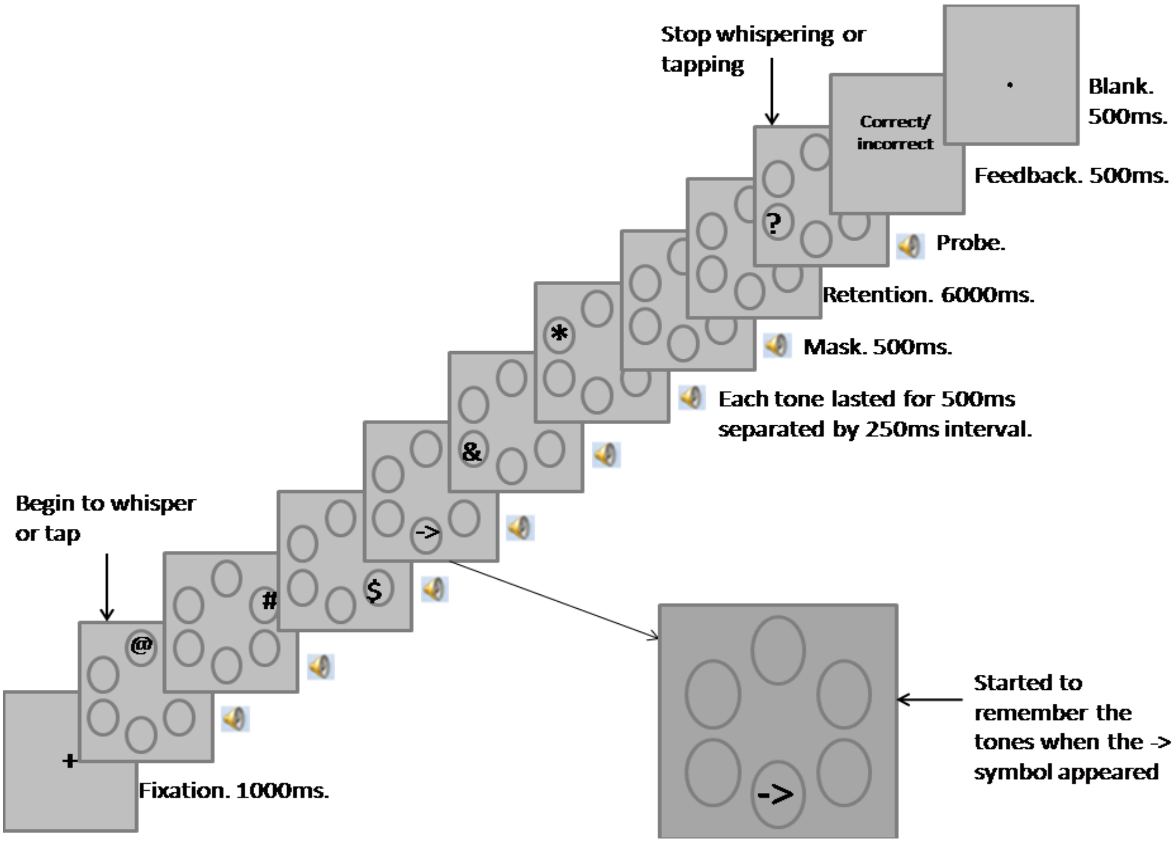
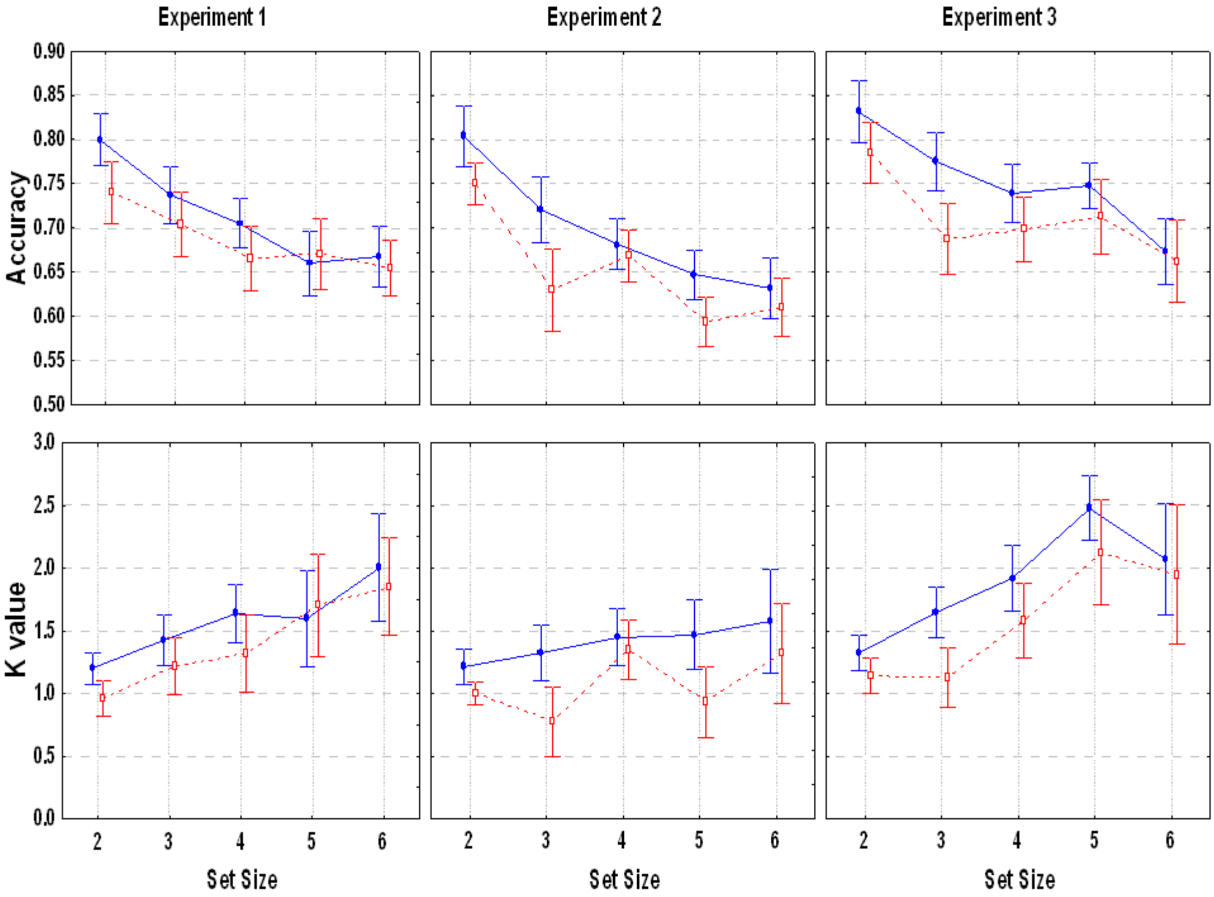

Stop whispering or tapping
Correct/ incorrect
Blank. 500ms.
Feedback. 500ms.
?
Probe.
Retention. 6000ms.
Mask. 500ms.
*
Each tone lasted for 500ms separated by 250ms interval.
&
->
Begin to whisper or tap
@
#
$
+
Fixation. 1000ms.
->
Started to remember the tones when the -> symbol appeared

Figure 2. Results for Presentation Method 1, in which a list of 6 tones was presented and an arrow cue indicated the first tone that was to be remembered. Top panels, accuracy rates; bottom panels, *k* value estimates. Results for Experiment 1, left-hand panels; Experiment 2, middle column; and Experiment 3, right-hand panels. The *k* values are from Cowan (2001). The solid curve refers to the "tap" trials, and the dashed curve refers to the "whisper" trials. Error bars represents 95% repeated-measure confidence intervals (Hollands & Jarmasz, 2010).


Experiment 1
Experiment 2
Experiment 3
Accuracy
K value
Set Size