Resource allocation in phonological working memory: Same or different principles from vision?

Christopher R. Hepner[a], Nazbanou Nozari[a,b]

[a] *Department of Neurology, Johns Hopkins University, Baltimore, United States*
[b] *Department of Cognitive Science, Johns Hopkins University, Baltimore, United States*

Correspondence may be addressed to:

Christopher R. Hepner (corresponding author)
Address: 1629 Thames Street, Suite 350, Baltimore, MD 21231, United States
E-mail: chepner3@jhu.edu

Nazbanou Nozari, MD, PhD
Address: 1629 Thames Street, Suite 350, Baltimore, MD 21231, United States
Phone: 443-287-1712
E-mail: nozari@jhu.edu

Abstract

The nature of working memory resources—in particular, their quantization (discrete vs. continuous)—has been studied extensively in the visual domain, with evidence supporting models with flexibly and continuously divisible resources. It remains unclear, however, whether similar mechanisms mediate the division of resources in phonological working memory. In three experiments, we show that, despite representational differences between visual and auditory domains, the principles of resource division are indeed similar in these domains. Exp. 1 tests slot vs. resource models, Exp. 2 gauges the effect of attention on resource division, and Exp. 3 investigates the influence of attention on different stages of working memory. Collectively, the results provide support for a resource model of phonological working memory and, more generally, point to similar computational principles governing the allocation of working memory resources.

Keywords: phonological working memory; cognitive resources; central executive; domain-generality; resource models; attention

Introduction

A central feature of working memory is its limited capacity (Baddeley, 2003). Despite the differences between the various theories of working memory, they all agree that people cannot hold an unlimited amount of information in working memory. This capacity limitation has given rise to questions like "How many items can be held in working memory, and how are they prioritized?", which has, in turn, led to the development of models of *resource allocation* in working memory. Given the very different nature of representations in visual and verbal domains of working memory (reflected in the division of the prominent multi-component model into separate visual and verbal subsystems: Baddeley & Hitch, 1974), a critical question is whether a single model of resource allocation is applicable to multiple domains—e.g., visual and verbal working memory. This paper answers this question.

The allocation of working memory capacity has been studied extensively in vision. Two types of models have been proposed: *slot models* (e.g., Cowan, 2001; Cowan, Rouder, Blume, & Saults, 2012) and *resource models* (e.g., Ma, Husain, & Bays, 2014). Slot models assume that there is a fixed upper bound *K* to the number of items that can be stored in working memory, that up to *K* items can be recalled with near-perfect accuracy, and that any additional items are lost completely and can only be guessed at random. Resource models, on the other hand, view working memory as a continuous resource with no fixed limit on the number of items that can be stored. Since this resource must be divided between items, any increase in set size, even from 1 to 2, will decrease the *precision* with which each of the items are remembered. Measuring the precision with which an item is remembered—and not simply whether it was remembered or not—requires a paradigm in which responses can be chosen from any point on a continuum, rather than selected from a set of discrete categories, so that the deviation of the response from the target can be measured. The visual domain offers a variety of features that fall on such a continuum: for example, the orientation of a line can range from 0 to 180 degrees from vertical, so the deviation between a response and a target can be measured as the absolute difference in orientations. For example, if a 45-degree target line is remembered as 50 degrees (response 1) and 55 degrees (response 2), the first response has less deviation (5 vs. 10 degrees) and is thus more precise. By contrast, linguistic stimuli are perceived more categorically (Liberman, Harris, Hoffman, & Griffith, 1957; but see Massaro & Cohen, 1983 for evidence of non-categorical perception in a different task): one can easily imagine the syllables /bɑ/ and /dɑ/, but not a whole range of /bɑ/-ish /dɑ/ and /dɑ/-ish /bɑ/ syllables in between. For this reason, the predictions of resource models remain, for the most part, untested in the domain of phonology. The current paper undertakes this challenge by developing a paradigm in which memory for auditory linguistic representations can be measured on a continuum, similar to the methods used to measure responses to visual stimuli on continua in visual working memory tasks.

We first confirm that participants are capable of non-categorical perception of phonological stimuli when asked to report them on a continuous scale. We then present three experiments closely mirroring those that have been used to answer questions about the mechanisms of resource allocation and item prioritization in visual working memory. Experiment 1 tests a key point on which the predictions of slot and resource models differ: the effect of set size on precision. To anticipate, the results show that the principles governing the allocation of phonological working memory capacity are in line with the predictions of resource models and similar to previous findings in visual working memory. Experiment 2 tests the influence of selective attention on resource allocation. The results, once again, show striking similarities between visual and phonological working memory. Finally, Experiment 3 investigates the stage at which attention influences the allocation of phonological working memory capacity.

Slot vs. Resource Models of Working Memory Capacity

Early investigations of working memory capacity focused on measurements of *span*, or the number of items that could be held in working memory at one time. In a typical experiment (e.g., Hayes, 1952), the participant heard a sequence of digits (e.g., "3, 2, 8"), letters (e.g., "M, Q, E"), or words (e.g., "cat, tree, wrench") and reported back as many items as they could remember. This was repeated with longer and longer sequences of items until the participant could no longer report all of them correctly. The results obtained in this fashion were remarkably consistent across a wide range of stimulus types: participants were able to recall a maximum of 7±2 items (Miller, 1956). Based on this observation, Miller (1956) concluded that working memory capacity was limited by the number of items, rather than their complexity or informational content.

This idea became the basis for slot models (e.g., Cowan, 2001), the simplest and most restrictive version of the limited resources account. In these models, working memory capacity is divided up into a fixed number of discrete slots, each of which can hold exactly one item. This "magic number" of slots—for example, Cowan's 4 (Cowan, 2001) or Miller's 7±2 (Miller, 1956)—places an upper bound on the number of items that can be stored in working memory. As long as the number of items N is less than the number of slots K, every item receives a slot. No matter which item is probed, it can be recalled with near-perfect accuracy and precision. However, when $N > K$, $N - K$ items do not receive slots; these items are lost completely and can only be guessed at random. If all items have an equal probability of being probed, the chance that an item that did not receive a slot will be probed, and thus the chance of guessing, is $(N - K)/N$. As N increases, the chance of guessing will also increase, resulting in decreased accuracy. Slot models therefore predict that accuracy will be at ceiling for small set sizes ($N \leq K$), but decrease as a function of set size beyond a certain threshold ($N > K$). This prediction was consistent with the findings of early working memory experiments that measured item-level accuracy: accuracy was perfect at small set sizes, then decreased linearly as a function of set size beyond a certain point (e.g., Sperling, 1960).

Unlike slot models, resource models (e.g., Ma et al., 2014) view working memory capacity as a continuous pool of resources that can be divided between an arbitrary number of items. The precision with which an item is stored in working memory is dependent on the amount of resources it receives. When there is only one item to remember, the entire pool of resources can be allocated to that item. When there are two items, the pool of resources must be divided in two, resulting in a decrease in the precision of the stored representations. In general, if resources are divided equally between items, then each item will receive $1/N$ of the available resources. Thus, all else being equal, the precision with which each item is stored will decrease (and deviation will increase) monotonically as a function of set size. Resource models can account for the older findings based on measures of binary accuracy in the following way: When the set sizes being compared are small (e.g., one vs. two items), the precision of the stored representations may still be sufficient to allow the correct response to be selected from a list of categories (e.g., red, orange, yellow, etc.) even at the larger of the two set sizes. Thus, resource models make the same predictions for the effect of set size on categorical accuracy as slot models do: accuracy should be at ceiling when set size is small but decrease as a function of set size when the precision of the stored representations is no longer sufficient to reliably select the correct response. Therefore, binary measures of accuracy (e.g., correct vs. incorrect) cannot distinguish between slot and resource models.

Slot and resource models can, however, be teased apart by measuring the precision instead of the accuracy of responses. As stated above, the prediction of the resource models is straight forward: a monotonic increase in response deviation (i.e. a decrease in response precision) with increasing set size. By contrast, retention in slot models is all-or-nothing: an item either gets a slot and is remembered with near-perfect precision, or it does not and must be guessed. Slot models thus do not predict variations in precision for stored items. However, when set size increases beyond K, the rate of random guessing increases, which increases the average deviation of the response from the target across trials, just like resource models. The critical test, then, is the effect of set size on precision at small set sizes ($N \leq K$). Slot models predict that response deviation should be negligible in all set sizes $N \leq K$, while resource models predict an increase in response deviation even from set size 1 to 2.

Response precision can be measured by having participants select a response from a continuum rather than from a set of discrete categories. A typical experiment (e.g., Wilken & Ma, 2004) employing this *continuous reproduction* (or *delayed estimation*) paradigm works as follows: an array of color swatches is briefly presented. After a short delay, a box corresponding to the location of one of the color swatches appears. The participant’s task is to select the color of the probed swatch on a continuous color wheel by clicking on the point which they believe corresponds to the exact hue of the probed color swatch. The dependent variable measured in this paradigm is the deviation of the participant’s response—in this example, the difference between the color the participant selected and the target color. This paradigm has been used extensively to study visual working memory. The findings have not been consistent with the predictions of classic slot models: across a wide range of studies, deviation has been found to increase monotonically as a function of set size, even from one to two items (Bays, Catalao, & Husain, 2009; Bays, Gorgoraptis, Wee, Marshall, & Husain, 2011; Gorgoraptis, Catalao, Bays, & Husain, 2011; Oberauer & Lin, 2017; Schneegans & Bays, 2016; van den Berg, Shin, Chou, George, & Ma, 2012; Wilken & Ma, 2004). In short, the empirical findings from visual working memory are compatible with both slot and resource models when binary accuracy is used as the dependent measure. However, the more sensitive measures of precision obtained using the continuous reproduction paradigm produce results that are better aligned with the predictions of resource models.

It is worth mentioning that a hybrid class of models, called *slots plus averaging models* (e.g., Zhang & Luck, 2008), has also been proposed. These combine the features of both slot and resource models. Since these models were developed primarily in response to the same criticisms against slot models that motivated the development of resource models, their predictions regarding the key findings that discriminate between them are very similar to resource models. For this reason, we will save further discussion of these models for the section of the General Discussion in which we discuss the findings of the experiments.

The Effect of Top-down Attention on Resource Allocation in Working Memory

To investigate the effect of top-down attention on resource allocation in working memory, the continuous reproduction paradigm can be altered such that one item in each set is given a higher priority than the others, e.g., by cueing one of the items to indicate that it is more likely to be probed. In a typical experiment (e.g., Gorgoraptis et al., 2011), an array of lines of various colors and orientations is briefly presented. After a short delay, one of the items is probed using a line of the same color. The participant's task is to rotate the line to match the orientation of the probed item. In some of the trials, the participant is instructed in advance to attend to a line of a particular color, since this line (the *cued* item) is more likely to be probed than the other lines in the same set (the *uncued* items). In the *cued* condition, the cued item is probed. In the *uncued* condition, one of the uncued items is probed. The deviation of the participant's responses in these two conditions can be compared to a *baseline* condition, in which none of the lines is cued. A variety of studies using variations of this paradigm have found lower deviation in the cued condition compared to the baseline and higher deviation in the uncued condition (Bays et al., 2011; Gorgoraptis et al., 2011; Oberauer & Lin, 2017; Pertzov, Bays, Joseph, & Husain, 2013), indicating a direct involvement of selective attention in the allocation of resources in visual working memory.

The next logical step is to probe the stage at which the allocation of resources in working memory is susceptible to the influence of selective attention. To this end, several studies manipulated the timing of the cues. *Pre-cues*, which appear before stimulus presentation, could potentially affect both encoding and maintenance in working memory: cued items could be encoded with higher precision, and the effects of decay and/or interference during the maintenance period could be reduced. *Retro-cues* (i.e., *retroactive* cues), which appear after the stimuli have disappeared, only have the potential to affect maintenance, since the encoding phase is over by the time a retro-cue is presented. Both pre-cues (Bays et al., 2011; Gorgoraptis et al., 2011; Oberauer & Lin, 2017) and retro-cues (Oberauer & Lin, 2017; Pertzov et al., 2013) have been found to affect deviation for cued and uncued items relative to baseline in visual working memory. To our knowledge, however, only one study (Oberauer & Lin, 2017) directly compared the effect of pre-cues to retro-cues on cued and uncued items; in that study, cued items benefitted much more from pre-cues than retro-cues (although the effect of retro-cues was still significant), while uncued items were affected equally negatively by both pre-cues and retro-cues. These findings can be accommodated by a model in which resource allocation remains flexible during the maintenance period, even after the stimuli have disappeared, such that resources can still be taken away from the non-prioritized items. Prioritized items, however, mainly benefit from receiving more resources during encoding, with a small advantage during the maintenance phase from the reassignment of the resources from the uncued items.

In short, a resource model in which resource allocation is influences by attentional cues and remains flexible even during the maintenance phase can account for the findings reported in visual working memory better than competing slot models. But can the same model be applied to phonological working memory? We discuss the challenges involved in the next section.

From Vision to Language

As the previous section makes clear, slot and resource models cannot be distinguished from each other on the basis of binary accuracy measures. Doing so requires stimuli with features that vary continuously. In the visual domain, finding such features is trivial. There are a wide variety of visual features (e.g., color, orientation, and spatial frequency) that can take any value along a continuum. For example, people can perceive blue, green, and a whole spectrum of colors in between. In the phonological domain, although the acoustic properties of speech sounds vary continuously, evidence from discrimination tasks (e.g., Liberman et al., 1957) suggests that people perceive phonemes categorically, e.g., as a /k/ or a /g/, but not as something in between. Categorical perception may serve a purpose: when processing linguistic input, the ability to reliably distinguish between phonemes is essential, since misidentifying a single phoneme can dramatically alter the meaning of a word (e.g., “back” vs. “bag”). It may be for this reason that humans are much better at detecting acoustic differences that cross category boundaries than differences of exactly the same magnitude that do not (Liberman et al., 1957).

At first glance, this seems problematic for the application of resource models to phonological working memory. If the perception of phonological stimuli is entirely categorical, then precision cannot be measured. However, the tendency to perceive phonemes categorically is not absolute. The degree to which participants perceive speech sounds categorically depends critically on the nature of the task they are performing (Gerrits & Schouten, 2004; Schouten, Gerrits, & van Hessen, 2003). For example, Massaro and Cohen (1983) have found that participants are capable of rating /k/-ish /g/ and /g/-ish /k/ sounds on a continuous scale between a prototypical /k/ and a prototypical /g/. In other words, they are capable of *non-categorical perception*. It is important to note that non-categorical perception is not simply a laboratory effect; it has consequences for language processing in real life. For example, gradient (i.e., non-categorical) acoustic information has been found to influence subsequent processing in spoken language comprehension (Brown-Schmidt & Toscano, 2017). Similarly, certain kinds of speech errors have been shown to involve co-production of target and error sounds, potentially reflecting gradient co-activation of the corresponding representations in the production system (Goldrick & Chu, 2014). Thus, despite surface differences in how people perceive visual and phonological information, both types of information are perceived and represented in enough detail to allow for fine-grained measurements of the deviation of a response from the target.

To adapt the continuous reproduction paradigm to verbal working memory, Joseph et al. (2015) used synthetic vowels with acoustic features that varied along a continuum. Participants heard a sequence of 1, 2, or 4 syllables consisting of a synthetic vowel followed by a consonant

(/d/). At the end of the sequence, one of the syllables was probed by presenting a number on a screen (1 for the first syllable, 2 for the second, etc.). Participants then heard a synthetic vowel from a random point on the continuum played continuously and responded by turning a dial to manipulate the vowel until it matched the one that had been probed. The deviation of each response was then recorded as the distance between the probed vowel and the participant's response on the continuum. Joseph et al. (2015) analyzed these deviations using a modified version of the Zhang and Luck (2008) mixture model designed to decompose responses into three categories: continuous responses, categorical responses, and random guesses. Critically, they found that the deviation of the continuous responses increased monotonically as a function of set size, even after categorical and random responses had been accounted for by the mixture model. Joseph et al. (2015) argue that this pattern is consistent with the predictions of resource models of phonological working memory.

In addition to replicating the set size effects found in visual working memory, Joseph et al. (2015) show that the continuous reproduction paradigm can be successfully adapted to linguistic items like speech sounds. However, their specific implementation of the paradigm is subject to some significant limitations. First, to make it possible to apply the modified Zhang and Luck (2008) mixture model, the authors needed their stimuli to come from a circular continuum. This is problematic because acoustic features form linear rather than circular continua; while colors, for example, naturally form a continuum that curves back on itself (i.e., red → orange → yellow → green → blue → purple → red again), phonological features do not. In other words, increasing the voice onset time (VOT) of /g/ from 0 ms to 60 ms will turn it into /k/, but continuing to increase the VOT will not turn /k/ back into /g/. The only way around this is to manipulate two acoustic features simultaneously in order to create an arbitrary circle in a two-dimensional feature space, as Joseph et al. (2015) did to create their synthetic vowels. The result is a continuum with variation along multiple dimensions simultaneously—variation which, in some portions of the continuum, may not be relevant in normal speech. Second, the use of a single stimulus continuum (also typical of experiments in visual working memory) introduces a potential confound. Since all of the stimuli presented in a trial vary along the same feature dimensions, it is difficult to disentangle the effects of increased interference (e.g., Oberauer & Kliegl, 2006; Oberauer & Lin, 2017) from the effects of dividing working memory resources into smaller shares. We will return to this point in the General Discussion.

The Current Study

In the current study, we implement a novel version of the continuous reproduction paradigm, adapted from the syllable rating task in Massaro and Cohen (1983), to measure phonological working memory performance. Our version of this paradigm incorporates three key differences from those used by Joseph et al. (2015) and in visual working memory:

1. *Linear response spaces*. Instead of the circular response spaces typically used in the continuous reproduction paradigm, we use the linear response spaces from Massaro and Cohen's (1983) syllable rating task: participants respond by indicating the position of the probed stimulus on a continuum between two extremes (e.g., between prototypical /k/ and prototypical /g/). The phonemes at the ends of the continuum differ in a single distinctive feature (e.g., [−voice] vs. [+voice]), so the acoustic properties of the stimuli at various point on the continuum vary along perceptually relevant dimensions (e.g., VOT).

2. *Multiple stimulus continua*. To minimize the potential for interference between the stimuli in a trial, we ensure that each of the stimuli comes from a different continuum. Specifically, we ensure that the stimuli come from continua that manipulate different distinctive features. There is thus no overlap in the relevant phonological or acoustic properties of the stimuli.

3. *Compensation for perceptual distortion*. Since distortions of perceptual space have been reported in both visual (Bae, Olkkonen, Allred, Wilson, & Flombaum, 2014) and verbal working memory (Joseph et al., 2015), we take these distortions into account in our analyses. To accomplish this, we included a baseline phase in the paradigm in which precision of perceptual judgments were assessed under minimal working memory load. We compare each participant's responses in the continuous reproduction task to their own responses in a baseline phase, rather than the actual position of the target.

Using this paradigm, we conducted three experiments to determine whether the key empirical findings in the visual domain could be replicated in phonological working memory. If these findings (i.e., the set size and cue effects discussed above) can be replicated, we can conclude that working memory capacity is allocated according to the same principles in both visual and verbal domains. If not, domain-specific principles of resource division must be proposed.

Experiment 1 tested the contrasting predictions of slot and resource models regarding set size effects. We manipulate the number of syllables presented in each trial and measure response deviation as a function of set size in sets containing 1, 2 and 4 stimuli. In Experiment 2, we manipulated attentional cues to determine their effect on the allocation of resources to cued and uncued items using a set size of 4. Finally, in Experiment 3, we replicated the results of Experiment 2 with a set size of 3 and a longer delay (2 s instead of 1 s). In addition, we manipulated the timing of the attentional cues (i.e., pre- vs. retro-cues) in order to determine whether or not resources can be reallocated after encoding is complete.

Experiment 1

Participants

A pre-defined target sample size of 48 was chosen, and participants who did not pass the exclusion criteria ($N = 2$; see Results) were replaced. Thus, 50 native speakers of American English (31 females, $M_{age} = 43.2$, age range: 24–65 years) participated for payment through Amazon Mechanical Turk (AMT; https://www.mturk.com). Consent was obtained under a protocol approved by the Johns Hopkins Medicine Institutional Review Board.

Materials

Stimuli were 28 syllables, seven from each of four acoustic continua: /bɑ/–/dɑ/, /kɑ/–/gɑ/, /ɹɑ/–/lɑ/, and /sɑ/–/ʃɑ/. The syllables at the ends of the continua (e.g., the most /bɑ/-like syllable, /ba/–/da/-1, and the most /dɑ/-like syllable, /ba/–/da/-7 on the /bɑ/–/dɑ/ continuum) were recordings of a native speaker of American English. The five intermediate syllables on each continuum were created by progressively changing the acoustic properties of the initial consonant in Praat (Boersma & Weenink, 2016) using scripts from Winn (2014) to create five equally-spaced consonants between the two recorded syllables while leaving the vowel (/ɑ/)

unchanged. To minimize interference, a different feature was manipulated in each continuum. We varied (1) place of articulation between [−coronal] and [+coronal] (rising vs. flat second formant transition; Liberman, Delattre, Cooper, & Gerstman, 1954) in the /bɑ/–/dɑ/ continuum and (2) manner of articulation between [−lateral] and [+lateral] (rising vs. flat third formant transition; O'Connor, Gerstman, Liberman, Delattre, & Cooper, 1957) in the /ɹɑ/–/lɑ/ continuum using the modified linear predictive coding (LPC) decomposition and re-synthesis procedure described by Winn and Litovsky (2015). (3) Place of articulation was varied between [+anterior] and [−anterior] (higher vs. lower spectral peak; Hughes & Halle, 1956) in the /sɑ/–/ʃɑ/ continuum by blending the initial fricative noises. (4) Finally, voicing was varied between [−voice] and [+voice] (longer vs. shorter voice onset time; Liberman, Delattre, & Cooper, 1958) in the /kɑ/–/ɡɑ/ continuum by decreasing the duration of the aspiration in /kɑ/ in 10 ms increments to create stimuli with voice onset times ranging from 60 ms (most /kɑ/-like) to 0 ms (most /ɡɑ/-like).

Procedures

The experiment was developed using jsPsych (de Leeuw, 2015), a JavaScript library for running behavioral experiments in a browser. PsiTurk (McDonnell et al., 2016) was used to integrate the experiment with AMT. To provide us with more data, each participant completed two sessions 24–72 hours apart with the same structure but a different trial order. Each session consisted of two phases: a *baseline phase* and a *working memory phase*.

The baseline phase. The baseline phase was adapted from Massaro and Cohen (1983), and was divided into four blocks, one for each acoustic continuum, the order of which was randomized across participants in each session. In each block, participants first completed an orientation in which all seven syllables along the continuum were played in order. As each syllable was played, its position was shown on a slider at the center of the screen visually representing the range between the most extreme syllables on the continuum (e.g., between the most /bɑ/-like syllable at the left end, labelled “B”, and the most /dɑ/-like syllable at the right end, labelled “D”). For example, when /bɑ/–/dɑ/-1 was played, the slider was set all the way to the left. When the /ba/–/da/-2 was played, the slider was set one sixth of the way from B to D, and so on, until the last syllable played and the slider was set all the way to the right (Figure 1a). This orientation procedure was repeated four times to give participants enough opportunities to learn the relationship between the syllables on the acoustic continuum and the corresponding positions on the visual slider. Their ability to map the syllables onto these positions—in the terminology used by Massaro and Cohen (1983), “rate” them on a scale, e.g., from most /bɑ/-like to most /dɑ/-like—was then tested.

Once the orientation was over, participants were tested on their ability to rate syllables (*baseline test*). They listened to the same syllables, this time presented in a random order, and indicated the position of each one on the continuum using the same slider (Figure 1a). Remember that, although we had seven syllables from each continuum, participants could adjust the slider continuously. Once participants had adjusted the slider to their satisfaction, they pressed a “submit” button and the position was recorded on a scale from 1 to 100 (participants did not see these numbers). If, for example, a participant had set the slider exactly at the correct position r upon hearing /bɑ/–/dɑ/-3, the response would have been recorded as 34. Only one syllable was played in each trial and there was no deadline for responding. The baseline test in each block consisted of 14 practice trials (two per syllable) followed by 56 experimental trials (eight per syllable). Thus, across the two sessions, participants completed a total of 448 experimental baseline test trials (16 for each of the seven syllables in each of the four continua).

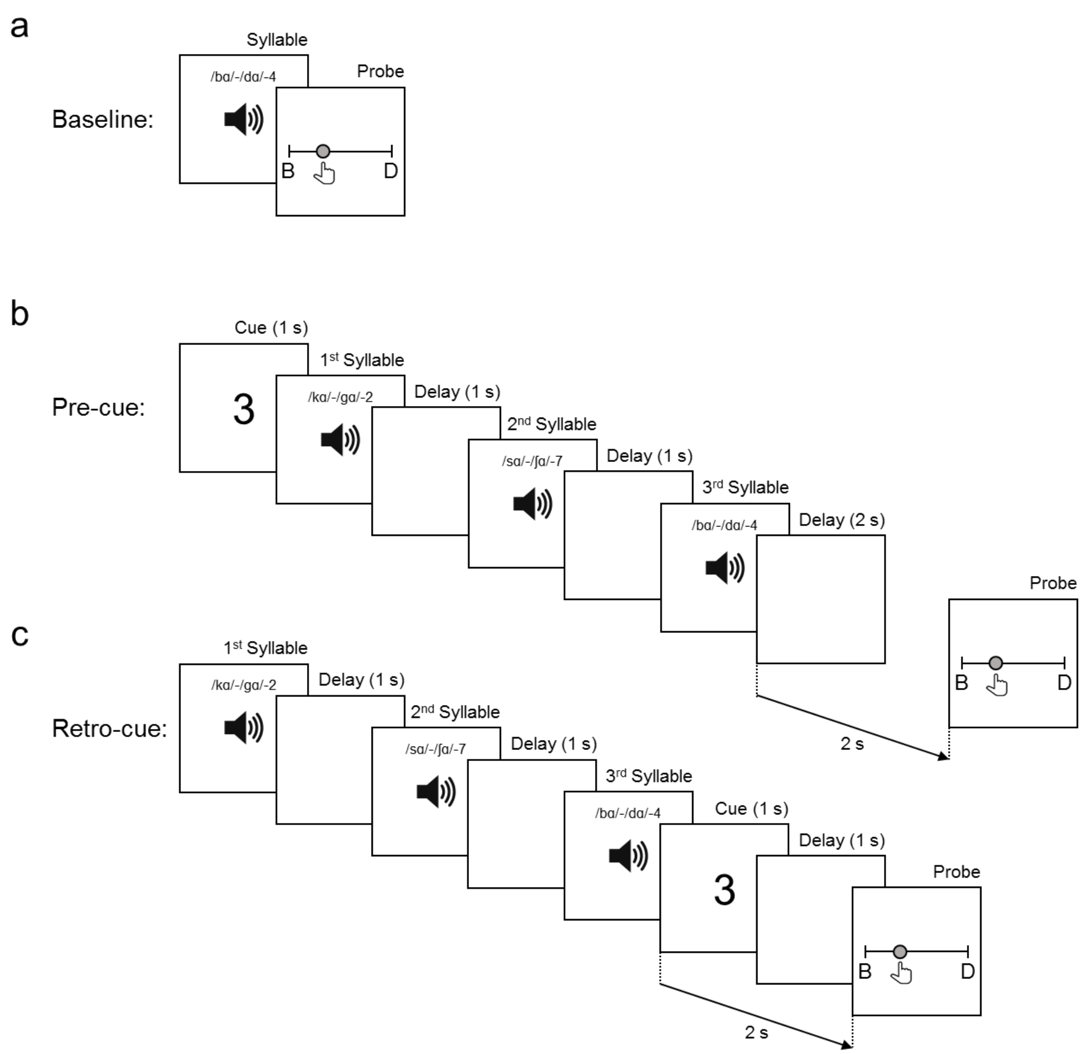

Figure 1. Examples of the trials in Experiments 1–3. The probed item in each of these trials is the second syllable on the /ba/-/da/ continuum. (a) The baseline test trial used in all three experiments. (b) Pre-cue and (c) retro-cue trials in Experiment 3. The trial structure in Experiment 2 was similar to the pre-cue condition, except that the delay was 1 second instead of

2 and 4 syllables were presented instead of 3. Experiment 1 followed the same format as Experiment 2 but with varying numbers of syllables (1, 2, or 4) and no cues.

The working memory phase. This phase tested the effect of set size on precision in working memory. On each trial, participants were presented with a sequence of one, two, or four syllables from different acoustic continua played at 1 s intervals. One second after the final syllable was played, the slider appeared, and participants had 1 s to rate the relevant syllable on the slider. Since two syllables from the same continuum were never played during the same trial, the labels on the slider unambiguously indicated which syllable to rate. In each session, there were 15 practice trials, followed by 12 blocks of 28 experimental trials with pseudorandomized order, such that no more than two consecutive trials had the same set size. Across sessions, participants completed a total of 672 experimental working memory trials (224 for each of the three set sizes). The design was fully counterbalanced, so each syllable was probed the same number of times (eight) in each set size for each participant. Within each set size, each syllable appeared the same number of times in each position.

Statistical Analyses

The exclusion criterion. To test the critical predictions of the resource model, it was necessary to establish that participants were attending to the task and had learned to indicate the positions of the syllables using the slider. At a minimum, there should be a reliable positive correlation between the position of a syllable within a continuum (e.g., two sixths of the way from /bɑ/ to /dɑ/) and the participant's ratings for that syllable (e.g., 34). To implement this criterion (the *correlation criterion*), we tested for a positive Spearman correlation between the positions of the syllables in each of the four continua and each participant's ratings for those syllables during the baseline test. Participants for whom the p-value of a positive correlation was greater than .05 for one or more continua were excluded.

The dependent variable. To measure the deviation (and thus the precision) of the responses in the working memory phase, we calculated the absolute deviations from the median ratings obtained in the baseline phase in three steps: (1) We calculated the median of the participant's 16 ratings for the same syllable in the baseline phase. (2) We then subtracted this baseline median from the response. If, for example, the median of the participant's ratings was 30, the results for the responses 33 and 29 in the working memory phase would be 33 - 30 = 3 and 29 - 30 = -1, respectively. (3) Finally, we took the absolute value of the number from (2) to get the deviation of the response. The deviations calculated in this way were the dependent variable in all three experiments.

Statistical models. The main analyses in this study were carried out with linear mixed-effects modeling (LMEM) using the lme4 package (version 1.1-14; Bates, Mächler, Bolker, & Walker, 2015) in R (version 3.4.2; R Core Team, 2017). We strove to include the maximal random effects structure tolerated by the model, in keeping with the suggestions of Barr, Levy, Scheepers, and Tily (2013). All numeric variables were centered and scaled, and the dependent variable (deviation) was log-transformed to approximate a normal distribution. The p-values were calculated based on Satterthwaite approximations using the lmerTest package (version 2.0-33; Kuznetsova, Brockhoff, & Christensen, 2016).

The data and analysis scripts for all experiments are available on the Open Science Framework at osf.io/943u7 (Hepner & Nozari, 2018).

Results

Two participants from the original sample of 48 did not pass the correlation criterion and were replaced to obtain the target sample size of 48.

Baseline test. Figure 2 shows (a) the distributions and (b) the medians of participants' ratings for each continuum in the baseline phase.

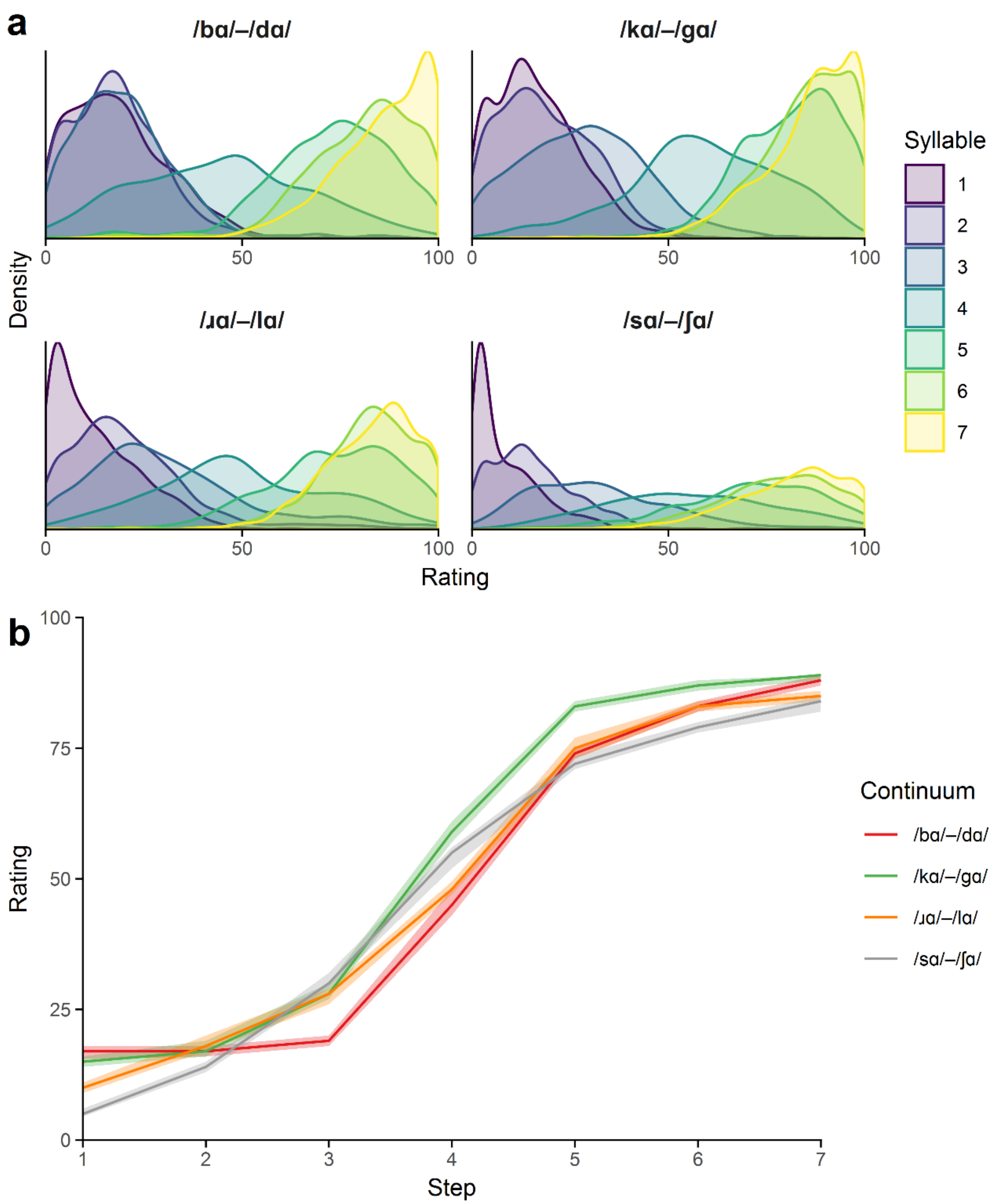

Figure 2. Participants' ratings for the syllables in the baseline test in Experiment 1. (a) Distributions and (b) medians of the ratings for the syllables on each of the four continua. The colors in (a) represent the seven acoustically-manipulated syllables (steps) on each continuum. The critical finding is the gradual progression of the peaks of the distributions for the syllables

from steps 1 to 7 on all continua, compatible with non-categorical perception, as opposed to only two peaks on the extremes. The lower panel shows the same effect in the median ratings for the steps along each continuum: a progression from steps 1 to 7 rather than an abrupt shift from 1 to 100 at step 4.

To determine whether participants had been able to rate the syllables continuously, rather than categorically, we analyzed the ratings using uninformed mixture modelling by means of the mclust package (version 5.3; Fraley & Raftery, 2002) in R (R Core Team, 2017). If participants were rating the syllables continuously, the overall distribution of the ratings should be a mixture of seven distributions centered on or near the "correct" rating for each syllable. For example, there should be a distribution centered on 1 (the lowest possible rating on the scale from 1 to 100) for the first syllable, another distribution centered 17.5 (one sixth of the way across the scale), and so on, up to the seventh distribution, which should be centered at 100 (the highest possible rating on the scale). Table 1 shows the relationship between the correct rating for each syllable and the means of the distributions fitted to participants' actual ratings. As can be seen, the differences are small, with a root-mean-square deviation (RMSD) of 4.27, compared to a distance of 16.5 between the target ratings for adjacent syllables on a continuum. To formally test whether seven distributions provided a better model for the data than two distributions near the ends of the rating scale (as would be expected if participants were rating categorically), we also fitted a model with only two distributions and compared the fit of the two models using the Bayesian information criterion (BIC), which penalizes for additional parameters. The BIC for the seven-distribution model (196,329) was much lower than the BIC for the two-distribution model (201,838; a difference of 5,509), providing very strong evidence against the two-distribution model according to the criteria given by Kass and Raftery (1995). These results indicate that the participants were able to perceive and rate the syllables continuously. Next, we test the effect of set size on ratings for the same syllables in the working memory phase.

Table 1. Comparison of an uninformed mixture model fitted to the ratings in the baseline phase of Experiment 1 to the target ratings for each syllable.

	Mean Rating		
Syllable	*Target*	*Model*	**Difference**
1	1.00	2.00	1.00
2	17.50	12.18	-5.32
3	34.00	27.10	-6.90
4	50.50	50.43	-0.07
5	67.00	72.34	5.34
6	83.50	88.05	4.55
7	100.00	98.83	-1.17

Working memory. Figure 3a shows the relationship between set size and deviation at each position. Before we analyze the effect of set size on working memory, we must first establish that the data indeed reflect working memory performance. A reliable index for this purpose is the presence of the well-known serial position effects: working memory performance is typically an inverted U-shape function of position with better performance for the first (primacy effect) and last (recency effect) items (e.g., Atkinson & Shiffrin, 1971; Caramazza, Miceli, Villa, & Romani, 1987; Healy, 1974; Murdock, 1968). We fitted a model to data from trials with set size four (smaller set sizes do not allow for clear testing of position effects) to test (a) whether the canonical position effects were obtained, and (b) which covariates needed to be included in subsequent models to control for the effects of nuisance variables.

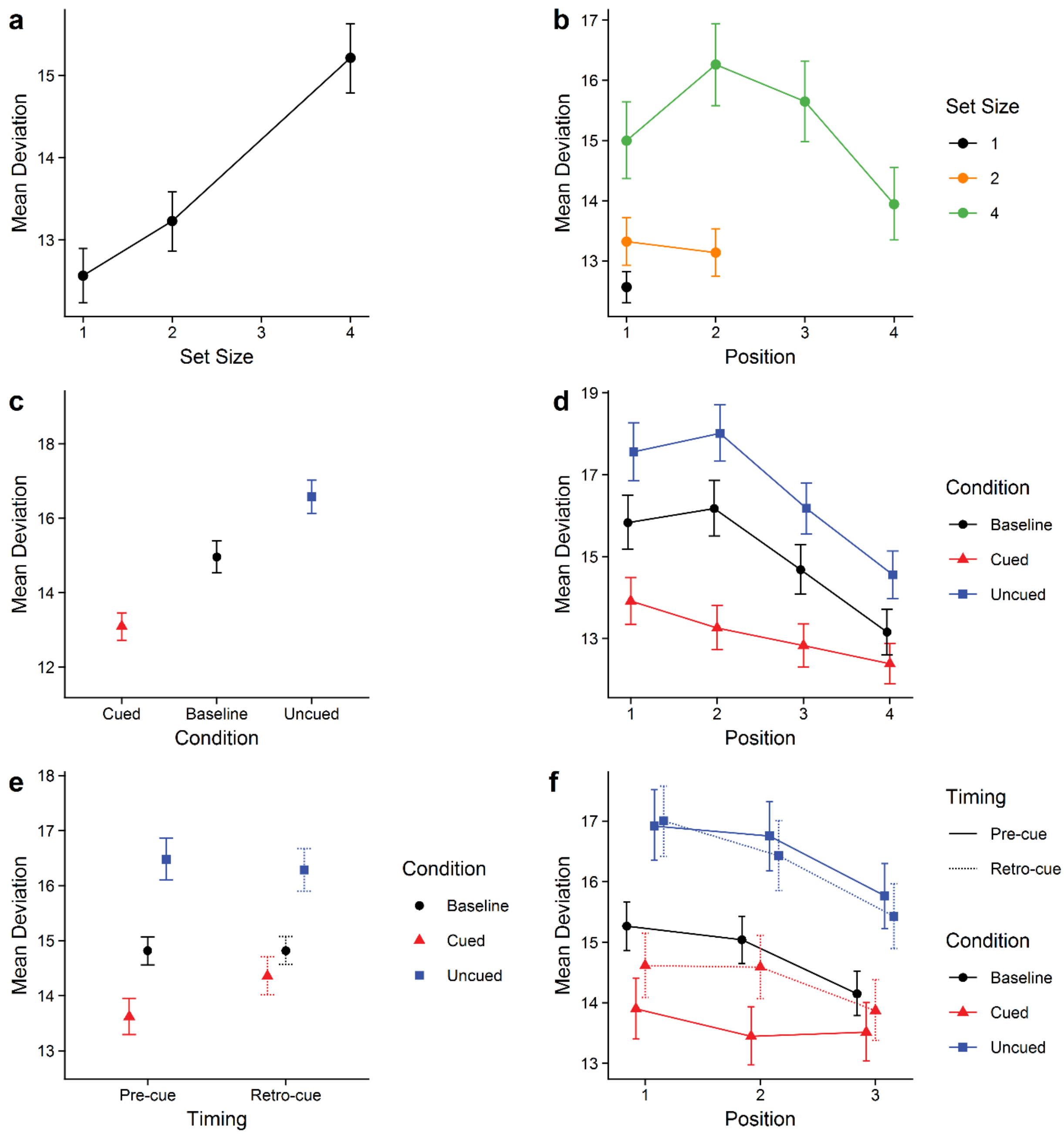

Figure 3. (a) The main effect of set size (1, 2, and 4) on deviation in Experiment 1, and (b) the interaction between set size and position. (c) The main effect of cue condition (*cued* and *uncued* vs. *baseline*) in Experiment 2, and (d) the interaction between cue condition and position. (e) The main effects of cue condition within each cue timing (*pre-* and *retro-cues*) in Experiment 3,

and (f) the interaction between cue condition, cue timing, and position. Error bars represent 95% within-subjects confidence intervals (Cousineau, 2005).

This model included position as a set of polynomial contrasts, with separate fixed effects for linear and quadratic (i.e., U-shaped) serial position effects, along with random slopes for each of these by subject and item. The model also included fixed effects for several nuisance variables: (a) *baseline median*, which was the absolute value of the distance between the participant's baseline rating for the syllable and the center of the rating scale, to account for the reduction in variability at the ends of the scale (see Figure 2a); (b) *baseline variability*, which was the standard deviation of the participant's baseline rating for the syllable; and (c) *session*, which was coded as a contrast between the first and second sessions. We also included random intercepts for participants and items, i.e., syllables. The results of this model can be seen in Table 2. Critically, there was a significant linear effect of position (t = -2.70, p = .012), indicating an overall decrease in deviation as a function of position, and a significant quadratic effect of position (t = -4.46, p < .001), indicating a decrease in deviation at either end of the sequence. There were also main effects of baseline median (t = -10.33, p < .001), corresponding to a decrease in deviation closer to the ends of the rating scale, and baseline variability (t = 10.17, p < .001), but not session (t = -0.66, p = .509). These findings suggest that (a) deviation indeed reflects working memory performance, and (b) both baseline median and baseline variability have significant influence on deviation. We thus included these covariates in all subsequent analyses. We also included serial position and its interaction with set size because serial position alone could create spurious set size effects (see Olson, Romani, & Caramazza, 2010).

Table 2. Results of the linear mixed effects model for serial position effects in set size 4 of Experiment 1.

Fixed effects	**Coefficient**	**SE**	**t**	**p-value**
Intercept	0.018	0.041	0.442	.659
Position-linear	-0.059	0.022	-2.701	.012
Position-quadratic	-0.102	0.023	-4.460	< .001
Baseline median	-0.136	0.013	-10.328	< .001
Baseline SD	0.122	0.012	10.173	< .001
Session	-0.012	0.018	-0.660	.509

Random effects

Subject	Variance
Intercept	0.0297
Position-linear	0.0012
Position-quadratic	0.0014

Syllable	Variance
Intercept	0.0066
Position-linear	0.0031
Position-quadratic	0.0043

The main prediction of the resource model investigated in this experiment—that deviation would increase as a function of set size—was tested using an LMEM with fixed effects for set size, serial position and the interaction between the two, along with the baseline median and baseline variability as covariates. The random effect structure included random intercepts for participants and syllables and random slopes for set size, position, and the interaction between the two by participant and syllable. Table 3 presents the results of this analysis. Most importantly, the model revealed a significant main effect of set size ($t = 5.32$, $p < .001$), with deviation increasing as a function of set size.

Table 3. Results of the linear mixed effects model for set size effects in Experiment 1.

Fixed effects	**Coefficient**	**SE**	**t**	**p-value**
Intercept	0.001	0.029	0.035	.972
Set size	0.080	0.015	5.318	< .001
Position	-0.021	0.015	-1.351	.187
Baseline median	-0.164	0.008	-20.662	< .001
Baseline SD	0.147	0.007	21.165	< .001
Set size × position	-0.002	0.010	-0.148	.883

Random effects

Subject	Variance
Intercept	0.0273
Set size	0.0050
Position	0.0004
Set size × position	0.0000

Syllable	Variance
Intercept	0.0059
Set size	0.0017
Position	0.0026
Set size × position	0.0009

There was no interaction between set size and position; however, to confirm that the effect of set size was robust across positions, we conducted additional post-hoc analyses. Four post hoc tests included comparisons between: (a) the initial positions of set sizes 1 and 2, (b) the final positions of set sizes 1 and 2, (c) the initial positions of set sizes 2 and 4, and (d) the final positions of set sizes 2 and 4. For each model, we compared the mean difference in deviation between the two set sizes across participants to the distribution generated by a Monte Carlo simulation with 1,000,000 permutations, resampling within participants. After Bonferroni correction for multiple comparisons, the set size effect was significant in all cases: (a) $M = 0.76$, 95% CI = [0.35, 1.18], $p < .001$, (b) $M = 0.57$, 95% CI = [0.18, 0.97], $p = .009$, (c) $M = 1.68$, 95% CI = [1.05, 2.31], $p < .001$ and (d) $M = 0.80$, 95% CI = [0.19, 1.44], $p = .018$.

We conducted an additional analysis to determine whether or not the increase in deviation as a function of set size reflected an upper bound on the number of items that could be stored in phonological working memory. Two post-hoc tests compared deviations at set sizes 2 and 4 to the minimum deviations predicted by slot models with 1 and 3 slots, respectively. If phonological working memory can store at most $K < N$ items, then $(N - K)/N$ of the responses at set size N will be random guesses. Assuming that the deviation of the remaining responses (i.e., those corresponding to stored items) is no larger than the deviation at set size 1 (i.e., assuming no interference from the additional items), the minimum deviation D_N predicted by a model with an item limit $K < N$ is a weighted average of the deviation at set size 1 (D_1) and the deviation of random guesses (D_G):

$$D_N = \frac{K}{N} D_1 + \frac{N - K}{N} D_G$$

For each test, the observed deviation of each participant's responses to each syllable at set size N were compared to the predicted D_N, using the participant's responses to the same syllable at set size 1 for D_1. D_G was obtained by a Monte Carlo simulation with 100,000 permutations, resampling the participant's responses at set size 1 within the same continuum. After Bonferroni correction for multiple comparisons, the differences between observed and predicted deviations were significant, with smaller than predicted deviations at both set sizes; set size 2 vs. a 1-item limit: $M = 10.57$, 95% CI = [10.34, 10.82], $p < .001$; set size 4 vs. a 3-item limit: $M = 2.81$, 95% CI = [2.69, 2.93], $p < .001$. We can thus rule out the possibility that phonological working memory capacity was less than the largest set size tested in this experiment (i.e., 4).

Discussion

Analysis of the baseline phase confirmed that participants were able to perceive and rate the syllables continuously, as previously reported by Massaro and Cohen (1983). This finding makes these materials appropriate for testing the predictions of a resource model of working memory. If such a models is appropriate for phonological working memory, then the precision with which the syllables are recalled should decrease as set size increases. Critically, this decrease in precision should be visible for any increase in set size, even from one to two syllables. The results of Experiment 1 confirmed this prediction. Deviation increased significantly as a function of set size. This increase was robust in position-specific analyses which also took primacy and recency effects into account, and could not be attributed to an increase in guessing caused by insufficient slots. Thus, despite the clear differences between visual and auditory stimuli, the results of Experiment 1 closely resembled those found in the visual domain (e.g., Bays et al., 2009; Wilken & Ma, 2004), and were in full accord with the predictions of a resource model of working memory.

In Experiment 2, we tested the influence of attentional cues on the allocation of resources to cued and uncued items. In visual working memory, cued items are recalled with greater precision compared to a baseline with no attentional cues, at the expense of uncued items in the same set (Gorgoraptis et al., 2011). Experiment 2 adapted this manipulation to phonological working memory.

Experiment 2

Participants

Similar to Experiment 1, a pre-defined target sample size of 48 was chosen, and participants who did not pass the exclusion criteria ($N = 4$; see Results) were replaced. Thus, 52 native speakers of American English (27 females, $M_{age} = 36.6$, age range: 21–58 years) participated for payment through AMT and the data from 48 of them were analyzed. Consenting procedure was the same as Experiment 1.

Materials

The materials were the same as Experiment 1.

Procedures

The same two-session design as Experiment 1 with a similar session structure was used. The baseline phase was unchanged.

The working memory phase kept the set size at four on all trials, but manipulated the presence/absence and validity of cues before the presentation of syllables. On 1/3 of the trials, no cue was presented. These *baseline* trials were identical to the set size 4 trials in Experiment 1. On the other 2/3 of the trials, a cue (a number between 1 and 4) was presented at the beginning of the trial. This cue indicated that the syllable in the corresponding position (1 to 4) had a 50% chance of being probed (1/3 of all trials; *cued* trials). The other syllables each had a 16.7% chance of being probed (1/3 of all trials; *uncued* trials).

On cued trials, the cue appeared for 1 s, after which participants pressed the corresponding number on the keyboard to confirm the identity of the cue. Failure to press the correct key on more than 5% of the trials was deemed an exclusion criterion (cue check 1). After 1 s, the four syllables were presented and one of them was probed in the same way as in Experiment 1. To rule out the possibility that errors might be due to forgetting the cues, we included check questions on 18% of the trials. The question asked whether or not the cued syllable was the one probed. Failure to respond significantly above chance (determined using an exact binomial test) was deemed an exclusion criterion (cue check 2).

Participants completed three blocks (one block with *baseline* trials and two blocks with a mixture of *cued* and *uncued* trials) in counterbalanced order. Each block consisted of 14 practice trials and 112 experimental trials, with breaks between sets of 28 trials, for a total of 224 trials in each cue condition (*baseline*, *cued*, and *uncued*) across both sessions. Each syllable was probed exactly once in each position in each block, resulting in a total of 8 samples for each syllable in each cue condition from each participant.

Results

Two participants did not pass the correlation criterion, and two did not pass the cue checks, and were replaced to meet the sample size goal of 48 participants. Deviation was calculated in the same manner as before. Figure 3b shows deviation as a function of attentional cueing.

To test for effects of cue condition, we used an LMEM with fixed effects for cue condition (contrast-coded as *cued* vs. *baseline* and *uncued* vs. *baseline*), position, the interaction between cue condition and position, and the same covariates as Experiment 1. The random effect structure included random intercepts for participants and syllables, along with random slopes for cue condition, position, and the interaction between the two by participant and syllable. Table 4 presents the results of this analysis. *Cued* trials had significantly lower deviation (t = -3.34, p = .002) and *uncued* trials had significantly higher deviation (t = 2.76, p = 0.008) compared to *baseline* trials. There was also a significant main effect of serial position, with deviation decreasing for more recent syllables (t = -4.52, p < .001). Interestingly, there was also a reliable interaction between the *cued* condition and position (t = 2.14, p = .037), but no such interaction between the *uncued* condition and position (t = -0.35, p = .729).

To further explore the influence of cueing on serial position effects, we fit separate polynomial models (with linear and quadratic terms) to the *cued* and *baseline* conditions. In the *baseline* model, there were both significant linear (t = -5.02, p < .001) and quadratic (t = -2.38, p = .023) effects of position, whereas only the linear effect was significant (t = -3.34, p = .001) in the cued model. See Tables A1 and A2 in Appendix A for the full results of these analyses.

Table 4. Results of the linear mixed effects model for cue condition effects in Experiment 2.

Fixed effects	**Coefficient**	**SE**	**t**	**p-value**
Intercept	-0.007	0.033	-0.195	.846
Condition-uncued	0.108	0.039	2.763	.008
Condition-cued	-0.089	0.027	-3.341	.002
Position	-0.061	0.014	-4.522	< .001
Baseline median	-0.176	0.009	-20.347	< .001
Baseline SD	0.184	0.007	26.668	< .001
Condition-uncued × position	-0.005	0.013	-0.347	.729
Condition-cued × position	0.032	0.015	2.135	.037

Random effects

Subject	Variance
Intercept	0.0294
Condition-uncued	0.0598
Condition-cued	0.0139
Position	0.0023
Condition-uncued × position	0.0009

Condition-cued × position	0.0015

Syllable	Variance
Intercept	0.0121
Condition-uncued	0.0038
Condition-cued	0.0073
Position	0.0017
Condition-uncued × position	0.0001
Condition-cued × position	0.0011

Discussion

The results of Experiment 2 closely mirrored those reported in visual working memory (Gorgoraptis et al., 2011). Manipulation of attention led to a significant decrease in deviation for *cued* items, with a corresponding increase in deviation for *uncued* items, relative to *baseline* trials. This effect mirrored a phonological working memory effect previously reported at the word level. Nozari and Dell (2012) asked participants to recite 4-word tongue-twisters from memory, and recorded the phonemic migrations between words, e.g., “mist wing whiff mink” turning into “mist wing *miff wink*”. In three experiments, they manipulated attention such that one of the four words in the sequence was singled out (it was to be attended to explicitly in Exp. 1, to be prosodically emphasized in Exp. 2, and to be whispered in Exp. 3). In all three experiments, the attentional cueing decreased the rate of phonemic migrations on the cued word (i.e., the attended word was produced more precisely), and increased the rate of phonemic migrations on the uncued words, compared to the baseline. These findings show that the effect of attention on verbal working memory exists in both perception and production, and at the level of both isolated phonemes and phonemes in the context of words. Together with findings from visual working memory, one can conclude broader generalization of the principle of flexible resource allocation based on attention.

Cueing also reduced the effect of serial position on recall, as evidenced by the significant interaction between the *cued* condition and both linear and quadratic effects of position. In particular, the prototypical decrease in accuracy (or in this case precision) for items in the middle of a list, observed in previous studies (e.g., Atkinson & Shiffrin, 1971) and in the *baseline* condition for this experiment, was eliminated by cueing. This result indicates that serial position effects may, at least in part, reflect uneven allocation of resources by default, which can be counteracted for cued items by allocating more resources through attention. In fact, one of the main explanations that has been offered for serial position effects is this kind of uneven distribution of resources. For example, Page and Norris (1998) proposed that an activation gradient is established during encoding to protect earlier items from the effects of decay and interference from subsequent items. This effectively means that some (i.e., earlier) items are allocated more resources than later items, explaining the primacy effect in serial position. The interaction between this gradient and decay/interference results in the classic U-shaped serial position curve. What cues seem to do is to change this default gradient in favor of allocating more resources to the cued items, thus eliminating the primacy effect.

In summary, the results of Experiment 2 showed that the distribution of resources is flexible and can be influenced by attention. It is, however, unclear which components of working memory processing are affected by these top-down attentional cues: encoding, maintenance, or both. Experiment 3 was designed to answer this question by manipulating the timing of the cues. Cues preceding stimulus presentation (*pre-cues*) can affect both encoding and maintenance, but cues following stimulus presentation (*retro-cues*) can only affect maintenance. If *retro-cues* have no effect on precision, we can conclude that attention operates exclusively on encoding. If, on the other hand, *retro-cues* are effective in changing precision, attention must be able to affect maintenance. It is also possible that *pre-* and *retro-cues* could affect *cued* and *uncued* items differently, which could potentially shed light on the nature of resource reallocation during working memory processing. Experiment 3 was designed to (a) replicate the results of Experiment 2 with a different set size (3 instead of 4) and longer recall time (2 s instead of 1 s), and (b) to explore the effect of cue timing on phonological working memory.

Experiment 3

Participants

Since the number of trials per factor combination was reduced from eight in Experiment 2 to three in this experiment, the target sample size was increased by the same ratio, i.e., N = 128. Participants who did not pass the exclusion criteria (N = 12; see Results) were replaced. Thus, 140 native speakers of American English (84 females, M_{age} = 36.1, age range: 21–68 years) participated for payment through AMT and the data from 128 of them were analyzed. Consenting was similar to the previous experiments.

Materials

The materials were the same as Experiments 1 and 2.

Procedures

We used the same two-session design as the previous experiments, with a session structure similar to Experiment 2. While the baseline phase was unchanged, there were a few differences in the structure of the working memory phase. First, the number of syllables was decreased from four to three so that a replication of Experiment 2 could be obtained at a different set size. Second, while the working memory phase still contained two blocks with cues and one block without, the two cued blocks now had different cue timings: in the *pre-cue* block, the cue appeared for 1 s immediately *before* the presentation of the syllables (Figure 1b); in the *retro-cue* block, the cue appeared for 1 s immediately *after* the presentation of the last syllable (Figure 1c). Third, the delay between the presentation of the last syllable and the appearance of the probe was increased from 1 s to 2 s to allow for a delay between the presentation of the retro-cue and the appearance of the probe. The structure of the working memory phase was otherwise similar to Experiment 2, with cued and uncued items in each of the *pre-cue* and *retro-cue* blocks. There were 14 practice trials and 84 experimental trials in each block, with breaks at 21-trial intervals. Over two sessions, participants completed 168 trials in the *baseline* condition and 84 in each combination of cue condition (*cued* or *uncued*) and timing (*pre-* or *retro-cue*).

Results

Eight participants failed the correlation criterion and four failed the cue checks. These were replaced to reach the target number of 128 participants. Deviation was calculated in the same manner as before. Figure 3c shows deviation as a function of cue condition and timing.

Given the complexity of this dataset, we analyzed the data in several steps: first, we constructed a model comparing the *pre-cue* trials to *baseline* trials in an attempt to replicate the results of Experiment 2. This model included cue condition (coded as *cued* vs. *baseline* and *uncued* vs. *baseline*), position, and the interaction between cue condition and position as fixed effects of interest, along with the same covariates used in all of the previous models in this study. The random effect structure included intercepts for participants and syllables, as well as the slopes of the two cue conditions, position, and the interactions between them over participants and syllables. Table 5 shows that the results replicate those of Experiment 2: *cued* trials had significantly lower deviation (t = -3.29, p = .002), while *uncued* trials had reliably higher deviation (t = 5.01, p < .001), relative to *baseline* trials. There was also a marginal interaction between the *cued* condition and position (t = 1.92, p = .055), but no reliable interaction between the *uncued* condition and position.

Application of separate polynomial models (with intercept and linear terms given the 3 data points in each line), revealed a significant linear effect in the *baseline* condition (t = -3.32, p = .001) but not the *cued* condition (t = -0.13, p = .894). See Tables A3 and A4 in Appendix A for the full results of these analyses.

Table 5. Results of the linear mixed effects model for cue condition effects after pre-cues in Experiment 3.

Fixed effects	**Coefficient**	**SE**	**t**	**p-value**
Intercept	-0.011	0.026	-0.427	.671
Condition-uncued	0.104	0.021	5.009	< .001
Condition-cued	-0.057	0.017	-3.292	.002
Position	-0.022	0.007	-3.227	.002
Baseline median	-0.176	0.007	-24.384	< .001
Baseline SD	0.140	0.006	23.995	< .001
Condition-uncued × position	-0.009	0.011	-0.777	.438
Condition-cued × position	0.021	0.011	1.923	.055

Random effects	
Subject	Variance

Intercept	0.0276
Condition-uncued	0.0336
Condition-cued	0.0085
Position	0.0002
Condition-uncued × position	0.0002
Condition-cued × position	0.0002

Syllable	Variance
Intercept	0.0122
Condition-uncued	0.0013
Condition-cued	0.0032
Position	0.0002
Condition-uncued × position	0.0001
Condition-cued × position	0.0000

Next, we repeated the same analysis to compare the *retro-cue* and *baseline* trials. Since the model did not converge with random slopes for position, they were removed from the model. Table 6 presents the results of this analysis. While *uncued* trials showed a reliable increase in deviation relative to *baseline* trials ($t = 4.25$, $p < .001$), the decrease in deviation for *cued* trials was not significant ($t = -1.20$, $p = 0.235$). Neither of the interactions between position and cue condition reached significance.

Table 6. Results of the linear mixed effects model for cue condition effects after retro-cues in Experiment 3.

Fixed effects	**Coefficient**	**SE**	**t**	**p-value**
Intercept	-0.017	0.026	-0.653	.517
Condition-uncued	0.089	0.021	4.245	< .001
Condition-cued	-0.021	0.017	-1.204	.235
Position	-0.022	0.006	-3.543	< .001
Baseline median	-0.189	0.007	-25.990	< .001
Baseline SD	0.141	0.006	24.165	< .001
Condition-uncued × position	-0.018	0.011	-1.635	.102
Condition-cued × position	0.014	0.011	1.314	.189

Random effects

Subject	Variance
Intercept	0.0275
Condition-uncued	0.0322
Condition-cued	0.0066

Syllable	Variance
Intercept	0.0113
Condition-uncued	0.0020
Condition-cued	0.0034

Next, we directly compared the effects of *pre-* and *retro-cues* on *cued* items, in a model with timing (*pre* vs. *retro*), position, the interaction between timing and position, and the same covariates as the other models. The random effect structure included intercepts for participants and syllables, as well as slopes for timing, position, and the interaction between the two by participant and syllable. Table 7 presents the results of this analysis. The model revealed a significant effect of timing ($t = 2.66$, $p = .009$), confirming that the *cued* condition affected performance differently depending on whether the cue appeared before or after syllables.

Finally, the last model compared the effects of *pre-* and *retro-cues* on *uncued* items. The model structure was identical to the previous analysis. Table 8 presents these results. There was no reliable effect of timing in this model ($t = -0.85$, $p = .398$).

Table 7. Results of the linear mixed effects model for cue timing effects in the *cued* condition in Experiment 3.

Fixed effects	**Coefficient**	**SE**	**t**	**p-value**
Intercept	-0.018	0.025	-0.713	.479
Timing-retro	0.037	0.014	2.662	.009
Position	-0.001	0.009	-0.133	.894
Baseline median	-0.245	0.010	-25.272	< .001

Baseline SD	0.141	0.008	17.651	< .001
Timing-retro × position	-0.007	0.013	-0.522	.604

Random effects

Subject	Variance
Intercept	0.0326
Timing-retro	0.0040
Position	0.0001
Timing-retro × position	0.0005

Syllable	Variance
Intercept	0.0085
Timing-retro	0.0003
Position	0.0000
Timing-retro × position	0.0005

Table 8. Results of the linear mixed effects model for cue timing effects in the *uncued* condition in Experiment 3.

Fixed effects	**Coefficient**	**SE**	**t**	**p-value**
Intercept	0.007	0.031	0.231	.818
Timing-retro	-0.014	0.016	-0.853	.398
Position	-0.031	0.010	-2.953	.005
Baseline median	-0.143	0.010	-14.268	< .001
Baseline SD	0.121	0.008	14.601	< .001
Timing-retro × position	-0.009	0.013	-0.671	.503

Random effects

Subject	Variance
Intercept	0.0583
Timing-retro	0.0131
Position	0.0001
Timing-retro × position	0.0002

Syllable	Variance
Intercept	0.0117
Timing-retro	0.0001
Position	0.0007

Timing-retro × position	0.0005

Discussion

Using a set size of three and a delay of 2 s, we replicated both findings of Experiment 2: (a) a reliable increase in the precision of recall for *cued* items that came at a significant cost to the *uncued* items, relative to *baseline* items; and (b) a reduction of serial position effects for *cued* items. The novel finding from Experiment 3 concerned the effect of cue timing on the effectiveness of the attentional cues. A dissociation was found here: while both *pre-* and *retro-cues* reliably decreased recall precision for *uncued* items to a comparable degree, only *pre-cues* significantly increased precision for *cued* items. This suggests that most of the benefit for *cued* items from attentional cueing can only be produced during the encoding phase. Once the stimulus is gone and no more information is available for encoding, a *retro-cue* does little to increase precision for *cued* items. On the other hand, the reliable decrease in the precision of *uncued* items regardless of cue timing suggests that releasing resources from these items is still possible during the maintenance period. This pattern is similar to what was found by Oberauer and Lin (2017) in the visual domain. Collectively, these results imply that attention can affect both encoding and maintenance phases of working memory processing through different mechanisms. More generally, they demonstrate that resource allocation in both visual and verbal domains is an ongoing process that affects working memory both at the encoding stage and at later stages of maintaining items until recall.

General Discussion

The study aimed to investigate whether the principles of resource allocation in working memory are shared between visual and verbal domains by undertaking the non-trivial challenge of creating an equivalent of the *continuous reproduction paradigm* in the phonological domain. After establishing that our participants were indeed capable of perceiving phonemes in a non-categorical fashion, our first critical finding was the decrease in response precision from set size 1 to 2, in keeping with the findings in visual working memory and in line with the predictions of resource, but not slot, models. This decrease in precision persisted after the potential confounds of serial position and random guessing had been ruled out. In agreement with the current findings, Joseph et al. (2015) also found a decrease in precision from set size 1 to 2, despite limitations that were discussed in earlier sections. Together, these findings provide strong evidence for similar principles of resource division in visual and verbal working memory (Bays et al., 2009; van den Berg, Awh, & Ma, 2014; van den Berg et al., 2012).

The second critical finding of the paper was the effect of attention on the precision of cued and uncued items compared to baseline. In both Experiments 2 and 3, we showed that attention significantly increased the precision with which cued items were remembered, but this came at the cost of reduced precision for uncued items. Similar findings have been reported in phonemic migrations between words recited from memory (Nozari & Dell, 2012; Nozari & Thompson-Schill, 2013), as well as in vision (Bays et al., 2011; Gorgoraptis et al., 2011; Oberauer & Lin, 2017). These findings once again show similar principles of resource allocation in visual and verbal domains. Finally, the last experiment showed that retro-cues, appearing after the stimuli have been presented and encoding is complete, can also affect resource allocation in phonological working memory, indicating that resource allocation remains flexible during the maintenance period. As has been reported in vision, the decrease in precision for uncued items was not affected by cue timing—participants were able to forget or ignore the non-prioritized items in response to both pre and retro-cues. On the other hand, the increase in precision for cued items was greatly reduced with retro-cues compared to pre-cues, similar to vision (Oberauer & Lin, 2017). Thus, across all three experiments, our findings support a resource model of phonological working memory, with resource allocation conforming to same principles as in visual working memory.

We have been discussing our results in relation to "phonological working memory", but one may object that the investigation involves a level below phonology (i.e., acoustic features), and as such may not be representative or even germane to language processing. Two sources of evidence speak against this. First, it has been proposed that speech perception may depend on units smaller than phonemes or syllables, i.e., on finer-grained acoustic features (Stevens, 2002). For example, although words and phrases like "probably" and "could have" are often produced in a highly reduced form (e.g., "prolly" and "coulda", respectively) in American English, these forms are easily understood by other native speakers. These highly reduced forms do not preserve the original syllable structure, but they still possess unique acoustic cues that, at least in certain contexts, lead to easy understanding of their meaning (Niebuhr & Kohler, 2011). Critically, recent studies have found that gradient acoustic information can influence language comprehension. For example, Brown-Schmidt and Toscano (2017) used sentences containing a word on an acoustic continuum between "he" and "she" to probe the effect of gradient acoustic information on the interpretation of the pronoun in an ambiguous context. They found that eye gaze and, subsequently, the amount of time needed to recover from an incorrect interpretation once a disambiguating word had been encountered were both sensitive to this information. In a similar vein, while classic speech error studies identified the critical unit to be phoneme in the slips of the tongue (e.g., Shattuck-Hufnagel & Klatt, 1979), recent evidence shows that such slips can be between-phoneme blends (Goldrick & Chu, 2014). Second, some of the principles demonstrated here such as the differential effects of attention on the attended and unattended items have also been reported in phonological migrations between lexical items in verbal working memory tasks that involve sequences of real words (Nozari & Dell, 2012; Nozari & Thompson-Schill, 2013), suggesting that, as computational principles go, the findings are pertinent to language processing, and hence verbal working memory.

Why a Resource Model?

The term *resource* is often used in the working memory literature to convey a vague sense of limited capacity. In the context of slot and resource models, however, the term *resource* refers to something with a set of clearly defined properties. First, this resource is limited in quantity. Second, it can be divided into parts—either continuously, in the case of a resource model, or into a fixed number of discrete slots, in the case of a slot or slots-plus averaging model. Third, for an item to be stored in working memory, a portion of this resource must be allocated to it. Finally, the precision with which the item can be recalled depends on the amount of the resource allocated to it. When *resource* is defined in this way, the claim that working memory is a limited resource becomes a testable hypothesis rather than a description, and it becomes possible to differentiate between models of resource allocation like the slot and resource models discussed in this paper.

When defined precisely in this way, it becomes possible to examine the biological basis of such a resource. Resource models have an advantage in this regard: they identify the limited resource simply as neural gain. Since the signals from individual neurons within these populations are noisy and imprecise, the overall pattern of activation in a population encodes a probability distribution over a particular feature dimension (Ma, Beck, Latham, & Pouget, 2006; Sanger, 1996; Zemel, Dayan, & Pouget, 1998). When the pattern is decoded, the resulting representation is drawn from this distribution. The precision of the decoded representation depends on the amplitude of the activation (i.e., the *gain*) of the corresponding population code (Ma et al., 2006). All else being equal, an item encoded with higher gain will be recalled with greater precision. An increase in gain is costly because increasing neural activity consumes significantly more energy (Attwell & Laughlin, 2001; Lennie, 2003). This places a limit on the total amount of gain that can be allocated to the items in working memory at a given time. As a result, there is a tradeoff between expected task performance and resource costs. The predictions of resource models are thus readily explained in terms of neural gain (Bays, 2014, 2015; Ma et al., 2006; Orhan & Ma, 2015; Schneegans & Bays, 2017; van den Berg & Ma, 2018).

The neural basis for discrete slots, on the other hand, is less clear. Though proponents of the slot model have suggested that the behavior they predict may arise from a need to lock the pattern of activity corresponding to each item to a limited number of distinct phases of neural oscillation in order to maintain the binding between features (e.g., Lisman & Idiart, 1995), there is no a priori reason why, given the wide range of potential frequencies that could be used for this purpose, the maximum number of distinct phases should correspond to the 3–4 slots that have been proposed (Bays, 2015); furthermore, other mechanisms for feature binding have been proposed which would not impose such constraints (e.g., Matthey, Bays, & Dayan, 2015).

Additionally, by identifying working memory resources with neural gain, resource models provide a natural explanation for the relationship between working memory resources and attention. On the one hand, past studies have suggested that attention and working memory are closely related; measures of working memory capacity and attentional control are correlated across individuals (e.g., Cowan, 1995; Kane, Bleckley, Conway, & Engle, 2001). Similarly, neuroimaging data suggests a link between working memory and the dorsal attention network (Majerus et al., 2016; Majerus, Péters, Bouffier, Cowan, & Phillips, 2017). On the other hand, the modulating effect of attention on neural gain has long been established in the literature (McAdams & Maunsell, 1999). A theory that views working memory resource as neural gain thus contains a natural mechanism for the modification of resource allocation by attention.

In sum, the biological plausibility of resource models, along with the natural link they provide between attention and working memory performance, makes them a particularly appealing class of model for explaining resource allocation in working memory.

Slots Plus Averaging as an Alternative to Continuous Resources

For the reasons outlined in the previous sections, we have interpreted the results of Experiment 1 as providing support for a resource model. However, as alluded to in the Introduction, these results are also compatible with an alternative model, the *slots plus averaging* (SA) model (e.g., Zhang & Luck, 2008). Like classic slot models, these SA models divide working memory capacity into a fixed number of slots, each of which can store exactly one item. However, the relationship between slots and items is not one-to-one: a single item can be stored in multiple slots. Each copy of the stored representation is subject to some amount of random error. By averaging over multiple copies of the representation, each stored in a different slot, precision can be increased (and deviation reduced). Thus, as in resource models, the precision with which a representation is stored in working memory is dependent on the amount of resources—in this case, the number of slots—it receives. This feature enables slots plus averaging models to account for many of the same empirical findings as resource models, making it difficult to distinguish between the two.

SA models have two critical parameters: σ, the deviation when an item is present in memory, and P_m, the probability that an item is present in memory. The model predicts a specific relationship between these parameters, set size (N), and the number of slots (K): both σ and the number of items stored in memory ($N \times P_m$) should increase as a function of N until they reach a plateau at $N = K$. Our experiments were not designed to test these predictions; however, we will review three lines of evidence that pose a problem for the SA model. First, the SA model fails to produce internally consistent parameter estimates when applied to empirical data. Although both σ and $N \times P_m$ should reach a plateau at the same set size ($N = K$), Bays (2018) has found that the actual correlation between the set sizes at which these two estimates reached a plateau was very low across participants in most studies. A more basic problem is that the estimates of σ and P_m obtained by fitting the model to empirical data cannot be interpreted if the true underlying distributions do not in fact correspond to the SA model; for example, the P_m parameter would not necessarily reflect the actual probability of an item being in memory (Ma, 2018). Similarly, it has recently been claimed that the random guessing component captured by the SA model may be an artifact of measuring deviations in physical stimulus space, rather than the perceptual space in which the contents of working memory are represented (Schurgin, Wixted, & Brady, 2018).

Second, key evidence that has been cited in support of the SA model is based on null results—the absence of statistically significant differences in σ (e.g., Zhang & Luck, 2008), or a neural signal linked to working memory load (e.g., Todd & Marois, 2004; Vogel & Machizawa, 2004; Xu & Chun, 2006), across different set sizes, taken as evidence that these measures had reached a plateau. Even if these plateaus were real, this test would not reliably discriminate between the SA model and a type of resource model called a *variable-precision resource model* (van den Berg et al., 2014), in which the division of resources, even in the absence of top-down attentional cues, is subject to stochastic noise. In fact, variable-precision resource models have been shown to provide better fit to the data than SA models across a range of studies (van den Berg et al., 2014).

Third, as explained in the previous section, slot models, including the SA model, lack the biological plausibility of resource models. Thus, although we did not directly test the predictions that differentiate the SA model from resource models, we conclude that our findings, taken together with these three lines of evidence, are better aligned with a resource model. It is worth mentioning that two other accounts, interference (e.g., Nairne, 1990; Oberauer & Kliegl, 2006; Oberauer & Lin, 2017) and decay (e.g., Baddeley, Thomson, & Buchanan, 1975; Barrouillet, Bernardin, & Camos, 2004; Burgess & Hitch, 1999) models, have proposed different mechanisms to explain capacity limits in working memory. However, unlike SA models, the mechanisms proposed by these accounts are not mutually exclusive with a limited resource of the kind proposed by resource models. Since the experiments presented in this paper were not designed to test these mechanisms, and direct comparisons of these models can be found elsewhere (Oberauer, Farrell, Jarrold, & Lewandowsky, 2016), we will refrain from contrasting these models with the resource model here.

Is the Allocation of Resources Domain-general or Domain-specific?

Debates over the domain-generality or domain-specificity of working memory and executive control resources are almost as old as the study of the topics themselves. However, the methodologies that have been used to test the extent to which these resources and the mechanisms responsible for their allocation are domain-general or domain-specific have been diverse, making the definitions of the terms somewhat opaque. To gain better traction on this question, it is helpful to break down the concept of domain-generality into multiple, clearly-defined aspects. For example, Nozari and Novick (2017) discussed the domain-generality of monitoring and executive control from three angles: (1) domain-generality of computational principles, (2) domain-generality of neural correlates, and (3) functional domain-generality, i.e., whether performance in one task affects performance in a different task when both require executive control. A similar framework can be applied to working memory. Note that these three meanings of "domain-generality" are distinct; for example, domain-generality of computational principles does not necessarily imply domain generality of the neural substrates performing those computations, or vice versa. The same computations could be performed by different populations of neurons in different domains, or different computations could be performed by the same population depending on the nature of the representations involved. Similarly, domain-generality of computational principles does not imply functional domain-generality: the same set of computational principles could govern the allocation of multiple pools of resources (e.g., corresponding to the domain-specific buffers in the multi-component model; Baddeley, 2000; Baddeley & Hitch, 1974). Different kinds of evidence are thus relevant to each aspect of domain- generality or specificity.

The latter two aspects—domain-generality of neural correlates and functional domain-generality—have been investigated extensively, with mixed results. Neuroimaging studies have identified some brain regions that are selectively associated with working memory in a specific domain (e.g., separate regions for phonological and semantic working memory: Martin, Wu, Freedman, Jackson, & Lesch, 2003; Shivde & Thompson-Schill, 2004) and others that are involved in maintaining information in working memory across domains (e.g., Chein, Moore, & Conway, 2011; Cowan et al., 2011; Majerus et al., 2009; Rottschy et al., 2012; Wager & Smith, 2003). The behavioral evidence for functional domain-generality is similarly mixed: some studies have reported little or no cross-domain interference (Cocchini, Logie, Sala, MacPherson, & Baddeley, 2002; Fougnie, Zughni, Godwin, & Marois, 2015) and a low correlation between verbal and visual capacities (Shah & Miyake, 1996), while others have found more significant cross-domain interference (Saults & Cowan, 2007). Taken together, these findings do not provide clear support for either complete functional and neural domain-generality or complete domain-specificity.

The first aspect, the domain-generality or specificity of computational principles, is the most important for computational models of working memory, and is the aspect addressed by this paper. If the principles of resource division depend on the nature of sensory stimuli, then different mechanisms would clearly be needed in computational models of visual and verbal working memory to account for the allocation of resources in each domain. The three experiments presented in this paper provide converging evidence against this possibility. Despite the different nature of visual and auditory stimuli, the close match between our findings in phonological working memory and those previously reported in visual working memory supports a model of working memory in which the principles of resource allocation are identical across these domains.

Conclusion

In summary, by adapting the continuous reproduction paradigm to phonological rather than visual materials, we were able to conduct a series of experiments testing the extent to which the principles of resource division are similar in visual and verbal working memory. In every case, our findings were consistent with those reported in visual working memory. Together, they provide evidence that the allocation of working memory resources in both visual and verbal domains is most consistent with a resource model, and more generally, that the computational principles governing the allocation of working memory resources are similar in the visual and verbal domains.

Acknowledgement

This research was supported by the Therapeutic Cognitive Neuroscience Fund.

Appendix A

Table A1. Results of the linear mixed effects model for serial position effects in the baseline condition of Experiment 2.

Fixed effects	**Coefficient**	**SE**	**t**	**p-value**
Intercept	0.000	0.033	0.000	1.000
Position-linear	-0.122	0.024	-5.023	< .001
Position-quadratic	-0.064	0.027	-2.383	.023
Baseline median	-0.190	0.014	-13.417	< .001
Baseline SD	0.175	0.012	14.755	< .001

Random effects

Subject	Variance
Intercept	0.0283
Position-linear	0.0061
Position-quadratic	0.0076

Syllable	Variance
Intercept	0.0114
Position-linear	0.0039
Position-quadratic	0.0069

Table A2. Results of the linear mixed effects model for serial position effects in the cued condition of Experiment 2.

Fixed effects	**Coefficient**	**SE**	**t**	**p-value**
Intercept	0.000	0.035	0.000	1.000
Position-linear	-0.061	0.018	-3.339	.001
Position-quadratic	0.003	0.021	0.155	.878
Baseline median	-0.253	0.014	-18.557	< .001
Baseline SD	0.208	0.011	18.303	< .001

Random effects

Subject	Variance
Intercept	0.0346
Position-linear	0.0000
Position-quadratic	0.0035

Syllable	Variance
Intercept	0.0116
Position-linear	0.0011
Position-quadratic	0.0023

Table A3. Results of the linear mixed effects model for serial position effects in the baseline condition of Experiment 3.

Fixed effects	**Coefficient**	**SE**	**t**	**p-value**
Intercept	0.001	0.026	0.023	.982
Position-linear	-0.038	0.012	-3.322	.001
Position-quadratic	-0.018	0.011	-1.609	.109
Baseline median	-0.178	0.010	-17.864	< .001
Baseline SD	0.155	0.008	18.992	< .001

Random effects

Subject	Variance
Intercept	0.0270
Position-linear	0.0000
Position-quadratic	0.0000

Syllable	Variance
Intercept	0.0114
Position-linear	0.0004
Position-quadratic	0.0001

Table A4. Results of the linear mixed effects model for serial position effects in the cued condition of Experiment 3 with pre-cues.

Fixed effects	**Coefficient**	**SE**	**t**	**p-value**
Intercept	0.000	0.024	0.016	.987
Position-linear	-0.002	0.015	-0.134	.894
Position-quadratic	0.015	0.018	0.854	.400
Baseline median	-0.257	0.013	-20.231	< .001
Baseline SD	0.144	0.011	12.994	< .001
Random effects				
Subject	Variance			
Intercept	0.0331			
Position-linear	0.0000			
Position-quadratic	0.0000			
Syllable	Variance			
Intercept	0.0072			
Position-linear	0.0001			
Position-quadratic	0.0024			

References

Atkinson, R. C., & Shiffrin, R. M. (1971). *The control processes of short-term memory*. Stanford, CA: Stanford University.

Attwell, D., & Laughlin, S. B. (2001). An energy budget for signaling in the grey matter of the brain. *Journal of Cerebral Blood Flow & Metabolism*, *21*(10), 1133–1145. https://doi.org/10.1097/00004647-200110000-00001

Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, *4*(11), 417–423. https://doi.org/10.1016/S1364-6613(00)01538-2

Baddeley, A. D. (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience*, *4*(10), 829–839. https://doi.org/10.1038/nrn1201

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *Psychology of Learning and Motivation* (Vol. 8, pp. 47–89). Academic Press. Retrieved from http://www.sciencedirect.com/science/article/pii/S0079742108604521

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*(6), 575–589. https://doi.org/10.1016/S0022-5371(75)80045-4

Bae, G.-Y., Olkkonen, M., Allred, S. R., Wilson, C., & Flombaum, J. I. (2014). Stimulus-specific variability in color working memory with delayed estimation. *Journal of Vision*, *14*(4), 7–7. https://doi.org/10.1167/14.4.7

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Barrouillet, P., Bernardin, S., & Camos, V. (2004). Time constraints and resource sharing in adults' working memory spans. *Journal of Experimental Psychology: General*, *133*(1), 83–100. https://doi.org/10.1037/0096-3445.133.1.83

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bays, P. M. (2014). Noise in neural populations accounts for errors in working memory. *Journal of Neuroscience*, *34*(10), 3632–3645. https://doi.org/10.1523/JNEUROSCI.3204-13.2014

Bays, P. M. (2015). Spikes not slots: Noise in neural populations limits working memory. *Trends in Cognitive Sciences*, *19*(8), 431–438. https://doi.org/10.1016/j.tics.2015.06.004

Bays, P. M. (2018). Failure of self-consistency in the discrete resource model of visual working memory. *Cognitive Psychology*, *105*, 1–8. https://doi.org/10.1016/j.cogpsych.2018.05.002

Bays, P. M., Catalao, R. F. G., & Husain, M. (2009). The precision of visual working memory is set by allocation of a shared resource. *Journal of Vision*, *9*(10), 7–7. https://doi.org/10.1167/9.10.7

Bays, P. M., Gorgoraptis, N., Wee, N., Marshall, L., & Husain, M. (2011). Temporal dynamics of encoding, storage, and reallocation of visual working memory. *Journal of Vision*, *11*(10), 6–6. https://doi.org/10.1167/11.10.6

Boersma, P., & Weenink, D. (2016). Praat: doing phonetics by computer (Version 6.0.21). Retrieved from http://www.praat.org/

Brown-Schmidt, S., & Toscano, J. C. (2017). Gradient acoustic information induces long-lasting referential uncertainty in short discourses. *Language, Cognition and Neuroscience*, *32*(10), 1211–1228. https://doi.org/10.1080/23273798.2017.1325508

Burgess, N., & Hitch, G. J. (1999). Memory for serial order: A network model of the phonological loop and its timing. *Psychological Review*, *106*(3), 551–581. https://doi.org/10.1037/0033-295X.106.3.551

Chein, J. M., Moore, A. B., & Conway, A. R. A. (2011). Domain-general mechanisms of complex working memory span. *NeuroImage*, *54*(1), 550–559. https://doi.org/10.1016/j.neuroimage.2010.07.067

Cocchini, G., Logie, R. H., Sala, S. D., MacPherson, S. E., & Baddeley, A. D. (2002). Concurrent performance of two memory tasks: Evidence for domain-specific working memory systems. *Memory & Cognition*, *30*(7), 1086–1095. https://doi.org/10.3758/BF03194326

Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorials in Quantitative Methods for Psychology*, *1*(1), 42–45. https://doi.org/10.20982/tqmp.01.1.p042

Cowan, N. (1995). *Attention and memory: An integrated framework*. New York, NY: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195119107.001.0001

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*(1), 87–114. https://doi.org/10.1017/S0140525X01003922

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), 2852–2863. https://doi.org/10.1162/jocn.2011.21625

Cowan, N., Rouder, J. N., Blume, C. L., & Saults, J. S. (2012). Models of verbal working memory capacity: What does it take to make them work? *Psychological Review*, *119*(3), 480–499. https://doi.org/10.1037/a0027791

de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods*, *47*(1), 1–12. https://doi.org/10.3758/s13428-014-0458-y

Fougnie, D., Zughni, S., Godwin, D., & Marois, R. (2015). Working memory storage is intrinsically domain specific. *Journal of Experimental Psychology: General*, *144*(1), 30–47. https://doi.org/10.1037/a0038211

Fraley, C., & Raftery, A. E. (2002). Model-Based Clustering, Discriminant Analysis, and Density Estimation. *Journal of the American Statistical Association*, *97*(458), 611–631. https://doi.org/10.1198/016214502760047131

Gerrits, E., & Schouten, M. E. H. (2004). Categorical perception depends on the discrimination task. *Perception & Psychophysics*, *66*(3), 363–376. https://doi.org/10.3758/BF03194885

Goldrick, M., & Chu, K. (2014). Gradient co-activation and speech error articulation: Comment on Pouplier and Goldstein (2010). *Language, Cognition and Neuroscience*, *29*(4), 452–458. https://doi.org/10.1080/01690965.2013.807347

Gorgoraptis, N., Catalao, R. F. G., Bays, P. M., & Husain, M. (2011). Dynamic updating of working memory resources for visual objects. *The Journal of Neuroscience*, *31*(23), 8502–8511. https://doi.org/10.1523/JNEUROSCI.0208-11.2011

Hayes, J. R. M. (1952). Memory span for several vocabularies as a function of vocabulary size. In *Quarterly Progress Report* (pp. 5–9). Cambridge, MA: Acoustics Laboratory, Massachusetts Institute of Technology.

Hepner, C. R., & Nozari, N. (2018). Resource allocation in phonological working memory: Same or different principles from vision? Retrieved from https://osf.io/943u7/

Hughes, G. W., & Halle, M. (1956). Spectral properties of fricative consonants. *The Journal of the Acoustical Society of America*, *28*(2), 303–310. https://doi.org/10.1121/1.1908271

Joseph, S., Iverson, P., Manohar, S., Fox, Z. V., Scott, S. K., & Husain, M. (2015). Precision of working memory for speech sounds. *The Quarterly Journal of Experimental Psychology*, *68*(10), 2022–2040. https://doi.org/10.1080/17470218.2014.1002799

Kane, M. J., Bleckley, M. K., Conway, A. R. A., & Engle, R. W. (2001). A controlled-attention view of working-memory capacity. *Journal of Experimental Psychology: General*, *130*(2), 169–183. https://doi.org/10.1037/0096-3445.130.2.169

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*(430), 773–795. https://doi.org/10.2307/2291091

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2016). lmerTest: Tests in linear mixed effects models (Version 2.0-33). Retrieved from https://CRAN.R-project.org/package=lmerTest

Lennie, P. (2003). The cost of cortical computation. *Current Biology*, *13*(6), 493–497. https://doi.org/10.1016/S0960-9822(03)00135-0

Liberman, A. M., Delattre, P. C., & Cooper, F. S. (1958). Some cues for the distinction between voiced and voiceless stops in initial position. *Language and Speech*, *1*(3), 153–167. https://doi.org/10.1177/002383095800100301

Liberman, A. M., Delattre, P. C., Cooper, F. S., & Gerstman, L. J. (1954). The role of consonant-vowel transitions in the perception of the stop and nasal consonants. *Psychological Monographs: General and Applied*, *68*(8), 1–13. https://doi.org/10.1037/h0093673

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, *54*(5), 358–368. https://doi.org/10.1037/h0044417

Lisman, J. E., & Idiart, M. A. (1995). Storage of 7 +/- 2 short-term memories in oscillatory subcycles. *Science*, *267*(5203), 1512–1515. https://doi.org/10.1126/science.7878473

Ma, W. J. (2018). Problematic usage of the Zhang and Luck mixture model. *BioRxiv*, 268961. https://doi.org/10.1101/268961

Ma, W. J., Beck, J. M., Latham, P. E., & Pouget, A. (2006). Bayesian inference with probabilistic population codes. *Nature Neuroscience*, *9*(11), 1432–1438. https://doi.org/10.1038/nn1790

Ma, W. J., Husain, M., & Bays, P. M. (2014). Changing concepts of working memory. *Nature Neuroscience*, *17*(3), 347–356. https://doi.org/10.1038/nn.3655

Majerus, S., Cowan, N., Péters, F., Van Calster, L., Phillips, C., & Schrouff, J. (2016). Cross-modal decoding of neural patterns associated with working memory: Evidence for attention-based accounts of working memory. *Cerebral Cortex*, *26*(1), 166–179. https://doi.org/10.1093/cercor/bhu189

Majerus, S., D'Argembeau, A., Martinez Perez, T., Belayachi, S., Van der Linden, M., Collette, F., … Maquet, P. (2009). The commonality of neural networks for verbal and visual short-term memory. *Journal of Cognitive Neuroscience*, *22*(11), 2570–2593. https://doi.org/10.1162/jocn.2009.21378

Majerus, S., Péters, F., Bouffier, M., Cowan, N., & Phillips, C. (2017). The dorsal attention network reflects both encoding load and top–down control during working memory. *Journal of Cognitive Neuroscience*, *30*(2), 144–159. https://doi.org/10.1162/jocn_a_01195

Martin, R. C., Wu, D., Freedman, M., Jackson, E. F., & Lesch, M. (2003). An event-related fMRI investigation of phonological versus semantic short-term memory. *Journal of Neurolinguistics*, *16*(4), 341–360. https://doi.org/10.1016/S0911-6044(03)00025-3

Massaro, D. W., & Cohen, M. M. (1983). Categorical or continuous speech perception: A new test. *Speech Communication*, *2*(1), 15–35. https://doi.org/10.1016/0167-6393(83)90061-4

Matthey, L., Bays, P. M., & Dayan, P. (2015). A probabilistic palimpsest model of visual short-term memory. *PLOS Computational Biology*, *11*(1), e1004003. https://doi.org/10.1371/journal.pcbi.1004003

McAdams, C. J., & Maunsell, J. H. R. (1999). Effects of attention on orientation-tuning functions of single neurons in macaque cortical area V4. *Journal of Neuroscience*, *19*(1), 431–441.

McDonnell, J. V., Martin, J. B., Markant, D. B., Coenen, A., Rich, A. S., & Gureckis, T. M. (2016). psiTurk (Version 2.1.6). Retrieved from https://github.com/NYUCCL/psiTurk

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*(2), 81–97. https://doi.org/10.1037/h0043158

Nairne, J. S. (1990). A feature model of immediate memory. *Memory & Cognition*, *18*(3), 251–269. https://doi.org/10.3758/BF03213879

Niebuhr, O., & Kohler, K. J. (2011). Perception of phonetic detail in the identification of highly reduced words. *Journal of Phonetics*, *39*(3), 319–329. https://doi.org/10.1016/j.wocn.2010.12.003

Nozari, N., & Dell, G. S. (2012). Feature migration in time: Reflection of selective attention on speech errors. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(4), 1084–1090. https://doi.org/10.1037/a0026933

Nozari, N., & Novick, J. (2017). Monitoring and control in language production. *Current Directions in Psychological Science*, *26*(5), 403–410. https://doi.org/10.1177/0963721417702419

Nozari, N., & Thompson-Schill, S. L. (2013). More attention when speaking: Does it help or does it hurt? *Neuropsychologia*, *51*(13), 2770–2780. https://doi.org/10.1016/j.neuropsychologia.2013.08.019

Oberauer, K., Farrell, S., Jarrold, C., & Lewandowsky, S. (2016). What limits working memory capacity? *Psychological Bulletin*, *142*(7), 758–799. https://doi.org/10.1037/bul0000046

Oberauer, K., & Kliegl, R. (2006). A formal model of capacity limits in working memory. *Journal of Memory and Language*, *55*(4), 601–626. https://doi.org/10.1016/j.jml.2006.08.009

Oberauer, K., & Lin, H.-Y. (2017). An interference model of visual working memory. *Psychological Review*, *124*(1), 21–59. https://doi.org/10.1037/rev0000044

O'Connor, J. D., Gerstman, L. J., Liberman, A. M., Delattre, P. C., & Cooper, F. S. (1957). Acoustic cues for the perception of initial /w, j, r, l/ in English. *WORD*, *13*(1), 24–43. https://doi.org/10.1080/00437956.1957.11659626

Olson, A., Romani, C., & Caramazza, A. (2010). Analysis and interpretation of serial position data. *Cognitive Neuropsychology*, *27*(2), 134–151. https://doi.org/10.1080/02643294.2010.504580

Orhan, A. E., & Ma, W. J. (2015). Neural population coding of multiple stimuli. *Journal of Neuroscience*, *35*(9), 3825–3841. https://doi.org/10.1523/JNEUROSCI.4097-14.2015

Page, M. P. A., & Norris, D. (1998). The primacy model: A new model of immediate serial recall. *Psychological Review*, *105*(4), 761–781. https://doi.org/10.1037/0033-295X.105.4.761-781

Pertzov, Y., Bays, P. M., Joseph, S., & Husain, M. (2013). Rapid forgetting prevented by retrospective attention cues. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(5), 1224–1231. https://doi.org/10.1037/a0030947

R Core Team. (2017). R: A language and environment for statistical computing (Version 3.4.1). Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Rottschy, C., Langner, R., Dogan, I., Reetz, K., Laird, A. R., Schulz, J. B., … Eickhoff, S. B. (2012). Modelling neural correlates of working memory: A coordinate-based meta-analysis. *NeuroImage*, *60*(1), 830–846. https://doi.org/10.1016/j.neuroimage.2011.11.050

Sanger, T. D. (1996). Probability density estimation for the interpretation of neural population codes. *Journal of Neurophysiology*, *76*(4), 2790–2793. https://doi.org/10.1152/jn.1996.76.4.2790

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. https://doi.org/10.1037/0096-3445.136.4.663

Schneegans, S., & Bays, P. M. (2016). No fixed item limit in visuospatial working memory. *Cortex*, *83*, 181–193. https://doi.org/10.1016/j.cortex.2016.07.021

Schneegans, S., & Bays, P. M. (2017). Neural architecture for feature binding in visual working memory. *Journal of Neuroscience*, *37*(14), 3913–3925. https://doi.org/10.1523/JNEUROSCI.3493-16.2017

Schouten, B., Gerrits, E., & van Hessen, A. (2003). The end of categorical perception as we know it. *Speech Communication*, *41*(1), 71–80. https://doi.org/10.1016/S0167-6393(02)00094-8

Schurgin, M. W., Wixted, J. T., & Brady, T. F. (2018). Psychophysical scaling reveals a unified theory of visual memory strength. *BioRxiv*, 325472. https://doi.org/10.1101/325472

Shah, P., & Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: An individual differences approach. *Journal of Experimental Psychology: General*, *125*(1), 4–27. https://doi.org/10.1037/0096-3445.125.1.4

Shattuck-Hufnagel, S., & Klatt, D. H. (1979). The limited use of distinctive features and markedness in speech production: Evidence from speech error data. *Journal of Verbal Learning and Verbal Behavior*, *18*(1), 41–55. https://doi.org/10.1016/S0022-5371(79)90554-1

Shivde, G., & Thompson-Schill, S. L. (2004). Dissociating semantic and phonological maintenance using fMRI. *Cognitive, Affective, & Behavioral Neuroscience*, *4*(1), 10–19. https://doi.org/10.3758/CABN.4.1.10

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs: General and Applied*, *74*(11), 1–29. https://doi.org/10.1037/h0093759

Stevens, K. N. (2002). Toward a model for lexical access based on acoustic landmarks and distinctive features. *The Journal of the Acoustical Society of America*, *111*(4), 1872–1891. https://doi.org/10.1121/1.1458026

Todd, J. J., & Marois, R. (2004). Capacity limit of visual short-term memory in human posterior parietal cortex. *Nature*, *428*(6984), 751–754. https://doi.org/10.1038/nature02466

van den Berg, R., Awh, E., & Ma, W. J. (2014). Factorial comparison of working memory models. *Psychological Review*, *121*(1), 124–149. https://doi.org/10.1037/a0035234

van den Berg, R., & Ma, W. J. (2018). A resource-rational theory of set size effects in human visual working memory. *ELife*, *7*. https://doi.org/10.7554/eLife.34963

van den Berg, R., Shin, H., Chou, W.-C., George, R., & Ma, W. J. (2012). Variability in encoding precision accounts for visual short-term memory limitations. *Proceedings of the National Academy of Sciences*, *109*(22), 8780–8785. https://doi.org/10.1073/pnas.1117465109

Vogel, E. K., & Machizawa, M. G. (2004). Neural activity predicts individual differences in visual working memory capacity. *Nature*, *428*(6984), 748–751. https://doi.org/10.1038/nature02447

Wager, T. D., & Smith, E. E. (2003). Neuroimaging studies of working memory: A meta-analysis. *Cognitive, Affective, & Behavioral Neuroscience*, *3*(4), 255–274. https://doi.org/10.3758/CABN.3.4.255

Wilken, P., & Ma, W. J. (2004). A detection theory account of change detection. *Journal of Vision*, *4*(12), 11. https://doi.org/10.1167/4.12.11

Winn, M. B. (2014). Praat Scripts. Retrieved October 4, 2016, from http://www.mattwinn.com/praat.html

Winn, M. B., & Litovsky, R. Y. (2015). Using speech sounds to test functional spectral resolution in listeners with cochlear implants. *The Journal of the Acoustical Society of America*, *137*(3), 1430–1442. https://doi.org/10.1121/1.4908308

Xu, Y., & Chun, M. M. (2006). Dissociable neural mechanisms supporting visual short-term memory for objects. *Nature*, *440*(7080), 91–95. https://doi.org/10.1038/nature04262

Zemel, R. S., Dayan, P., & Pouget, A. (1998). Probabilistic interpretation of population codes. *Neural Computation*, *10*(2), 403–430. https://doi.org/10.1162/089976698300017818

Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, *453*(7192), 233–235. https://doi.org/10.1038/nature06860