Auditory Versus Visual Stimulus Effects on Cognitive Performance During the N-back Task

Mary Jean Amon (mjamon@iu.edu)
Department of Psychological & Brain Sciences, Indiana University Bloomington
Bloomington, IN 47405 USA

Bennett I. Bertenthal (bbertent@indiana.edu)
Department of Psychological & Brain Sciences, Indiana University Bloomington
Bloomington, IN 47405 USA

Abstract

The n-back task is one of the most popular methods for studying working memory, and it is tested with either auditory or visual stimuli. Previous research comparing stimulus modalities has demonstrated that auditory and visual tasks often elicit differential responding and, potentially, different underlying cognitive processes. In this study, performance accuracy and response time were measured during an n-back task that varied in terms of stimulus modality and difficulty. Findings demonstrate that participants respond faster but less accurately during a visual as compared to an auditory condition where participants are more accurate but slower to respond. These results are discussed in terms of dual coding and feature binding. Implications for the presentation of n-back tasks in studies of working memory are discussed.

Keywords: n-back; auditory memory; visual memory; working memory; dual-coding; feature binding

Introduction

Visual and auditory stimuli are processed differently at the input stage, but it is unclear whether these differences result in a downstream effect in cognitive processing. For example, research suggests that visual information, as compared to auditory information, may constrain performance during simple and sequential cognitive tasks (Greene, 1992; Penney, 1989). A number of explanations have been offered for performance differences based on sensory modality. One theory suggests that differences in cognitive performance may occur as visual stimuli, but not auditory stimuli, are spontaneously named such that both a visual and auditory representation of the stimuli is retained (Paivio, 1990; Snodgrass, Wasser, Finkelstein, & Goldberg, 1974). While richer information may be produced through visual stimuli, dual-coding may also create interference during tasks involving serial recall or memorization of lists (Allen, Baddeley, & Hitch, 2006; Burgess & Hitch, 1999; Maybery, Clissa, Parmentier, Leung, Harsa, Fox, & Jones, 2009; Nairne, 1990). In addition, some theories (e.g., Precategorical Acoustic Storage theory) propose that information presented aurally may have an advantage because stimuli are retained in an echoic buffer storage while new ones are presented (Crowder & Morton, 1969; cf. Beaman, 2002). Given that there are conflicting findings regarding the effect of stimulus modality on performance, it is possible that performance differences based on auditory and visual presentations are dependent on additional details of the experimental protocol and not generalizable across working memory tasks.

The present study compares auditory and visual n-back task performances in order to test the hypothesis there are significant differences as a function of stimulus modality, and also to test whether these differences vary with the complexity of the task. Findings from this study should have implications for protocols involving n-back tasks, as well as the cognitive processes that underlie n-back task performance.

In the next section, we review evidence from cognitive performance research indicating that auditory and visual stimuli are represented differently in working memory and how their representation may facilitate or impede cognitive performance. We then present a study that examines n-back task performance and response times based on auditory and visual stimuli, as well as task difficulty. In addition to differences in performance accuracy, the findings indicate response time differences based on auditory and visual presentation. These findings are discussed in terms of dual coding and feature binding. Lastly, implications for the presentation of n-back tasks in cognitive studies on working memory are discussed.

Auditory and Visual Processing

Working memory describes the cognitive process through which information is held for a short period of time in the absence of external cues in order to guide behavior (Fuster, 1995; Goldman-Rakic, 1996; Owen, McMillan, Laird, & Bullmore, 2005). Through this process, perceptual and linguistic information can be temporarily stored, manipulated, and selectively recalled. Given the significance of this broadly applicable cognitive process, numerous studies have measured various aspects of working memory in order to understand its fundamental characteristics.

The n-back task is a paradigm that has been used extensively in cognitive and neuroimaging studies focused on working memory. This requires participants to view a sequence of stimuli, one at a time, and respond whenever the current stimulus matches the stimulus that was previously presented *n* stimuli back. Participants must

attempt to monitor, store, update, and manipulate information over time (Owen et al., 2005). In this way, the n-back task is a particularly accessible method for manipulating and measuring various aspects of working memory. A number of n-back task variants are utilized, which differ in terms of content (auditory or visual) and process (location and identity monitoring; Owen et al., 2005). Auditory and visual task stimuli are often used interchangeably. However, research suggests that tasks that superficially appear to be very similar may elicit different underlying cognitive processes.

Paivio's (1971) dual-coding theory suggests that visual and auditory information are processed through different channels that yield distinct mental representations. This model is compatible with Baddeley and Hitch's (1974) working memory model, which identifies separate subsystems for short-term retention of auditory and visual information. In this model, the components of working memory have a limited capacity and are relatively independent of one another. According to these theories, coding both visual and auditory representations improves the chances of successful retrieval, as both subsystems can be used to recover information.

The presence of separate but related auditory and visual subsystems in working memory is supported by performance on a number of commonly used cognitive tasks. During digit span tasks, for example, participants tend to recall auditory numbers more readily than number images (Greene, 1992; Penney, 1989). On the other hand, mental arithmetic performance is enhanced during visual tasks (Klingner, Tversky, & Hanrahan, 2011; Logie, Gilhooly, & Wynn, 1994). These apparently contradictory findings may occur as visual stimuli generate more complex mental representations, which may facilitate performance on more complicated and non-sequential tasks (e.g., mental arithmetic, schema learning, and pattern-finding; Chen, 2004; Clark & Paivio, 1991; Klingner et al., 2011). It is also noteworthy that visual stimuli are associated with dual coding, as people tend to automatically name visual stimuli, creating both a visual and auditory representation of the stimuli (Paivio, 1990). Dual coding is less likely to occur during auditory tasks, as people do not necessarily generate visual representations of auditory stimuli (Snodgrass et al., 1974). In addition to augmenting memory, dual coding may allow for enhanced information processing as stimuli are stored on the "visuospatial sketchpad," freeing the "articulatory loop" to have more access to information processing resources (Klingner et al., 2011).

In contrast, findings on enhanced performance in auditory tasks suggest that auditory stimuli have more durable feature binding as compared to visual stimuli (Allen et al., 2006; Maybery et al., 2009). Feature binding refers to the association of object features within or across perceptual modalities (Treisman, 1999). Maybery and colleagues (2009) reported greater feature binding in auditory stimuli, such that the specific features of these stimuli were retained for longer periods of time compared to visual stimuli.

Moreover, Allen and colleagues (2006) noted that feature binding in more recently presented visual stimuli disrupts the feature binding of earlier stimuli. Taken together, research suggests that visual stimuli are associated with the encoding of richer representations, while auditory stimuli are associated with more durable feature binding and longer lasting representations.

Notably, there are exceptions to the finding that simple and sequential task performance is enhanced with auditory versus visual stimuli. For example, Klingner and colleagues (2011) found improved performance on visual versus auditory tasks with both digit span and mental arithmetic. Improved performance with visual presentation in Klingner and colleagues' digit span study may be related to lower effort required during their visual condition. Participants demonstrated a decrease in pupil diameter in the visual condition versus auditory condition, with smaller diameter being a common indicator of lower cognitive effort (Klingner et al., 2011). It is therefore possible that some procedural detail of the experimental protocol may have made the auditory condition more difficult than the visual condition, thus reversing the effect seen in other working memory studies. However, given their convincing matching of experimental protocols between auditory and visual presentations, it is not at all clear what factor may have contributed to this contradictory finding. Clearly, more research is needed to establish whether this finding is specific to the protocol used or instead represents a more general challenge to the previous literature.

The present study utilizes a series of n-back tasks varying in terms of stimulus modality (auditory and visual) and difficulty (zero-back, one-back, two-back, and three-back) in order to examine the extent to which auditory and visual tasks influence cognitive performance. Unlike previous studies that utilized either auditory or visual stimuli, the current study holds the experimental procedure constant except for features relevant to the experimental manipulation. It was hypothesized that participants would demonstrate greater performance accuracy in the auditory condition, as the dual coding of visual information may interfere with selective attention and the retention of multiple, sequential items. It was also hypothesized that response times would be faster in the auditory versus visual condition, as an easier task might reduce the amount of time needed for cognitive processing.

Method

Participants

Twenty-five Psychology students with normal or corrected-to-normal vision were enlisted from an online recruitment system in exchange for course credit. Participants ranged in age from 18 to 23 years (*Mdn* = 19). Ten participants were women and fifteen were men. Four participants identified as African-American, three as Asian, three as biracial, two as Hispanic, and 13 as Caucasian.

Stimuli and Apparatus

Participants completed auditory and visual conditions of n-back tasks that ranged in difficulty (zero-back, one-back, two-back, and three-back). For each n-back, a sequence of letters was presented one letter at a time. Letters were presented verbally in the auditory condition and visually in the visual condition. Participants were instructed to press the "Z" key when the letter presented was the target and the "/" key when the letter presented was a nontarget. For example, during the zero-back task participants pressed the "Z" key when the target letter "X" was presented and pressed the "/" key when a nontarget letter (i.e., not an "X") was presented. During the one-back task, participants were instructed to refer "one letter back" to find the target stimuli. That is, a letter was considered a target if it matched the letter that came before it. During the two-back task, letters were considered a target if they matched the letter that was presented two letters back in the sequence, and so on. One-third of the stimuli were targets.

A PC computer controlled stimulus presentation and data collection using Eprime software (Sharpsburg, PA). Participants were seated approximately 66 cm away from a 1920 x 1080 computer monitor with a stabilizing chin- and forehead-rest. Each trial block began with the presentation of a square. The boundary of the square was delineated with a thin dark gray border. The square remained on the screen across each trial block as letters were presented, and participants were instructed to look toward the center of the square during both auditory and visual conditions. For each trial in the visual condition, an individual letter (approx. visual angle 1.38°) was displayed in the center of the square for 500 ms. After the letter disappeared, participants had an additional 3000 ms to indicate whether the letter was a target or nontarget via a keypress. In the auditory condition, letters were presented verbally over speakers while the square was displayed on the screen. Each trial began with the presentation of a letter, which lasted between 549 and 780 ms ($M = 690$, $SD = 68$), and participants had a total trial length of 3500 ms to respond. Participants were instructed to respond as quickly and accurately as possible after each letter was presented. Letters and the boundary of the square were shown in dark gray (95 cd/m^2) and the background color was light gray (200 cd/m^2).

Procedure

All research was carried out in accordance with a protocol approved by Indiana University Bloomington's Institutional Review Board. After obtaining informed consent, participants were seated in front of a computer monitor in a stabilizing headrest. Participants received verbal and written instructions prior to the presentation of each n-back task and condition. To become acquainted with the task, participants completed 15 practice trials prior to the auditory and the visual conditions of each n-back. Participants received feedback on their performance and had an opportunity to ask the experimenter questions after completing the practice trials. They then proceeded to complete 30 block 1 trials of

the corresponding n-back and condition. After completing practice and block 1 trials for all n-backs and conditions, participants completed block 2 trials. For each trial block, one third of stimuli presented were targets and two thirds were nontargets. The order of target and nontarget letters was random. Lures (e.g., n = 1 foils in 2-back or 3-back tasks) were not included. The order of n-backs was randomized and the order of auditory and visual conditions was counterbalanced across participants. Participants completed a demographic survey following the n-back tasks.

Results

Performance Accuracy

A 2 (condition: auditory vs visual) × 3 (block: practice, block-1, block-2) × 4 (n-back: 0, 1, 2, 3) repeated-measures ANOVA was used to assess the effect of condition and block on performance accuracy across n-back tasks (Figure 1). Greenhouse-Geisser adjustments were applied when the assumption of heterogeneity was violated. There was a statistically reliable main effect of condition ($F(1,24) = 33.12$, $p < .001$, $\eta_p^2 = .58$). Percent correct was significantly higher during the auditory condition ($M = .95$, $SD = .02$) than the visual condition ($M = .92$, $SD = .03$).
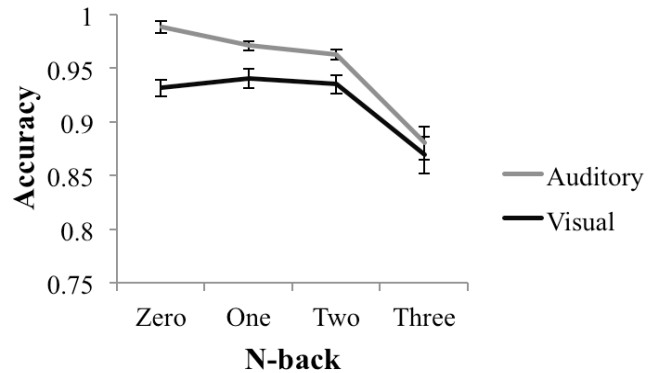

Figure 1: Average performance accuracy during auditory and visual conditions across n-back tasks.

There were also statistically significant main effects of block ($F(2,48) = 7.12$, $p = .002$, $\eta_p^2 = .23$) and n-back task ($F(1.88,44.99) = 26.36$, $p < .001$, $\eta_p^2 = .52$; see Table 1). The main effects were superseded by a significant interaction between block and n-back, $F(3.13,75.13) = 3.03$, $p = .03$, $\eta_p^2 = .11$. Simple effects analyses revealed that participants had a significantly lower percent correct on the three-back ($M = .87$, $SD = .07$), as compared to the other n-back conditions ($M_{zero-back} = .96$, $SD_{zero-back} = .03$, $M_{one-back} = .96$, $SD_{one-back} = .03$, $M_{two-back} = .95$, $SD_{two-back} = .03$). This effect was statistically significant across practice ($F(1.57,37.75) = 15.66$, $p < .001$, $\eta_p^2 = .40$), block 1 ($F(2.13,51.13) = 13.36$, $p < .001$, $\eta_p^2 = .36$), and block 2 trials ($F(2.24,53.68) = 9.57$, $p < .001$, $\eta_p^2 = .29$). Interactions between condition and block, condition and n-back, and

condition, block, and n-back based on performance accuracy were non-significant, p > .05.

Table 1: Means and standard deviations for percent correct and response time based on n-back task.

	Accuracy (out of one)		Response Time (ms)	
N-back	Auditory	Visual	Auditory	Visual
Zero-back	.99 (.03)	.93 (.04)	817 (107)	492 (75)
One-back	.97 (.02)	.94 (.04)	896 (123)	581 (115)
Two-back	.96 (.02)	.93 (.04)	1107 (194)	671 (180)
Three-back	.88 (.08)	.87 (.08)	1221 (259)	787 (240)

Note. $M(SD)$

D-prime was also calculated to examine performance accuracy based on n-back, trial block, and stimulus modality. Unlike percent accuracy, d-prime is a sensitivity index that reflects signal detection while controlling for biases in responding. Using d-prime as the dependent variable, a 2 (condition) × 3 (block) × 4 (n-back) repeated-measures ANOVA revealed reliable main effects of block ($F(2,48) = 12.96$, $p < .001$, $\eta_p^2 = .35$) and n-back task ($F(3,72) = 28.75$, $p < .001$, $\eta_p^2 = .55$). However, the main effect of condition (stimulus modality) was non-significant, $F(1,24) = 1.01$, $p = .33$.

In addition, there was a significant interaction between block and n-back ($F(6,144) = 4.54$, $p = .006$, $\eta_p^2 = .12$). Simple effects analyses revealed significant differences in d-prime based on n-back across practice ($F(6,144) = 14.70$, $p < .001$, $\eta_p^2 = .38$), block 1 ($F(3,72) = 20.14$, $p < .001$, $\eta_p^2 = .46$), and block 2 trials ($F(2.40,57.65) = 10.13$, $p < .001$, $\eta_p^2 = .30$). Across blocks, d-prime was significantly lower in the three-back, as compared to the other n-backs. Interactions between condition and block, condition and n-back, and condition, block, and n-back based on performance accuracy were non-significant, $p > .05$.

Response Time

Response time was calculated based on the time between trial onset and participant key press. A 2 (condition) × 3 (block) × 4 (n-back) repeated-measures ANOVA was used to examine differences in response time based on condition, block, and n-back task (see Figure 2). This analysis revealed a reliable main effects of condition ($F(1,24) = 332.19$, $p < .001$, $\eta_p^2 = 93$, block ($F(2,48) = 20.57$, $p < .001$, $\eta_p^2 = .46$), and n-back ($F(1.60,38.49) = 55.09$, $p < .001$, $\eta_p^2 = .70$). In addition, there were statistically significant interactions between condition and n-back ($F(1.92,46.17) = 10.47$, $p < .001$, $\eta_p^2 = .30$), as well as block and n-back $F(2.97,71.33) = 3.61$, $p = .02$, $\eta_p^2 = .13$). The two-way interaction between condition and block and the three-way interaction between condition, block, and n-back task was non-significant, $p > .05$.

Two repeated-measures ANOVAs were used to examine the simple effects of n-back on response time for each condition. Simple effects analyses of the condition and n-back interaction indicated there was a significant difference in response time based on n-back in the auditory condition,

$F(1.41,33.71) = 57.29$, $p < .001$, $\eta_p^2 = .71$ and in the visual condition, $F(1.86,44.51) = 35.33$, $p < .001$, $\eta_p^2 = .60$. All contrasts between zero-back, one-back, two-back, and three-back conditions were significant, $p \leq .005$. Participants responded fastest in the zero-back task ($M = 654.30$, $SD = 84.74$), followed by the one-back task ($M = 738.48$, $SD = 111.77$), the two-back task ($M = 888.89$, $SD = 168.84$), and the three-back task ($M = 1003.98$, $SD = 232.19$). In addition, a series of paired-samples t-tests were used to examine the simple effects of condition on response time for each n-back task. Simple effects analyses revealed statistically reliable differences between response times based on condition across zero-back ($t(24) = 21.74$, $p < .001$), one-back ($t(24) = 19.60$, $p < .001$), two-back ($t(24) = 13.39$, $p < .001$), and three-back tasks ($t(24) = 11.73$, $p < .001$; see Table 1). Across n-back tasks, participants responded more slowly during the auditory ($M = 1009.99$, $SD = 146.59$) than visual condition ($M = 632.83$, $SD = 137.16$).
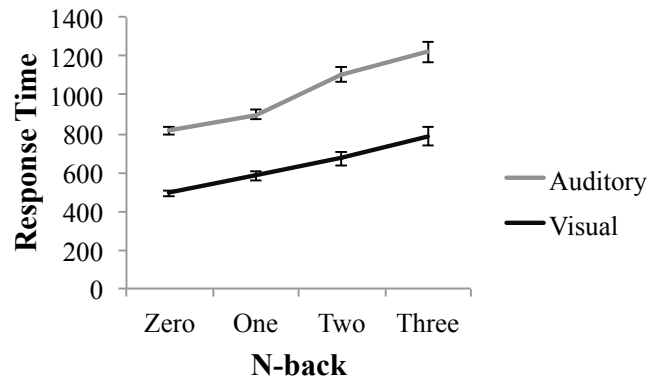

Figure 2: Average response time during auditory and visual conditions across n-back tasks.

A series of repeated-measures ANOVAs were used to examine the simple effects of n-back on response time for each block. Response time varied as a function of n-back across practice ($F(1.84,44.14) = 38.22$, $p < .001$, $\eta_p^2 = .61$), block 1 ($F(1.78,42.77) = 46.95$, $p < .001$, $\eta_p^2 = .66$) and block 2 trials ($F(1.57,37.62) = 37.43$, $p < .001$, $\eta_p^2 = .61$). Across trial blocks, participants responded fastest in the zero-back task, followed by the one-back task, the two-back task, and the three-back task, $p \leq .005$ (see Table 1). An exception was during practice trials, where response time was statistically equivalent during two-back and three-back tasks, $p > .05$.

Discussion

Collapsed across auditory and visual conditions, performance accuracy was equivalent for zero-back, one-back, and two-back tasks. Changes in task difficulty were only reflected in performance accuracy during the three-back task, during which performance accuracy was significantly lower than the other n-backs. On the other hand, performance differences based on task difficulty were reflected in response time across n-back tasks. Participants

responded fastest during the zero-back task, followed by the one-back task, the two-back task, and the three-back task. The effect of the n-back task on performance was more evident for response times than for accuracy, perhaps because of a ceiling effect.

Performance accuracy was greater during the auditory n-back than the visual n-back condition. The advantage of auditory stimuli was most apparent during easier versions of the task. In the zero-back, for example, participants performed approximately six percent better on the auditory than on the visual n-back task. The performance advantage afforded during the auditory condition was less pronounced with more difficult n-back tasks, such that performance on the auditory condition was only one percent better than on the visual condition during the three-back (Figure 1). It is important to note, however, that this advantage for auditory stimuli was no longer significant when the data were analyzed with a more conservative d' signal detection measure suggesting the tentativeness of this finding.

Contrary to the hypothesis that participants would respond faster during the auditory than visual condition, participants responded slower during the auditory condition. Response times increased from an average of 633 ms in the visual condition to 1,010 ms in the auditory condition. This may have occurred as visual stimuli are available immediately upon viewing, while auditory stimuli unfold over time. That is, visual stimuli were presented on the screen and maintained for 500 ms during each trial. In contrast, auditory stimuli took between 549 and 780 ms (M = 690, SD = 68) to present as they were said aloud over a speaker. Given that responses could occur during the presentation of the auditory stimuli, response time was calculated based on the time between trial onset and the participant's key press.

The difference in response time during the auditory versus visual condition was not directly proportional to the difference in time the stimuli were presented. There was a non-significant correlation between response time and time required for letter to be read aloud, $r = .11$. In addition, the auditory stimuli were presented an average of 190 ms longer than the visual stimuli, given the time it took for a letter to be read aloud. Participants, on average, had response times 377 ms longer in the auditory than the visual condition. This difference could be attributable to the auditory stimuli not being processed until read to completion, whereas the visually-presented letter's features are available to be processed immediately upon viewing. Although this hypothesis is rational, the evidence discussed next suggesting that the differences are a function of memory processing as opposed to perceptual encoding contradicts it.

In the present study, the zero-back task provides a baseline response time difference based on perceptual encoding between auditory and visual conditions. That is, in the zero-back tasks participants respond as fast as possible any time the letter "X" is presented. This is different from the other n-back tasks, during which participants have to monitor and store sequences of information to respond

accurately. Average response times in the zero-back task were 462 ms in the visual condition and 817 ms in the auditory condition, such that participants took an average of 325 ms longer to perceptually encode information in the auditory condition. The difference in response time between conditions increased with task difficulty, such that three-back response times were an average of 433 ms longer in the auditory than visual condition. Using zero-back as a baseline for perceptual encoding, the widening gap in response times between conditions with increased task difficulty can be attributed to differences in memory processing, as opposed to perceptual encoding.

As such, there was dissociation in the results as a function of stimulus condition. The response times suggested that visual stimuli were processed more quickly, but the accuracy measure suggested that the auditory stimuli were processed more accurately. One reason for the visual stimuli resulting in lower accuracy is that feature binding in more recently presented visual stimuli may have disrupted the feature binding of earlier stimuli (Allen et al., 2006).

These findings lend credibility to the idea that performance during sequential working memory tasks is facilitated by auditory as opposed to visual presentation of stimuli. As previously reported, the results from a number of studies demonstrate enhanced performance during auditory conditions when tasks are "simpler" or sequential (Chen, 2004; Clark & Paivio, 1991; Greene, 1992; Penney, 1989). One exception to this finding is the evidence reported by Klingner and colleagues (2011) suggesting more accurate responding to visual than auditory stimuli in three different working memory tasks. Intriguingly, they also report that the magnitude of the pupil dilation was less for the visual as opposed to auditory stimuli. This suggests that the cognitive load of the participants was lower for the visual than the auditory stimuli. If we assume that faster response times are also associated with less cognitive load, then our results are not altogether inconsistent with those of Klingner and colleagues. Instead, it appears that our results converge with regard to cognitive load, but diverge with regard to accuracy.

These findings have implications for the presentation of n-back tasks in experimental settings. Based on the discrepancies between our results and those of Klingner and colleagues (2011), it is important to specifically consider the selected measure when assessing the effects of stimulus conditions on working memory. In addition, the complexity and difficulty of the task should be considered given that these factors interact with cognitive processes, such as dual coding and feature binding. As emphasized by the growing body of literature generated from the replication debate, researchers should take seriously the fact that numerous experimental details have the potential to influence study outcomes (e.g., Amon & Holden, 2016).

Conclusions

The current study compares performance during auditory and visual n-back conditions with an otherwise identical

experimental protocol. Participants responded less accurately during a visual presentation condition, as compared to an auditory condition where participants were significantly more accurate. This may occur as auditory stimuli yield more durable representations. By contrast, they responded more quickly to the visual than to the auditory stimuli. As such, these findings suggest that stimulus modality influences encoding and cognitive processing on working memory tasks. Future research will benefit from incorporating additional measures of cognitive load, including pupillary response, in order to gain a better understanding of auditory and visual effects on performance. In addition, a condition with both auditory and visual information can be incorporated into future work to examine the extent to which richer mental representations influence working memory performance during simple and sequential cognitive tasks.

Acknowledgments

References

Allen, R. J., Baddeley, A. D., & Hitch, G. J. (2006). Is the binding of visual features in working memory resource-demanding? *Journal of Experimental Psychology: General, 135*, 298-313. doi:10.1037/0096-3445.135.2.298

Amon, M. J., & Holden, J. G. (2016). Fractal scaling and implicit bias: A conceptual replication of Correll (2008). In. A. Papafragou, D. Grodner, D. Mirman, & J. C. Trueswell (Eds.), *Proceedings of the 38th Annual Conference of the Cognitive Science Society* (pp. 1553-1558). Philedelphia, PA: Cognitive Science Society.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G.H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory*. New York: Academic Press.

Beaman, C. P. (2002). Inverting the modality effect in serial recall. *The Quarterly Journal of Experimental Psychology, 55*, 371-389.

Burgess, N., & Hitch, G. J. (1999). Memory for serial order: A network model of the phonological loop and its timing. *Psychological Review, 106,* 551-581.

Chen, C. (2004). *Information visualization: Beyond the horizon*. London: Springer Verlag.

Clark, J., & Paivio, A. (1991). Dual coding theory and education. *Educational Psychology Review, 3*, 149-210.

Crowder, R. G., & Morton, J. (1969). Precategorical acoustic storage (PAS). *Perception and Psychophysics*, *5*, 365–373.

Fuster, J. M. (1995). *Memory in the cerebral cortex: An empirical approach to neural networks in the human and nonhuman primate*. Cambridge, MA: MIT Press.

Glickman, M. E., Gray, J. R., & Morales, C. J. (2005). Combining speed and accuracy to assess error-free cognitive processes. *Psychometrika, 70,* 405-425.

Goldman-Rakic, P. S. (1996). The prefrontal landscape: Implications of functional architecture for understanding human mentation and the central executive. *Philosophical Transactions: Biological Sciences, 351*, 1445-1453.

Greene, R. L. (1992). *Human memory*. Mahwah, NJ: Erlbaum.

Klingner, J., Tversky, B., & Hanrahan, P. (2011). Effects of visual and verbal presentation on cognitive load in vigilance, memory, and arithmetic tasks. *Psychophysiology, 48*, 323-332.

Logie, R. H., Gilhooly, K. J., & Wynn, V. (1994). Counting on working memory in arithmetic problem solving. *Memory & Cognition, 22*, 395-410.

Maybery, M. T., Clissa, P. J., Parmentier, F. B. R., Leung, D., Harsa, G., Fox, A. M., & Jones, D. M. (2009). Binding of verbal and spatial features in auditory working memory. *Journal of Memory and Language, 61,* 112-133.

Nairne, J. S. (1990). A feature model of immediate memory. *Memory and Cognition*, *18*, 251-269.

Owen, A. M., McMillan, K. M., Laird, A. R., & Bullmore, E. (2005). N-back working memory paradigm: A meta-analysis of normative functional neuroimaging studies. *Human Brain Mapping, 25,* 46-59.

Paivio, A. (1971). *Imagery and verbal processes*. New York: Holt, Rinehart, and Winston.

Paivio, A. (1990). *Mental representations: A dual coding approach*. Oxford, UK: Oxford University Press.

Penney, C. G. (1989). Modality effects and the structure of short-term verbal memory. *Memory & Cognition, 17*, 398-422.

Snodgrass, J. G., Wasser, B., & Finkelstein, M. (1974). On the fate of visual and verbal memory codes for pictures and words: Evidence for a dual coding mechanism in recognition memory. *Journal of Verbal Learning and Verbal Behavior, 13,* 27-37.

Treisman, A. (1999). Solutions to the binding problem: Progress through controversy to convergence. *Neuron, 24*, 105-110.