Performance of a Sonification Task in the Presence of Verbal, Visuospatial, and Auditory Interference Tasks

Michael A. Nees[1] & Bruce N. Walker[2]
[1]Lafayette College, Easton, PA, neesm@lafayette.edu
[2]Georgia Institute of Technology, Atlanta, GA, bruce.walker@psych.gatech.edu

An experiment examined performance with sonifications—a general term for nonspeech auditory displays—as a function of working memory encoding and the demands of three different types of interference tasks. Participants encoded the sonifications as verbal representations, visuospatial images, or auditory images. After encoding, participants engaged in brief verbal, visuospatial, or auditory interference tasks before responding to point estimation queries about the sonifications. Results were expected to show selective impact on sonification task performance when the interference task demands matched the working memory encoding strategy, but instead a pattern of general working memory interference emerged in addition to auditory modal interference. In practical applications, results suggested that performance with auditory displays will be impacted by any interference task, though auditory tasks likely will cause more interference than verbal or visuospatial tasks.

INTRODUCTION

Interference task paradigms have been used extensively to examine information processing conflicts (Ogden, Levine, & Eisner, 1979). If performance of a primary task declines in the presence of a secondary task, then the two tasks are inferred to draw upon the same cognitive structures or resources at some stage of processing. The notion of multiple distinct pools of resources was detailed by Wickens (1984) . Resources are finite. Resources are also allocatable; they can be strategically split between concurrent tasks. Multiple separable resource pools were operationalized as a series of dichotomies—bifurcations that described gross interference phenomena in dual-task time-sharing. These included auditory and visual modalities, verbal and spatial processing codes in working memory, and manual and vocal response modalities. The dichotomies resulted in “either/or” guidelines. Concurrent tasks either compete for the divisible resources of the same member of a dichotomous resource pair, or each constituent task utilizes a different member of a resource pair (and thus the tasks enjoy no interference) for that resource dichotomy.

Recent research, however, has offered conflicting evidence regarding the extent to which working memory functions as a single general resource or multiple domain-specific resources. Some studies have suggested that concurrent verbal and visuospatial tasks can interfere with each other (Morey & Cowan, 2004; Vergauwe, Barrouillet, & Camos, 2010), whereas others (Cocchini, Logie, Della Salla, MacPherson, & Baddeley, 2002) have continued to provide evidence for domain-specific processes for verbal and visuospatial information. Further, although the modality (visual or auditory) of the stimulus and the working memory code (verbal or visuospatial) are theoretically independent (e.g., both auditory speech and visual orthography are processed in verbal working memory), the relationship between modal processing and working memory encoding has not been explained well in the current literature.

These issues are particularly critical for predicting the best use of both speech and nonspeech sounds in systems, as auditory displays increasingly have been deployed in multitasking scenarios—examples include in vehicles (Nees & Walker, 2011a) and at medical workstations (e.g., Sanderson, 2006)—where other (often visual) stimuli must be processed concurrently. Theories like MRT have not explained how nonspeech sounds are registered in working memory, and research (Nees & Walker, 2008, 2011b, 2013) has suggested that the encoding of nonspeech sounds in working memory may not be fully explained by the verbal and visuospatial dichotomy. These studies, however, have not established the presence or absence of meaningful task interference that would warrant consideration of nonspeech auditory imagery as a distinct pool of working memory resources for nonspeech sounds. In fact, the establishment of heuristics for avoiding task interference at the level of processing codes in MRT has proven difficult altogether (Sarno & Wickens, 1995, pp. 127).

Research, however, has suggested that processing codes can be malleable by simple instructions to use particular encoding strategies (Mathews, Hunt, & MacLeod, 1980; Nees & Walker, 2011b, 2013). If working memory interference can be either induced or avoided by encoding instructions, a heuristic for improving performance with auditory displays in multitasking scenarios could be articulated.

Despite the likelihood that auditory displays will be deployed in multimodal, multitasking scenarios, the field of auditory display has produced few empirical results to inform the use of auditory displays in the presence of other stimuli and tasks. This experiment used a mixed design to examine performance for understanding the information contained in sonifications—the general term for nonspeech audio representations of data (Kramer et al., 1999)—in the presence of three different types of interference tasks. Participants heard brief sonifications that represented the price of a stock. In a between-subjects manipulation, they were instructed to use a verbal strategy, a visuospatial imagery strategy, or an auditory imagery strategy to encode the stock prices. Participants estimated the stock price in dollars at a queried time of day (a point estimation task) during a single task block. Three additional blocks of trials were presented in a within-subjects manipulation that required performance of the sonification task in the presence of each of three interference tasks: a verbal interference task, a visuospatial interference task, and a non-verbal auditory interference task. Participants were predicted to experience greater disruption to performance of the sonification task and higher perceived workload when the interference task drew upon the same theoretical processing code resources as the prescribed encoding strategy for the sonification. For example, participants in the verbal encoding group were predicted to show the highest error on the sonification task and highest perceived workload during the verbal interference task trials.

METHOD

Participants

Participants (N = 55, 21 females, M age = 19.9, SD = 1.5 years) were recruited from undergraduate psychology courses and received either course credit or monetary compensation ($10 per hour) for their participation in the study.

Apparatus and Sonification Stimuli

A program written with Macromedia Director 2004 software presented stimuli and collected data. Visual stimuli were shown on 43.2 cm Dell LCD computer monitors. Sounds were presented through Sennheiser HD 202 headphones.

Sonifications were auditory graphs that depicted the price of a fictional stock at opening, mid-morning, mid-afternoon, and closing over the course of a trading day. The prices were represented with 200 ms notes in the MIDI piano timbre, and 300 ms separated each note. Each sonification was 1700 ms in length. Stock prices ranged between a low value of 6 dollars (MIDI C4) to a high value of 106 dollars (MIDI C7). Participants heard the low and high values for mapping dollars to frequency in the instructions. The data for each stimulus were randomly chosen from values within the range. A set of sonification stimuli was constructed, and each stimulus was used in four different point estimation trials (each querying the value of the stock price at one of the four times of day represented in the sonification).

Interference tasks

For the *visuospatial interference task*, participants made judgments about block stimuli (Shepard & Metzler, 1971) from a library of three-dimensional block figures (Peters & Battista, 2008). A block figure was presented as a standard stimulus. After a 5000 ms delay (a blank grey screen), a comparison visual block figure appeared. The comparison stimulus was either a rotated version of the standard stimulus, or a rotated mirror image of the standard. Participants judged whether the comparison stimulus was the standard or its mirror image. Difficulty also was manipulated by varying the angle of rotation of the comparison stimulus (relative to the standard): 15 degrees (easy) and 45 degrees (difficult).

For the *verbal interference task*, participants viewed a 1200 ms presentation of a set of upper case consonants in a modified version of the Sternberg (1966) memory scanning task. After a 5000 ms delay, participants saw a single lower case consonant and determined whether or not the lower case consonant was a member of the original set. Difficulty of the task was manipulated by varying the number of consonants in the initial set: 4 consonants (easy) and 8 consonants (difficult).

The *auditory interference task* was modeled after the task described by Deutsch (1970). Participants heard a standard tone from the equal-tempered musical scale in the octave above middle C followed by a series of to-be-ignored tones. A 1000 ms delay followed the interference tones, and a final comparison tone was heard. Participants' task was to judge whether the comparison tone was the same as or different from the standard. For "different" trials, the comparison tones were one semi-tone different from the standard. All tones were 200 ms in duration with 10 ms onset and offset ramps, and stimuli used the MIDI saxophone instrument. The initial standard tone and each interference tone were all separated by 300 ms of silence. Difficulty of the auditory interference task was manipulated by varying the number of intervening tones between the standard and comparison: 4 tones (easy) and 8 tones (difficult).

Procedure

Participants were randomly assigned to one of three encoding strategy conditions for the sonification task. Participants in the *visual imagery condition* were instructed to encode and rehearse the sounds as a visuospatial image, like a visual graph that represented higher stock prices as higher up on the visual Y-axis and time of day on the visual X-axis. Participants in the *verbal encoding condition* were instructed to encode and rehearse the sounds as a verbal list—a list of values, one for each tone, which named the stock prices from the beginning to the end of the day. Participants in the *auditory imagery* condition were instructed to encode and rehearse the sounds as auditory images of the sonifications, whereby the sensory experience of hearing the tones was to be retained in working memory.

Following instructions for their respective encoding strategies, participants completed 30 single-task point estimation sonification trials. Trials began with a screen that allowed the participant to hear the lower and upper reference tones or the actual sonification stimulus for the trial as many times as needed before proceeding. They pressed either the "Z" or the "?" key to indicate that they had encoded the stimulus according to the assigned strategy. After a 10 s delay (a blank grey screen), participants saw the query for the point estimation task. Participants estimated the price of the stock at given time of the trading day (e.g., "What was the price of the stock at mid-afternoon?"). Across trials all four possible times of day were queried. Across all blocks, the particular sonification stimulus and the queried time of day were randomly ordered. Participants received feedback that gave the correct answer following every point estimation trial with the sonification task throughout the study.

Participants then performed three additional blocks of the sonification task in the presence of interference tasks. The computer program introduced each interference task with instructions. Participants completed 10 practice trials for each interference task alone and then began a block of 20 experimental trials of the sonification task paired with the interference tasks. Participants were instructed to respond as quickly and as accurately as possible to the interference tasks. The interference tasks were inserted during the 10 s delay (see Figure 1). Responses for all interference tasks were made using the "Z" and "?" keys on the computer keyboard, and mappings of responses to keys were counterbalanced across participants. The time between the beginning of the interference task and the query for the sonification task was not permitted to exceed 10 s. Participants who did not log a response within the 10 s window had the trial terminated. Terminated interference trials were repeated. The order of presentation of the interference task blocks was counterbalanced. Participants were instructed to divide their mental resources equally between both tasks.

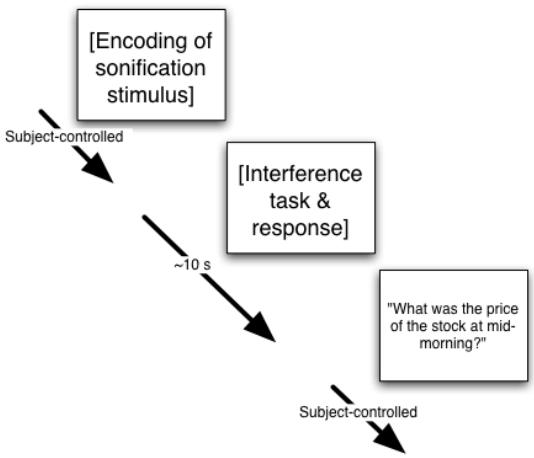

Figure 1. Flow of an interference task trial.

The primary dependent variable was the root mean square (RMS) error of participants' responses to the sonification task for each of the four (single task plus three interference task) blocks of the study. Accuracy data also were collected for each interference task. Participants also completed the NASA-TLX (Hart & Staveland, 1988) as a measure of subjective workload following each of the four blocks of the study.

RESULTS

Four participants (two in the visuospatial encoding condition and one each in the auditory and verbal encoding conditions) reported complete strategy noncompliance for one or more blocks of the study and were excluded case-wise from further analyses. Four additional cases of partial missing data were the result of a computer crash, a dropout, and two instances where participants did not complete all study tasks in the time allotted and had to leave the study. When possible, participants who gave partial data for reasons unrelated to strategy compliance were included in analyses. Data for 11.6% of trials across the study for the remaining 51 participants were excluded because a participant indicated she or he did not use the instructed encoding strategy for the sonification task. For all within-subjects variables, Greenhouse-Geisser corrections were used in analyses where sphericity assumptions were violated.

The analysis of primary interest was a three (encoding strategy: verbal, visuospatial imagery, or auditory imagery) by four (block: single task, verbal interference, visuospatial interference, or auditory interference) mixed ANOVA on the RMS error dependent variable for the sonification task. Forty-eight participants gave complete data across all conditions of the study. Results showed a significant main effect of interference block, F (3,135) = 6.25, p = .001, partial η^2 = .12. The effect of strategy and the interaction of interference block with strategy were not statistically significant, ps > .05. Follow-up pairwise comparisons for the block manipulation (see Figure 2) revealed that performance of the sonification task during the single task block (M = 20.94, SE = 0.77) was better than the sonification task paired with either the visuospatial interference task (RMS error M = 22.75, SE = 1.06), p = .047, or the auditory interference task (M = 25.15, SE = 1.24), p = .001. Performance of the sonification task with the verbal interference task (M = 22.53, SE = 0.92) was not different from the single task block, p > .05. Performance of the sonification task with the auditory interference task was significantly worse as compared to both the visuospatial, p = .026, and verbal, p = .014, interference blocks. There was no difference in performance of the sonification task during the verbal as compared to visuospatial interference task, p > .05.

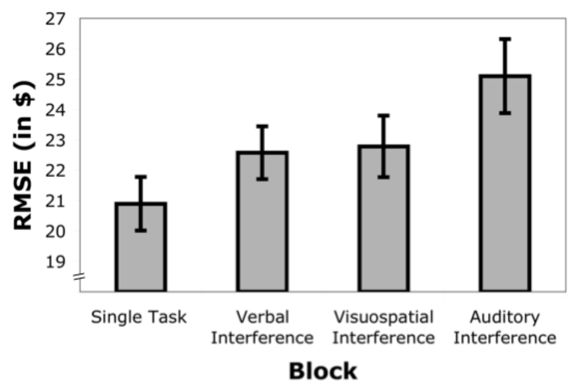

Figure 2. RMS error on the sonification task as a function of block of the study. Error bars represent standard error.

A three (encoding strategy: verbal, visuospatial imagery, or auditory imagery) by three (interference task type: verbal, visuospatial, or auditory) by two (interference task difficulty: easy or hard) mixed ANOVA examined the percent of correct response recorded from each of the interference tasks. Significant main effects were shown for the strategy, F (2,45) = 3.30, p = .046, partial η^2 = .13, the interference task type, F (1.67,75.10) = 40.08, p < .001, partial η^2 = .47, and the difficulty of the interference task, F (1,45) = 7.10, p = .011, partial η^2 = .14. The interaction of interference task type with interference task difficulty was also significant, F (2,90) = 7.82, p = .046, partial η^2 = .15. Nonsignificant effects included the interaction of interference task type with strategy, the interaction of task difficulty with strategy, and the three way interaction, ps > .05.

Follow-up comparisons for the effect of strategy showed that participants using the verbal encoding strategy (percent correct $M = 66\%$, $SE = 2.2\%$) performed significantly worse across interference tasks than participants using the visuospatial imagery encoding strategy ($M = 74\%$, $SE = 2.1\%$), $p = .02$. Participants using the verbal strategy did not differ from participants using the auditory imagery encoding strategy ($M = 68\%$, $SE = 2.0\%$), $p > .05$. The difference between the visuospatial and auditory imagery conditions did not reach statistical significance, $p > .05$.

Follow-up comparisons on the effect of interference task type showed that participants were better at the verbal interference task ($M = 83\%$, $SE = 1.4\%$) than the visuospatial interference task ($M = 61\%$, $SE = 2.1\%$), $F(1,45) = 102.86$, $p > .001$, partial $\eta^2 = .70$ or the auditory interference task ($M = 64\%$, $SE = 2.2\%$), $F(1,45) = 55.12$, $p > .001$, partial $\eta^2 = .55$. The difference between the visuospatial and auditory tasks was not significant, $p > .05$. For the difficulty independent variable, participants did better on easy interference tasks ($M = 72\%$, $SE = 1.3\%$) as compared to difficult interference tasks ($M = 67\%$, $SE = 1.6\%$). Both main effects should be interpreted cautiously in light of the significant interaction of difficulty with interference task type, which was followed up with a simple effects analysis. For the verbal interference task, performance was better with the easy ($M = 90\%$, $SE = 1.4\%$) as compared to the difficult ($M = 76\%$, $SE = 2.5\%$) version of the task, $p < .001$. For the visuospatial interference task, performance was not significantly better with the easy ($M = 64\%$, $SE = 2.9\%$) as compared to the difficult ($M = 58\%$, $SE = 2.8\%$) version of the task, $p > .05$. For the auditory interference task, performance also was not better with the easy ($M = 61\%$, $SE = 2.2\%$) as compared to the difficult ($M = 67\%$, $SE = 3.3\%$) version of the task, $p > .05$. The interaction showed that the main effect of difficulty, then, was carried only by the large difference in performance for easy versus difficult verbal interference tasks.

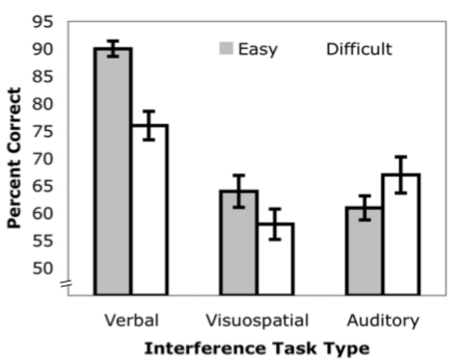

Figure 3. Percent correct on the interference tasks as a function of interference task type and interference task difficulty. Error bars represent standard error.

A 3 (encoding strategy: verbal, visuospatial imagery, or auditory imagery) by 4 (block: single task, verbal interference, visuospatial interference, or auditory interference) mixed ANOVA on the TLX composite scores examined the possibility that the encoding strategy selectively affected perceived workload for interference tasks. Forty-seven participants gave complete TLX data across all conditions of the study. Results showed a significant main effect of interference block, $F(3,132) = 8.86$, $p < .001$, partial $\eta^2 = .17$. The main effect of encoding strategy, and the interaction of interference block with encoding strategy were not significant, *p*s > .05. Follow-up comparisons showed that perceived workload was significantly lower for the single task block (TLX $M = 12.28$, $SE = .40$) as compared to the verbal ($M = 13.61$, $SE = .40$, $p = .023$), visuospatial ($M = 14.17$, $SE = .32$, $p < .001$), and auditory interference ($M = 13.99$, $SE = .37$, $p = .001$) blocks. Workload was significantly higher during both the visuospatial, $p = .047$, and auditory, $p = .032$, interference task blocks as compared to the verbal block; these blocks were not different from each other, $p > .05$.

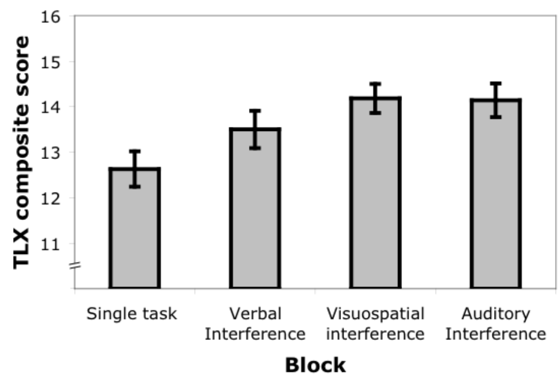

Figure 4. TLX composite score as a function of block. Error bars represent standard error.

DISCUSSION

One of the most common arguments in favor of the use of auditory displays (e.g. in computers systems, vehicles, medical monitoring stations, etc.) posits the benefits of dividing information across modalities (e.g., Bonebright & Nees, 2009; Stokes, Wickens, & Kite, 1990). These arguments frequently have been justified with reference to MRT, and indeed, the modalities dimension of MRT offers a useful design heuristic that was confirmed here. The processing codes dimension of MRT, however, has been more difficult to use to guide design, in part due to problems with making a priori designations of processing codes (e.g., orthography is visual but should assume a verbal processing code). This is especially true of nonspeech sounds—a class of stimuli for which the working memory code is indeterminate (though Nees & Walker, 2011b, 2013, found evidence suggesting nonspeech sounds may assume visual, verbal, or auditory codes). The current study suggested that the combination of a sonification task with an auditory task did indeed, as the modalities dimension of MRT would predict, result in worse performance of the sonification task. Simply diverting the second task to the visual modality, however, did not alleviate interference relative to the single task baseline. The visuospatial and verbal interference tasks both used only visual stimuli, and the visuospatial task also resulted in worse performance of the sonification task, as very likely would a more difficult version of the verbal task used here. These results suggested that designers should assume that performance with auditory displays will be impacted by the mere presence of a second task—regardless of its modality—although our findings suggest that a second task in the auditory modality may impact performance with auditory displays to a greater extent.

The attribution of the source of interference (or a lack thereof) in resource models can be difficult. Clearly more research is needed to determine the role played by working memory processing codes in dual-task interference. Other research has suggested that different encoding strategies can result in considerably different performance on some tasks (Logie, Della Salla, Laiacona, Chalmers, & Wynn, 1996; MacLeod, Hunt, & Mathews, 1978). Working memory codes are inherently difficult to establish empirically. A priori determination of encoding strategies frequently attempt to simply instruct participants to use a particular strategy (as in the sonification task of the current study) or induce task demands that assume the task can only be accomplished with a particular working memory code (as in the interference tasks of the current study).

In cases where information must first be encoded and then maintained/rehearsed in the presence of an intervening task, the results here suggested that the qualitative format of the processing code (i.e., the encoding strategy) offers no heuristic value in predicting interference. General interference is to be expected regardless of the qualitative combinations of encoding strategies (i.e., processing codes) and qualitative (i.e., visuospatial, verbal, etc.) task demands on working memory, and this interference appears to be a function of the combined (quantitative) difficulty of both tasks rather than their domain-specific (qualitative) characteristics. This interpretation is entirely consistent with the notion of a central executive in working memory that behaves as a general resource during multitasking. This finding may be attributable to the sequential-but-overlapping interference task design, and may further suggest that the processing codes dimension of MRT is most relevant during actual perception (i.e., true time-sharing of tasks as opposed to maintenance and rehearsal of a previously perceived stimulus). For practical applications of sound in systems, our results suggest performance with auditory displays that require delayed (post-perceptual) responses from working memory will be impacted by any intervening task regardless of its qualitative properties, though auditory interference tasks will likely cause more interference than verbal or visuospatial tasks. Further research is needed to examine, for example, scenarios in which the auditory display task receives secondary emphasis.

Acknowledgments

This work partially fulfilled requirements for the first author's doctoral dissertation. We thank Amy Gress and Courtney Lynch for collecting data for this experiment.

REFERENCES

Bonebright, T., & Nees, M. A. (2009). Most earcons do not interfere with spoken passage comprehension. *Applied Cognitive Psychology*, *23*(3), 431–445.

Cocchini, G., Logie, R. H., Della Salla, S., MacPherson, S. E., & Baddeley, A. (2002). Concurrent performance of two memory tasks: Evidence for domain-specific working memory systems. *Memory & Cognition*, *30*(7), 1086–1095.

Deutsch, D. (1970). Tones and numbers: Specificity of interference in immediate memory. *Science*, *168*(3939), 1604.

Hart, S. G., & Staveland, L. E. (1988). Development of the NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human Mental Workload* (pp. 239–250). Amsterdam: North Holland Press.

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

Kramer, G., Walker, B. N., Bonebright, T., Cook, P., Flowers, J., Miner, N., … Tipei, S. (1999). *The Sonification Report: Status of the Field and Research Agenda. Report prepared for the National Science Foundation by members of the International Community for Auditory Display*.

Logie, R. H., Della Salla, S., Laiacona, M., Chalmers, P., & Wynn, V. (1996). Group aggregates and individual reliability: The case of verbal short-term memory. *Memory & Cognition*, *24*(3), 305–321.

MacLeod, C. M., Hunt, E. B., & Mathews, N. N. (1978). Individual differences in the verification of sentence-picture relationships. *Journal of Verbal Learning and Verbal Behavior*, *17*(5), 493–507.

Mathews, N. N., Hunt, E. B., & MacLeod, C. M. (1980). Strategy choice and strategy training in sentence-picture verification. *Journal of Verbal Learning and Verbal Behavior*, *19*(5), 531–548.

Morey, C. C., & Cowan, N. (2004). When visual and verbal memories compete: Evidence of cross-domain limits in working memory. *Psychonomic Bulletin & Review*, *11*(2), 296–301.

Nees, M. A., & Walker, B. N. (2008). Encoding and representation of information in auditory graphs: Descriptive reports of listener strategies for understanding data. In *International Conference on Auditory Display (ICAD 08)*.

Nees, M. A., & Walker, B. N. (2011a). Auditory displays for in-vehicle technologies. In P. Delucia (Ed.), *Reviews of Human Factors and Ergonomics* (pp. 58–99). Thousand Oaks, CA: Sage Publishing/Human Factors and Ergonomics Society.

Nees, M. A., & Walker, B. N. (2011b). Mental scanning of sonifcations reveals flexible encoding of nonspeech sounds and a universal per-item scanning cost. *Acta Psychologica*, *137*, 309–317.

Nees, M. A., & Walker, B. N. (2013). Flexibility of working memory encoding in a sentence-picture-sound verification task. *Journal of Cognitive Psychology*, *25(7)*, 800–807.

Ogden, G. D., Levine, J. M., & Eisner, E. J. (1979). Measurement of workload by secondary tasks. *Human Factors*, *21*(5), 529–548.

Peters, M., & Battista, C. (2008). Applications of mental rotation figures of the Shepard and Metzler type and description of a mental rotation stimulus library. *Brain and Cognition*, *66*(3), 260–264.

Sanderson, P. M. (2006). The multimodal world of medical monitoring displays. *Applied Ergonomics*, *37*, 501–512.

Sarno, K. J., & Wickens, C. D. (1995). Role of multiple resources in predicting time-sharing efficiency: Evaluation of three workload models in a multiple-task setting. *The International Journal of Aviation Psychology*, *5*(1), 107–130.

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, *171*(3972), 701–703.

Stokes, A., Wickens, C. D., & Kite, K. (1990). Auditory displays. In *Display Technology: Human Factors Concepts* (pp. 47–64). Warrendale, PA: Society of Automotive Engineers.

Tsang, P., & Wilson, G. (1997). Mental workload. In G. Salvendy (Ed.), *Handbook of Human Factors and Ergonomics (2nd ed.)* (pp. 417–489). New York: Wiley.

Vergauwe, E., Barrouillet, P., & Camos, V. (2010). Do mental processes share a domain-general resource? *Psychological Science*, *21*(384-390).

Wickens, C. D. (1984). Processing resources in attention. In R. Parasuraman & D. R. Davies (Eds.), *Varieties of Attention* (pp. 63–102). New York: Academic Press.

Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science*, *3*(2), 159–177.

Wickens, C. D., & Liu, Y. (1988). Codes and modalities in multiple resources: A success and a qualification. *Human Factors*, *30*(5), 599–616.