Encoding of Information in Auditory Displays: Initial Research on Flexibility and Processing Codes in Dual-task Scenarios

Michael A. Nees and Bruce N. Walker
School of Psychology
Georgia Institute of Technology
Atlanta, GA

Interest in the use of sound as a means of information display in human-machine systems has surged in recent years. While researchers have begun to address issues surrounding good auditory display design as well as potential domains of application, little is known about the cognitive processes involved in interpreting auditory displays. In multi-tasking scenarios, dividing concurrent information display across modalities (e.g., vision and audition) may allow the human operator to receive (i.e., to sense and perceive) more information, yet higher-level conflicts in the encoding and representation of information may persist. Surprisingly few studies to date have examined auditory information display in dual-task scenarios. This study examined the flexibility of encoding of information and processing code conflicts in a dual-task paradigm with auditory graphs—a specific class of auditory displays that represent quantitative information with sound. Results showed that **1) human listeners can flexibly encode the information in auditory graphs; and 2) dual-task interference may result at the level of cognitive information representation, even when concurrent information is separated by modality**.

INTRODUCTION

As high fidelity audio has become increasingly cheaper and easier to implement in systems, interest in the viability of sound as an alternative to traditional (i.e., visual) information displays has grown accordingly (Flowers, Buhman, & Turnage, 2005). Auditory displays researchers and designers have identified a number of scenarios where sound may be beneficial, including instances where vision is overtaxed or otherwise not a viable mode of information display (e.g., on a mobile device with a small screen, see Brewster, 2002). Recent studies have examined auditory displays in operating rooms (Watson & Sanderson, 2004) and automobile cockpits (McKeown & Isherwood, 2007). Intuitive logic supports the notion that separating concurrent displays of information by modality should be beneficial. A person's eyes (or even more specifically, one's foveal vision) can only focus on a small area of the visual field at any given moment in time, thus the simultaneous display of detailed information by multiple visual displays will likely ensure that the human operator of a system will miss at least part of the displayed information (see Wickens, 2002). Likewise, an all-audio mode of presentation for multiple streams of information can incur problems with masking that will impair the operator's ability to perceive all of the intended message (see, e.g., Durlach et al., 2003).

A number of theoretical perspectives from both basic psychological research (e.g., Baddeley's model of working memory, see Baddeley, 2002) and human factors psychology (e.g., multiple resources, see Wickens, 2002; Wickens & Liu, 1988) have formulated theoretical mechanisms to account for the apparent advantage of dividing simultaneous information presentation across modalities. Baddeley's phonological loop was posited to manage the internal processing of "acoustic and verbal information," while the visuospatial sketchpad was available for processing "visual and spatial information" (p. 86). Wickens and colleagues similarly suggested that separation of information by modality was one way for system designers to avoid information processing conflicts, but their work made explicit yet another level of processing (that was strongly implied in Baddeley's account)—that of the internal code of information representation. According to Wickens and Liu, processing codes represent a continuum anchored by verbal codes and spatial codes. In this view of information processing, a system designer should be concerned not only with the modality of input of information for the human processor, but also with the format of the internal representation assigned to information by the human operator during the performance of cognitive operations. This model offered perhaps a more flexible account of information processing than the Baddeley model, as modality and representation are not necessarily wedded (i.e., acoustic information need not be verbal).

Despite a wealth of research on auditory displays, basic auditory perceptual processes, and music perception, relatively little empirical attention has been paid to the internal representation and processing of sounds at the cognitive level. Likewise, while the advantages of spreading information display across modalities have been a central justification for including sound where possible in a system (e.g., Kramer, 1994), researchers have yet to examine the potential for multimodal conflicts at the level of processing codes. If sound is included in a system to alleviate visual overload by diverting some of the information processing burden from the eyes to the ears, the potential remains for information processing bottlenecks to arise due to overlap in internal information representation, even when the information from both modalities can be sensed and perceived. In most practical applications, the auditory display user will be performing some task *in addition to listening to the auditory display*, yet surprisingly few studies have examined performance with auditory displays under dual-task or other conditions where

the precise nature of conflicts between incoming auditory and visual information can be explored.

The current study examined processing codes—the internal format of information representation—for auditory graphs—a class of auditory displays that use frequency mappings to represent quantitative data (see, e.g., Brown, Brewster, Ramloll, Burton, & Riedel, 2003; Flowers et al., 2005; Nees & Walker, 2007). Typically, the visual Y-axis is mapped to frequency in auditory graphs, with higher frequencies corresponding to higher Y-axis spatial position and therefore higher quantity or "more" (see Walker, 2002, 2007). The visual X-axis is represented by time in an auditory graph, whereby the presentation of data points in time corresponds to scaling of the visual X-axis in some meaningful fashion (e.g., 1 second in the auditory graph = 1 unit on the X-axis, etc.).

Very little is known about how people internally represent the information in nonspeech auditory displays such as auditory graphs. Seminal accounts of cognition and information processing (e.g., Baddeley, 2002) have generally treated the auditory modality as a vehicle for the processing of speech and other material represented in a verbal format (e.g., text presented visually). Verbal labeling is but one of several information encoding possibilities for frequency. Mikumo's (1997) research, for example, suggested that auditory frequency can be encoded in at least 4 different formats: 1) with a verbal label, as in when a note is identified and encoded by its musical name (e.g., A#, etc.); 2) with a visuospatial image that captures the contours (i.e, the ups and downs) of frequency changes in a picture-like format; 3) with sensory-musical codes, whereby the listener attempts to maintain an isomorphic representation of frequency changes by whistling, humming, singing etc.; and 4) with motor codes, whereby tapping (e.g., for rhythmic stimuli) and perhaps even motor codes associated with instrument fingering (e.g., in trained musicians) are used to encode frequency information.

The primary research questions approached here are: 1) To what extent is the internal representation of information in auditory graphs malleable by instruction? and 2) Do secondary tasks aimed at disrupting specific representational formats (i.e., verbal and spatial) impact performance with auditory graphs as a function of encoding format? As such, participants were instructed and trained to encode auditory graph stimuli as either visuospatial mental images (like pictures in the mind) or as verbal lists (like a table of values). After a period of practice with feedback, participants from each training condition experienced both verbal and spatial secondary interference tasks while encoding the information in auditory graphs. We hypothesized that encoding could be accomplished flexibly as either a visuospatial image or as a list. We further expected to find a dissociation such that a spatial secondary task would be more disruptive to performance with auditory graphs than a verbal secondary task for participants who were instructed to encode the information as a visuospatial image; we predicted the converse pattern of interference for participants who encoded the auditory graphs as verbal lists.

METHOD

Participants

Participants (N = 66; 36 males and 30 females; mean age = 19.6 years, SD = 1.8) were recruited from undergraduate psychology classes at the Georgia Institute of Technology and were compensated with course extra credit.

Apparatus

Visual presentations were made on a 17 in. (43.2 cm) Dell LCD computer monitor, while auditory presentations were delivered via Sennheiser HD 202 headphones. All presentations of stimuli and data collection were accomplished with the Macromedia Director MX 2004 software package.

Auditory Graph Stimuli

Data sets for stimuli. All auditory graph stimuli depicted the price of a stock over the course of an 8-hour trading day, from 8 am to 4 pm. One data point representing each hour on the hour was used, thus a total of 9 discrete data points were present in each data set. A variety of different data sets for stimuli were created. For all data sets, the minimum price of the stock over the course of the trading day was 6 dollars, while the maximum price was 106 dollars. These constraints ensured consistent scaling of the data to frequency mapping across stimuli.

Data sets were constructed to have 0, 1, or 2 trend reversals. Stimuli with 0 trend reversals were either simple linear increasing or decreasing data sets, while 1 trend reversal data sets were parabolic. Data sets with 2 trend reversals were pseudo-sinusoidal. From these basic data contour patterns, data points (excluding the fixed minima and maxima, which were achieved in each data set) were given variability by randomly adding or subtracting values between 0 and 5 dollars from each data point. The data sets, therefore, were systematically constructed, but at the same time offered variety and varying levels of complexity (see Nees & Walker, in press). Two data sets were created for practice trials, 6 data sets were created for the extended practice set with feedback, and 6 additional data sets were created for the dual-task trials. The complexity of data sets was balanced across all experimental trials (e.g., the extended practice and dual-task trials), such that all participants experienced equal numbers of graphs with 0, 1, or 2 trend reversals for a given block of the experiment (see Procedure below).

Sonification of auditory graph data. Data sets were sonified using the Sonification Sandbox software package, version 4.2.1 (Davison & Walker, 2007). The MIDI piano timbre was used, and one second of time in the auditory graph was mapped to one hour in the trading day. Therefore, one tone was played roughly every second, and each auditory graph was 8 seconds in length. Data were mapped to frequency such that MIDI note 43 (97.99 Hz) represented the lowest stock price of the trading day (6 dollars) and MIDI note 91 (1567.8 Hz) represented the highest stock price of the trading day (106 dollars). Data points that fell in between

notes on the equal-tempered musical scale were rounded to the nearest note on the musical scale. For a more detailed discussion of auditory graph design, see (Brown et al., 2003; Nees & Walker, 2007)

Primary Tasks

For a given trial, participants were asked to perform one of two tasks with the auditory graph stimulus: point estimation and local trend identification. While these are but two of many possible graph reading tasks, they offer a representative starting point for investigating the research questions of the current study.

Point estimation task. The point estimation task asked participants to identify the price of the stock at a given hour of the trading day (e.g., "What was the price of the stock at 10 am?"). Performance data and task analyses of point estimation tasks with sonified displays of quantitative data have been reported extensively in previous research (e.g., Nees & Walker, in press; Smith & Walker, 2005). The primary dependent variable for the point estimation task with auditory graphs was the root mean squared (RMS) error of responses.

Local trend identification task. The trend identification task asked participants to identify the direction of the stock price's change between two successive hours of the trading day (e.g., "Was the price of the stock increasing or decreasing between 10 am and 11 am?"). Participants' responses were limited to either "stock price was increasing" or "stock price was decreasing." For the trend identification task, percent correct was scored for each block of the study.

Procedure

Following informed consent, participants were randomly assigned to either the visuospatial imagery encoding condition or the verbal encoding condition. Participants experienced a brief (~20 min) training paradigm that featured a short presentation followed by 16 training practice trials:

Visuospatial imagery encoding. Participants were instructed to visualize an image of the data points on a visual graph in their minds as the auditory graph stimuli unfolded. During the initial 16 training trials, participants saw a visual graph of the data unfold as the auditory graph stimulus was played on each trial. The visuospatial imagery encoding emphasized that participants were only to use the imagery strategy to accomplish the study tasks.

Verbal training. Participants were instructed to assign a verbal label to each data point in the auditory graph stimuli and use the verbal labels to perform the study tasks. In other words, participants were encouraged to think of the stimuli as an auditory table and to encode the data as a verbal list. During the initial 16 training trials, participants were prompted with a list of values that populated as the auditory graph stimulus unfolded. This condition emphasized that participants were to use only the verbal list strategy to encode the data in the auditory graphs.

Following the initial training sessions, participants experienced 96 single task trials with auditory graphs (48 point estimation trials and 48 trend identification trials). These 96 trials used the same set of auditory graph stimuli for all participants, regardless of their encoding condition. For a given trial, participants listened to an auditory graph (with no visual graph or table), then answered either a point estimation or trend identification question about the stimulus. Participants were told the opening price of the stock as a reference, and they were only permitted to listen to an auditory graph once during each of these trials. Following each trial, feedback was provided about the correct answer. At three separate times during the 96 trials, participants were reminded about using only their respective encoding strategies (visuospatial imagery or verbal labeling) to accomplish the tasks. Six different auditory graph stimuli were used during this extended session: two each with 0, 1, or 2 trend reversals. Note that each of these graphs (with 0, 1, or 2 trend reversals) could be constructed with an initially increasing or decreasing trend; half of the auditory graph stimuli initially increased while the other half initially decreased. Trials during this extended block were randomly interleaved.

Following the longer block of single task trials, participants next experienced two additional blocks of 16 trials each (8 point estimation trials and 8 local trend identification trials during each block) with a concurrent interference task. The order of presentation of these blocks was counterbalanced across participants.

Spatial interference secondary task: Participants were required to make judgments about Shepard-Metzler type block stimuli (see Peters & Battista, in press; Shepard & Metzler, 1971). Two visual block figures were presented concurrently. The left figure was a standard stimulus, while the right figure was a comparison. The comparison stimulus was either the standard stimulus rotated 160 degrees, or a mirror image of the standard rotated 160 degrees. During each spatial interference trial, participants judged whether the comparison stimulus was a rotated depiction of the standard or a rotated mirror image of the standard.

Verbal interference secondary task: Participants viewed a brief 1200 ms presentation of 6 upper case consonants in a modified version of the Sternberg task (Sternberg, 1966). After a 2500 ms delay (i.e., a blank screen), participants saw a single lower case consonant. Their task was to determine whether or not the lower case consonant was a member of the original 6 consonant set.

For dual-task trials, presentation of the auditory graph began simultaneously with the beginning of the secondary task trial such that they were required to listen to the auditory graph while also attending to the visual stimuli of the secondary task. Participants were instructed to log a response for the secondary task with the mouse while the auditory graph was still playing. If no response was logged before the completion of the auditory graph stimulus during a dual-task trial, both secondary task and primary task data were excluded from analyses for that trial. For both secondary tasks, accuracy and reaction time were recorded for each trial, and data for trials where an incorrect secondary task response was logged were excluded from final analyses. As the secondary tasks were speeded response tasks, reaction time was the dependent variable of primary interest.

RESULTS

For the between group manipulation of encoding format, no significant differences in mean performance across the 96 single-task practice trials were found for either the point estimation task [visuospatial encoding M =21.6 dollars *RMS* error, SD =7.3; verbal encoding M = 22.6 dollars *RMS* error, SD = 7.6; t (64) = -0.53, p = .60] or the local trend identification task [visuospatial encoding M = 85% correct, SD =12%; verbal encoding M = 83% correct, SD = 11%; t (64) = 0.94, p = .37].

For the dual-task portion of the study, a pair of 2 (encoding condition: visuospatial imagery versus verbal list) x 2 (interference condition: visuospatial secondary task versus verbal secondary task) mixed ANOVAs were performed, one for each of the auditory graphing tasks. For the *point estimation task*, a significant effect of interference task was found [F(1,64) = 4.31, p = .04, partial η^2 = .06], with worse average performance (higher *RMS* error) evident for performance of the auditory graph point estimation task in the presence of the verbal interference task for both encoding conditions (see Figure 1). The main effect of encoding condition and the interaction between encoding condition and interference condition were both nonsignificant [F(1,64) = 0.66, p = .42, and F(1,64) = 0.52, p = .48, respectively]. There was no systematic relationship between mean point estimation task performance (*RMS* error*)* and mean reaction time for either the visuospatial interference task (r = .03, p = .83) or the verbal interference task (r = .16, p = .19).

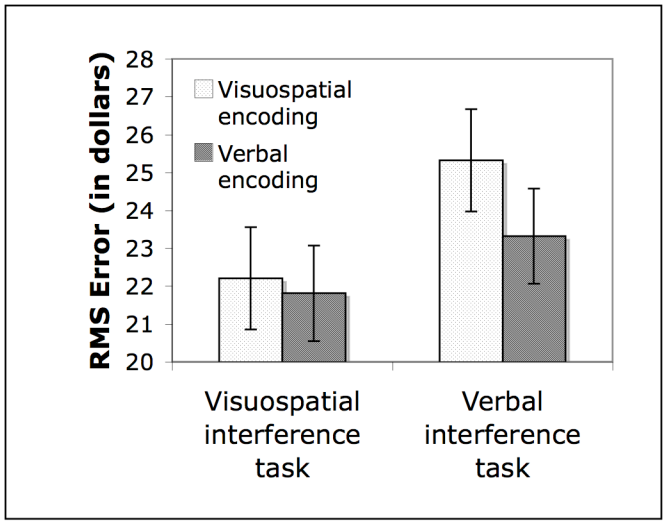

Figure 1. RMS error in dollars on the point estimation auditory graph task as a function of encoding condition and interference task. Higher RMS error means worse performance; error bars represent standard error.

For the local *trend identification auditory graphing task* performance during dual-task trials, neither the main effect of encoding condition nor the main effect of interference task were significant [F(1,64) = 0.98, p = .33, and F(1,64) = 1.46, p = .23, respectively]. A significant interaction of encoding condition with interference task was found [F(1,64) = 4.84, p = .04, partial η^2 = .07], however, and is depicted in Figure 2. Participants in the visuospatial encoding condition performed better than in the verbal encoding condition *only* in the presence of a visuospatial interference task. There was no systematic relationship between mean trend identification task performance (*RMS* error*)* and mean reaction time for either the visuospatial interference task (r = -.07, p = .59) or the verbal interference task (r = -.11, p = .38).

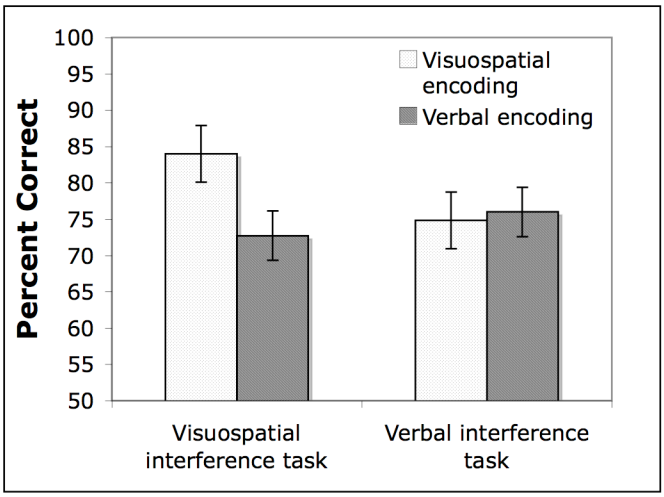

Figure 2. Percent correct on the auditory graph task as a function of encoding condition and interference task. Error bars represent standard error.

A final exploratory pair of within-subjects ANOVAs compared performance on the point estimation and local trend identification auditory graphing tasks in the presence of another task (i.e., collapsing across interference tasks) to performance during the single-task practice trials. No significant dual-task decrement was found for the point estimation task [F(1,65) = 0.99, p = .32], but the addition of a second task did significantly decrease performance of the trend identification task [single task percent correct M = 84%, SD = 11%; dual-task percent correct M = 77%, SD = 17%; F(1,65) = 15.33, p < .01, partial η^2 = .19].

DISCUSSION

The data failed to confirm the hypothesized patterns of dual-task interference based on processing code conflicts, but a number of interesting findings emerged. The attempted manipulations of encoding format (visuospatial imagery or verbal encoding) did not produce significantly different performance during the extended practice session. Indeed, the encoding manipulation's only significant effect was in the interaction depicted in Figure 2. A number of possible explanations for the general failure to confirm the malleability of processing codes are possible. The internal format of information representation, or processing code, may be an immutable property of the stimulus or an individual difference characteristic of the listener rather than a property that is malleable by instructions or practice. In other words, despite instructions and training emphasizing either visuospatial imagery or verbal list-making, participants may have had difficulty employing the prescribed cognitive strategy.

Another possibility is simply that both encoding strategies allowed for equivalent performance of the auditory graphing tasks across the conditions studied here. The introduction of a second task did not confirm the effectiveness of the encoding manipulation, and the examination of internal cognitive representations remains a difficult empirical enterprise that will require more research to understand as theoretical constructs.

Interestingly, the verbal interference task was more disruptive than the visuospatial interference task to performance of the point estimation auditory graphing task across encoding strategies. These data suggested that performance with auditory displays that employ frequency mapping may be less prone to interference from concurrent visuospatial tasks than concurrent verbal tasks, even when the interfering verbal task is entirely visual (i.e., reading text, etc.). Although the observed effect was small, this finding warrants further investigation.

Perhaps the most curious finding was the interaction such that participants who used visuospatial encoding strategies for the auditory graph trend identification task actually performed better in the presence of a visuospatial task than all other conditions. Our hypotheses predicted the precise opposite, given that a visuospatial encoding strategy for auditory graphs and a concurrent visuospatial secondary task should impact the same pool of cognitive resources at the level of representation or processing code. Again, the participants in our study may not have been able to use a prescribed encoding strategy and may have been relying on other strategies (e.g., verbal, motor, etc.), either alone or in concert with a visuospatial strategy, to accomplish the trend identification task. Another possibility is that the combination of tasks used in the current study did not sufficiently tax the intended theoretical pools of mental resources, although anecdotal reports during debriefing suggested that participants found the dual-task blocks to be very difficult.

The current study offered an initial exploration of both the flexibility of encoding of information in auditory displays that use frequency mappings, as well as an examination of the potential for conflicts at the cognitive level of information representation in multimodal information processing scenarios. We found little evidence to corroborate our hypothesis that internal representations in auditory graphs are malleable, at least under the instructional paradigm employed here. In general, performance with auditory graphs in the current study suffered more during a concurrent verbal task than during a concurrent visuospatial task, a finding that has implications for the appropriate use of frequency mapped auditory displays in applied scenarios. Interestingly, however, the addition of either secondary task did not significantly impair performance as compared to a single task condition for the point estimation auditory graphing task, a finding that suggests the point estimation task may be readily accomplished during performance of another visual task. The same pattern of results did not hold for the trend identification task, which suffered under dual-task conditions.

The current study's findings, then, suggest a complex interplay between task dependencies, display modalities, and internal formats of information representation in multimodal, multi-tasking scenarios. The explanatory power of current theory in both cognitive psychology and human factors research has not been established with respect to nonspeech auditory displays, and more research will be needed to clarify the mechanisms to account for performance with nonspeech auditory displays in complex task environments.

REFERENCES

Baddeley, A. D. (2002). Is working memory still working? *European Psychologist*, 7(2), 85-97.

Brewster, S. (2002). Overcoming the lack of screen space on mobile computers. *Personal and Ubiquitous Computing*, 6(3), 188-205.

Brown, L. M., Brewster, S. A., Ramloll, R., Burton, M., & Riedel, B. (2003). Design guidelines for audio presentation of graphs and tables. *Proceedings of the International Conference on Auditory Display (ICAD2003)* (pp. 284-287), Boston, MA.

Davison, B. K., & Walker, B. N. (2007). Sonification Sandbox reconstruction: Software standard for auditory graphs. *Proceedings of the ICAD 07 - Thirteenth Annual Conference on Auditory Display* (pp. TBD), Montreal, Canada (26-29 June).

Durlach, N. I., Mason, C. R., Kidd, G., Arbogast, T. L., Colburn, H. S., & Shinn-Cunningham, B. (2003). Note on informational masking. *Journal of the Acoustical Society of America*, 113(6), 2984-2987.

Flowers, J. H., Buhman, D. C., & Turnage, K. D. (2005). Data sonification from the desktop: Should sound be part of standard data analysis software? *ACM Transactions on Applied Perception*, 2(4), 467-472.

Kramer, G. (1994). An introduction to auditory display. In G. Kramer (Ed.), *Auditory Display: Sonification, Audification, and Auditory Interfaces* (pp. 1-78). Reading, MA: Addison Wesley.

McKeown, D., & Isherwood, S. (2007). Mapping candidate within-vehicle auditory displays to their referents. *Human Factors*, 49(3), 417-428.

Mikumo, M. (1997). Multi-encoding for pitch information of tone sequences. *Japanese Psychological Research*, 39(4), 300-311.

Nees, M. A., & Walker, B. N. (2007). Listener, task, and auditory graph: Toward a conceptual model of auditory graph comprehension. *Proceedings of the ICAD07 - Thirteenth International Conference on Auditory Display* (pp. 266-273), Montreal, Canada (26-29 June).

Nees, M. A., & Walker, B. N. (in press). Data density and trend reversals in auditory graphs: Effects on point estimation and trend identification tasks. *ACM Transactions on Applied Perception*.

Peters, M., & Battista, C. (in press). Applications of mental rotation figures of the Shepard and Metzler type and description of a mental rotation stimulus library. *Brain and Cognition*.

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

Smith, D. R., & Walker, B. N. (2005). Effects of auditory context cues and training on performance of a point estimation sonification task. *Applied Cognitive Psychology*, 19(8), 1065-1087.

Sternberg, S. (1966). High-speed scanning in human memory. *Science*, 153(3736), 652-654.

Walker, B. N. (2002). Magnitude estimation of conceptual data dimensions for use in sonification. *Journal of Experimental Psychology: Applied*, 8, 211-221.

Walker, B. N. (2007). Consistency of magnitude estimations with conceptual data dimensions used for sonification. *Applied Cognitive Psychology*, 21, 579-599.

Watson, M., & Sanderson, P. M. (2004). Sonification helps eyes-free respiratory monitoring and task timesharing. *Human Factors*, 46(3), 497-517.

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.