RICE UNIVERSITY

Implications of differences of echoic and iconic memory for the design of multimodal displays

by

Daniel Shields Glaser

A THESIS SUBMITTED

IN PARTIAL FULFILLMENT OF THE

REQUIREMENTS FOR THE DEGREE

Doctor of Philosophy

APPROVED, THESIS COMMITTEE:

David Lane, Associate Professor

Psychology, Statistics, and Management

Michael Byrne, Associate Professor

Psychology and Computer Science

Philip Kortum, Professor in the Practice, Faculty Fellow

Psychology

H. Albert Napier, Professor Management and Psychology

HOUSTON, TEXAS

FEBRUARY, 2012

ABSTRACT

Implications of differences of echoic and iconic memory for the design of multimodal displays

by

Daniel Shields Glaser

It has been well documented that dual-task performance is more accurate when each task is based on a different sensory modality. It is also well documented that the memory for each sense has unequal durations, particularly visual (iconic) and auditory (echoic) sensory memory. In this dissertation I address whether differences in sensory memory (e.g. iconic vs. echoic) duration have implications for the design of a multimodal display. Since echoic memory persists for seconds in contrast to iconic memory which persists only for milliseconds, one of my hypotheses was that in a visual-auditory dual task condition, performance will be better if the visual task is completed before the auditory task than vice versa.

In Experiment 1 I investigated whether the ability to recall multi-modal stimuli is affected by recall order, with each mode being responded to separately. In Experiment 2, I investigated the effects of stimulus order and recall order on the ability to recall information from a multi-modal presentation. In Experiment 3 I investigated the effect of presentation order using a more realistic task. In Experiment 4 I investigated whether manipulating the presentation order of stimuli of different modalities improves humans' ability to combine the information from the two modalities in order to make decision based on pre-learned rules.

As hypothesized, accuracy was greater when visual stimuli were responded to first and auditory stimuli second. Also as hypothesized, performance was improved by not presenting both sequences at the same time, limiting the perceptual load. Contrary to my expectations, overall performance was better when a visual sequence was presented before the audio sequence. Though presenting a visual sequence prior to an auditory sequence lengthens the visual retention interval, it also provides time for visual information to be recoded to a more robust form without disruption. Experiment 4 demonstrated that decision making requiring the integration of visual and auditory information is enhanced by reducing workload and promoting a strategic use of echoic memory. A framework for predicting Experiment 1-4 results is proposed and evaluated.

Acknowledgements

I would like to acknowledge all of those who assisted in the completion of my dissertation. I give thanks to my advisor Dr. David Lane who spent many hours providing thoughtful feedback on my research and writing. I would also like to give thanks to Drs. Michael Byrne and Phil Kortum for their time an advice offered as my committee members. My wife Yi was quite generous putting up with me and supporting me over the last couple years. She continually and generously agreed to offer her thoughts during the development and writing of this research.

Table of Contents

List of Tables

Introduction

It is common for humans to monitor and respond to multiple sources of information (e.g., driving while conversing on the phone). However, it has been repeatedly demonstrated that doing so often results in error. A critical body of human performance research has demonstrated that, in many cases, monitoring and responding to multiple sources of information often results in less error than if the information sources are presented in different modalities (e.g., visual & auditory). It has also been well demonstrated that the persistence of sensory memories vary in length. The aim of this research is to explore whether the differences in visual and auditory sensory memory have implications on how to present and respond to visual and auditory information when presented in close temporal proximity. Knowledge gained from this research is thought to be potentially useful in the design of auxiliary multimodal displays. In particular, cases when the operator is using the display to assist with primary task by using it to receive concise and discrete updates.

Iconic vs. Echoic Memory

It is well documented that auditory stimuli are retained in sensory memory longer than are visual stimuli. In a pioneering study, Sperling (1960) estimated that the number of items available from visual memory diminishes rapidly throughout the course of a second. The estimated number of digits available was 10 (of 16) items when participants were allowed to respond immediately to a 4 x 4 array of stimuli. The estimated number of digits available were 7 or 4 items with a 400ms or 1000ms delay, respectively. Sperling (1960) believed that recall in the no-delay condition was superior to the delay conditions because with no delay or very short delay participants were able to "read" directly from iconic memory. In longer-delay conditions an iconic representation of the event was not available to assist recall.

Additional evidence of the brief persistence of iconic memory comes from Haber and Standing (1969) who measured the duration of visual sensory (iconic) memory by presenting a circle intermittently and asking participants to report whether they perceived a continuous stimulus or whether they perceived the stimulus to completely disappear before reappearing. These authors were interested in how long the circle could be absent and still be perceived as present, a direct measure of stimulus persistence. They found that this value lies between 250 and 300ms. In an analogous experiment Darwin, Turvey, and Crowder (1972) presented three simultaneous three-digit lists, one to the right ear, another to the left ear, and one to both the left and right ear (center). With no recall delay, the estimated number of digits available was 4.9, with a 4000ms recall delay the estimated number of digits available was 4.4. The relative drop in estimated-items-available indicate clearly that auditory stimuli are far more persistent than visual stimuli, with iconic memory lasting milliseconds and echoic memory lasting seconds.

Conrad and Hull (1968) compared the duration of iconic and echoic memory in a serial recall paradigm; one group recalled read words and another group recalled spoken words. The authors found that recall of the final three items for the read group was only 25% as accurate as recall of the final three items from the read aloud group. This difference in recency was coined the modality effect.

Though memory research on sensory memory have generally shown that auditory information is available longer than visual information, there is ample evidence that echoic memory is comprised of two separate memory stores working in sequence. The first store is a short auditory store which may be thought of as a literal auditory store that decays within a fraction of a second following a stimulus. Two types of evidence for this store derive from persistence-of-auditory-sensation paradigm studies that have investigated how stimulus integration can only occur during short intervals consistent with short auditory storage. Auditory persistence evidence comes from a series of studies performed by Efron (1970a) who examined the extent that a persistence of a tone differed from the actual length of the tone duration. Efron (1970a) noted that when tones were shorter than 180ms the perception of the tone length was the same as when the tone was actually 180 ms. Efron (1970a) concluded that 180 ms is an estimate of the minimal duration of the short audio store. In a similar investigation Plomp (1964) found that two tones can be separated by a 200ms soundless gap but still be perceived as a continuous tone. This result suggests that the amount of time that the sound is sustained perceptually in pure physical form has an upper limit of approximately 200ms.

The short auditory store is due to persistent memory traces in the primary echoic cortex. A method for assessing the state of the memory trace is to assess mismatch negativity (MMN) which is a component of the event-related brain potential elicited by infrequent, physically deviant stimuli in a sequence of auditory stimuli. Researchers posit that the MMN reflects a neural process that is activated when incoming stimuli do not correspond to a well-formed echoic memory trace. The amplitude of the MMN is thought to be a proxy for the strength of the trace. Methods assessing echoic memory using MMN have found that MMN signals are present 400ms after the original tone suggesting that a neural signature of the tone remains present after the physical tone has ceased and are a likely explanation of why audio tones are perceived to persist after the offset of the stimulus.

The second component to echoic memory, and the component of greatest interest in this research, has been termed the Long Auditory Store. The primary phenomenon that led researchers to hypothesize that such a store exists is the modality effect. Though the sensory component of auditory memory lasts for only a few hundred milliseconds, why is it that words at the end of a list are remembered more if they are spoken then if they are seen? Which aspects of the auditory stimulus are being sustained? Crowder and Morton (1969) suggested that auditory information held is pre-categorical such that the information in the store remains unprocessed for meaning and is much like an echo of the raw auditory stimulus. The most common method for testing this hypothesis is the suffix paradigm where a not-to-be recalled auditory stimulus (or suffix) is presented after the final list item. The degree to which the suffix attenuates the modality effect is thought to provide information on the nature of the Long Auditory Store. In a series of 17 suffix experiments Morton, Crowder, and Prussin (1971) reported that the disruptive effect of the suffix depended upon the acoustic similarity of the suffix, such as pitch and voice quality, but not semantic similarity. This result is consistent with the pre-categorical acoustic store hypothesis. Inconsistent with the pre-categorical acoustic store hypothesis, however, was the finding that manipulating the description of an ambiguous "wa" sound as either a human voice or a musical instrument could modulate the modality effect. When the participant was told that the "wa" sound was a human voice the modality effect was attenuated. In contrast, when the participant was told that the "wa" sound was an instrument the modality effect was not affected. If the Long Auditory Store was purely acoustic, the perception of the suffix should not affect the degree to which it interfered with information in the auditory store. Another finding suggesting that the auditory store is not a literal store is that suffixes can be introduced via silently articulated words indicating that auditory sensory memory may be confounded with a kind of articulatory memory (Narine & Crowder, 1982).

Though the exact nature of the Long Auditory Store is not entirely clear, there is ample evidence that without the presence of an interfering suffix auditorily presented words remain far more available than visual items many seconds after the offset of the stimulus. Two result patterns should be considered when considering how to most effectively leverage echoic memory when present audio and visual information together. First, although the Long Auditory Store often holds information for many seconds, it is prone to masking from latter stimuli. Hence, when presenting auditory information, the time interval should not be so long such that the latter sounds replaces (pushes out) interferes with the beginning sounds. This kind of interference is common with middle-list items in modality effect paradigms where the recall benefit of auditorily presented words are observed for the latter list items only. Second, this body of research has demonstrated that presenting language in both the auditory and visual domain will likely diminish the effectiveness of echoic memory since both semantic content from the visual list and the act of articulating the visual list words to keep them active will interfere with the long echoic store. Since this research is focused on how to leverage echoic memory, this investigation will restrict the auditory and visual information to non-verbal stimuli.

Sonification

The form of auditory communication used in these experiments is commonly referred to as sonification. Sonification is the transformation of data relations into perceived relations in an acoustic signal for the purposes of facilitating communication. Research efforts have elucidated the type of information that can and cannot be communicated well via sonification.

Sonification Success Stories - Data Exploration: Since sonification is the mapping of data to non-verbal sound components, it has been examined as a tool for exploring data sets. Perhaps the most common sonification method for data exploration is the formation of auditory graphs which use changes in auditory frequencies over time to communicate data values. Nees and Walker (2007) argue that advantages of visual graphs, such as the emergence of patterns, can be preserved in auditory representations similar to how individual data points combine to form cohesive patterns in visual graphs.

Efforts have been made to examine auditory versions of traditional display formats such as scatterplots, box-whisker plots, and histograms (Bonebright et al., 2001; Peres & Lane, 2003; Peres & Lane, 2005). These efforts have commonly relied on variation of the pitch-time display format described before. These displays suggest common visual graphical displays can be perceived in an analogous auditory fashion. Though evidence suggests that sonified scatter plots, box-whisker plots, and histograms may not often lead to more accurate perception for educated sophomores with normal vision, there may be advantages to these sonified graphs for visually impaired students and researchers, and young students who have been shown to respond to immersive and multimodal educational experiences (Upson, 2001).

Auditory perception is particularly sensitive to changes over time, providing humans with a marked ability to discriminate between periodic and aperiodic events and minor changes in the frequency of a continuous signal (Kramer, 1994; Flowers & Hauer, 1995; Kramer et al., 1999). Hence, sonification has proven to be useful in making realizations about complex data sets that are elapsed over a time interval. During the Voyager 2 space mission the spacecraft experienced malfunction when traversing the rings of Saturn. Mission operators were unable to identify the malfunction cause using visual displays which seemed to output merely noise. When the same data was sonified through the use of a music synthesizer, a machine gun sound elucidated that the problem was electromagnetically charged micrometeoroids (Kramer, 1994). In a related scientific scenario, Kramer et al. (1999) notes:

> After months of unsuccessful study of visual oscilloscope traces for evidence of an oscillation predicted by quantum theory, Davis and Packard decided to listen to their

> experiment instead. What they heard was a faint whistling-the first evidence that these oscillations actually do occur. (p. 13)

These cases illustrate the ability of the auditory system to extract underlying structure and temporal aspects of complex signals that are often important in scientific exploration and discovery. Sonification remains a prominent exploratory data method for scientific agencies like NASA and the Design Rhythmics Sonification Research Lab that have used sonification to help discover and communicate ecological research about ice cores.

Sonification Success Stories - Assistance for the Visually Impaired: Those poised to gain the most from sonification may be the visually impaired. Within this community there is a need to receive non-visual navigational information of the surrounding environment. A specific technology poised to address these questions is vOICe (v Oh I see e), a vision-replacement technology for the blind which offers image-to-sound renderings from live camera views. The developers of this system claim that such renderings could lead to synthetic vision through cross modal sensory integration by exploiting the existing multisensory processing and neural plasticity of the human brain through training. The vOICe system is comprised of a head mounted camera, a mobile computing device loaded with vOICe software, and headphones. The vOICe software transposes the digital image in the following three ways.

1. *Left/Right Visual /Left/Right Audio*: An image is scanned from left to right. Objects in the right visual field will be heard most fully in the right audio channel, objects in the left visual field will be heard mostly in the left visual channel.

2. *Lightness /Loudness*: The lighter the image the louder sound

3. *Height /Frequency*: The higher the image in the visual field, the higher the frequency of the sound

Merabet et al. (2009) demonstrated that trained blind users of the vOICe system can identify objects (e.g. a toy horse and a sail boat) with 65% accuracy providing evidence that sonification can lead to object recognition. When repetitive transcranial magnetic stimulation (rTMS) was delivered to specific areas in the occipital cortex, these same blind users were significantly impaired in their ability to identify the same kinds of objects (30%). In an fMRI study it was found that the lateral-occipital tactile-visual area (LOtv) a brain region activated when objects are recognized by vision or touch is also activated by trained sighted and blind participants when hearing specific soundscapes and objects. Activation was not present when these participants were exposed to natural sounds associated by objects such as mooing (Amedi et al., 2007).

Sonification Success Stories - Medical Practitioner Assistance: Sonification/audio alerts play an important role in the medical field. The Pulse-oximeter, became standard equipment in medical operating theaters in the United States during the mid 1980s. The Pulse-oximeter produces a tone that varies in pitch in accordance to the level of oxygen in a patient's blood, allowing the doctor to monitor this critically important information while visually concentrating on surgical procedures. The idea was extended to a six-parameter medical workstation by Fitch and Kramer, (1994). Medical students working with this workstation in a simulated operating room scenario were able to identify emergency situations more quickly with the audio display than with a visual display or a combined audio + visual display (Fitch & Kramer, 1994). Similar results have been demonstrated with other kinds of medical monitoring. These findings have clear real-world implications because it shows that medical professionals are more responsive to medical emergencies when using sonified alerts, which has a clear impact on their ability to respond quickly in threatening situations. Such results show that sonification has the potential to be sufficiently effective as a monitoring tool that adding a visual element such as a heads up display may hurt monitoring performance by introducing distracting redundancy.

Sonification Design Considerations

Though there is evidence that sonification may be useful both alone and in conjunction with other senses (not just vision), it is important to identify broad issues that a designer must consider when incorporating sonification and other non-speech audio elements to a user interface. To date specific sonification design guidelines grounded in literature are generally not available. However, work by de Campo, Frauenberger, Stockman, and Bourget's (2007) audio design survey, together offer some insight into the broad critical issues, that with careful consideration, will help ensure that an interface will be provide an enhance experience for many populations of users.

Delectability and Discriminability: First and foremost an audio interface (or any sonified element) has no value unless the user can detect the sound. Designers should be aware that the highly controlled environment(s) where a device is tested will likely share few resemblances to many locations that the device is actually used. It is recommended that designers be aware of minimum thresholds for sound detection along the relevant auditory dimensions. Hartmann (1997) has provided a useful review of these dimensions and their thresholds. Assuming that the user can hear the auditory elements in his chosen environment, it is equally as important that any individual sound and its meaning can be discriminated from other sounds and their meanings; if two sounds carry different pieces of information, they must be distinguished to ensure the successful communication of the message. As with detection, much work has been dedicated to understanding how well listeners can discriminate sounds using its many dimensions including loudness, pitch, duration, and tempo. For information on determining how to use dimensions to promote highly discriminate auditory elements, it is recommended that the reader review the work such as tempo (Boltz, 1998) and duration (Jeon & Fricke, 1997).

Annoyance: Auditory elements have the potential to annoy users. Even in cases when a sound helps performance, if the sound is annoying, there is a possibility that the sound will be silenced. Ramloll et al. (2001) among others have gathered evidence suggesting that musical non-speech sounds with a rich harmonic and acoustic features are easier and thus more pleasant to perceive than pure tones and simple waveforms (which may have been an issue in my study). Sounds may be piloted with a representative user group as to eliminate particular sounds that are considered displeasing or annoying. Annoying sounds, however, may be appropriate if tied to a critical event.

Mapping: As mentioned, the tenant of sonification is the transformation of data relations into perceived relations in an acoustic signal. Mapping is the grouping of sound dimension to data. These sound dimensions include

- Location: The origin of a sound
- Loudness: The magnitude of a sound
- Pitch: The highness or lowness of the frequency
- Register: The relative location of a pitch in a given range of pitches

- Timbre: A the quality or characteristic of a sound
- Duration: The length of time a sound is heard
- Rate of Change: The varying of the duration of a sound over time
- Order: The sequence of sounds over time
- Attack/Decay: The time it takes for a sound to reach its minimum or maximum Considerable work has been done to identify the sound dimensions which are particularly effective mapping agents. For example, Flowers (2005) have found that pitch generally is effective at representing quantity. In contract, however, loudness often is not effective for both practical and perceptual reasons (Walker & Kramer, 2004). Additional efforts of determining appropriate acoustic dimensions for a given data type comes from Walker (2002, 2007) who examined three data dimensions (e.g., temperature, pressure, and danger) along with three acoustic dimensions (e.g., pitch, tempo, and brightness). It was found that pitch matches to changes in temperature while tempo does not.

Peres and Lane (2003, 2005) tested whether using redundant acoustic dimensions to represent data leads to more accurate judgments. Results from these investigations showed that redundancy often may hurt performance or leave it unaffected. Peres and Lane noted (2003, 2005), however, that redundancy gains can be expected in the case of symmetrically integral dimensions (e.g., pitch and loudness). Specifically, dimension groupings such that changes in one dimension will effect perceptions of the other. In the case of pitch and loudness, the higher the pitch the more loud the pitch will be perceived. Dimensions that are not integral will likely not have a redundancy benefit (Peres & Lane, 2005). Though redundancy did not help performance in many cases, it did lead to greater task satisfaction. Future research should employ similar research methods discussed to expand upon this knowledge base.

Individual Differences: Individual differences should be considered because of the potential for the same sounds to be interpreted with different meanings. Potentially relevant individual difference variables include cognitive abilities, musical abilities, listening skills, perceptual abilities, and learning styles. Walker and Mauney (2004) attempted to identify how individual differences in cognitive abilities may affect auditory magnitude estimations. Some evidence was found that cognitive abilities affect the interpretation of auditory displays. Specifically, participants with a better working memory and nonverbal reasoning measures performed better on the magnitude estimation tasks than those participants with lower working memory and non-verbal reasoning test scores (Walker and Mauney, 2004). It has long been predicted that having musical training would endow a listener with a superior performance with auditory displays in comparison to participants with no musical training. Though some researchers have found this trend to be true (e.g., Neuhoff , Kramer, & Wayand, 2002; Lacherez, Seah, & Sanderson, 2007) many found this trend to be weak to non-existent (Bonebright et al., 2001; Walker, 2002). A potential explanation to this non-benefit proposed is that one's ability to successfully interface with an auditory display comes from a more fundamental perceptual acuity which would be expected to be distributed similarly across both musicians and non-musicians. Beyond absolute performance, audio display designers should be aware that different populations have different intuitions regarding the directional relationship between the data and acoustic dimension. For instance Walker and Lane (2001) identified that sighted individuals intuited a positive polarity when mapping frequencies to the conceptual dimension dollar amount whereas visually impaired subjects intuited a negative polarity. Designers of audio interfaces should be aware of their users and the potential of such conflicts when creating a display.

Sonification Limitations

Though using the auditory modality along with the visual modality may help to increase the amount of information that can be communicated, the kind of information that auditory information can communicate is restricted. Perhaps the biggest constraint is that auditory information is not persistent and may need to be repeated. Another constraint is that the task of decomposing rich non-verbal messages may be difficult for some operators.

Multiple Resources: A History and Critique

Why is perceiving and accurately responding to multiple sources of information more difficult than a single source? Further, why is there often a performance benefit when presenting information using different modalities? An early explanation of why monitoring multiple sources of information typically leads to errors was developed by Broadbent (1958) who proposed that humans are similar to communication channels in that attentional processes occur within a single channel of limited capacity (or bandwidth). Since Broadbent (1958) believed humans must process all information within a single limited channel, there are instances in which information must be filtered to keep the system from becoming overloaded. Broadbent asserted that errors occur when critical information is filtered and prohibited from being processed.

Broadbent's filter model of selective attention (1958) was inspired by his famous dichotic listening experiment in which participants were presented with different 3-digit numbers simultaneously in the left and right ears. Broadbent noticed that participants were most successful in recalling information from a single ear, suggesting that participants can only attend to one "channel" at a time. Broadbent claimed that information from the originally unattended ear could be revisited by accessing a short-term memory store that holds perceptual information in a pre-filtered state.

Triesman (1964) argued against the idea of a filter by which information is completely blocked and, as an alternative, suggested that non-attended information is "attenuated." Triesman (1964) claimed that attenuated information typically is generally not selected for attention unless it is somehow semantically related to the subject of the human's primary attention or of constant particular relevance to the human like his name or a distress call. Both Broadbent and Triesman's models of attention can be criticized, however, for failing to explain why some perceptual-response pairings lead to greater levels of performance than others. For instance, pointing in response to a spatial task is faster than speaking, and conversely, speaking in response to a verbal task is faster than pointing (Brooks, 1968). Moray (1967) asserted that humans have more in common with a limited capacity processor than a fixed, limited capacity channel. Mental operations on perceptual inputs use capacity and are available to different processors when divided attention is required. It should be noted that both Triesman and Broadbent were early selection theorists. Early selection theories posit that the only processed components of the unattended to stimuli perceived in a dichotic listening task are physical properties such as the gender of the speaker or language being spoken. Late selection theories, claim that both semantic and physical properties of information are evaluated. Consistent with late-selection theory, Mackay (1973) found that semantic information from the "un-attended" channel would act to disambiguate meaning in the attended channel. For example when river bank was used in the unattended message, participants were more likely to interpret ambiguous "bank" as river bank as opposed to a place to keep one's money.

Kahneman (1973) contributed to the resource metaphor with the notion that there is a single pool of limited capacity available for a variety of tasks, and that performance is positively correlated with the degree to which capacity is allocated to the task. Kahneman's (1973) theory was augmented by Norman and Bobrow (1975), who also used the analogy of "resources" to human performance and attention. They distinguished resource-limited from data-limited processes. Resource-limited processes improve as more resources are allocated. Data-limited processes, in contrast, are not improved by an increase in resource allocation because there is a lack of data to process and a processing ceiling has already been met. The same task may have data-limited and resource-limited regions. In data-limited regions the allocation of additional resources does not lead to an improvement in performance.

In the single-resource view, resources are treated as a central homogenous body. Apt analogies noted by Navon and Gopher (1979) include a common currency, energy in a physical system, and general intelligence. According to this view, when all available resources are required to support performance on a primary task, performing a secondary task will recruit primary task resources and necessarily diminish primary task performance.

An underlying assumption of a single-resource theory is that if two tasks use attentional capacity, performing a third task should interfere with the two tasks comparably. Moreover, comparable interference should occur regardless of the nature of the tasks. The credibility of single-resource theory was called into question by a series of contradictory findings. The following are examples of scenarios where such did occur.

Brooks (1968) demonstrated that vocal responses interfere more with the recall of a sentence then a spatial diagram. Baddeley, Grant, Wight, and Thomson (1975) followed up on

Brooks' result by demonstrating that performance in a pursuit rotor task was more negatively affected by the recall of a line diagram then in a verbal recall task. These results are evidence that tasks that use overlapping systems for processing and responses are more likely to interfere with each other than when multiple systems are used. In a similar vein Mowbray (1964) showed that presenting a to-be-remembered word auditorily impairs the human's ability to attend to a message being played in the alternate ear more than a visual a presentation of a word. And perhaps of most direct relevance to this paper is the finding by Triesman and Davies (1973) that monitoring tasks of a single modality interfered with each other more than monitoring tasks of multiple modalities.

Navon and Gopher (1979) revised the resource framework as a response to these findings. This revised framework accounted for the finding that introducing a second task will often negatively affect primary task performance in addition to findings that cannot be explained by single-resource theory (Figure 1). Specifically, the observation that a third task (task C) may differentially affect tasks A and B. Navon and Gopher (1979) argued that the human is a multiple capacity processor (multiple structures). Each processor may have its own capacities and each capacity may be shared by several processes. Hence, the impact of changes in task difficulty on resources may not be qualitative (i.e., structural) and quantitative. Thus, if Task A and B require resources from two separate resources, and a third task is introduced that requires the same resources as Task A, then task A's performance would degrade whereas task B performance would remain relatively unchanged.

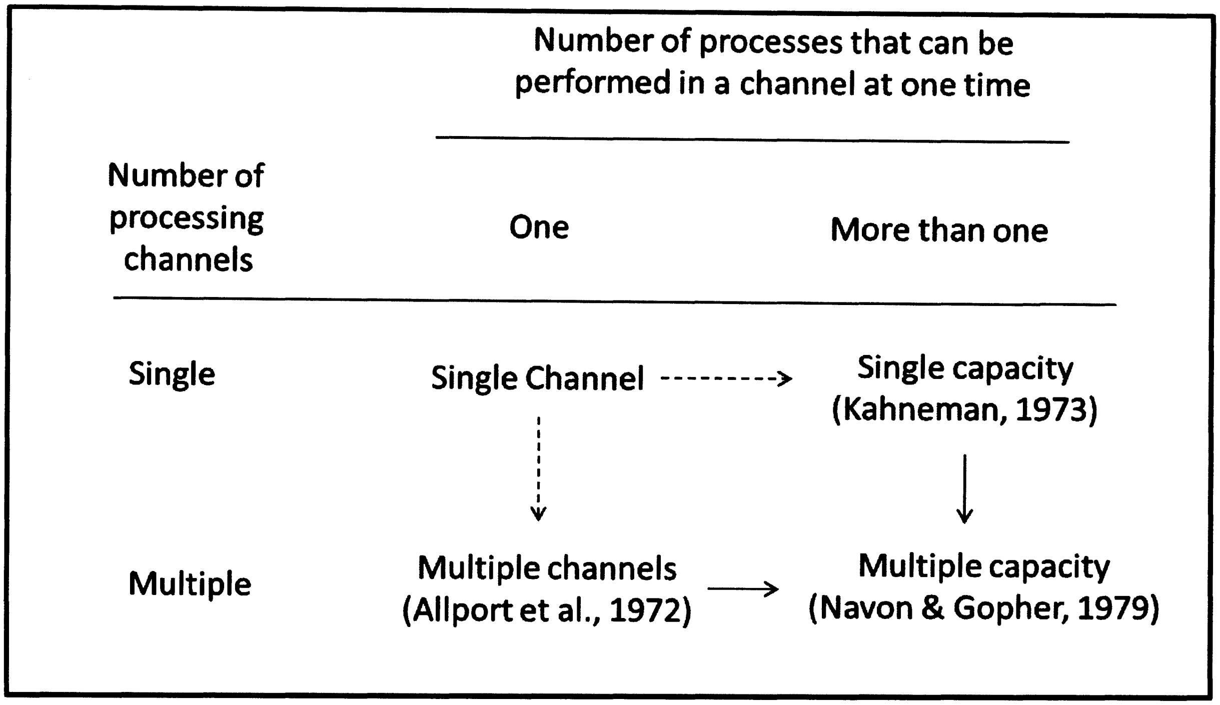

Figure 1. A adoption from Navon & Gopher (1979). Dashed arrows are orthogonal expansions of the single channel hypothesis. Solid arrows merge the once-thought opposed expansions.

Wickens (1984) introduced a particularly influential addition to the resource theory in which he applied three dichotomies: stages of processing (perceptual versus task selection), processing codes (spatial versus verbal), and modalities (auditory versus visual). Wickens (1984) posited that the more two tasks use different levels along these three dichotomies, the easier it will be to perform the task simultaneously.

A clear example of the separate pools notion can be highlighted by imagining the relative ease of a task that does not use different pools at different stages of processing as compared to those that do. For instance, imagine trying to simultaneously listen to an audio book and a friend's story about her day. Through experience we know that making sense of both auditory sources is difficult. Multiple resource theory would predict that this scenario would be difficult to process because the only information code used is verbal and the only modality used is auditory. This would likely overtax both the information code resource and perceptual resource. In contrast, imagine a scenario in which an operator is driving and listening to an audiobook. Multiple resource theory would predict that this would be a relatively easy task since there is no overlap in information code or modality.

Wickens (2002) noted, however, that the often-observed advantage of cross-modal (auditory-visual) time-sharing over intra-modal time sharing (auditory-auditory and visual-visual) may more be the result of structural interference. In the case of a visual-visual task, if two visual information streams are far enough apart in physical space there will be the additional cost of visual scanning; if the two channels are too close there may be physical masking. Auditory-auditory tasks are susceptible to masking as well. Wickens (2002) noted instances where no cross-modal benefit was found when visual scanning was controlled, suggesting that a cross-modal benefit may be purely due to physical constraints. The question of whether the benefit of using multiple modalities is accounted for more by resources of physical factors remains largely unresolved. Researchers have noted, however, that instances in which a modality is physically interfered with are intrusive enough such that a benefit will likely come from "off-loading" a portion of an intra-modal visual or auditory task to the auditory or visual modality respectively (Rollins & Hendricks, 1980; Seagull et al., 2001).

Wickens Computational Model

In order to increase the applicability of Multiple Resource Theory to human factors design, Wickens (2002) devised a computational model designed to predict the level of performance of two of more time-shared tasks. Wickens (2002) argues that the model is useful as a way of predicting task interference in multi-task environment such as that of a vehicle driver, busied secretary, or a commander in an emergency.

Modeling task performance

An important aspect of the model is that it calculates the degree to which two tasks may interfere with each other. This calculation is largely dependent upon the degree to which two tasks share resources. The basic idea is that periods in which multiple tasks are performed at the same time are more likely to be high workload periods. Wickens (2002) defines workload as the time required to complete all tasks divided by the time available to complete all tasks. When this ratio exceeds 1.0 within a specified time interval, then "overload" has occurred and according to the model, one or more of the tasks will suffer. Conclusions from this kind of analysis may help a task designer decide when/if a task should be rescheduled or altered. The model of multiple tasks may be constructed as follows: (1) a task is represented on both a quantitative and qualitative level. On a qualitative level the modeler identifies which resources are required for a task. On a quantitative level the modeler identifies how much of each resource is being used. (2) The difficulty multiplier for each resource is context dependent. Consider a driving task as an example. As visual conditions become cloudier, the visual difficulty multiplier will increase. Similarly as the roads become more slippery the manual difficulty multiplier will increase. (3) Performance loss is increased in cases where both tasks demands are high and both tasks draw from a common resource. Drawing from a common resource occurs when the same level within a dichotomous resource dimension is utilized.

These operations are completed computationally by generating a value based on the task difficulty (i.e., demand component) and then a value based on how much the tasks will interfere with each other based on multiple resource theory (i.e., conflict component). The sum of these values is the total interference expected. The interference value is designed to be a ratio of task performance if the tasks were to be performed alone to task performance if the tasks were to be performed together. Examples of performance are reaction time and response accuracy. Thus, an interference value of 2 means that a task will take twice as long and/or inaccuracy will be twice as high as if the tasks were performed separately.

The demand and conflict components are calculated as follows: The demand component is calculated by first generating a resource vector where the modeler identifies which level of each dimension is used. The task is then given an ordinal value equal to or greater than one. For an example consider a situation where one is monitoring a crowded radar of planes moving through a restricted area and responses are made with a button press while at the same time the operator is conversing with pilots with a text messaging application. The vector for the radar task might be perceptual-visual-spatial: 2, cognitive-spatial: 1, and response-spatial: 1. The perceptual-visual-focal component received a multiplier of 2 rather than 1 because the crowded nature of the radar is thought to be especially taxing to the visual-spatial resource. The conversation task has a resource vector of perceptual-visual-verbal: 1, cognitive-verbal: 1, and response-spatial:1. The demand component is the average of the vector values. In this example the demand component for the radar task is 2+1+1/4 = 4/3 = 1.33. The demand component for the conversation task is (1+1+1)/3 = 1. The combined demand component score is 2.33.

The conflict component is calculated by generating a conflict matrix specific to the tasks. This matrix helps to identify the degree to which resources will be used conflictingly between the two tasks (i.e., perceptual-visual). Wickens (2002) argues that in some instances a resource cannot be shared. The response-verbal resource is an example in which only a single utterance can be executed at a time. In these instances the conflict value is 1. In contrast, if a resource can be perfectly shared as with response-verbal and perceptual-visual-spatial, the conflict value is minimal. Table 1 is a conflict matrix proposed by Wickens (2002). Each value in the matrix calculated in the following way a conflict matrix in which each level of the modality, code, and stage (and focal/ambient if necessary) is used in both Task A and task B. The numbers within each cell of the matrix are derived as follows.

Every channel pair will have an executive processor conflict of 0.2. Additional conflict of 0.2 is added to this value for each instance when a dimension resource overlaps. For example, if Task A uses a spatial perceptual resource and Task B uses a spatial response resource the conflict values for the two resources will be .2 (general) +.2 (spatial) = .4. In contrast, if the response code was verbal, the conflict value would be just .2. Wickens (2002) notes that since cognitive resources do not have a modality distinction, the average of 0 and .2 (e.g., .1) is added to the conflict value. Moreover Wickens (2002) notes that the only resource that cannot be time shared to any degree is a verbal response which results in a conflict value of 1. Using that radar (Task A) and conversation task (Task B) as an example, a conflict matrix as shown below would be yielded (Table 1).

Table 1. Example MRT conflict matrix. The sum of this conflict matrix is .4+.3+.2+.3+.5+.2+.4+.5+.6 = 3.4.

	Perceptual-Visual-Spatial	Cognitive-Spatial	Response-Spatial
Perceptual-Visual-Verbal	.4	.3	.2
Cognitive-Verbal	.3	.5	.2
Response-Spatial	.4	.5	.6

The total interference score is the demand score (2.33) + 3.4 = 5.73

The model can be utilized in a casual/intuitive and formal manner. Casually the model can serve to guide designers in making decisions such as when it is better to use voice rather than manual control, when it is better to use auditory rather than visual displays, and to use spatial rather than verbal information aids. Wickens (2002) notes that when employing multiple resource theory to guide such dichotomous categorical design decisions, certain modalities should be considered as better equipped to deliver specific kinds of information. A notable example is the difficulty in presenting audio-spatial information as a geographical space.

Multiple efforts have been made to test the interference model. Horrey and Wickens (2006) evaluated whether Wickens' interference model could account for driving performance variations when coupled with a secondary number memorizing task. Three primary simulated driving tasks were evaluated: City Driving, Rural Straight Driving, and Rural Curved Driving. Each of these tasks was assigned a unique demand vector. For example city driving was given a demand vector of: Visual-Focal: 2; Visual-Ambient: 1; Cognitive-Spatial: 2; Response-Spatial:1

= 2 + 1 + 2 + 1 = 6 demand scalar, whereas rural straight driving was assigned a demand vector of: Visual-Focal: 1; Visual-Ambient: 2; Cognitive-Spatial: 1; Response-Spatial: 2 = 1 + 1 + 1+1 = 4 demand scalar. The secondary task was to memorize a number string which varied in length from trial to trial. To increase the external validity of this experiment, numbers were presented using common technologies. The three technologies used were the following: (1) a head up display positioned 7 degrees below the horizon line; (2) a head own display 38 degrees below the horizon and positioned near the mid-console; (3) and auditorily using the car speakers. As with the driving tasks, each of the number delivery methods was assigned a demand vector. The head down display was assigned a vector of: Perceptual-Visual-Focal: 2; Perceptual-Audio-Verbal: 0; Cognitive-Verbal: 2; Response-Verbal: 1 = 2 + 0 + 2 + 1 = 5 demand scalar. In contrast the Speaker method was assigned a demand scalar of Perceptual-Visual-Focal: 0; Perceptual-Audio-Verbal: 2; Cognitive-Verbal: 2; Response-Verbal: 2 = 2 + 0 + 2 + 2 = 6 demand scalar. Conflict components were calculated using the method described earlier in this paper. Three performance criteria were measured and predicted: Lane Keeping, In-Vehicle Response Time, and Hazard Response Time. Overall RT was predicted with high accuracy with 85% of IVT RT accounted for and 98% of the Hazard RT accounted for. In contrast only 2% of lane keeping performance was accounted for. The authors noted that the low variance accounted for suggests that drivers are adept at allocating resources needed for controlling the vehicle. Performance costs, rather, were manifested in IVT tasks and hazard detection. Since the consequences of poor lane keeping are much greater than longer RTs IVT tasks, there should be a mechanism for the modeler to account for a probable resource allocation strategy. The inability for the MRT model to account for strategic allocation of resources is a significant limitation of the current version of MRT.

General Motors, Ford, and others (Angell et al. 2006) assessed the Multiple Resource Theory Model's accuracy in predicting how quickly a driver can complete secondary in-vehicle technology tasks while driving in a simulator, on a test track, and on a road. Secondary tasks included visual-manual tasks, tasks that involved visual and auditory perception, and auditory verbal tasks. Example tasks from the visual manual tasks include such changing CD tracks, entering destinations in a navigation unit, and dialing of a telephone number. Examples from the Auditory-Visual tasks include a voice dialing task and ordering a plane ticket. Examples of Auditory-Verbal tasks include answering biographical questions and recalling details from a sports broadcast. Metrics examined were secondary task time and driver performance such as the standard deviation of lane position. The model was able to predict between .52 and .84 of the RT variance of visual manual tasks from the simulator and track conditions. The Multiple Resource Theory was overall not successful in predicting Auditory-Verbal tasks however. Furthermore, the authors noted that the Multiple Resource Theory models did not discriminate well between high and low workload tasks. Angell et al. (2006) noted that Horrey et al. (2006) had greater success predicting driver performance by engaging in considerable model fitting. Overall the authors believed that the MRT model data was not sufficiently predictive to replace driving data. The authors found that modeling efforts were more successful when a more detailed modeling method was employed. The authors noted that detailing tasks by individual steps provided flexibility in differentiating between shorter and longer more complex tasks.

In a similar study Sarno and Wickens (1995) had pilots perform a simulated flight task concurrently with a verbal cognitive task. The authors addressed the value added in terms of predicting performance by including and excluding various aspects of multiple resource theory. The notable conclusions were that using multiple rather than a single resource improved the accuracy of prediction. A variety of multiple resource models accounted for 60 to 70% of the performance variance. Including a timeline component to the model did not significantly alter the variance that was accounted for by the model. A similar result was found when including a demand scalar.

Wickens (2002) notes three challenges of the MRT model. First is that greater parsimony between the "resource" and "multiple" aspect in which there is a greater balance between variance accounted for and demand coding. Second, there needs to be a better understanding of circumstances in which an operator becomes so engrossed in a single task that an alternative task is ignored. In these cases any potential benefit of multiple resources are eliminated. Wickens notes that in some instances adding a secondary task that theoretically should have minimal disruption to a primary task may draw all attention away from the primary task whereas a secondary task that theoretically should be more interfering with the primary task. A specific example of this comes from Wickens and Liu (1988) in which it was found that auditory messages during a flight task caused greater disruption than a visual message. Third, is the challenge of identifying the degree to which a saccade or head movement when orienting between multiple visual channels should be reflected in the conflict matrix. Overall, MRT theory has proven useful in being able to predict relative performance between different task combinations. In the research reviewed this usefulness has applied most consistently to relatively unpracticed laboratory visual manual tasks. The model does not have an inherent ability to account for the potential benefits of expertise and strategic allocation of resources between tasks. In addition the model does not account for performance in data-limited areas of resource-performance curves where changes in resource interference will have little to any effect on performance. For the purpose of this investigation the model is limited in that it does not account for the potential benefits of the Long Auditory Store.

Unified Theory of Cognition

The most comprehensive approaches to predicting human performance are unified theories of cognition. These theories consist of models equipped to account for many of the behavioral phenomena for which MRT is not. Examples include performance consequences of strategically allocating attentional resources and the beneficial performance benefits of task expertise and practice. Though most models include an echoic buffer module intended to account for properties of the Long Auditory Store, there is no precedence for modeling how an echoic buffer may be strategically used to improve performance in tasks that use a visual and auditory display and have a big memory component. The following is a history of UTC.

Allen Newell (1973) first outlined a major complaint of early cognitive research in a chapter entitled "You Can't Play 20 Questions with Nature and Win." Newell believed that the current state of experimental psychology time was overly focused on isolated and narrow empirical research that studied simple binary hypotheses (e.g., early vs. late attentional selection, serial vs. parallel memory search, and imaginal vs. propositional knowledge). Newell (1973) estimated that research to date uncovered approximately 3000 behavioral regularities but how the pieces fit together remained a puzzle.

To solve the puzzle a unified theory of cognition was advocated, or more specifically, a unified system of mechanisms that operate in concert to produce the full range of human cognition. Newell (1973) specified that a unified theory of cognition should have a detailed information-processing architecture with well-defined interconnected components that implement elementary symbolic computational processes for perception, cognition, and motor behavior. A unified theory of cognition is successful if it is able to simulate covert mental processes and overt behavior associated with learning, memory, perceptual skills, language comprehension, decisions making, problem solving, and other complex functions. Meyers & Kieras (1999) list pioneering efforts of a unified theory of cognition, their strengths and shortcomings.

Model Human Processor

One of the first tests of the Model Human Processor was to predict the speed and accuracy at which people could perform a text editing task. To predict performance the model human processor was endowed with memory stores as well as a perceptual, cognitive, and motor processor. Qualities of these memory codes and processing units were consistent with past and current research findings. Using a flowchart to integrate processing units and memory stores, the author noted that he was able to derive acceptable predictions from and engineering perspective of speed and accuracy (Card et al, 1983, as cited in Meyers & Kieras, 1999). The successes of this model was seen as supporting evidence that available data was sufficient to make significant strides towards a unified theory of cognition. Successes notwithstanding, the model human processor had notable limitations including the absence of motor and sensory processing. As such, the Model Human Processor was not as flexible or as powerful as a Unified Theory of Cognition is expected to be.

Handling of Auditory Sensory Memory: For the purposes of this investigation it is important that a modeling method contains a sensory module that acts similarly to the Long Auditory Store. MHP uses a sensory buffer that is constantly passing information to short-term memory and being overwritten. Continual access to auditory memory is dependent upon refreshing auditory chunk information in working memory. If performance in multi-modal tasks may at times be improved by leveraging the Long Auditory Store, the MHP is not equipped to explain such advantages.

ACT-R and Executive Process Interactive Control (EPIC)

ACT-R is more comprehensive system than Card's Model Human Processor. Anderson (1983) contributed to the human processor metaphor by making a distinction between knowledge types. Procedural knowledge in ACT-R acted as a production system where actions are produced and tasks are performed via if-then rules. ACT's rule interpreter deals with conflict-resolution by comparing production rules with the state of working memory. Subsequent rules and actions are connected to the outcomes of these comparisons.

The sophistication of ACT allowed for accurate predictions in accuracy data from a broad array of cognitive tasks including comprehension and reasoning tasks. In addition ACT defines/explains algorithms for compiling and tuning procedural knowledge and cognitive-skill acquisition phenomena which has been able to explain phenomena in the area of perception, attention, learning, memory, problem solving, and language processing. These described qualities endow ACT with more inherent potential than MHP to enable computer simulations and to become a true UTC. In more recent versions of ACT, treatment of ocular, manual, and articulatory motor control has been included. And further, a framework of addressing complex problem solving of the sort addressed previously by Newell and Simon's (1972) General Problem Solver (GPS).

Handling of Auditory Sensory Memory: ACT-R has an auditory buffer called the audicon which holds auditory information for a time that can be set by the modeler. Information in the audicon is deleted after the set amount of time. Such a buffer makes ACT-R a capable tool for generating a model of multimodal tasks in which strategic use of the Long Auditory Store is expected to impact performance.

EPIC is similar to ACT-R in that both theories utilize multiple modules to correspond to separate perceptual and motor modalities and to "central cognition." These modules, including a perceptual, motor, and cognitive processor are considered encapsulated and independent. Where these two models differ most fundamentally is the production system within the cognitive processor. Toward one end of the spectrum, ACT-R posits that only a single production rule can fire at a time. However, ACT-R enables parallelism in other fashions specifically through "asynchronous parallelism among its perceptual and motor modules; parallel retrieval of information from declarative memory; and parallel production-rule matching and selection" (Holyoak & Morrison, 2005, p. 405).

EPIC differs in that it is the only system with a fully parallel production-rule firing. EPIC uses task-coordination strategies that impose ordering constraints when necessary to manage multiple threads of central cognition. Two reasons Meyer and Kieras (1999) notes that a fully parallel production rule system is superior to a single firing rule system is that (1) a maintenance of architectural simplicity is achieved. Meyer and Kieras (1999) notes that a serial production rule unnecessarily limits the architectural structure and that a serial production rule lacks "neuropsychological plausibility" considering the brain's distributed and parallel infrastructure. A fundamental goal for EPIC was to be able to explain Psychological Refractory Period (PRP) procedure data. A PRP procedure involves presenting two stimuli in close temporal proximity where each stimulus is responded to separately. PRP has gained attention because it is thought to mimic real-world performance challenges. The PRP procedure by measuring RTs to the two stimuli is thought to be an effective way to investigate how humans perform perceptual, cognitive, and motor tasks concurrently. Meyer and Kieras (1999) explain that a fully parallel production system is more capable of explaining PRP phenomena.

It has been found that within-modality PRP tasks generally yields larger costs than between-modalities PRP tasks (Pashler, 1998) which is consistent with MRT. EPIC which assumes no limitations in the simultaneous execution of central operations (Meyer & Kieras, 1997) includes mechanisms that can modulate dual-task costs. EPIC, however, does not specify why these mechanisms are engaged differently by the different modality pairings.

Meyer and Kieras (1999), unlike the authors of contemporary unified theory of cognition models makes a pointed argument that limitations on multiple-task performance provide especially informative clues in the process of developing a unified theory of cognition. Meyer and Kieras (1997) asserts that a UTC with a fully parallel production rule system promotes a greater understanding of goal related behavior, executive processes, and strategy which has the potential to improve the flow of multifaceted tasks such as line operations and interacting with a cockpit.

Handling of Auditory Sensory Memory: EPIC does not have a module akin to a Long Auditory Store. Auditory information flows through an auditory processor into auditory working memory where a rehearsal process is needed to keep the information active.

MHP, EPIC, and ACT-R can account for benefits of using multiple modalities in part by the utilization of modality specific buffers and modules that store and process information for different sensory modalities. Since buffers and modules are finite in their ability store and process incoming information, there may be a benefit to distributing storage and processing demands between different modalities' buffers, and modules.

Experiments 1-4: Tasks investigated

The tasks investigated in Experiments 1 through 4 were specifically designed to determine if changes in presentation and response order to multiple stimuli from both the visual and auditory modality may lead to substantial changes in response quality. Participants in Experiments 1 and 2 were given a recall test. In Experiment 3 participants were given a recognition test. In Experiment 4 participants were asked to make one of three decisions based on memory of the sequences in conjunction with a pre-learned rule set. Visual and Auditory stimuli in Experiments 1 & 2 were analogous sequences that varied in a number of dimensions including where the presentation occurred (e.g., pitch). In the visual domain this manipulation was executed by altering where information was presented on the computer screen. In the auditory domain this manipulation was executed by altering the pitch range in which the sequence was presented. In Experiments 3 & 4 auditory sequences were similar to Experiments 1 & 2 while visual stimuli were changed to an array of static images. In all experiments stimuli were presented within a 4000 millisecond window, thus this investigation is limited to displays of brief and discrete information. In all experiments response quality was contingent upon retaining information from both stimuli. These investigations are designed to exaggerate the cognitive challenges operators may face in tasks such as a surgeon monitoring information from visual and auditory monitors.

What is the best method for generating a parsimonious account of Experiments 1-4 data? One option would be to use one of the aforementioned modeling techniques. MRT accounts for many of the perceptual and working memory interference costs inherent in dual-task scenarios, however, MRT does not account for any differences in the duration of echoic and iconic memory. Neither EPIC nor MHP have a module analogous to the Long Auditory Store. Since the utilization of such a store is central to the central thesis of the following investigations, these techniques are not considered an appropriate framework. ACT-R does include an echoic buffer similar to the Long Auditory Store, however, the buffer does not act as a gradient. The information in the auditory buffer is either present or not present when an availability gradient is more plausible. Further, modeling using ACT-R and generating a virtual test environment is beyond the scope of this dissertation. For the purposes of this investigation Experiments 1-4 are considered exploratory. However, data from Experiments 1 and 2 data assist in the development of a framework for making predictions regarding tasks in which reliance on the Long Auditory Store is expected. Experiments 3 and 4, which are believed to involve similar cognitive processing as the tasks in Experiment 1 and 2, will be used to challenge the prediction framework. This framework is described and evaluated after Experiment 4.

The remainder of this dissertation is organized into five sections. The first section is Experiment 1 which examines the fundamental thesis that manipulating the use of the Long Auditory Store may impact recollection of the event. Results from Experiment 1 were consistent with my hypothesis that, overall, responding to an auditory stimulus after responding to a visual stimulus will affect overall accuracy of both responses. Section 2 describes Experiment 2 which was designed to determine whether manipulating both the presentation and response order leads to better overall retention than manipulating response order only. Results from Experiment 2 demonstrated retention benefits of presenting visual and auditory information in sequence rather than together. It was also found that presenting visual information before audio information led to better retention then presenting auditory information before visual information. This result was surprising because in the presentation phase, unlike the response phase, lengthening the time between the visual presentation and visual response lead to better retention. An explanation of this affect is proposed and additional phenomena are considered in a proposed prediction framework discussed in section 6. Section 3 describes Experiment 3 in which visual-auditory task of Experiments 1 and 2 were modified to more closely resemble a real-world scenario. Additionally, the investigation extended to how presentation order affects a participant's natural spontaneous response order. Results from Experiment 3 demonstrated that presenting audio information before visual information leads participants to most often recall visual information before audio information. Results from Experiment 3 demonstrated that, like Experiments 1 and 2, a visual then auditory recall order leads to the better overall retention. Section 4 describes Experiment 4 which is designed to assess how the presentation order of a visual-auditory display affects decisions based on rule knowledge. Results from Experiment 4 are considered to have particular real-world relevance because responses to displays are typically decisions rather than responses to a memory test. Experiment 4 demonstrated that presenting visual information before audio information leads to the best decision making. In section 5 a framework for predicting performance in tasks similar to those explored in Experiments 1-4 is proposed. The proposed framework leverages observations from Experiment 1 and 2 and seeks to provide a parsimonious account of the results from each of the four experiments. To conclude I recapitulate findings from Experiment 1 through 4 and expand on their implications.

Experiment 1

Experiment 1 is designed to assess the idea that overall performance on a visual-auditory memory task may be affected by how heavily the Long Auditory Store is leveraged. To address this hypothesis Experiment 1 utilizes a scenario in which participants are presented with non-verbal visual and auditory information simultaneously. Both the visual and auditory information contained five analogous dimensions that were to be recalled when the information was prompted. The task is organized such that information from one modality must be retained while responding to information from the competing modality. Thus, one modality will have a stimulus to response interval that is longer than the other. As discussed in the introduction, auditory sensory information persists longer than visual information. Hence, a crude prediction could be that overall performance will be best if auditory information is responded to after the visual information because auditory information is more likely to be successfully re-accessed through the visual response interval than vice-versa. Figure 2 illustrates the assumed interaction between modality and stimulus response interval for recall errors.

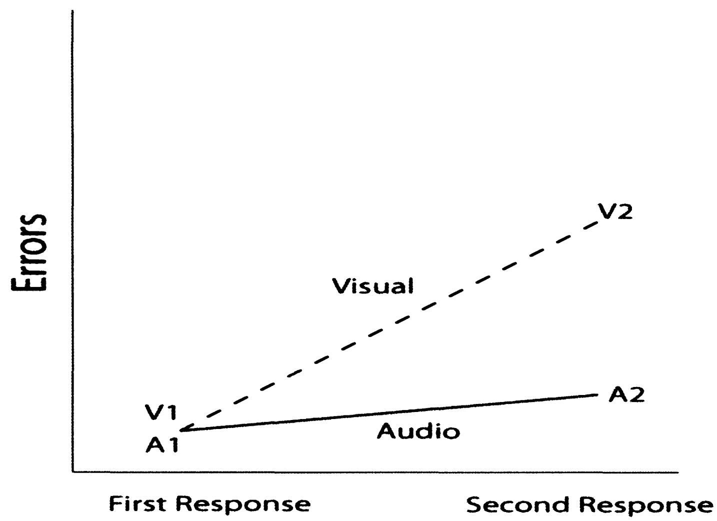

Figure 2. A representation of the assumed interaction between modality and stimulus response interval (SRI) (first response = short SRI; second response = long SRI).

Method

Participants

Forty-eight undergraduates (26 females) participated as research subjects. The ages ranged from 17-22 with an average of 19.3. Each participant was awarded credit towards a psychology class's requirement.

Design

On each trial a tone sequence, a dot sequence, or a simultaneous tone sequence and dot sequence was presented. Trials during which a tone or dot sequence was presented separately were considered baseline trials. The simultaneous trials were part of a 2 (modality: visual and audio) x 2 (response order: visual first, visual second) within-subjects factorial design. The tone (auditory modality) and dot (visual modality) sequences varied on five dimensions: Orientation, Range, Direction, Numerosity, and Sequence item complexity. These 5 dimensions were used for both modalities in an effort to keep the amount of information within each sequence constant.

Stimuli

One hundred and eight sequences were developed for each modality. Details on the stimuli follow.

(1) Orientation (Left, Center, or Right): Tone sequences were sounded from either the left speaker, both speakers, or the right speaker. Analogously, dot sequences were shown on the left, center, or right side of the screen.

(2) Range (Low, Medium, or High): Tone sequences were sounded in either a high, medium, or low range. Dot sequences were shown in the top third, middle, or bottom third of the screen.

(3) Direction (Up, Down, or the Same): Tone sequences went up, down, or stayed the same in pitch. Dot sequences either went up, down, or maintained its vertical position.

(4) Numerosity (2 or 3): Tone and Dot sequences were composed of two or three items.

(5) Sequence item complexity (Simple or Complex): Tone sequences were either a simple sinusoidal tones or the same sinusoidal tone along sounded with a percussive beat. Dot sequences were either a black circle or a black circle with a white inner circle. The complex tone and dot are considered complex because they include elements of the simple counterpart (i.e., tone: same frequency but with an overlaid drumbeat; dot: same black circle but with a white circle overlaid.) but are made more complex by adding a component.

For auditory sequences, tones would start at 60, 240, or 960 Hz. If the sequence had ascending tones, the pitch's hertz followed the following formula $\textit{starting pitch} * 2^{(\textit{sequence order of tone1})/\textit{number of tones in sequence}}$. This formula was chosen so that the sequence would stay within a single octave and not overlap in pitch with other pitch ranges. If the sequence was flat, all items were the same frequency as the first item. If the sequence was descending then the following items in the sequence was $\textit{starting pitch} / 2^{(\textit{sequence order of tone1})/\textit{number of tones in sequence}}$. If there were three tones in a series, than the sequences frequency range would span an octave. If only two tones were presented, then the tones' frequencies would span less than an octave. Simple tones consisted of simple sine wave forms. Complex tones were created by adding a snare drum wave form to the simple wave form. The snare drum wave form was taken from the Rock Kit in GarageBand '09 ©.

Dots were circles with a 25 pixel radius. Simple dots were completely black. Complex dots were the same black circle with an added white circle with a 15 pixel radius centered in the middle of the black dot. Visual sequences were presented within a 3 x 3 matrix. Each cell was

250 pixels wide and 250 pixels tall (Figure 3). Within each cell there were five rows. Each row was 50 pixels tall. Dot sequences were restricted to a single cell and always started in the middle row (Figure 4). If the sequence went up, the dots would appear sequentially higher in the cell matrix. If sequences went down dots would appear sequentially lower in the cell matrix (Figure 5). Visual sequences were presented on a 15 inch eMac using the Firefox browser. Auditory sequences were displayed with headphones. On simultaneous trials tone and visual sequences were generated independently.

a.  simple dot b.  complex dot

Figure 3: Panel a. is an image of a simple dot; Panel b. is an image of a complex dot

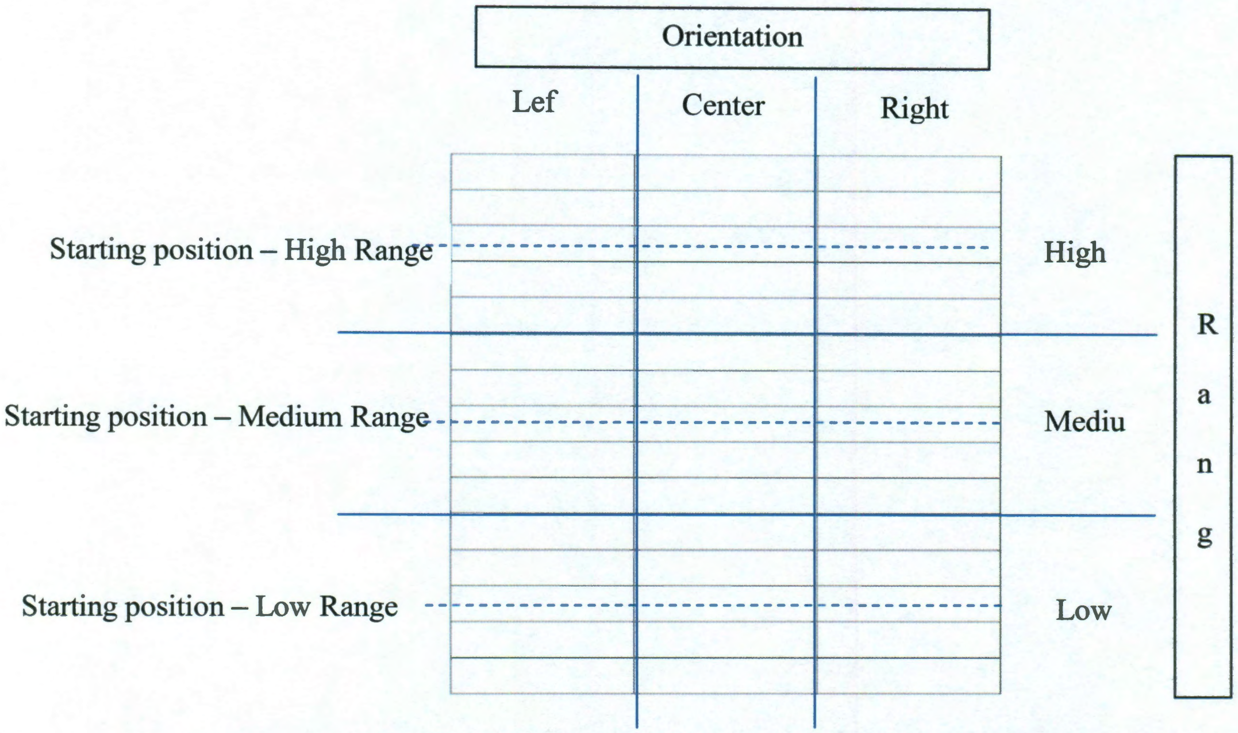

Figure 4. The matrix structure in which dots were presented. Any one sequence would be restrained to one of the 9 primary cells (e.g., Orientation = Center, Range = High). All sequences would start on the middle row of a primary cell (shown with the dashed lines). From the starting position sequences would stay in the same place, move up 1 or 2 rows, or move down 1 or 2 rows.

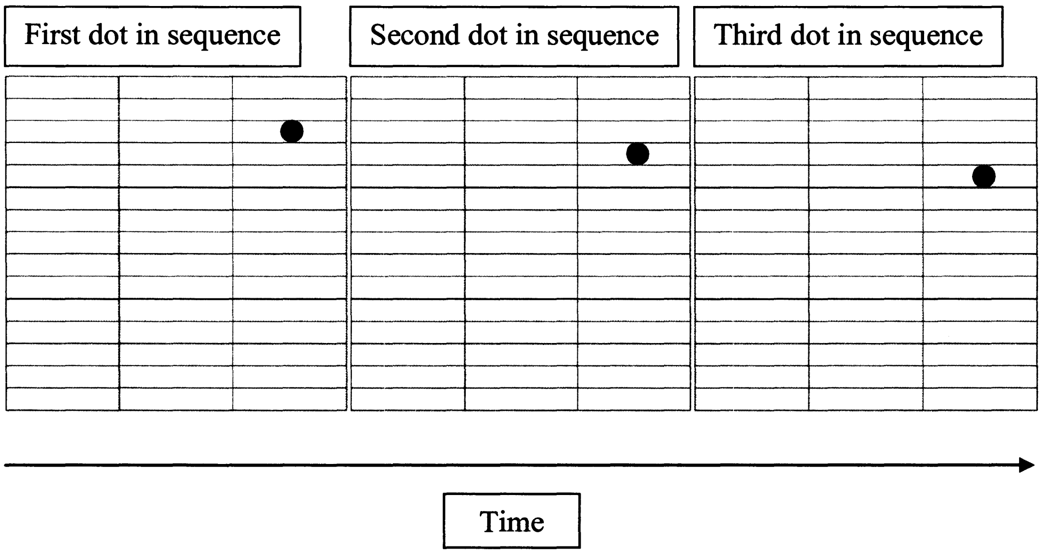

Figure 5. A visual sequence with complexity: simple, numerosity: three, direction: down, range: high, orientation: right.

Procedure

Participants were told that they would be completing a memory experiment in which they will be recalling aspects of visual and auditory sequences. Participants were then informed that each sequence involved five dimensions including pitch, direction, orientation, numerosity, and complexity. Participants were told that for the audio sequences pitch refers to the pitch range; orientation refers to the direction that the sound originates; direction refers to the pitch direction of the individual tones in the sequence; numerosity refers to the number of tones in a sequence; and, complexity refers to the sound of the tone.

Participants were informed that for the visual sequence the dot's pitch is in reference to how high the dot is on the display matrix; in contrast, the dot's orientation refers to where the dot appears on the screen horizontally; direction refers to the vertical direction of the dot sequence; numerosity refers to the number of dots in a sequence; and, complexity refers to the visual quality of dot. After the verbal explanation of the two kinds of sequences, the participant heard examples of tone sequences and was asked to identify the five dimensions. Participants were corrected after making a mistake. After participants were able to complete 5 sequences correctly, the participant was considered to have sufficient mastery of audio sequence decoding. Next, the participant was shown examples of the visual sequences and asked to recall the five visual dimensions. Once a participant was able to recall 5 sequences accurately, the participant was thought to have sufficient mastery over visual sequence decoding. Directly after, the participant began experimental trials.

Participants started experimental trials by pressing “begin” 250 ms after either a visual, auditory, or visual-auditory sequence was displayed. Immediately after the sequence, participants were prompted to answer questions about the dimensions with radio button selections. In half the simultaneous trials, participants were prompted to answer the visual sequence first. In the remaining half of the simultaneous trials, participants answered the auditory sequence first. After answering all dimensions, participants were prompted to start the next trial with the appearance of the “Start” button.

Results and Discussion

As can be seen in Figure 6, the prediction that the difference in errors from the first response to the second response would be greater for the visual responses than for the auditory responses was confirmed. Error rate here was the proportion of trials a participant did not answer all five dimensions were not answered correctly. A significant Modality x Response-order interaction was found for error rate, $F(1,47) = 19.5$, $p < .01$.

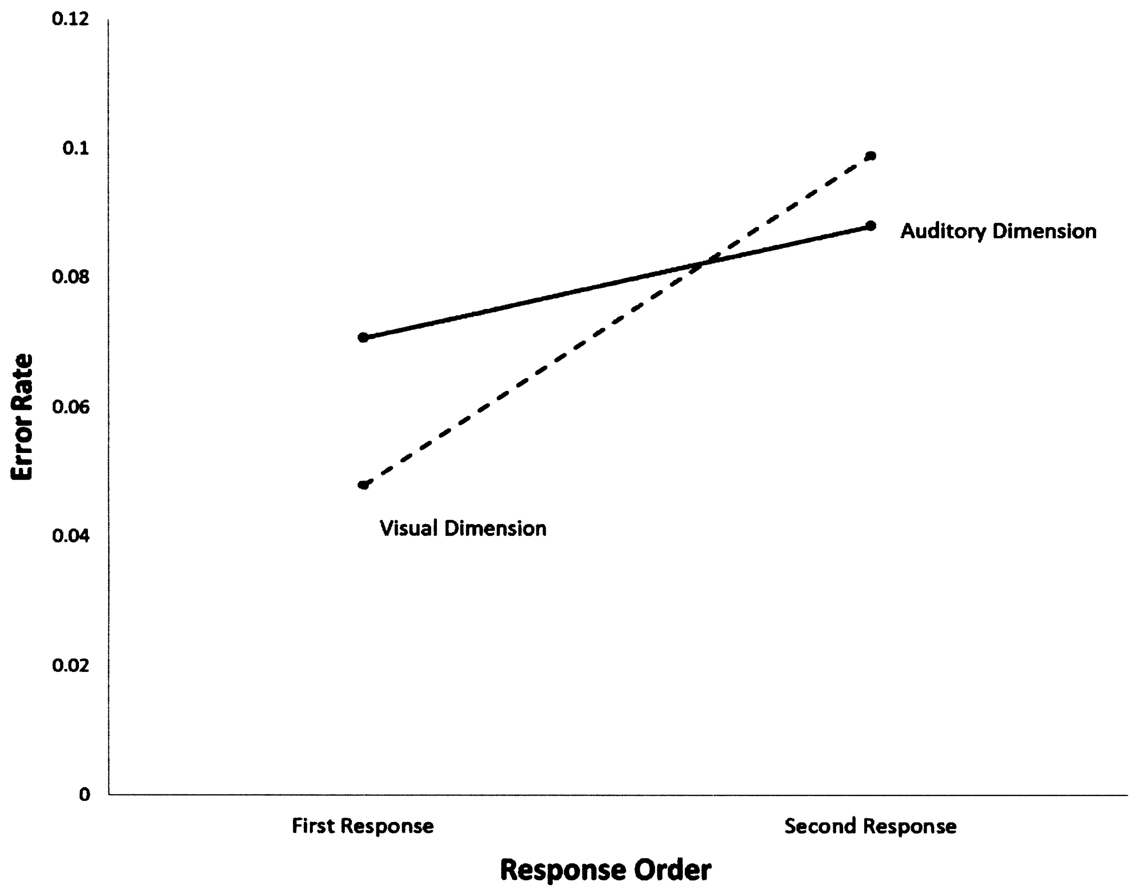

Figure 6. The error rate for visual and auditory sequences as a function of response order.

The probability of recalling all 5 sequence dimensions of an audio sequence without error was less affected by response order than visual sequences indicating a significant Modality x Response-order interaction $F(1,47) = 24.32, p < .01$ (Figure 7). Specifically, as the delay between presentation and response lengthened, dot-sequence errors increased more than for tone-sequence errors.

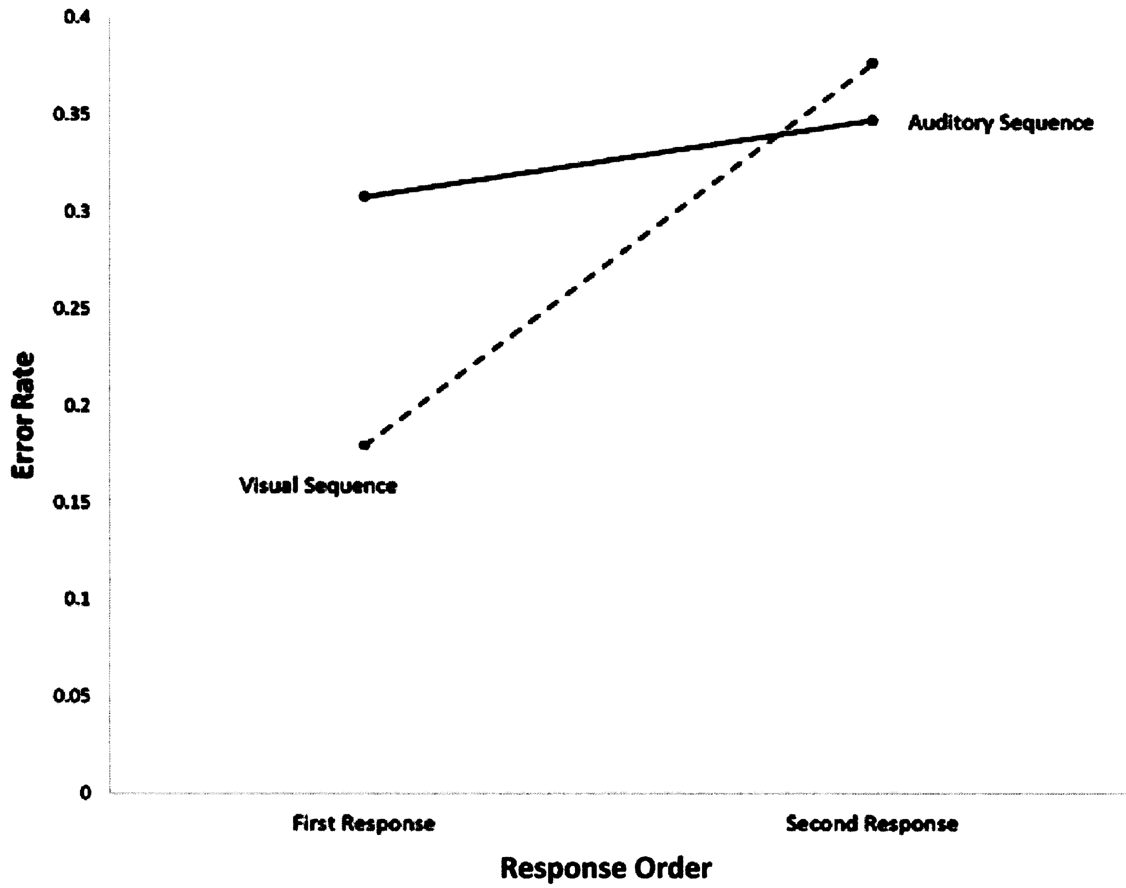

Figure 7. The probability of not recalling all five dimensions from visual and auditory sequences across the first and second response condition

For each dimension the error rate increment between the first and second response was greater for visual items than for auditory items. This result is consistent with the common finding that echoic memory is longer lasting than iconic memory. The numerosity advantage, however, was very slight with the visual error rate increasing only.0008 more than the auditory error rate between the first and second response. Figures 8-12 present the performance for each of the 5 dimensions separately for each response type.

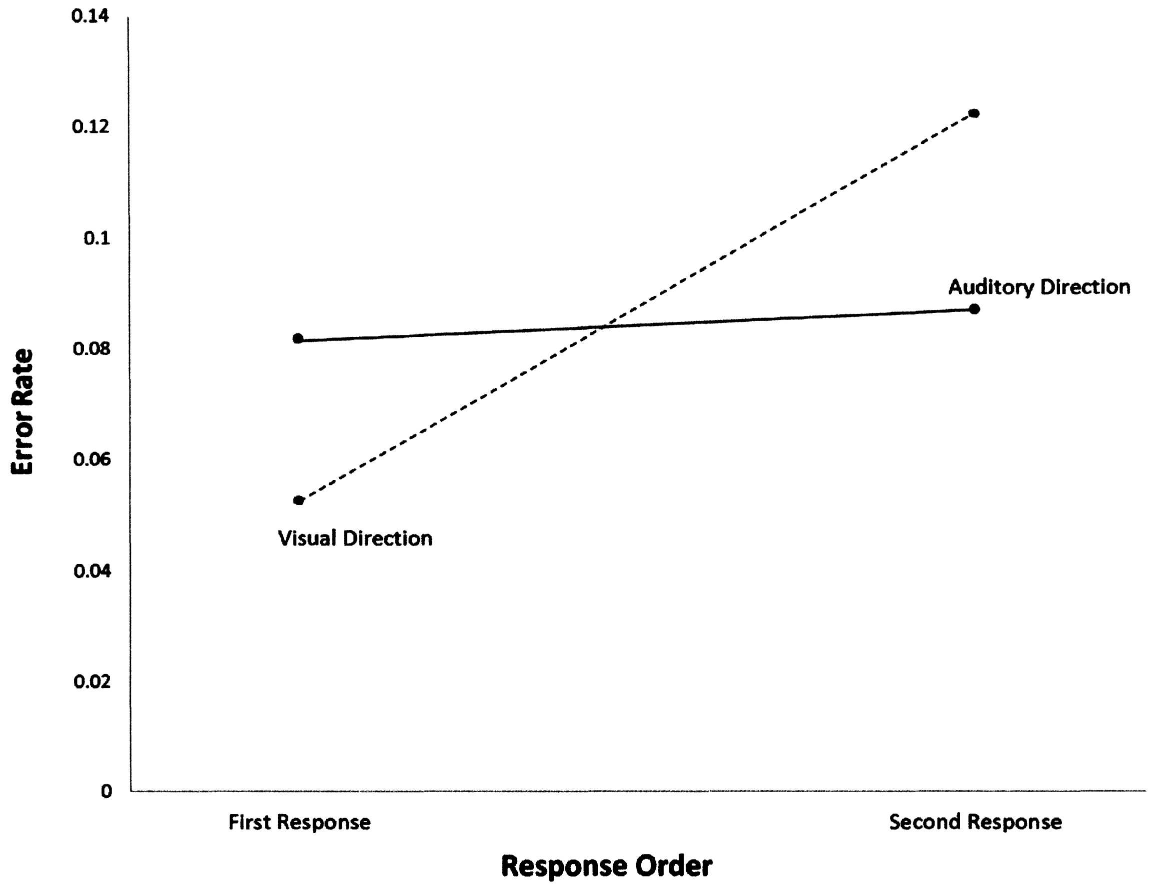

Figure 8. The mean error rate for direction as a function of dimension (visual and auditory) and response order (first and second).

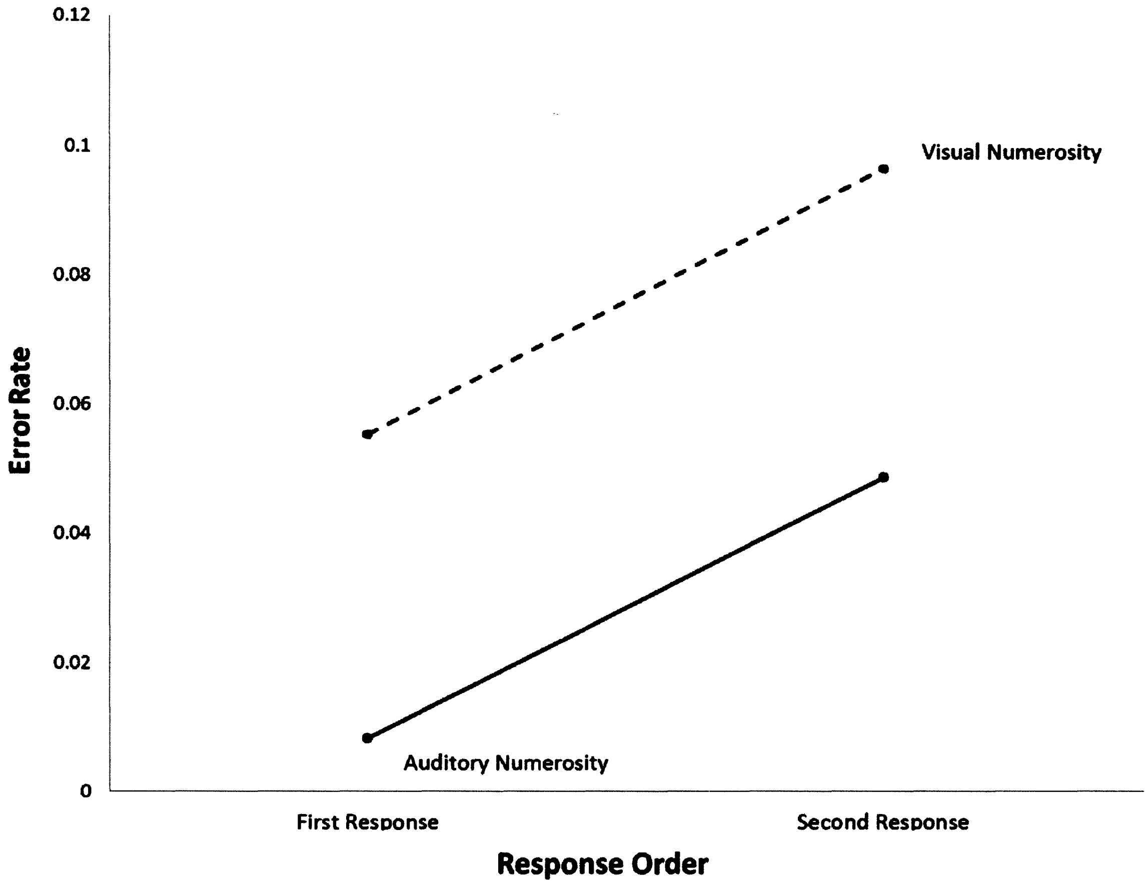

Figure 9. The mean error rate for direction as a function of numerosity (visual and auditory) and response order (first and second).

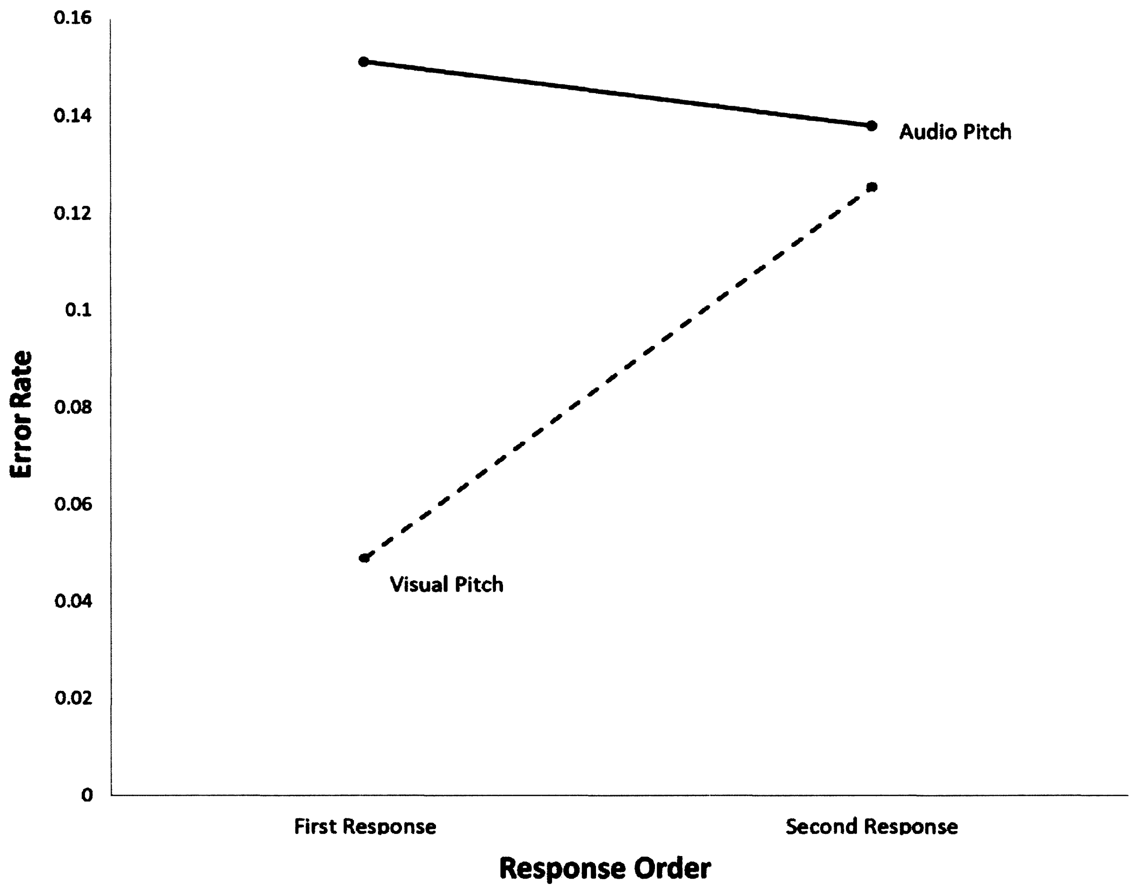

Figure 10. The mean error rate for direction as a function of pitch (visual and auditory) and response order (first and second).

Figure 11. The mean error rate for direction as a function of orientation (visual and auditory) and response order (first and second).

Figure 12. The mean error rate for direction as a function of complexity (visual and auditory) and response order (first and second).

Taken together, the relatively flat auditory error curve compared to the visual error curve is consistent with the hypothesis that auditory information is more persistent. Though the auditory curve was relatively flat compared to the visual, auditory errors were more prevalent in the second response condition than in the first suggesting that auditory decoding is more difficult than visual decoding.

Of greatest practical interest in this experiment is how overall accuracy is affected by response order. Overall accuracy was the proportion of trials within a condition that all ten dimensions across the two modalities were recalled correctly. For the following analysis each participant provided one data point (e.g., proportion) for each condition. As expected and shown on Figure 13, overall accuracy was greater when dots were responded to first (M = .53) than when tones were responded to first (M = .44). This difference was significant, (M_{diff} = .09, CI_{95} = [.05, .13]) $t(47) = 4.3, p < .01$.

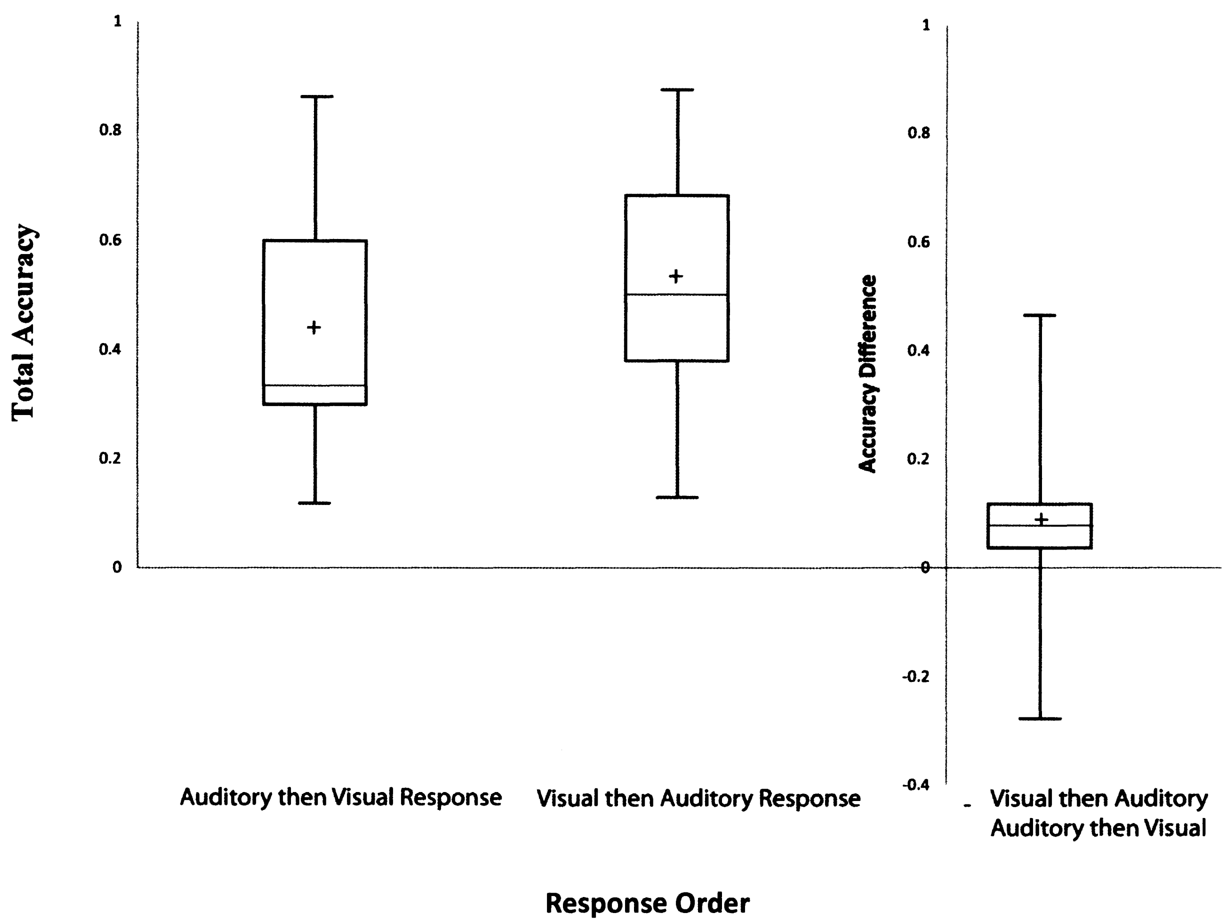

Figure 13. The proportion of time participants answered all 10 dimensions of a simultaneous sequence correctly. The right-most box plot represents the within subject difference accuracy score distribution between the first response: visual and first response: auditory condition

The mean proportion of trials that participants answered all 5 dot dimensions correctly (M = .77) was greater than the mean proportion of time all five tones dimensions were recalled (M = .71). This difference is significant, (M_{diff} = .07, CI_{95} = [.04,.10]) $t(47) = 4.04$ $p < .001$. This pattern was similar when sequences were presented alone (Dots Baseline: .88, Tones Baseline: .76; CI_{95} = [.08, .16]) $t(47) = 5.82$, $p < .001$ and in the first response condition when sequences were presented together (Dots: .82; Tones: 69; CI_{95} = [.09, .17]) $t(47) = 5.84$, $p < .001$. The difference in accuracy in the second response condition between dots (M = .62) and tones (M = .65) was not significant (M_{diff} = .02 CI_{95} = [.04, .02]) $t(47) = 1.19$, p = .25.

Baseline and First Response Accuracy

As can be seen in Figure 14, baseline performance was better than first response performance for both the visual and auditory modality. This difference was significant for both the visual, (M_{diff} = .08, CI_{95} = [.05, .11]) $t(47) = 4.86$, $p < .001$ and auditory modality (M_{diff} = .06, CI_{95} = [.03, .09]) $t(47) = 4.54$, $p < .001$. This difference occurred even though the retention interval between the presentation and response was the same for baseline and the first response.

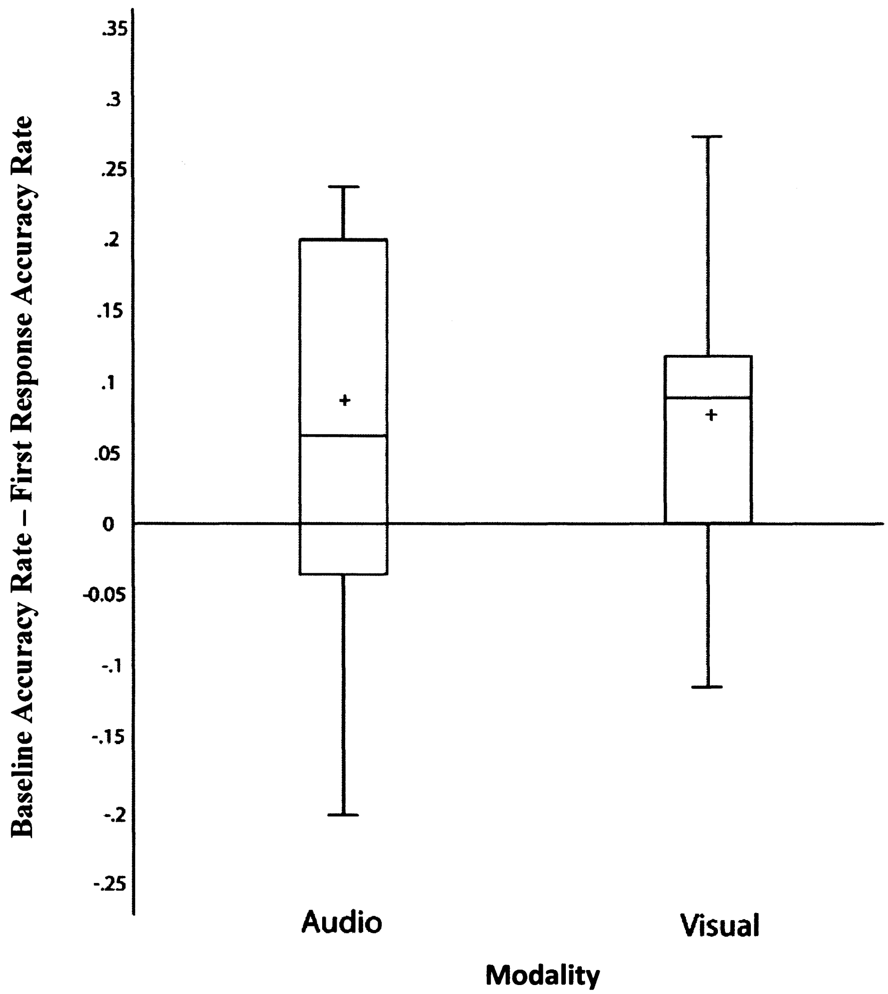

Figure 14. The mean differences between audio baseline and audio first response and the mean difference between visual baseline and visual first response.

Visual and Auditory sequence baseline performance

Even though auditory accuracy worsened at a slower rate than visual accuracy, in the first response condition participants were better at identifying visual dimensions than audio dimension in baseline trials. To a large extent this difficulty may be explained by the sequence dimensions that were chosen. Visual stimuli could be chosen that were more difficult to discriminate making the visual baseline lower. Similarly, different auditory stimuli could have been chosen to make the auditory baseline higher.

Two dimensions that were particularly difficult for some participants during auditory discrimination were pitch range and pitch direction. Though during training almost all participants could distinguish between a high and low tone, several participants confused a low range for a medium range (and vice versa) and a medium range for a high range (and vice versa). In addition, it was observed that select participants had difficulty identifying whether a tone sequence traveled up or down at times stating "I have no idea" when asked to choose. Difficulty in making a direction judgment was exacerbated when direction and range were opposed. For example, it was more difficult for a participant to identify when a tone traveled higher in pitch when in a low range than when in a high range.

In a real-life scenario perhaps a better choice than pitch direction would have been sequence tempo which has shown to lead to particularly high discernment. Another method to increase the number of discernible sequences would be to increase the number of tone qualities. Patterson (1982) provided a guide to making timbres with maximum discriminability. Though abstract tones require training to impart meaning on them, they can convey information quickly and are not masked by speech-rich environments (Meredith & Edworthy, 1991; Montahan, Hetu, & Tanskey, 1993; Patterson & Milroy, 1980). Though there are clues in auditory discrimination research on how to make audio tones more memorable, additional targeted research on how to optimize audio sequences for the purpose of auditory displays is needed.

Even though auditory baseline performance did not match visual baseline performance, the use of auditory stimuli was advantageous. Firstly, overall performance would have likely been far worse if both sequences were visual because, as mentioned earlier, humans have cognitive limitations restricting their ability to process simultaneous events of the same modality (Triesmann & Davies, 1973; Navon & Gopher, 1979). Secondly, eyes can only be fixated on one event at a time necessitating the need to continually move eyes between sequences making the processing of any one sequence more difficult. And lastly, audio sequences can be optimized to more naturally match humans' auditory discernment.

It should be noted that even if baseline performances were matched, visual performance would likely still decrease more than auditory performance across the two response conditions. As such, the primary result that overall performance is best when visual items are responded to before auditory items would be the same.

In Experiment 1 it was confirmed that visual accuracy decreases more than auditory accuracy between the first and second response conditions. This accuracy pattern led to the best overall performance when visual information was recalled before auditory information. It was also observed that first response recall accuracy was significantly lower than baseline response accuracy. This result is somewhat surprising because the SRI for first response and baseline conditions is the same. However, as mentioned in the introduction, there is cause to believe that even when different resources are used a small amount of interference is likely to persist. To improve overall performance, it is worthwhile to investigate alternative methods of presenting visual and auditory information in a way that may reduce this disruption. In Experiment 2 I investigate whether this disruption is attenuated by presenting visual and auditory sequences separately rather than together. Results from such an investigation are expected to assist in the development of a parsimonious prediction framework discussed after Experiment 4.

Experiment 2

Experiment 1's results confirmed that requiring participants to respond to the visual event (i.e., dots) before the auditory event (i. e., tones), leads to greater retention. This finding is expected and is posited to be attributed to the relative longevity of echoic memory over iconic memory. Of interest in Experiment 2, however, is whether retention of a visual-auditory event may be further improved by serializing the visual-auditory presentation. My investigation of visual-auditory sequence ordering is not merely a fishing expedition. I propose that the simultaneous display of visual and auditory information (in Experiment 1) challenged humans' natural attentional limitations and that such challenges would be reduced by presenting the visual and auditory information in a serial rather than parallel fashion.

To investigate these questions Experiment 2 expands upon the Experiment 1 design by adding a presentation sequence variable. To recapitulate, in Experiment 1 there were four conditions: Baseline Audio, Baseline Visual, Audio Response First, and Visual Response First. In the latter two conditions, the visual and auditory stimuli were presented together. Experiment 2 contains conditions in which the visual and auditory stimuli are presented (a) simultaneously, (as in Experiment 1), (b) sequentially with the auditory stimulus presented first, and (c) sequentially with the visual stimulus presented first.

Method

Participants

Forty-eight undergraduates (28 females) participated as research subjects. The ages ranged from 17-21 with an average of 18.9. All participants participated as one way to fulfill a class requirement.

Materials

The sequences were presented using the same equipment as Experiment 1. In an attempt to make baseline audio task more accessible and closer in accuracy to visual performance, sequences were simplified by reducing the number of dimensions and number of levels in the pitch dimension. Though the three pitch levels used were are well below the 7+2 items, George Miller's famous apprehension number (Miller, 1956), many participants were highly inaccurate in classifying pitch range. Since the audio task was designed to be performed at a near 100% accuracy when performed alone, a decision was made to restrict the number of pitch levels to high and low.

Additional research suggested that orientation dimension is not an auditory feature stored in echoic memory. Darwin, Turvey, and Crowder (1972) noted in their famous echoic memory experiment that participants reported hearing alphanumerics in the middle of their head. Since the focus of these experiments is on how to leverage echoic memory, the decision was made to remove orientation as a dimension and have all sequences presented to both ears (i.e., central). These dimension restrictions were also applied to the visual sequences to keep the sequences analogous as in Experiment 1. Figure 17 is a representation of the visual matrix in which visual sequences were presented.

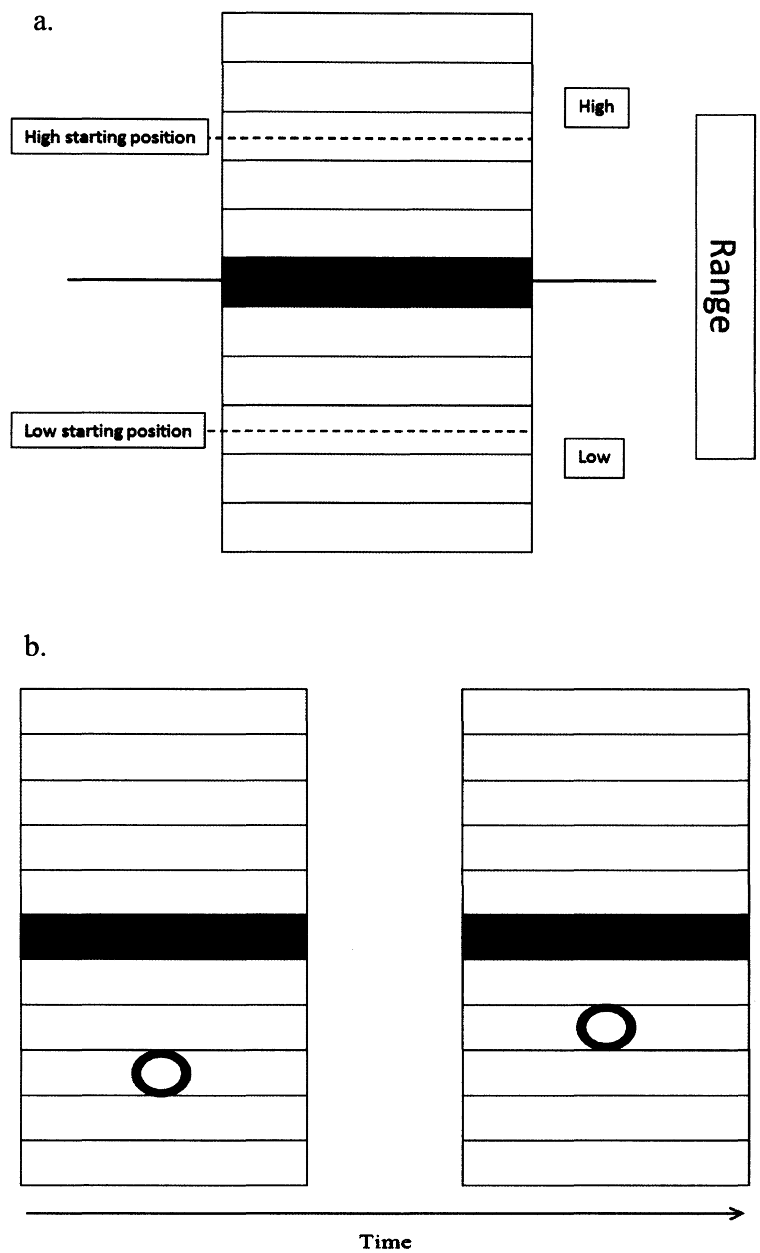

Figure 17. (a) The visual presentation matrix used in Experiment 2. Notable differences include the restricted pitch range which is now just low and high and exclusion of orientation. (b) A representation of a sequence with a numerosity: 2, complexity: complex, range: low, direction: up.

Design

Experiment 2 had eight conditions: two baseline performance conditions in which tones and dots were presented separately; and 6 conditions as part of a Presentation Order (3) x Response Order (2) factorial repeated measures design. The three levels of the presentation variable were, (1) Auditory/Visual Simultaneous Presentation, (2) Auditory-then-Visual-Separate Presentation, (3) and Visual-then-Auditory Separate Presentation. The two levels of the response variable were the same as in Experiment 1, (1) Auditory-then-Visual Response and (2) Visual-then-Audio Response.

Procedure

The procedure was the same as Experiment 1 with the exception that visual and auditory information was at times presented separately as opposed to simultaneously. Responses were executed in the same fashion. As in Experiment 1, each condition received 25 trials.

Results and Discussion

Figures 18-20 are a descriptive view of the probability of recalling all dimensions correctly for visual dimensions only, audio dimensions only, and both visual and auditory dimensions together. As in Experiment 1, accuracy was the proportion of trials that all dimensions were recalled correctly. In Figures 18 and 19 recall accuracy worsens between the first and second response. As expected this difference is more exaggerated for visual items (Figure 17) then auditory items (Figure 18). Also as expected, first responses in the presented separate conditions were answered at an accuracy level more closely resembling baseline accuracy than first responses in the auditory/visual simultaneous presentation conditions.

In the auditory-then-visual response condition, visual accuracy was appreciably greater in the visual-then-auditory presentation condition. This result directly contradicts the hypothesis that extending the visual presentation-to-response interval will have a negatively impact on visual performance. An explanation for this result is discussed after Experiment 4.

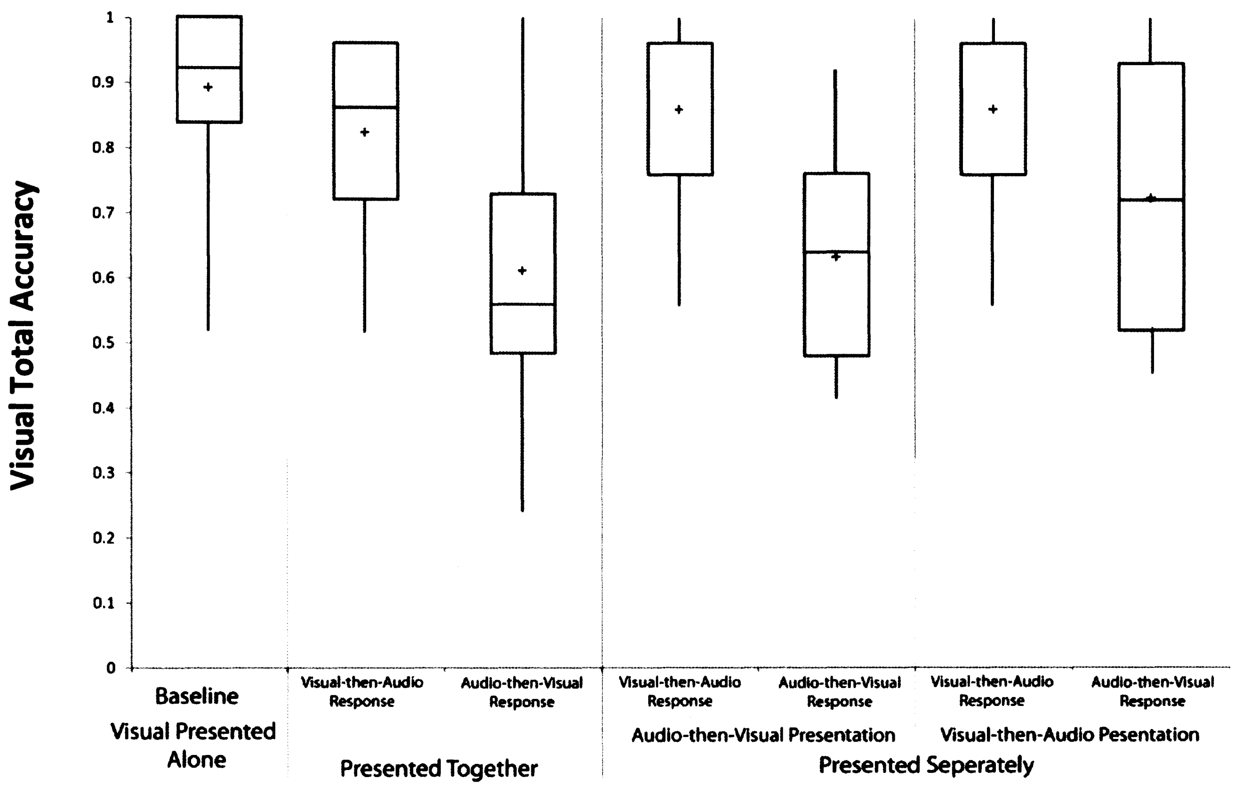

Figure 17. The proportion of trials all visual sequences were answered correctly.

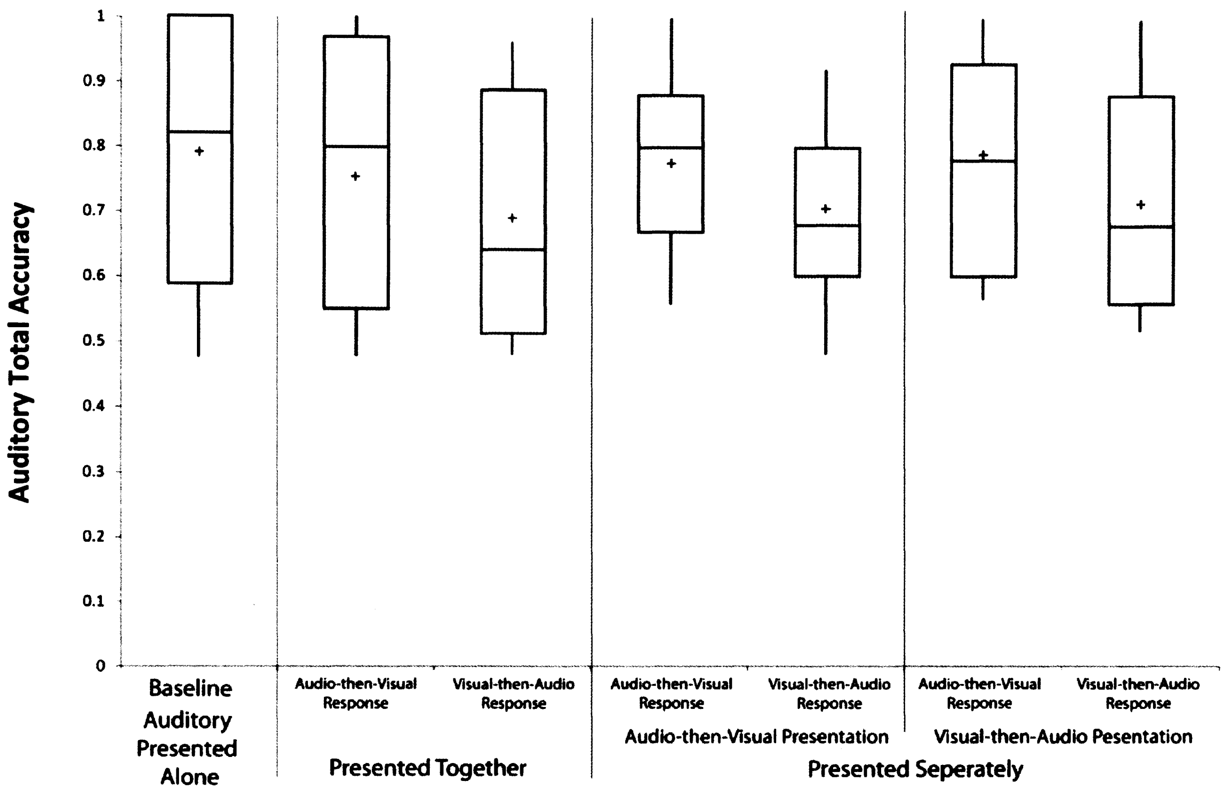

Figure 18. The proportion of trials all auditory dimensions were answered correctly.

An accuracy advantage of responding to visual items first is clearly demonstrated in Figure 19. This finding is in accordance with Experiment 1 findings. Of particular interest in Experiment 2, however, is whether presenting items separately will improve overall accuracy. As can be seen in Figure 19, the mean accuracy in the auditory/visual-simultaneous presentation is less than the mean accuracy in the presented-separate conditions. This accuracy gain was predicted and is likely explained by a reduction in perceptual load resulting from not perceiving both information streams simultaneously. Of note is that the accuracy difference between the two visual-then-auditory presentation conditions is smaller than both the auditory-then-visual presentation condition as well as the auditory/visual-simultaneous presentation. The direction of

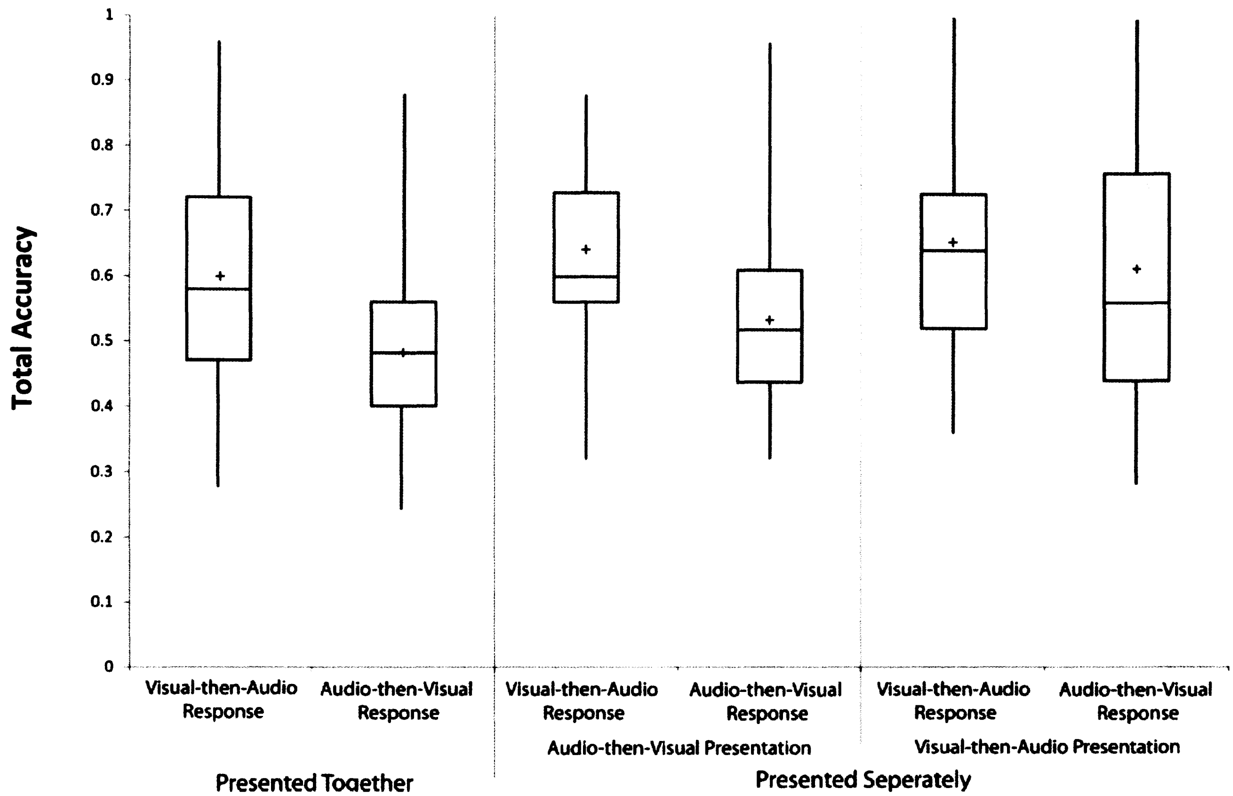

Figure 19. The proportion of trials that all 8 attributes from both the visual and auditory were recalled correctly.

this interaction was unexpected and is largely driven by the relatively high auditory-response-first visual accuracy in the visual-presented-first condition.

To reaffirm the finding that visual accuracy mean changes more than the auditory accuracy mean between a response-first scenario response-second scenario, a pairwise t-test was conducted. As in Experiment 1, participants mean accuracy for a given conditions was used as the data points for the inferential tests. The mean difference in visual accuracy conditions (M_{diff} = .20) and auditory accuracy conditions (M_{diff} = .11) between the first and second response conditions were found to be significant (M_{diff} = .09, CI_{95} = [.05, .13]) $t(47) = 5.41, p < .001$.

The result that visual accuracy varies more over time then auditory accuracy is consistent with Experiment 1 results. To reconfirm the primary Experiment 1 finding that response order affects overall accuracy, a within-subjects t-test was run comparing the probability of recalling all dimensions correctly (8 of 8) when sequences between the auditory-then-visual response (M = .54) and visual-then-auditory response (M = .63) conditions. A significant difference score was found between the two conditions indicating that participants were again more accurate when responding to the visual sequence first (M_{diff} = .09; CI_{95} = [.05, .13]) $t(47) = 4.8$, $p < .001$. This difference can be clearly seen in Figure 20.

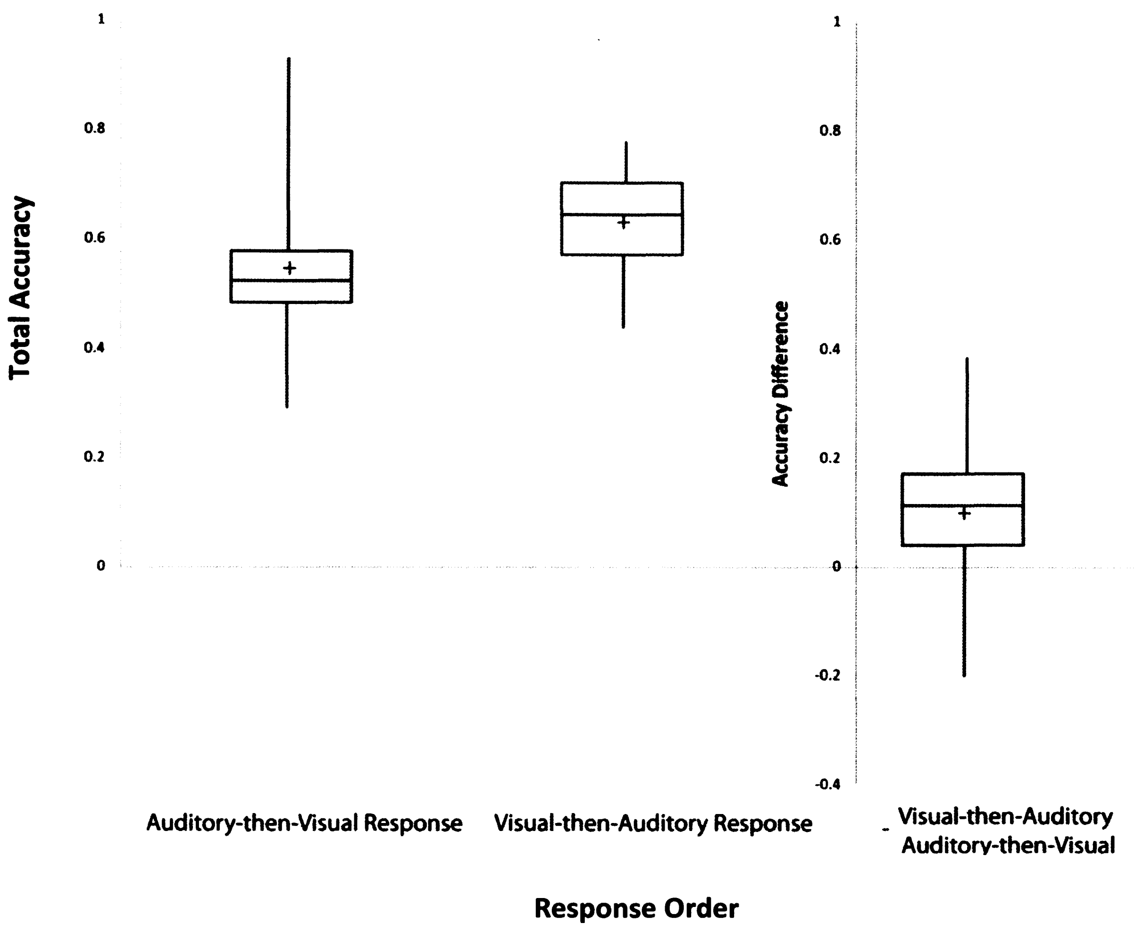

Figure 20. The response accuracy of the two response conditions and the difference between the two conditions.

A key goal of Experiment 2 was to determine whether presenting visual and auditory information separately would lead to first-response accuracy that more closely matches baseline performance than responses to sequences where information was presented together. In a planned comparison it was found that the difference between the modalities respective baseline performance and first response accuracy when visual and auditory information are presented together is significant, but not significantly greater than the difference between baseline performance and first response accuracy when the sequences were presented separately (M_{diff} =

.03 CI_{95} = [.01, .07]) $t(47) = 1.43$, p = .08. This result indicates that, on average, first responses were approximately 3% more accurate when the visual and auditory information were not presented simultaneously. It is not clear, however, whether this difference can be trusted because a conventional significance level was not met. Though presenting items separately does not lead to single-modal accuracy that is significantly closer to its respective baseline performance than presenting items together, of greater practical importance is whether presenting sequences separately improves overall performance? As can be seen in Figure 21, overall accuracy was higher with serial presentation than with parallel presentation. The difference was significant (Md_{iff} = .07, CI_{95} = [.03, .11]) $t(47) = 3.40$ $p < .001$.

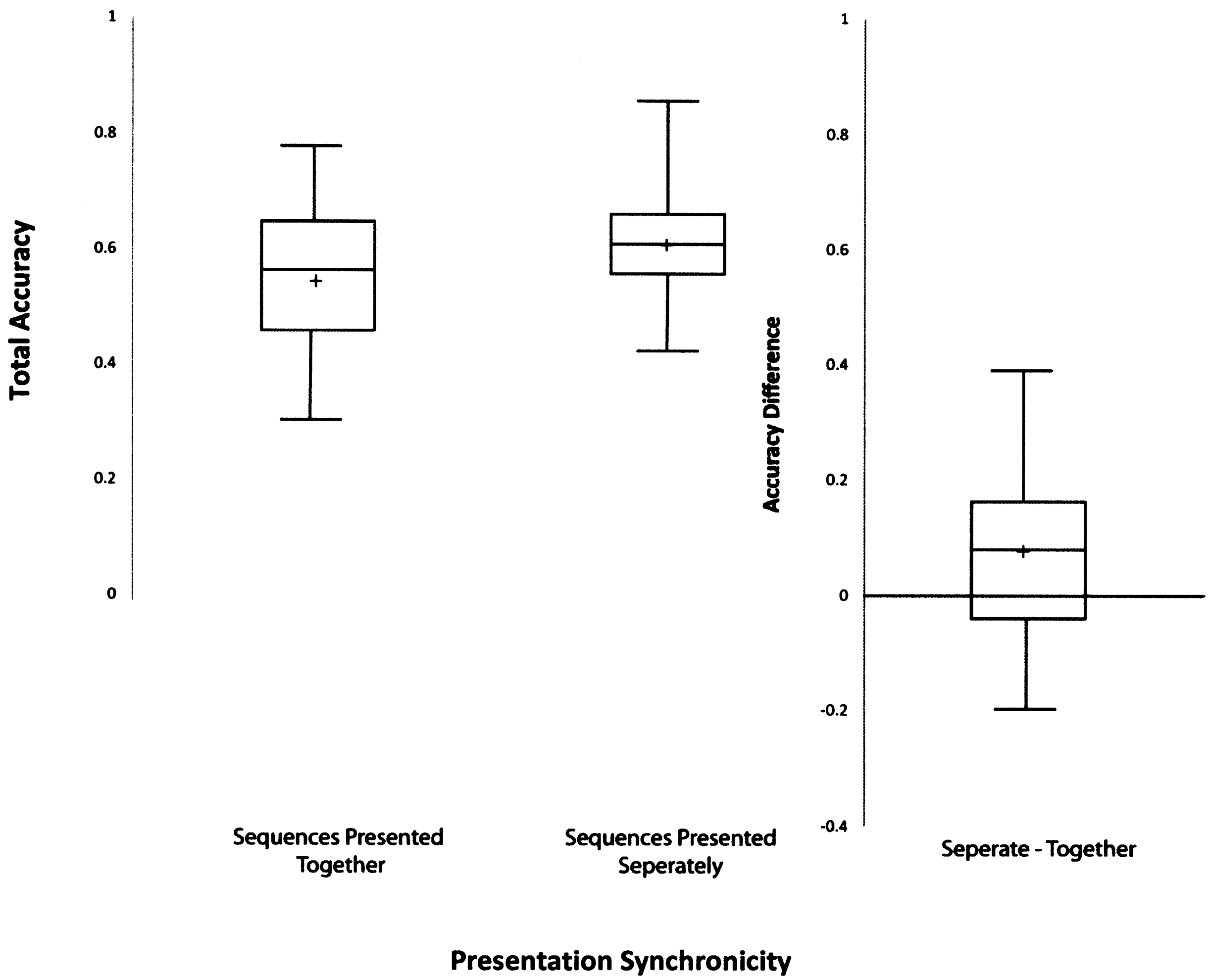

Figure 21. The response accuracy of the presented together condition, presented separately conditions, and the accuracy difference between the two.

These results are consistent with Wickens (2002) that presenting two modalities together incurs a small (yet present) resource cost that would not be present if the two modalities were presented separately. Moreover, the cost of elongating the average retention interval, if present, was far outweighed by the perceptual benefit.

To test the hypothesis that presenting the visual sequence first would lead to better accuracy than when the visual sequence is presented first because of the benefits of an uninterrupted recoding interval, a paired sample t-test was run comparing the mean performance in the auditory-then visual presentation trials to the visual-then-audio presentation trials. The expectation was that if any difference between the two groups exists, the audio-then-visual presentation would have greater accuracy than the visual-then-audio presentation.

Contrary to the core thesis of this paper, it was found that the longer visual retention interval condition (visual-then-audio presentation) lead to significantly greater accuracy then shorter visual retention interval condition (i.e., auditory-then-visual presentation) (M_{diff} = .05, CI_{95} = [.03, .07]) t = 2.07, p = .044 (Figure 22).

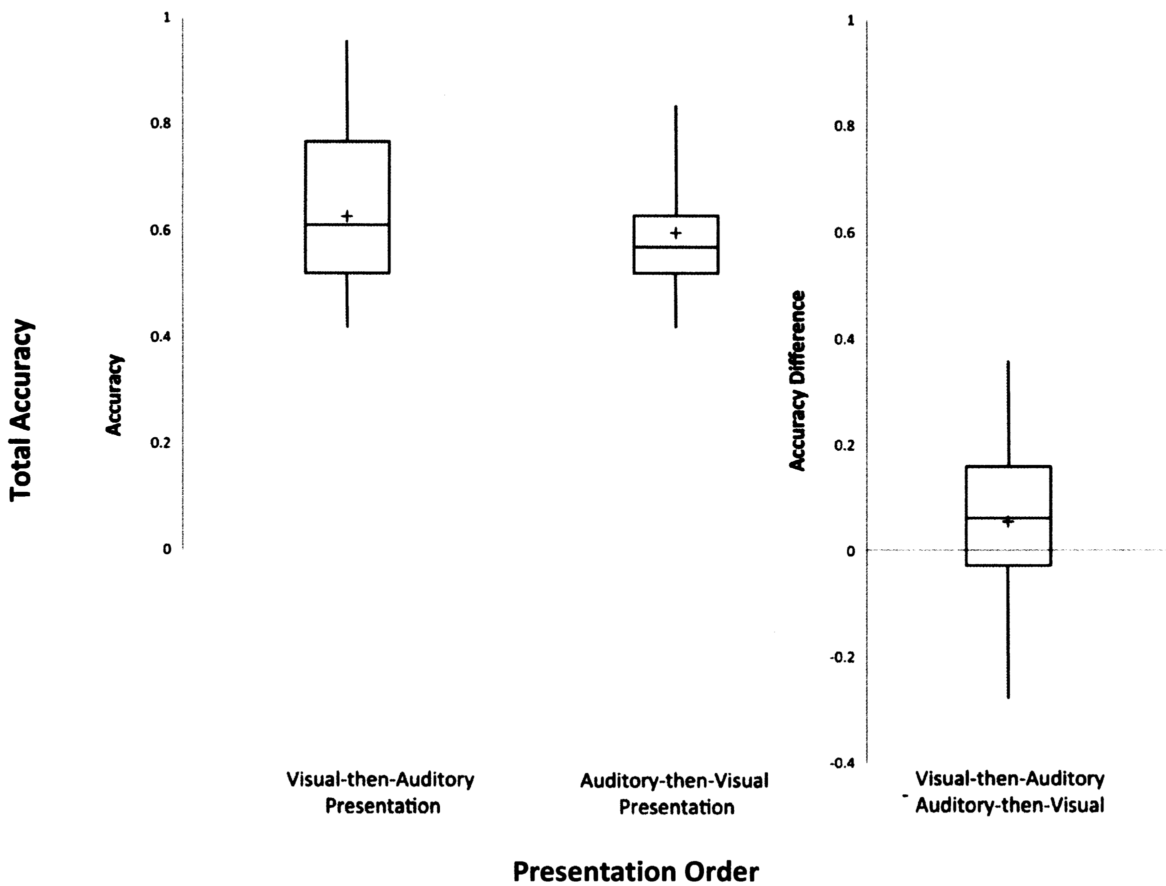

Figure 22. The response accuracy of the Visual-then-Auditory presentation condition, the Auditory-then-Visual presentation, and the accuracy difference between the two.

Figure 23 represents the error rate of the two presented separately conditions across the two response conditions. In accordance with the explained beneficial effects of a recording interval, there is less difference between the two response conditions when in the visual-presented-first condition (built in visual recoding interval) than in the audio-presented-first condition (no built in visual recoding interval).

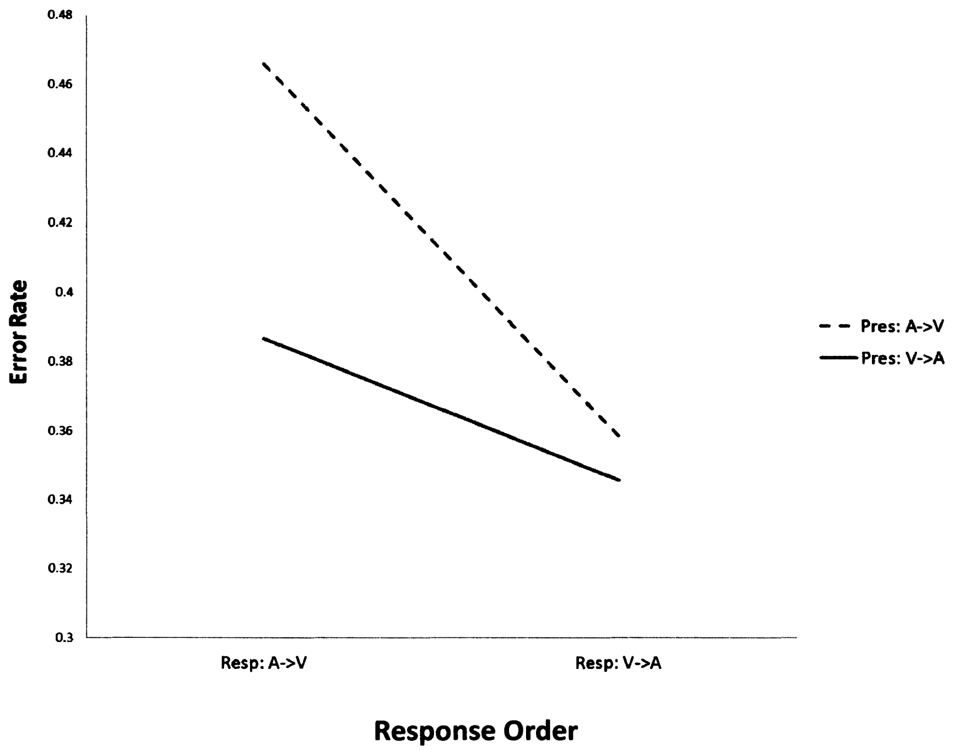

Figure 23. The error relationship between the two sequential presentation conditions crossed with the two response conditions. Error is defined as missing at least one of the eight dimensions.

To test whether the relationship between presentation and response sequence is as proposed, a paired sampled t-test was run comparing the average accuracy difference between visual-then-auditory presentation and auditory-then-visual presentation across the two response conditions. This difference was found to be significant (M_{diff} = .07 CI_{95} = [.01, .15]) $t(47) = 1.44$ p = .077. It should be noted that this interaction may have more theoretical than practical significance because presentation order has almost zero benefit in the more retention friendly visual-then-audio response condition. This interaction indicates that the benefit of the visual then response condition was contingent upon an auditory then visual response order. More specifically, this interaction suggests that while a long visual retention interval has a performance cost, the cost is exaggerated when an auditory response follows directly after a visual presentation. To examine this hypothesis, conditions that had an auditory response directly after a visual presentation was compared to conditions where there was not an auditory response directly after a visual presentation. Figure 24 is a simplified pictorial representation of these two kinds of conditions with their respective recall accuracy. Using a paired-sample t-test it was found that accuracy of conditions where an auditory response did not follow a visual presentation (M = .63) and accuracy of conditions where auditory responses did directly follow a visual presentation (M = .51) had a significant mean difference (M_{diff} = .12, CI_{95} = [.08, .16]) $t(47)$ = 6.08, p < .001 (Figure 25).

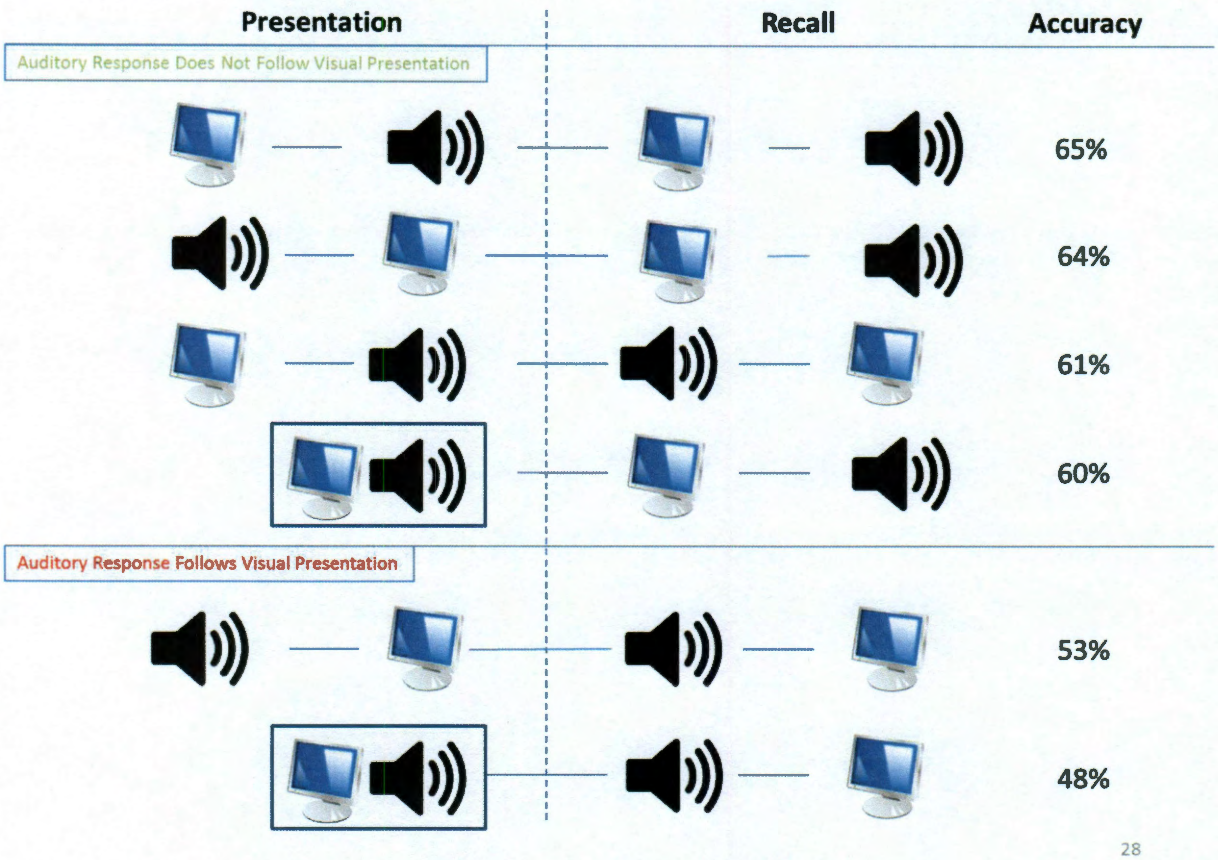

Figure 24. Presents a pictorial representation of the six dual-modality trials of Experiment 2. The four trials on the top have auditory responses that do not follow a visual presentation. The two trials on the bottom have auditory responses that do follow a visual presentation.

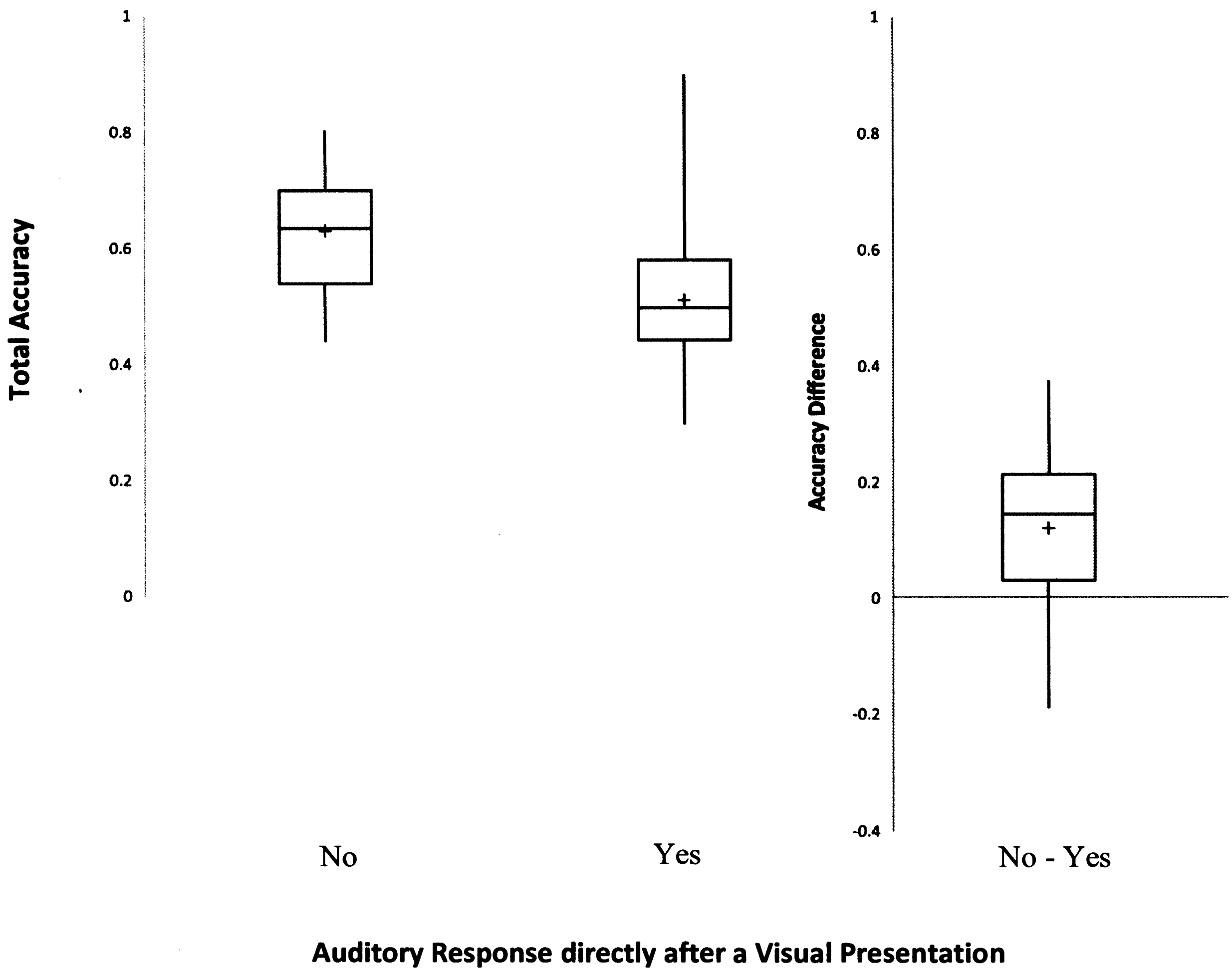

Figure 25. A representation of the response accuracy of the conditions with a recoding interval, conditions without a recoding interval, and the accuracy difference between the two.

The primary impetus of Experiment 2 was to test the hypothesis that presenting sequences serially leads to more accurate responses than when presenting sequences in parallel (as in Experiment 1). This hypothesis was confirmed. When items are presented sequentially the tax on perception and working memory resource will less likely be overloaded then if items are presented at the same time, increasing the likelihood that the same perceptual and working memory processes are occurring at the same time.

Perhaps the most surprising and impactful finding of this Experiment, however, is that presenting the visual sequence first leads to greater accuracy then presenting the audio sequence first. The expectation was that it would be better to present the audio sequence first such that, on average, the visual retention interval would be shorter and less likely “forgotten.” Contrary to this seemingly intuitive prediction, it was found that presenting the visual sequence first leads to greater response accuracy. Though not significant, there is some evidence that the benefit of presenting V before A is greater if the first response is to the A sequence. A theoretical grounding for this finding is evaluated in the prediction framework section after Experiment 4.

Experiment 1 and 2 addressed how manipulating the recall and response order of visual and auditory sequences affects a participant’s ability to recall abstract sequences. In these first two experiments participants were given no task context beyond the fact that they were taking part in a memory experiment. Though I have demonstrated that recall accuracy of a multi-modal display can be increased by presenting visual information before audio information, the real-world usefulness of these findings is limited to real-world tasks. Experiment 3 and 4 addresses this shortcoming by investigating whether the manipulations of Experiment 1 and 2 impacts how well participants can complete task more similar to real-world memory and decision-based monitoring tasks.

Experiment 3

During a real-world task, operators expect information from a display to have relevance to the task they are performing. In some instances it may be difficult for the operator to continually reference a display. This kind of scenario is thought to be most likely if the display is assisting/informing the primary task. In these instances it may be assumed that the easier the displayed information is to remember from a single viewing, the more helpful the display will be at assisting in the task for which the display is designed. The primary aim of Experiment 3 was to assess whether the findings of Experiment 2 generalize to a more complex and somewhat more realistic task. In response to results from Experiment 2, it is expected that performance will be best when the task allows for a recoding interval. In Experiment 3 the participants' task was to monitor (display) and verify (recall) reconnaissance information of pilots and planes flying through a restricted zone. This monitoring task is considered to be more complex and more realistic than recalling tone and dot sequences.

The second aim of Experiment 3 was to assess the hypothesis that the sequencing of visual and auditory information and the resulting performance patterns are specific to multi-modal memory tasks. Since ordering effects are hypothesized to be due to differences in sensory memory, such effects are not expected when only a single modality is used. For instance, imagine both plane and pilot information are presented visually. Here performance is not expected to be affected by presentation order because iconic memory will be used in remembering both kinds of information. Rather, if pilot information is presented visually and plane information auditorily, both kinds of sensory memory will be employed. As a result, order manipulations are expected to affect performance.

The third aim of Experiment 3 was to examine participants' response order strategy. Experiments 1 and 2 forced participants to respond in either a visual-then-audio or audio-then-visual order. As such, remains unclear which response order participants will naturally choose. In Experiment 3 participants were allowed to choose their recall order. Of interest was whether participants are naturally drawn to the same response order that led to the best performance in Experiment 1 and 2. Also of interest was whether a visual-then-audio response strategy maintains a recall benefit over an audio-then-visual recall strategy when the order is freely chosen. These results may have implications on how operators should be trained to respond to a multi-modal display.

Method

Participants

Thirty-six undergraduates (20 females) participated as research subjects. The ages ranged from 18-22 with an average of 19.5. Each participant was awarded credit towards a psychology department requirement.

Monitoring Task

Participants were instructed that they would be monitoring information about planes and pilots traveling through a restricted area. On each trial participants were presented with four images each showing a unique piece of information about the pilot. On any one trial, plane information was presented either visually, in a small grid as shown in Figure 26a, or auditorily using tones similar to Experiment 1 and Experiment 2. Pilot information was always presented visually as a row of images as shown in Figure 26b. The plane and pilot information are shown in greater detail in the Stimuli section. Since the plane information was presented either visually or auditorily and pilot information was always presented visually, the plane-pilot display was sometimes multimodal (pilot images and audio planes) and sometimes single modal (pilot images and visual planes). After being presented with plane and pilot information, a participant answered four accept-or-reject multiple-choice items (Figure 27). The four items consisted of two questions about pilot information and two questions about plane information. These questions were presented vertically and in a randomized order. A plane question was a written statement (e.g., "Number of Planes: 2") whereas a pilot question was an image. If the participant believed the item matched the presented information, the participant would choose "Accept". If an item did not match the presented information, the participant would choose "Reject". Once the participants were satisfied with their four responses they were asked to submit their answers by pressing a "Submit" button.

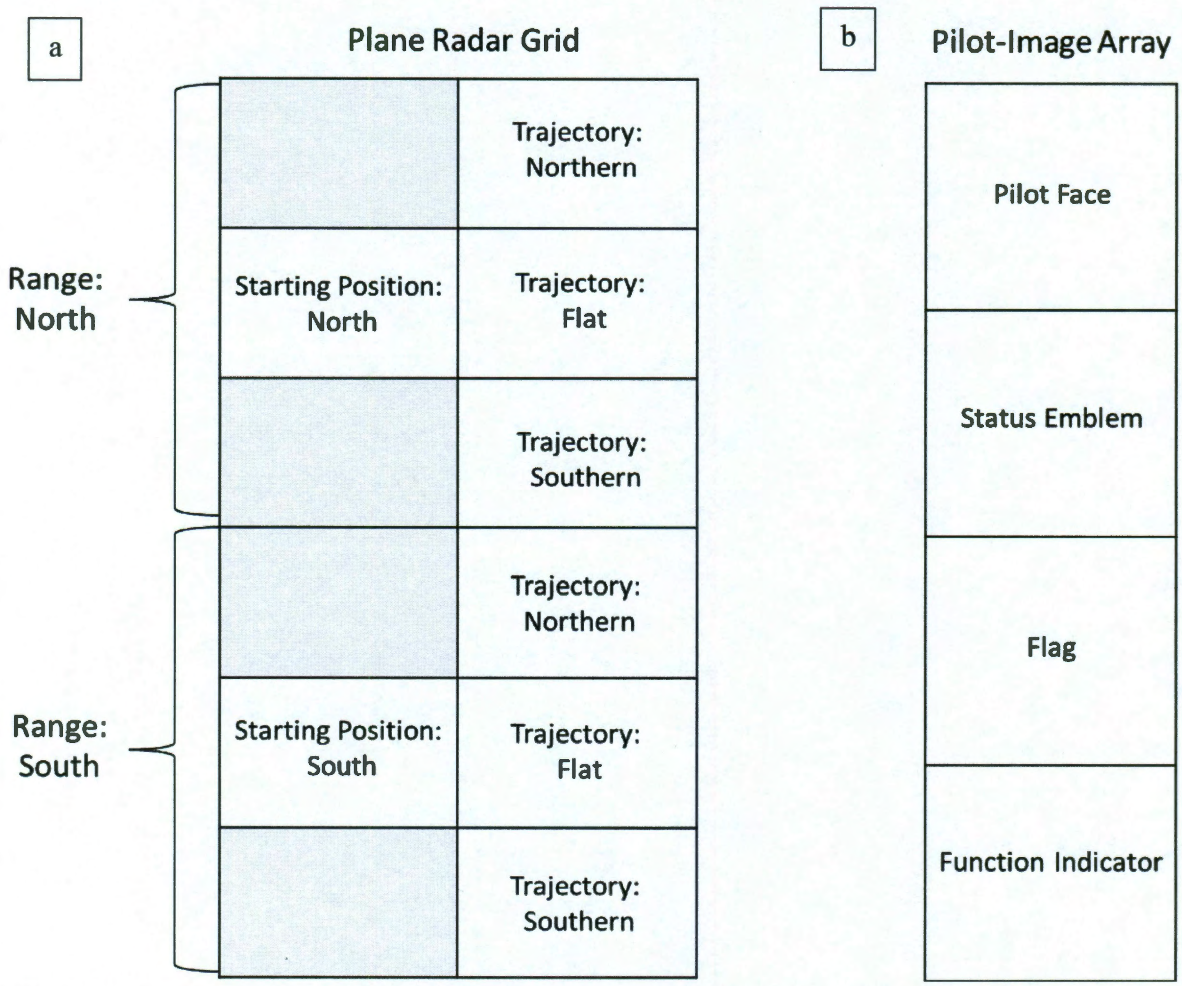

Figure 26. A representation of the (a) plane grid and (b) pilot-image array.

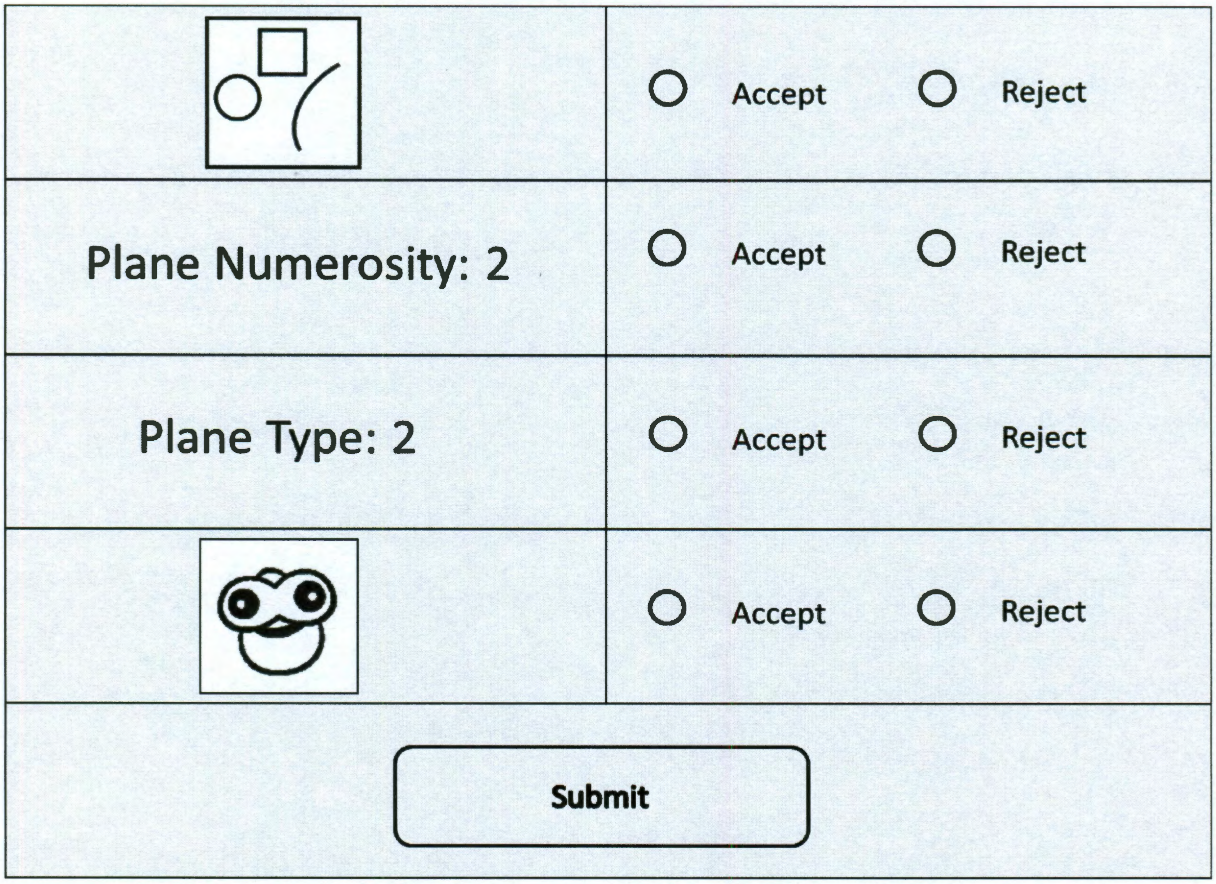

Figure 27. Represents a sample reconnaissance checklist. The order of the items from top to bottom is randomized. The participants can record the information in the order they wished.

Stimuli

The pilot-image array was composed of four images stacked vertically in the left half of the computer monitor (Figure 26b). Each image was 150px by 150px. The four images each represented one of four dimensions of the pilot. Each image was framed as being either a picture of the pilot directly (e.g., face) or identifiers on the pilot's person. Each image was selected independently of the other three images. The four pilot dimensions were the pilot's face (4 types), the pilot's status emblem (3 types), the pilot's flag (2 types), and the pilot's function indicator (2 types). In all there were 48 (4x2x2) possible image combinations. Examples of each of these images are shown in Figures 28 - 31.

The Plane information likewise varied in four dimensions. The dimensions were numerosity or number of planes (4 types: 1-4 planes), vertical screen location (2 types: top or bottom of screen), direction (3 types: up, flat, or down), and plane type (2 types). Numerosity was represented visually as the number of plane images and auditorily as the number of sequential tones (i.e. one plane: "beep", four planes: "beep. beep. beep . beep"). The combined duration from the start of a beep to the end of the last beep was 800ms. If there was one beep the tone duration was the entire 800ms. If there were multiple beeps the duration of each beep was 600ms/number of beeps. Range was either North or South. Range was represented visually as the plane being in the top or bottom of the grid and auditorily as either being a high (Experiment 1 high range tone starting pitch) or low tone (Experiment 1 low range tone starting pitch). Direction was northern, flat, or southern. Direction was represented visually as a plane or planes moving upward, flatly, or downward as the plane(s) moved right across the grid. Direction was represented auditorily as a series of tones that became higher in pitch, retained the same pitch, or became lower in pitch. Pitch differences were the same as in Experiment 1. For example, an auditory plane Type was represented visually as a jet plane (type 1) or a propeller plane (type 2). See figure 32 and 33 for pictures of the two plane types. Type 1 planes were represented auditorily as a pure wave tone (i.e., a simple tone from Experiment 1). Type 2 planes were represented as a sine wave tone accompanied by a drum beat (i.e., a complex tone from Experiment 2). Visual stimuli were presented on an iMac 15-inch CRT monitor. Auditory stimuli were presented with Koss headphones.

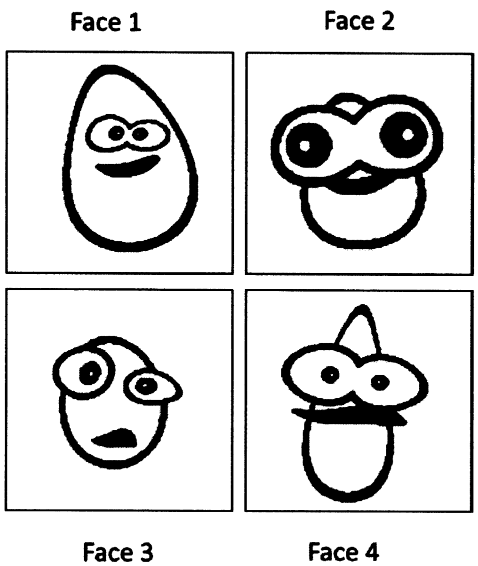

Figure 28. The four pilot face images.

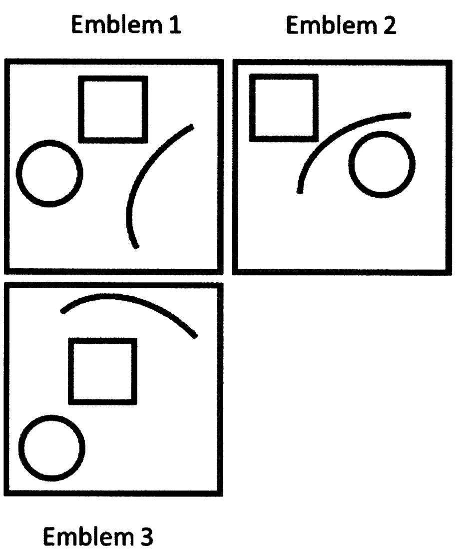

Figure 29. The three pilot emblem pictures

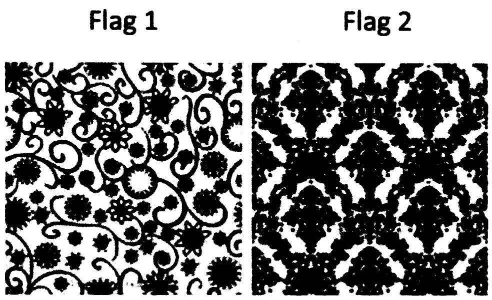

Figure 30. The two flag images.

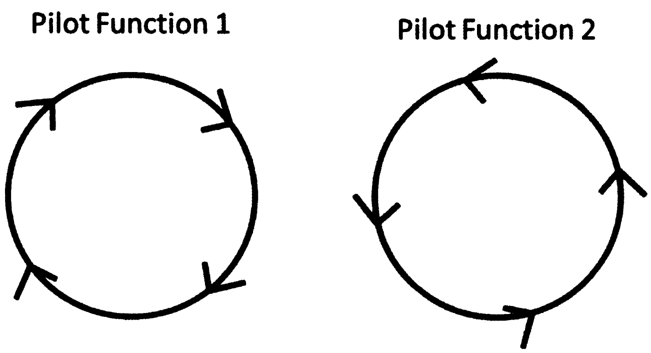

Figure 31. The 2 pilot function images

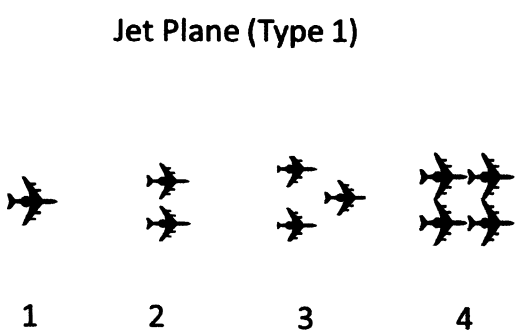

Figure 32. The 4 jet plane numerosities.

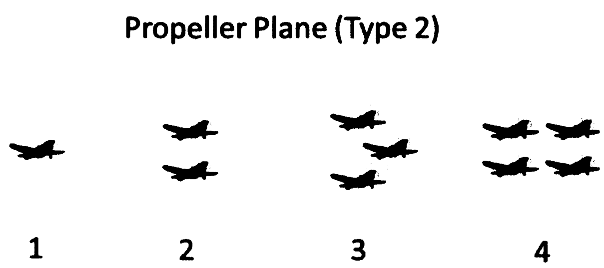

Figure 33. The 4 jet propeller plane numerosities.

Design

The design was a Modalities (2) x Presentation order (2) factorial design with repeated measures on both factors. The levels of modalities were dual (planes: audio/pilot visual) and single (planes: visual/pilot visual). The presentation order levels were plane-then-pilot presentation and pilot-then-plane presentation. The dependent variables were the accuracy of each checklist-item response (correct = 1, incorrect = 0) and the order that the responses were made. Since each checklist consisted of four items, a response order for any one item was 1, 2, 3, or 4.

Procedure

Each participant was told they would be monitoring radar information of planes flying through a restricted zone and images of the pilots flying the plane. The participant was asked to monitor the information coming in and subsequently confirm and or reject information on an intelligence report checklist. After the participant was told that they would be monitoring pilot and plane information, the participant was given training on how to decode plane and pilot information.

Plane Mapping Training: To start, the participant was told that a high starting pitch corresponded to a north range and a low starting pitch corresponded to a south range. Next the participant was told that pitches that get higher correspond to a northern trajectory and that pitches that get lower correspond to a southern trajectory. Next the participant was told that the number of beeps corresponds to the plane numerosity. And lastly, participants were told that a regular tone corresponds to plane type 1 and tone accompanied by a snare drum corresponds to plane type 2. Participants would then complete an audio plane assessment where they were presented with an audio plane example and asked to identify the four dimensions. A participant would continue this assessment stage until he or she correctly identified the three plane dimensions in 10 contiguous audio plane examples. The same training and assessment procedure was completed for visual-planes.

Pilot Mapping Training: Next, participants were shown how to map a type to the pilot images. Participants were shown pilot faces, emblems, flags, and type on a screen with the type printed to the left. Participants were then given a pilot assessment in which they were asked to identify the type of each of the four pictures. The participant would continue the assessment until he or she correctly matched the three pictures with its corresponding number in 10 contiguous trials.

After the participant demonstrated mastery of plane and pilot mapping the experimental monitoring task was introduced. The participant was told that he or she would be monitoring both pilot and plane information and completing a four item list. The participant was told that for each item they must discern whether the information in the list matches what they just monitored. If the information matches the participant is to select "Accept". If the information does not match the information is to select "Reject".

A monitoring trial occurred as follows. A trial began with a plane radar matrix on the left of the screen, four blank squares stacked vertically on the right, and a "Start" button below the matrix. Upon pressing "Start" the participant was shown a pilot-image array for two seconds. The pilot image included pilot face 1, emblem 1, flag 2, and function 2. After the two seconds the images would disappear and the plane sequence began. In this example scenario the planes were auditory. Two planes were presented (sound: "(start position) beep .beep … (second position) beep . beep"), planes were of type 1 (sound: regular tone), the panes were in the north (sound: high pitch range), and the planes had a northern trajectory (sound: the second "beep" cluster had a higher pitch than the first cluster). The response phase started immediately after the offset of the plane sounds. The entire presentation phase lasted exactly 4000 milliseconds. During the response phase the participant was shown a gray box with four items. Two of the lines were pilot-items (images) and the remaining two lines were plane items (statements). Next to each item was an "Accept" and "Reject" radio button toggle. Assuming the participant had a perfect memory, if the pilot or plane information in the gray box matched the information from the presentation phase, the participant would select "Accept." If the supplied information did not match the pilot or plane information then the participant would select "Reject." After accepting or rejecting all four pieces of reconnaissance information, the participant was instructed to press the "Submit" button. Upon pressing "Submit" the gray box disappeared and participants were prompted to begin the next trial by pressing the "Start" button. Participants repeated this procedure 110 times. There were 10 practice trials and 100 experimental trials. The one-hundred practice trials were divided evenly among the four conditions. The conditions were ordered randomly. After the 100th experimental trial, subjects were informed that the experiment had concluded and that they should notify the experimenter. See Figure 34 for a timeline of the procedure.

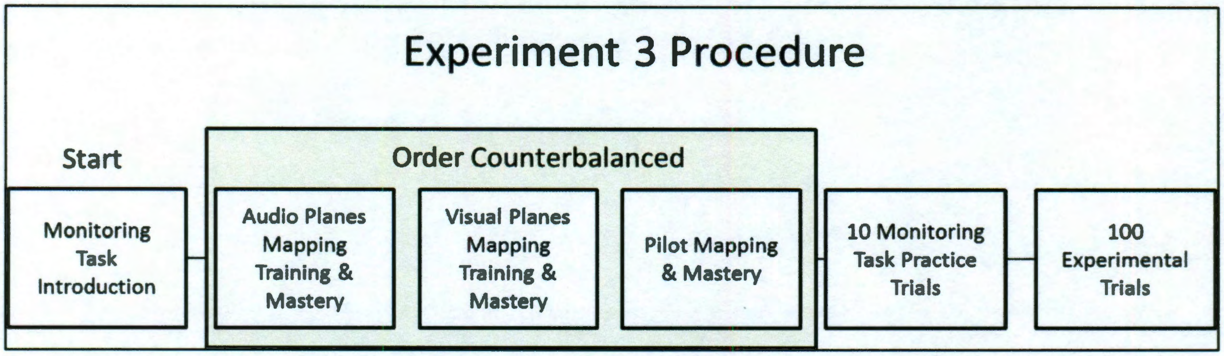

Figure 34. Experiment 3 procedure from left to right.

Results and Discussion

Plane recall error

As in the previous two experiments, participants' average accuracy within conditions was used as data points for inferential tests. Using a within subjects t-test it was found that the plane response error difference between the two presentation order conditions (planes-then-pilots presentation – pilots-then-planes presentation) in the audio-planes condition (M_{diff} = .02) was significantly less than the difference between the two presentation order conditions in the visual-planes condition (M_{diff} = .09) (M_{diff} = .07, CI_{95} = [.02, .11]), $t(35)$ = 2.67, p = .011. Using a within-subjects t-test a main effect of presentation order was found with a plane-then-pilot presentation order (M = .19) leading to more plane errors than a pilot-then-plane presentation order (M = .14) (M_{diff} = .05, CI_{95} = [.03, .08]) $t(35)$ = 3.60, p < .001. Using a within-subject's t-test, a significant main effect of modality was also found. An audio planes presentation (dual modality) (M = .13) resulted in fewer errors than a visual planes presentation (single modality) (M = .19) (M_{diff} = .06, CI_{95} = [.03, .09]), $t(35)$ = 4.50, p < .001 (Figure 35).

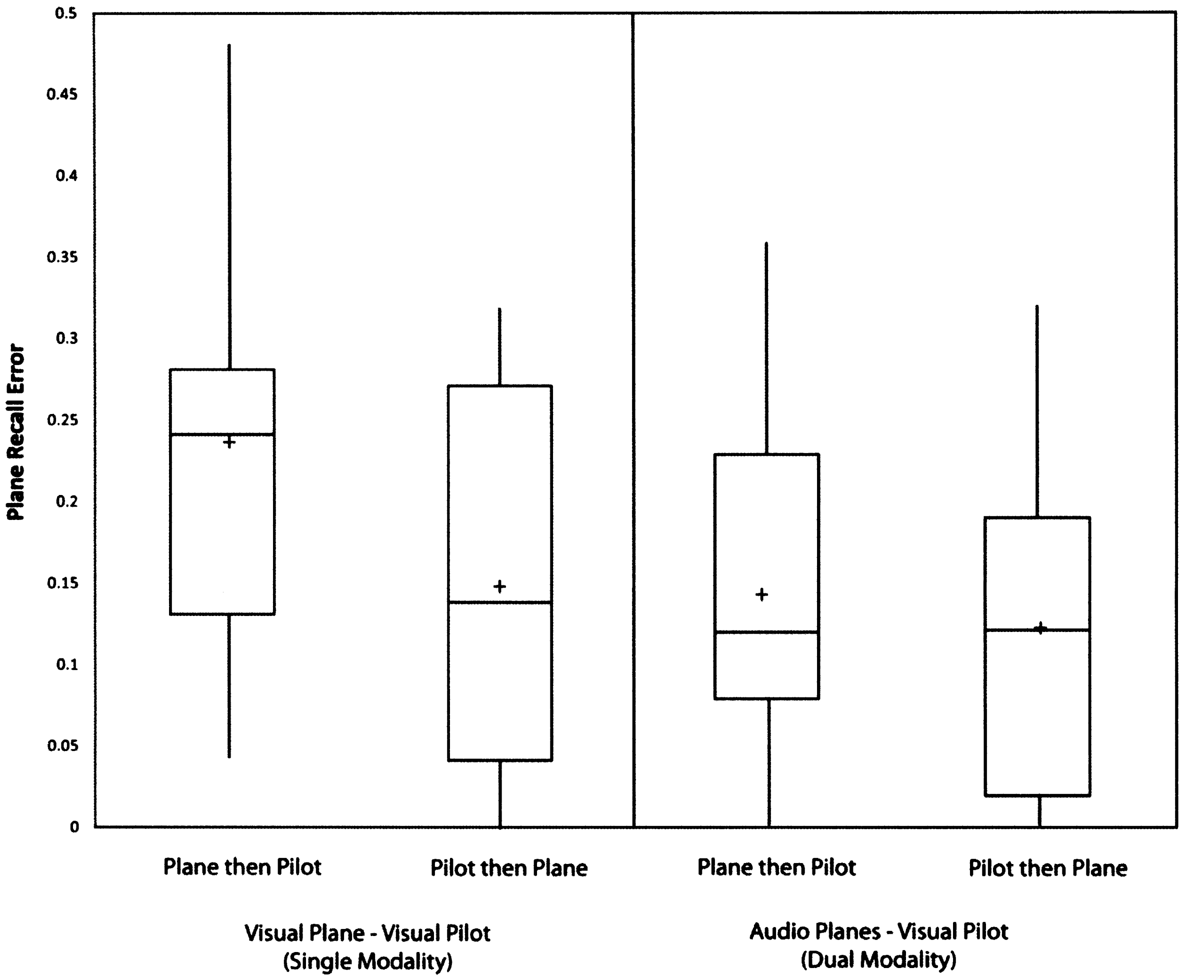

Figure 35. The proportion of the time that at least one of the two prompted plane dimensions was answered incorrectly.

Pilot recall error

Figure 36 refers to the proportion of time that one of the two prompted pilot dimensions were answered incorrectly. As can be seen, the effect of presentation order was approximately the same for visual planes as it was for audio planes. The pilot error difference between the two presentation order conditions between the visual and auditory plane conditions was not appreciable.

The difference in error rate between the two presentation order conditions were not significantly different in the audio-planes condition (M = .07) than in the visual-planes condition (M = .07) (M_{diff}= .00, CI_{95} = [.05, .05]), $t(35)$ = 0.08, p = .937. Using a within-subject's t-test a main effect of presentation order was found with the plane-then-visual presentation order (M = .26) leading to fewer errors than a pilot-first presentation order (M = .32) (M_{diff}= .07, CI_{95} = [.03, .09]), $t(35)$ = 3.88, $p < .001$. As can be seen in Figure 36, the errors with the visual-planes (M = .30) were slightly higher than with audio-planes (M = .28). However, the difference was not significant (M_{diff} = .02, CI_{95} = [.03, .07]), $t(35)$ = 1.32, p = .193.

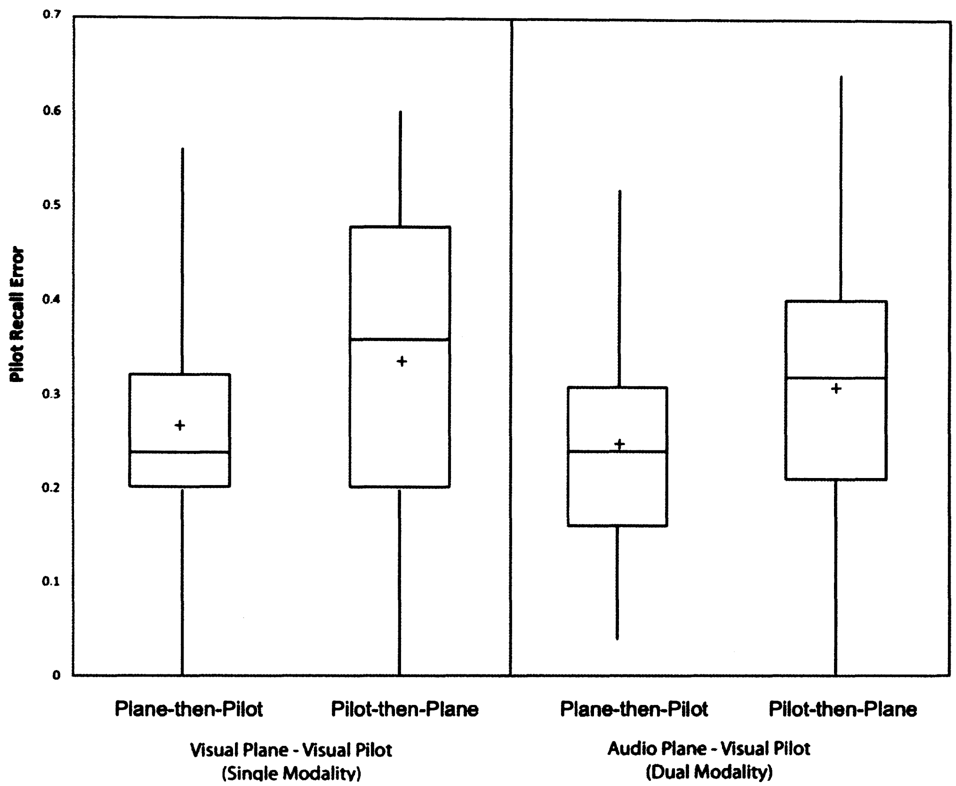

Figure 36. The proportion of the time that at least one of the two prompted pilot dimensions was answered incorrectly.

Signal Detection Results

Table 2 presents the average hit rate, false alarm rate, d', and criterion value for each of the four conditions. Participants were more sensitive in the plane then pilot audio plane-visual pilot condition then in all other conditions ($p < .01$). Overall participants had a liberal bias indicated by negative criterion values. This indicates that, all else being equal, participants would rather accept information than reject it. Though C and d' are theoretically orthogonal, there is a strong correlation between sensitivity and a liberal bias. As sensitivity increased the liberal bias increased. As with sensitivity, C was the smallest in the plane-then-pilot audio-plane, visual-pilot condition. One explanation for the general liberal bias may be that "Accept" was presented to the left of the "Reject" thus, in cases when participants were unclear of the correct response, the participants simply selected the first option.

Forced-Choice Paradigm (two choices) for a single dimension					
Order	Modality Pairing	H	FA	d'	criterion c
Plane then Pilot	Visual Plane - Visual Pilot	0.960	0.150	1.970	-0.357
Pilot then Plane	Visual Plane - Visual Pilot	0.956	0.148	1.949	-0.333
Plane then Pilot	Audio Plane - Visual Pilot	0.994	0.112	2.621*	-0.636*
Pilot then Plane	Audio Plane - Visual Pilot	0.971	0.137	2.117	-0.401

Table 2. The hit rate, false alarm rate, sensitivity, and criterion for the four conditions in Experiment 3

Overall recall error

Figure 37 represents the proportion of time that participants responded to all four items correctly across all four conditions. Of note is the overarching accuracy benefit of a dual-modality presentation. Also of note is that the planes-first response order led to a greater accuracy benefit in the audio-planes condition (dual modality) than in the visual-planes condition (single modality).

The difference in error rate between the two audio-planes conditions (M_{diff} = .05) was slightly greater than the difference in error rate between the two visual-planes conditions (M_{diff} = .01) (M_{diff} = .07, CI_{95} = [0.0, .13]) $t(35)$ = 2.03, p = .050. A significant main effect was observed with audio-plane trials (dual modality) resulting in fewer errors (M = .38) than visual-planes trials (single modality) (M = .43) (M_{diff} = .06, CI_{95} = [.03, .10]) $t(35)$ = 3.6, $p < .001$. There was little difference in error rate between pilot-then-plane condition (M = .40) and the plane-then-pilot condition (*M* = .42) *t(*35) = .61, p = .71. These data are consistent with the assertion that the benefit of display order occurs only when more than one modality is used. A simple main effect was observed in error rate between the pilot-then-plane (*M* = .40) and plane-then-pilot (*M* = .35) condition of the audio-planes condition (M_{diff} = .05, CI_{95} = [0.0, .10]) *t(*35) = 2.08, p = .045. No such simple main effect was found between the between the pilot-then-plane (*M* = .44) and plane-then-pilot (*M* = .43) conditions of the visual-planes condition, *t(*35) = .61, p = .54.

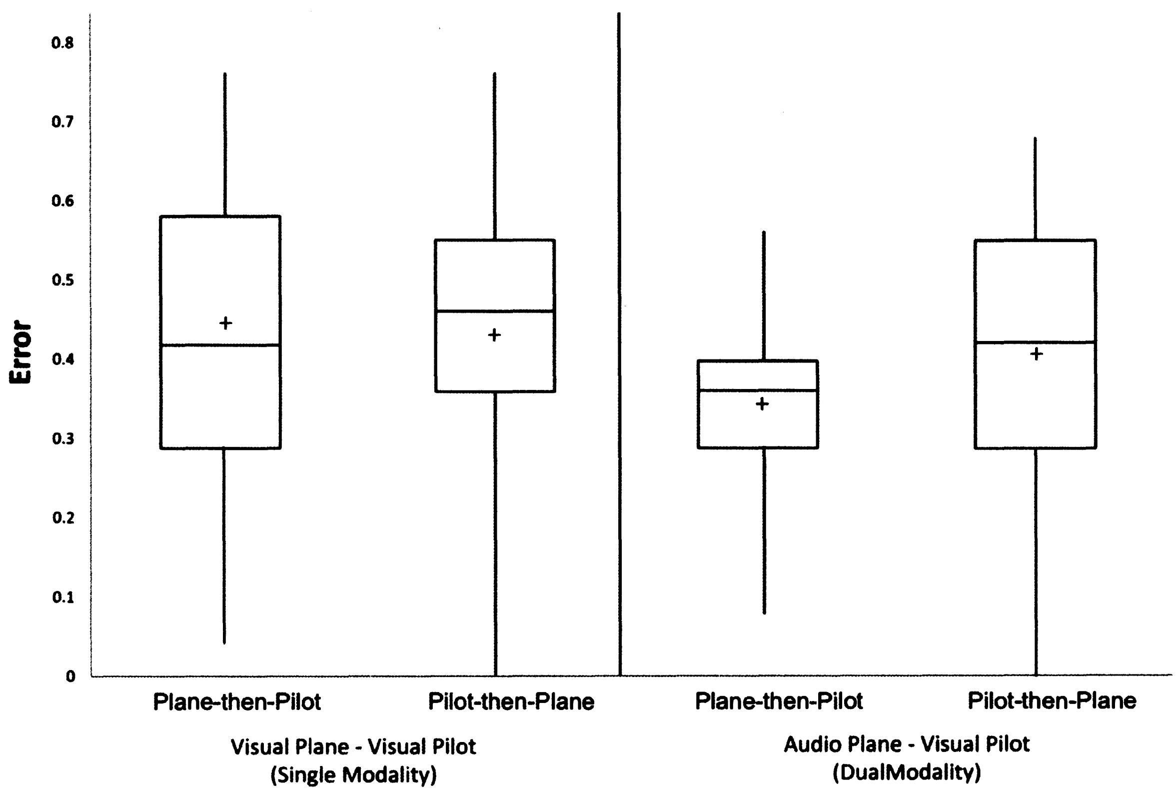

Figure 37. The proportion of time all plane and image information was answered correctly across the four presentation conditions.

Performance in the audio-planes condition was better when the visual pilot information was presented first than when the audio-planes information was presented first. This result is in line with the original hypothesis because an audio-then-visual presentation will shorten the visual retention interval. On the surface, however, this result is not consistent with the presentation order results from Experiment 2. In Experiment 2 it was demonstrated that a visual-then-audio presentation leads to the best performance. The beneficial effect of a visual-then-audio presentation in Experiment 3 was observed only when participants were forced to respond first to auditory items. Since participants were free to choose their own response order, it is possible that participants did not respond in a similar order across both presentation conditions. Specifically, if participants tended to respond to the last-presented sequence first, the two audio-plane conditions would have a similar presentation/response sequence as the auditory-then-visual presentation/ visual-then-audio response condition (Experiment 2 accuracy: 64%) and visual-then-audio presentation/ audio-then-visual response condition (Experiment 2 accuracy: 61%) of Experiment 2.

The effect of presentation order on response order

To identify if presentation order affected response order in Experiment 3, the order that the second presented sequence information was either accepted or rejected was calculated. Since participants had the choice to switch between audio and visual responses, for instance AVAV rather than AAVV or VVAA, the response order to the second presented sequence was not binary. In the context of this experiment response order was calculated by averaging the order that second-presented information were responded to. If second-presented information was always responded to first, then these responses would always take up the response 1 and response 2 spots and the average response order would be 1.5. In contrast, if responses to the secondly presented sequence always came last, then the responses would always be response 3 and response 4 and with an average response order of 3.5. Since a scale from 1.5->3.5 is somewhat arbitrary, response order averages were transformed to a 0 -> 1 scale. On this transformed scale a 0 reflects always responding first to the second-presented sequence, 1 corresponds to a strategy of always responding to the second-presented sequence last, and .5 reflects no response order difference. The transformation was calculated as follows:

RO = Response Order (using raw order values)

ARO = Adjusted Response order (to fit between 0 and 1)

HP = Highest possible response order average

LP = Lowest possible response order average

$$ARO(RO) = \frac{RO - LP}{HP - LP}$$

If a participant changed a response on an item, the item was tagged as the latest response. For example, if the response order was B, A, C, D, and then C again (because it was changed), the recorded order was B (1^{st}), A (2^{nd}), D (3^{rd}), C(4^{th}).

As can be seen in Figure 38, results indicate that there is a strong tendency to respond to the second-presented sequence information prior to that of the first-presented sequence. The average adjusted response order for secondly presented items ($M = .32$ $CI_{95} = [.30, .33]$) was significantly lower than the neutral .5 value, $t(35) = 34.4, p < .001$.

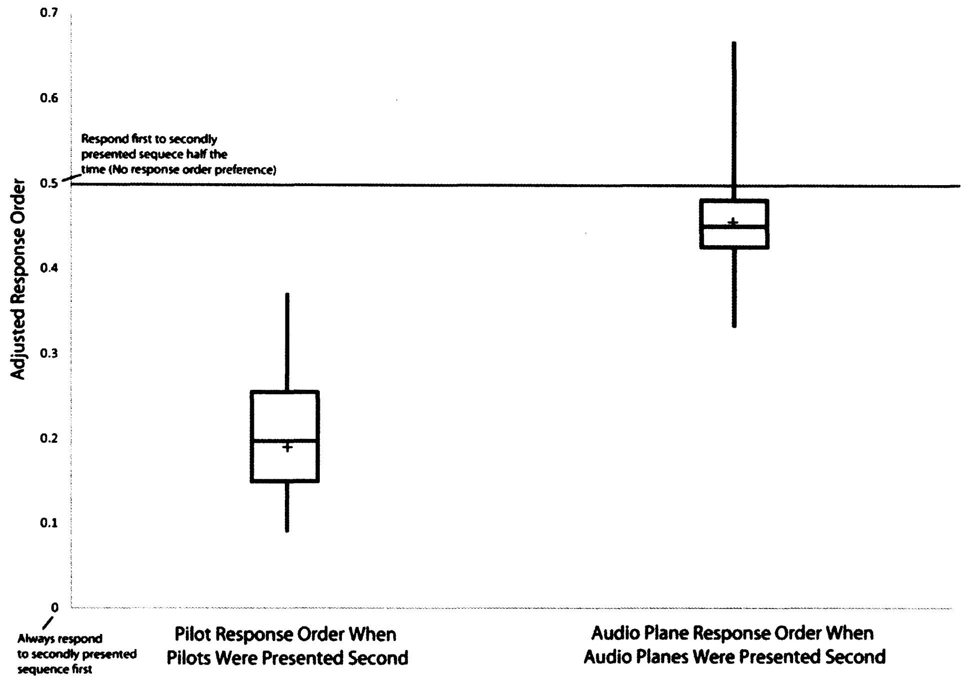

Figure 38. The adjusted response order for pilots and audio planes when each was the second presented sequence.

Experiments 1 and 2 demonstrated that responding to visual information prior to auditory information leads to more accurate recall. Since participants tended to respond first to the secondly presented information, an auditory-then-visual presentation most often led to the better response order. Experiment 2 demonstrated a visual-then-audio presentation led to a recall accuracy advantage. In the current experiment, however, the advantage of a visual-then-audio presentation was tempered because it biased the participants towards an audio-then-visual response strategy systematically increasing the visual response interval. Overall, the response order strategy closely resembles recall strategies from a free recall test. In these tests participants are typically presented with a list of words and then asked to recall the words presented. The typical strategy is for participants to recall the last presented words first and then the first (Deese

& Kaufman, 1957; Murdock, 1962). Though presentation order biased response order, it was found that pilot images were generally answered before audio planes suggesting some natural concern that visual information should be "offloaded" earlier than audio information. There was a significant difference in adjusted image response order ($M = .19$) and the adjusted audio response order ($M = .44$) ($M_{diff} = .25$, $CI_{95} = [.21, .29]$), $t(35) = 12.5$, $p < .001$.

The effect of response order on overall error rate

To gain clues on whether a participant's spontaneous response strategy leads to similar performance patterns as when a response order strategy is forced, the accuracy of each response profile was calculated. Figure 39 shows the six response profiles and their adjusted average visual response order.

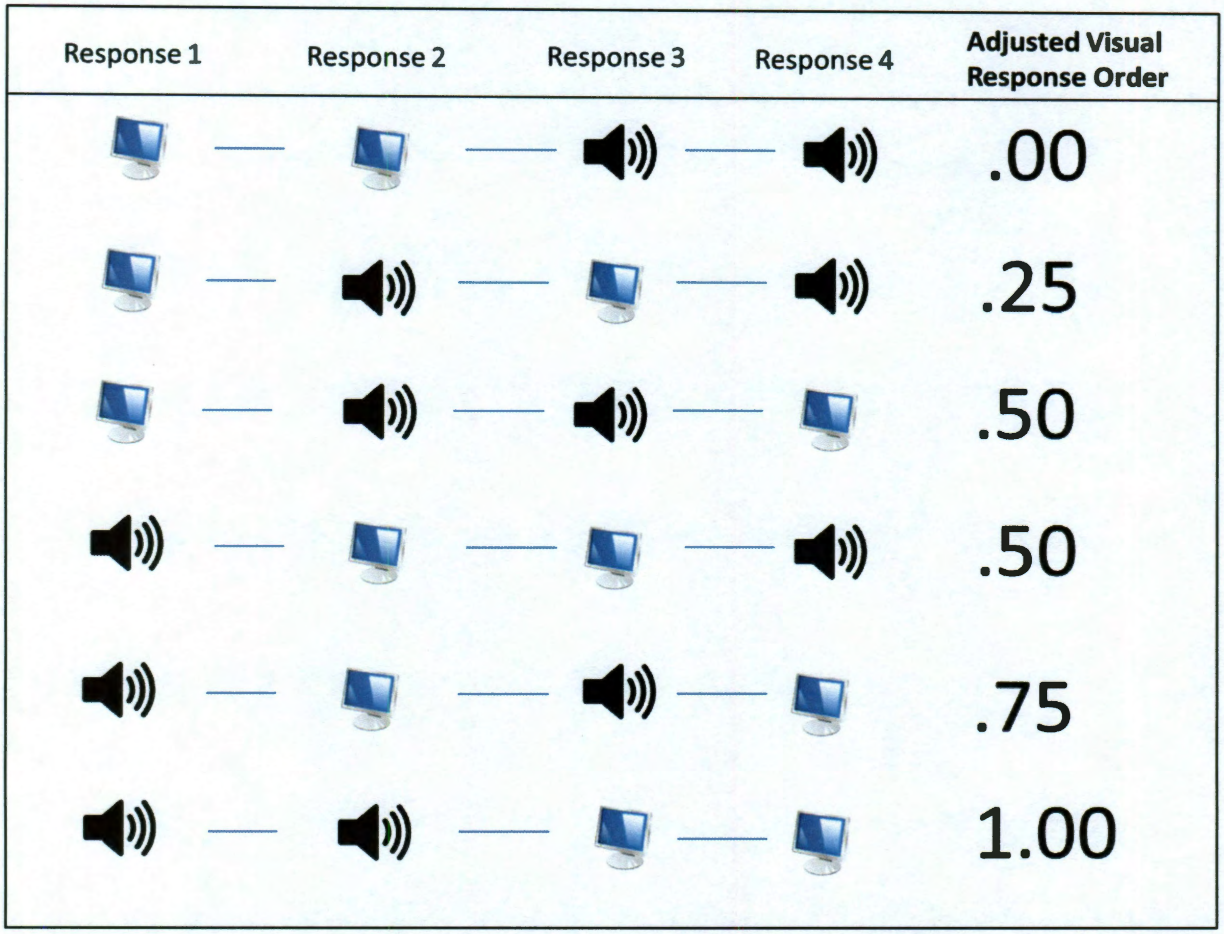

Figure 39. The six response profiles and their respective adjusted visual response order.

If at all times it is it is better to respond to visual information first then a negative linear relationship between average visual response order and accuracy would exist. In contrast, if performance is flat across the six profiles then this is evidence that a visual-first strategy may not necessarily lead to better performance when participants are choosing the response order. A comparison of the six response profiles demonstrates that as the average visual response order increases, errors increase. A linear contrast *(2 (vvaa_error) -1(vava_error) + 0 (vaav_error)+0 (vaav_error)+1 (avav_error) + 2 (aavv_error))* across the six response patterns is significant $t(35) = 4.85, p < .001$.

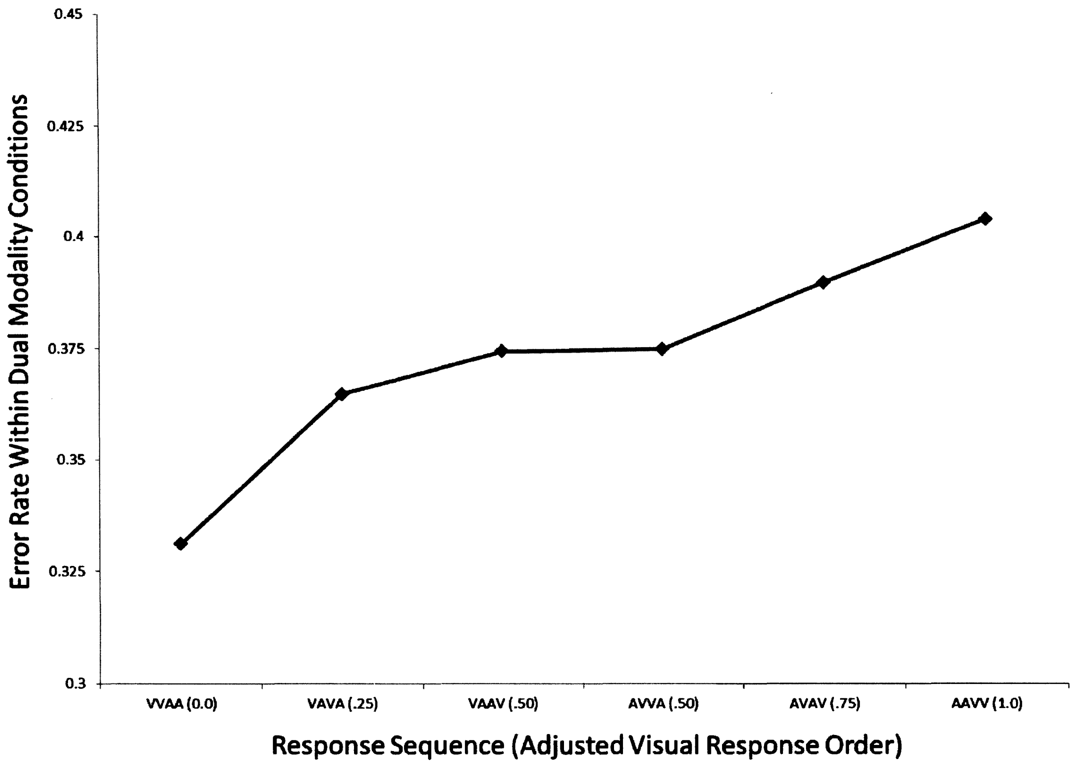

Figure 40. The mean error from all six possible response sequences.

There is evidence that performance in Experiment 3 may have been better if a response order had been forced rather than spontaneous. In Glaser (2008) it was demonstrated that participants are able to recall items from within a category (e.g., "tools") faster if they are forced to move between categories rather than able to choose themselves. This loss in performance was termed an "Executive-Decision Cost." This cost was posited to share the same resource used when engaging working memory. Though not tested in Experiment 3, there remains a possibility that a similar cost globally affected recall performance. Results are evaluated using the proposed framework described after Experiment 4.

Experiment 4

In Experiment 4 this investigation was extended to decision making tasks. Of particular interest was whether findings from the first three experiments generalize to decision making. Addressing decision making is important because in the real world people are rarely called upon only to remember information but rather to use the information to make a decision and take action. Experiment 4 addressed rule-based decisions. In rule-based decisions, the primary decision is based on whether circumstances meet the conditions for one of a number of pre-defined choices. This type of decision making was a focus because it is amenable to study in a laboratory setting. In addition, it is representative of many kinds of decisions professionals have to make. And lastly, decision correctness is objectively defined.

A classic example of a rule-based decision is when a pilot must decide to continue (response 1) or not continue (response 2) on to a runway. To address empirically how the ordering of visual and auditory information affects decision making, participants were asked to complete a task similar to that of Experiment 3. As in Experiment 3, participants were told that they would be monitoring plane and pilot information. The primary difference in Experiment 4 is that they would be asked to make one of three rule-based decisions rather than recall the events. Also, since the rule set was designed be taught in approximately 10 minutes, fewer dimensions and levels (and thus combinations) were presented.

Method

Participants

Thirty undergraduates (17 females) participated as research subjects. The ages ranged from 18-21 with an average of 19.2. All participants participated as one way to fulfill a class requirement.

Monitoring Task

The monitoring task was similar to the task used in Experiment 3. Participants were instructed that they would be monitoring spy information about planes traveling through a restricted area along with and images of pilots and their encrypted jacket symbols. A critical difference in Experiment 4, however, was that after the presentation of plane and pilot information, participants were instructed to use the presented information to make one of three decisions. The three possible decisions were "Ignore", "Report", and "Evacuate." Ignore was described as a decision in response to a harmless situation. Report was described as a decision in response to a potentially dangerous situation. "Evacuate" was described as a response to a scenario where the threat was imminent.

Decisions Making Rules

The rules for making the decisions were as follows: Participants were to choose "Report" if the plane(s) was/were south or traveling south or if the number of planes was/were two AND if the plane is piloted by pilot 2 or showing emblem 2 or showing flag 2.

Scenarios where on should decide on "Report" were as follows …

If the (1) plane and the (2) pilot criteria were both satisfied:

(1) Plane criterion for a warning: range was south or trajectory was southern or numerosity was 2.

(2) Pilot criterion for a warning: face was Pilot 2, or emblem was emblem 2, or flag was flag 2.

Participants are to choose Ignore if Report is not satisfied AND the plane(s) is north or traveling north or if the number of planes is one AND if the plane is piloted by pilot 1 or showing emblem 1 or showing flag 1. In scripting language the Ignore rule would be expressed as …

Scenarios where on should decide on "Ignore" were as follows …

If the (1) plane and the (2) pilot criteria are both satisfied and the criteria for (3) Report are not met:

(1) Plane criterion for a warning: range was north or trajectory was northern or numerosity was 1.

(2) Pilot criterion for a warning: face was Pilot 1, or emblem was emblem 1, or flag was flag 1.

(3) Report criterion was not met.

And lastly, participants are not to choose "Evacuate" if both the (1) Report and (2) Ignore criteria are not met:

(1) Report criteria not met

(2) Ignore criteria not met

Table 3 lists all presentation configurations and the appropriate decision. Note that participants learn the response rules, not all the possible configurations separately.

Table 3. A list of all the presentation configurations and the relative correct decision.

Plane Information			Pilot Information			
Range	Trajectory	Numerosity	Pilot	Emblem	Flag	Decision
North	Northern	1	1	1	1	Ignore
North	Northern	1	1	1	2	Ignore
North	Northern	1	1	2	1	Ignore
North	Northern	1	1	2	2	Ignore
North	Northern	1	2	1	1	Ignore
North	Northern	1	2	1	2	Ignore
North	Northern	1	2	2	1	Ignore
North	Northern	1	2	2	2	Evacuate
North	Northern	2	1	1	1	Ignore
North	Northern	2	1	1	2	Report
North	Northern	2	1	2	1	Report
North	Northern	2	1	2	2	Report
North	Northern	2	2	1	1	Report
North	Northern	2	2	1	2	Report
North	Northern	2	2	2	1	Report
North	Northern	2	2	2	2	Report
North	Southern	1	1	1	1	Ignore
North	Southern	1	1	1	2	Report
North	Southern	1	1	2	1	Report
North	Southern	1	1	2	2	Report
North	Southern	1	2	1	1	Report
North	Southern	1	2	1	2	Report
North	Southern	1	2	2	1	Report
North	Southern	1	2	2	2	Report
North	Southern	2	1	1	1	Ignore
North	Southern	2	1	1	2	Report
North	Southern	2	1	2	1	Report
North	Southern	2	1	2	2	Report
North	Southern	2	2	1	1	Report
North	Southern	2	2	1	2	Report
North	Southern	2	2	2	1	Report
North	Southern	2	2	2	2	Report
South	Northern	1	1	1	1	Ignore
South	Northern	1	1	1	2	Report
South	Northern	1	1	2	1	Report
South	Northern	1	1	2	2	Report

Plane Information			Pilot Information			
Range	Trajectory	Numerosity	Pilot	Emblem	Flag	Decision
South	Northern	1	2	1	1	Report
South	Northern	1	2	1	2	Report
South	Northern	1	2	2	1	Report
South	Northern	1	2	2	2	Report
South	Northern	2	1	1	1	Ignore
South	Northern	2	1	1	2	Report
South	Northern	2	1	2	1	Report
South	Northern	2	1	2	2	Report
South	Northern	2	2	1	1	Report
South	Northern	2	2	1	2	Report
South	Northern	2	2	2	1	Report
South	Northern	2	2	2	2	Report
South	Southern	1	1	1	1	Ignore
South	Southern	1	1	1	2	Report
South	Southern	1	1	2	1	Report
South	Southern	1	1	2	2	Report
South	Southern	1	2	1	1	Report
South	Southern	1	2	1	2	Report
South	Southern	1	2	2	1	Report
South	Southern	1	2	2	2	Report
South	Southern	2	1	1	1	Evacuate
South	Southern	2	1	1	2	Report
South	Southern	2	1	2	1	Report
South	Southern	2	1	2	2	Report
South	Southern	2	2	1	1	Report
South	Southern	2	2	1	2	Report
South	Southern	2	2	2	1	Report
South	Southern	2	2	2	2	Report

Materials

The stimuli and display method were similar to Experiment 3 (Figures 41-43). In Experiment 3 both the pilot and plane information varied on four dimensions. In Experiment 4, plane and pilot information varied on only three dimensions. In another change, each plane and pilot dimension varied on just two. The three plane dimensions were pole (north and south), direction (northern and southern), and numerosity (1 and 2). The three pilot dimensions were face (Pilot 1 and Pilot 2), flag (Flag 1 and Flag 2), and emblem (Emblem 1 and Emblem 2). Visual stimuli and sound were both delivered using the same headphones and computer screen as in Experiment 2.

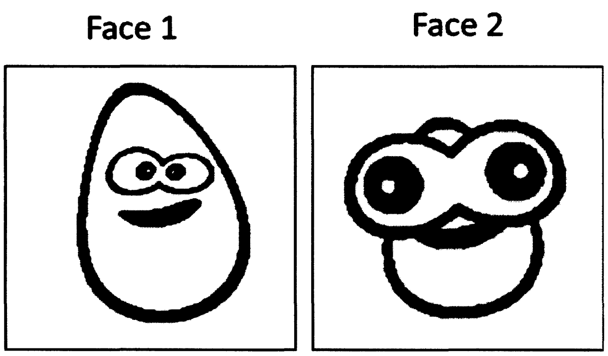

Figure 41. The two face images

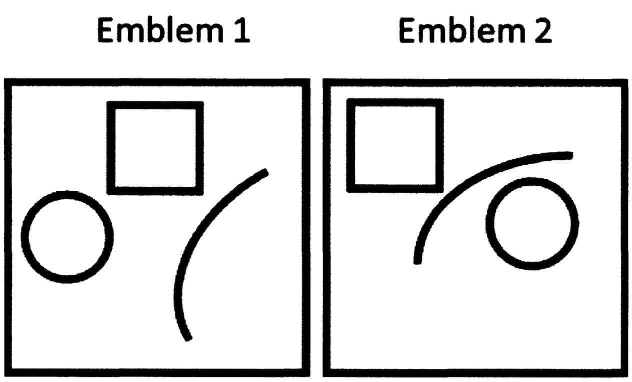

Figure 42. The two pilot emblem pictures.

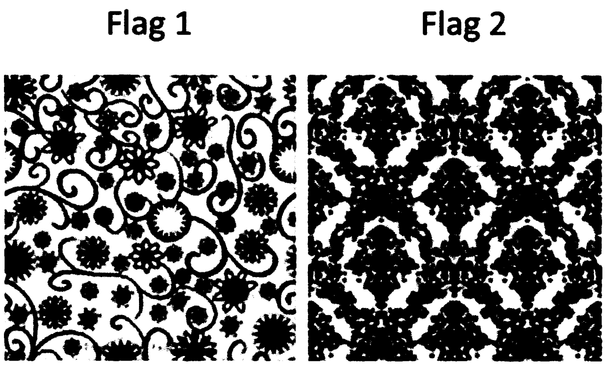

Figure 43. The two flag images

Design

The experimental design was a Plane (2) x Presentation Order (3) factorial design with repeated measures on both factors. The two levels of plane modality were (1) visual planes and (2) auditory planes. The three levels of presentation order were (1) simultaneous, plane-first (plane → pilot), and pilot-first (pilot → plane). In each of the six conditions "Report", "Ignore", and "Evacuate" was each the correct decision eight times.

Training

Training began with a discussion of the monitoring task. During this discussion the participant was told that he or she would be inspecting spy information and making decisions based on the spy information. Decision making training was separated into four steps: (1) Decision rule training, (2) plane mapping, (3) pilot mapping, and (4) practice trials.

Decision Rule Training: The rules were described with similar wording as written in the "Decision Making Rules" section. Participants were encouraged to ask questions throughout the explanation. After the rules were explained the experimenter covered a few written scenarios presented on the computer screen (Table 4) where the participant was asked to identify the correct decision and describe the reasoning behind the decision. The experimenter would correct the participant when mistakes were made.

Table 4. An Example of information scenarios presented on a computer screen.

Plane Information			Pilot Information			
Location	Direction	Number	Pilot	Emblem	Flag	Decision
South	Northern	1	1	2	2	
North	Northern	1	2	2	2	
North	Northern	1	2	1	2	

Once the experimenter felt the participant understood the decision making rules, the participant was given a rule-knowledge assessment with 10 information scenarios. Table 5 represents a sample decision-rule assessment. A decision-rule assessment included plane and pilot information scenarios presented in a table-like format. Participants were asked to complete the row by writing the correct decision. A participant continued receiving instruction and completing assessments until an assessment was completed with 100% accuracy. Each test included an information scenario for each of the three decisions.

Table 5. An Example of a ten-item assessment designed to judge the participants' understanding of the decision-rules.

Plane Information			Pilot Information			
Location	Direction	Number	Pilot	Emblem	Flag	Decision
South	Northern	1	1	2	2	
North	Northern	1	2	2	2	
North	Northern	1	2	1	2	
South	Northern	2	1	2	2	
South	Southern	1	1	1	1	
South	Southern	2	1	1	1	
South	Southern	2	2	1	1	

Plane Mapping: Next, a participant was given training on how to map audio planes to plane range, trajectory, and numerosity. Participants were told that a high pitches corresponded to a north range and the low pitches corresponded to a south range. Next participants were told that pitches that get higher correspond to a northern trajectory and that pitches that get lower correspond to a southern trajectory. Participants were then instructed that a single beep corresponds to a single plane and two beeps correspond to two planes. Participants would then complete an audio plane assessment where they were presented with an audio plane example would identify the three dimensions. A participant would continue this assessment stage until correctly identifying the three plane dimensions in 10 contiguous audio plane examples. The same training and assessment procedure was completed for visual-planes.

Pilot Mapping: Next, participants were shown how to map a type to the pilot images. Participants were shown pilot faces, emblems, and flags on a screen with the type printed to the left. Participants were then given a pilot assessment in which they were asked to identify the type of each of the three pictures. The participant would continue the assessment until he or she correctly matched the three pictures with its corresponding number in 10 contiguous trials.

Practice Trials: Next, the participant completed 10 practice trials while being watched by an experimenter. If a trial was answered incorrectly the participant was asked to explain why he chose the given answer. If the error was based on the forgetting of decision rules, then the rules were reviewed and a new practice trial began. If the error was based on a forgetting of how to apply type to a dimension, types were reviewed. As with the previous training assessments, participants continued practice trials until ten contiguous trials were completed without a rule-based error.

Procedure

An experimental session began with a training session which lasted approximately 30 minutes. After the participant successfully completed the training session, twenty-four trials for each of the six conditions (144 trials) were completed in a random order. The participant was debriefed after the conclusion of the last monitoring task.

Results & Discussion

Figure 44 represents the decision error rate from all six conditions. The error rate was calculated for each participant by identifying the proportion of trials within each condition that resulted in an incorrect decision. The fewest errors were made in the pilot-first audio-planes-second condition.

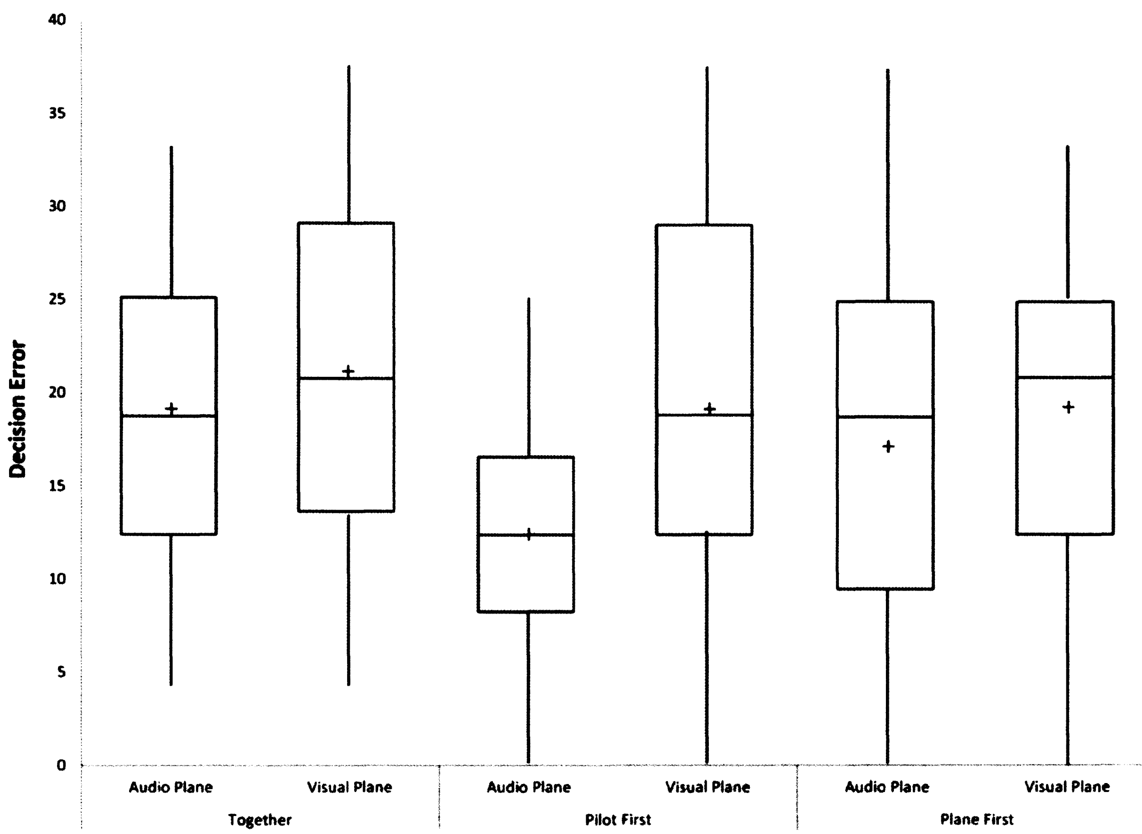

Figure 44. The mean decision error rate across all six presentation conditions.

To test differences among the means of all six conditions for significance, fifteen pair-wise comparisons were made. After the Bonferroni adjustment the new alpha for rejecting the null hypothesis (i.e., that two conditions have equal means) was .0034. Table 5 is a matrix of all condition pairings and their probability values. The pilot-first/audio-planes condition had the fewest errors of all the conditions. The p-value for this condition met the Bonferroni corrected value for all comparisons except for the plane-first/audio-plane condition. This comparison was significant at the conventional .05 level. Overall, there is strong empirical support that the pilot-first audio-plane condition is the best for decision making.

Table 5. The mean proportion error differences between all six conditions. * denotes conventional level of significance. ł denotes significance after the Bonforroni adjustment.

Row-Column	Pilot-First/Audio-Plane	Pilot-First/Visual-Plane	Plane First/Audio-Plane	Plane-First/Visual-Plane	Together/Audio-Plane	Together/Visual-Plane
Pilot-First/Audio-Plane		0.075ł	0.043*	0.076ł	0.067ł	0.099ł
Pilot-First/Visual-Plane	-0.075ł		-0.032	0.001	-0.008	0.024
Plane-First/Audio-Plane	-0.043*	0.032		0.033	0.024	0.056ł
Plane-First/Visual-Plane	-0.076ł	-0.001	-0.033		-0.010	0.022
Together/Audio-Plane	-0.067ł	0.008	-0.024	0.010		0.032
Together/Visual-Plane	-0.099ł	-0.024	-0.056	-0.022	-0.032	

Modality Pairing

Figure 45 represents the error rate difference between the presented-together conditions and presented-separately conditions. The mean error rate in the presented-together conditions was greater than in the presented separately conditions.

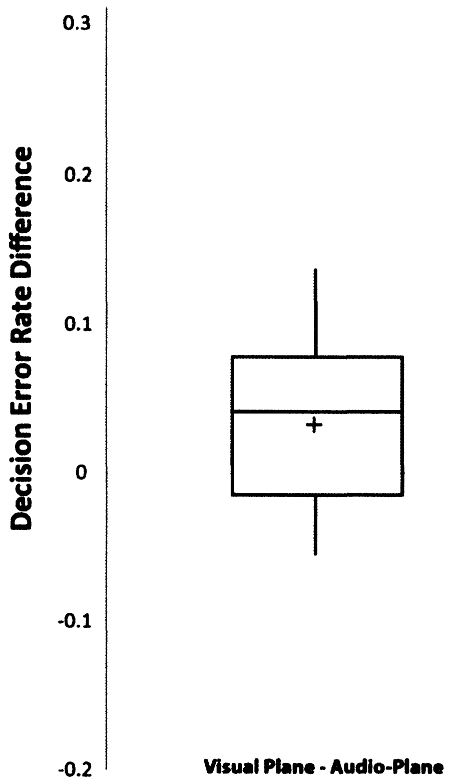

Figure 45. The error rate difference between the visual and audio plane conditions.

To test whether errors were more common when multiple modalities were used, a pair-wise comparison was conducted between the audio-plane and visual-plane mean error rate. A significant difference in error rate between the audio-plane (M = .16) and visual-plane (M = .21) was found (M_{diff} = .05, CI_{95} = [.03, .07]), $t(29)$ = 4.81, p < .001.

Presentation Synchronicity

Figure 46 represents the difference in error rate between the audio-planes and visual-planes conditions. Consistent with MRT, fewer errors were made when multiple modalities were used (audio planes condition). This pattern was consistent across all three presentation order conditions.

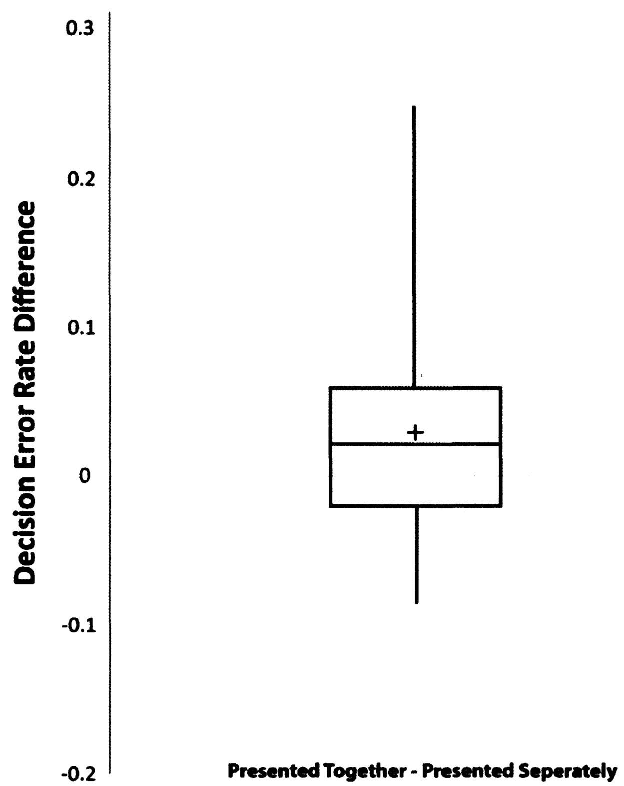

Figure 46. The error rate difference between the two conditions in which plane and pilot information were presented together and the four conditions in which plane and pilot information were presented separately.

To test the effect of information synchronicity, a mean error rate pair-wise comparison was conducted between presented-separately and presented-together conditions. A significant difference in error rate between the presented-together (M = .16) and presented-separately (M = .21) conditions was found (M_{diff} = .05, CI_{95} = [.02, .07]), $t(29)$ = 3.09, p = .012.

Two consistencies were present between the benefits of visual-then-audio presentation and results from the previous three experiments. First is the finding that presenting information separately leads to better performance than presenting information together. Second is the finding that using multiple modalities leads to better performance than the use of a single modality. The finding that performance is enhanced when visual information is presented prior to auditory information is consistent with Experiment 2 but not Experiment 3.

General Discussion

This research was based on the idea that since echoic memory persists for seconds in contrast to iconic memory which persists only for milliseconds, responses to multimodal presentations may be benefited if the visual presentation to response interval is minimized. To a certain degree evidence from these investigations suggests this may be true. There were, however, results that on the surface seemed counter to this fundamental assumption. Specifically, the finding that in some cases elongating the visual stimulus response interval improved performance which was found in Experiments 2 and 4. In the following sections potential contributing perceptual and working memory factors are considered in concert with Experiment 1 and 2 findings to assist in the development of a proposed framework for a parsimonious account of Experiments 1-4 data.

Framework for explaining Experiments 1-4 data

For the purposes of this investigation I posit that performance predictions may be derived using a point system similar to MRT with an additional mechanism for utilizing the Long

Auditory Store. The point value reflects how display and response order affects utilization of the Long Auditory Store and may contribute to perceptual and working memory interference. The purpose of the framework is to provide a parsimonious account from data collected in Experiments 1-4 while being congruous with well documented sensory, perceptual, and working memory phenomena. The framework proposed is specifically designed to be a prediction framework for the kind of tasks evaluated in the following four experiments. As a way to challenge the framework, Experiment 1 and 2 data only are used to elucidate potential perceptual and working memory costs, and sensory memory strategy. Experiments 3 and 4, which utilized a varied stimulus set and required a different kind of responses (recognition and decision making rather than recall) than Experiments 1 and 2, were also evaluated using the framework. Tasks in Experiments 1-4 are similar in that they involve a brief discrete display of information, each are structured such that leveraging the long echoic store is predicted to improve response quality, and each task has a working memory component.

This proposed framework predicts relative performance between conditions by assigning each with a value related to the expected quality of the response. The final value is based on the sum of the expected perception and working memory cost. Conditions with a higher value are predicted to result in more accurate performance than conditions with lower values. The model draws parallels to Wickens' (2002) MRT Model to the extent that the more resource common tasks are completed in close temporal proximity, the higher the expected performance cost. Unlike MRT, the proposed framework gives consideration to how presentation and response orders may lead to cognitive flows that affect utilization of the Long Auditory Store. Before explaining how the values are calculated, consideration is given to cognitive costs that are expected in tasks with multiple sources of information.

Visual vs. Auditory Modality

Results from experiments 1-2 demonstrated that dimensions from visual information were more easily identified than dimensions from auditory information. This result is consistent with findings from auditory research where it was found that is difficult for some to perceive auditory dimension s such as pitch and timbre due to perceptual interactions (Grau & Kelmer 1988; Melara & Marks, 1990).

Sensory Memory Considerations

The Long Auditory Store is expected to hold auditory information in a form that can be accessed for many seconds after the event. Thus, it is expected that long auditory SRIs will be less negatively impactful on overall response quality then long visual SRI. The relatively long persistence of auditory information acts as a robust recovery mechanism for reinforcing auditory working memory processes.

Perceptual/Working Memory Interference Considerations

Humans are more adept at dividing attention between a visual and auditory stimulus than two visual stimuli. Wickens (2002) recounts many examples when this has been true. Perceptual resources are structured such that parallel multi-modal stimuli is less perceptually taxing then parallel single-modal stimulus input. Since these resources are finite, perceiving two things at once will likely result in a greater performance cost than perceiving one thing at a time.

In Wickens and Hollands (2000) several studies are reviewed which suggest a common resource for perception and working memory (Figure 47). Their review found evidence that increases in perceptual-cognitive difficulty often has little effect on the performance of a concurrent task which is centered on response selection and execution. In contrast, performance on concurrent tasks with a working memory component was degraded.

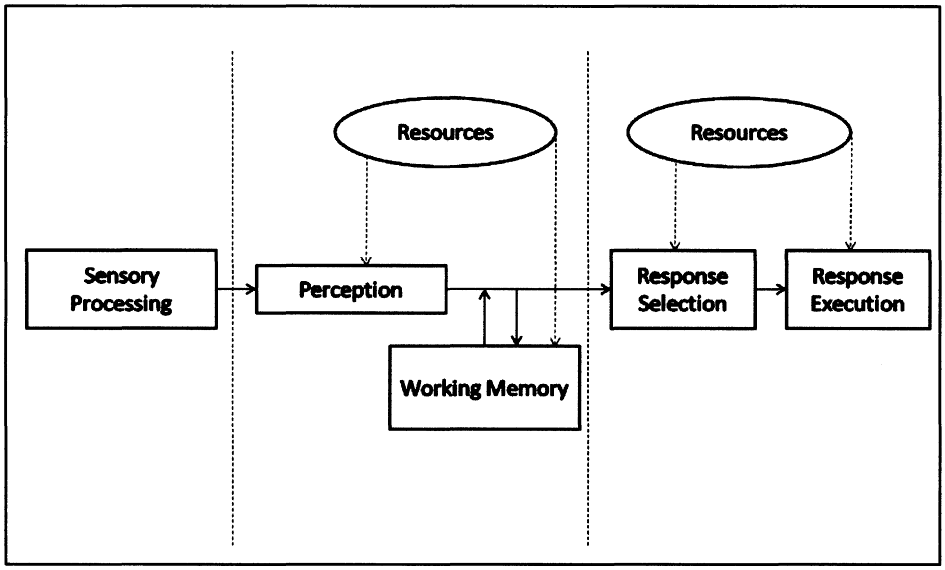

Figure 47. Wickens and Hollands (2000) information processing resource model

This resource model is organized such that increasing the perceptual load will have negative impacts on the working memory task of storing the decoded stimulus attributes. If the simultaneous display of similar stimuli is overtaxing the Perceptual-Working Memory resource than it is reasonable to assume that having participants perceive the two modalities serially would ease the task of perceiving and storing the attributes of each sequence because it promotes a situation in which much of the chunking work for the first sequence may already be completed before the second sequence is presented.

When a participant is presented with the multimodal sequence, he or she must at some point determine the dimensions and keep them available over the short term for the eventual response. In the case of the dot sequence, this short term retention required a form of chunking (likely through sub-vocal rehearsal) because the sensory persistence to the visual sequence is fleeting. Thus some sort of recoding interval may be advantageous if dot information is to be successfully stored and recalled at a later time. Moreover, the period of time in which the dots can be recoded is more restricted than with the auditory sequence because iconic memories are less available after a delay than echoic memories (Sperling, 1960; Conrad, 1964; Baddeley, 1966). Hence, it may be beneficial to have time separating the end of the visual sequence and the start of the auditory response if the auditory response acts as a kind of distracter task. This assertion is based on the assumption that recoding the visual stimuli and recalling auditory sequences will interfere with each other since both are working memory tasks. Results from Experiments 2 and 3 support this assertion. As such, it is believed that these two processes cannot be completed in parallel. Though participants may choose to ignore the auditory response in some cases, prompting participants to recall audio sequence dimensions directly after the visual presentation may to lead to a scenario in which time-sensitive visual recoding processes may be delayed to the point that visual information is no longer readily accessible. If instead visual information is presented prior to auditory information, there would be a time buffer between the end of the visual presentation and the start of the auditory response in which a participant may be encouraged to complete recoding of visual information before shifting to the auditory recall task. As a consequence, the visual chunking process may be more successful.

Framework Calculations

The following modeling approach is semi-qualitative and is designed to for predicting relative accuracy among similar tasks. The framework proposed may provide useful information on how to potentially improve user performance by strategically utilizing the Long Auditory Store or to a person wishing to model UTC production logic.

The model yield is the sum of the expected perceptual and working memory cost. Perception cost is a proxy for the quality of the initial impression of the stimulus. The working memory cost value is a reflection of how effectively the participant is expected to be able to store and sustain information from the initial impression. A value of 0 is the lowest possible value. A value of 0 is yielded if no perceptual or working memory costs are expected. If perceptual and working memory costs are expected, values of greater than 0 were assigned. The response cost is the sum of the perception and working memory cost. The higher the response cost value the lesser the predicted response quality.

-Framework Prediction-

Response_Cost = Perception_Cost + Working_Memory_Cost

The absolute cost of values is not considered to be predictive of actual performance. Rather, values applied to costs are posited to be representative of the relative expected cost to performance accuracy. Costs are increased at multiples of .25. Since the tasks evaluated are expected to be most benefited if the Long Auditory Store is utilized strategically, the highest cost in the following tasks models occurs when the stimulus to response interval for visual information is long.

Tasks from Experiments 1-2 typically involve 2 stimuli, perception cost is the sum of perception cost of stimulus 1 (e.g., visual) and stimulus 2 (e.g., auditory). Similarly, working memory cost is the sum cost of stimulus 1 and stimulus 2. In cases when only one stimulus is presented, the second stimulus' cost values will be set to 0.

-Perception and Working Memory Cost-

$$Perception\_Cost = Perception\_Cost\_S1 + Perception\_Cost\_S2$$

$$Working\_Memory\_Cost = Working\_Memory\_Cost\_S1 + Working\_Memory\_Cost\_S2$$

The quality of a perception is expected to be the combination of how effectively a participant is expected to be able to decode dimensions from a stimulus (baseline cost) along with potential interference from the additional stimulus.

$$Perception\_Cost\_S(i) = Baseline\_Cost\_S(i) + Interference\_Cost\_S(i)$$

Baseline differences were observed in Experiments 1 and 2. Specifically it was observed that when a visual stimulus was presented alone, recall of the visual dimensions was near perfect. In contrast recall of auditory dimensions was substantially less accurate. Hence, a baseline visual cost for these tasks was 0 while the baseline auditory cost was set to 1. The modeler may use discretion to adjust the baseline value based on the difficulty of the task.

-Perceptual Baseline Cost-

$$Perceptual\_Baseline\_Cost\_Auditory = 1$$

$$Perceptual\_Baseline\_Cost\_Visual = 0$$

Interference between multiple stimuli reflects the degree to which the presentation order of the two stimuli is expected to interfere with each other. Perceptual interference between stimuli is increased only when stimuli are presented simultaneously. The interference cost is conditionally augmented depending on the physical similarity between stimulus 1 and 2. If S1 and S2 are both visual and presented simultaneously then the interference cost for both stimuli is 2. Instead if stimulus 1 and 2 are auditory and visual respectively, then the interference cost for each is .25. This difference in interference cost is consistent with MRT which predicts that, all else being equal, Visual/Visual trials will lead to poorer performance than Visual/Auditory trials.

-Perceptual Interference Cost Logic-

If S1 and S2 are presented simultaneously

If S1 and S2 are of a different modality

Interference_Cost_S1 = .25

Interference_Cost_S2 = .25

Else

Interference_Cost_S1 = 1

Interference_Cost_S2 = 1

Else

Interference_Cost_S1 = 0

Interference_Cost_S2 = 0

Working memory cost is expected to be affected by the stimulus response interval (SRI) and the level of working memory chunking interference from the competing stimulus. Both of these factors are believed to influence how well the information from a presentation can be maintained and utilized when needed.

Working_Memory_Cost_S(i) = SRI_Cost_S(i) + Interference_Cost_S(i)

For information to be successfully recalled, dimensions from each stimulus must be successfully identified and stored in working memory. It is expected that such storing processes will be most effective if performed during or in close temporal proximity to the stimulus presentation. When S1 and S2 are presented simultaneously, chunking processes may be more temporally removed from the initial presentation than if S1 and S2 were presented separately. Since auditory sensory information is persistent, this potential delay or interference cost applies only to visual information. Though working memory interference is expected to be highest when multiple stimuli are presented at the same time, interference is expected to a certain extent within a modality since recoding a stimulus requires multiple chunking processes. Hence, the working memory interference cost for visual items is 1 for when visual and auditory items are presented together and .5 when not presented together. Due to the persistence of the Long Auditory Store the auditory chunking processes are not expected to be negatively affected by multiple chunking processes. If initial chunking is interfered with, the participant may use the Long Auditory Store to re-chunk information at a later time. Thus, working memory interference costs for auditory items is set to 0.

-Working Memory Interference Cost Logic-

If S1 and S2 are presented at the same time

Interference_Cost_Visual = 1

Interference_Cost_Auditory = 0

Else

Interference_Cost_Visual = .5

Interference_Cost_Auditory = 0

Since auditory sensory information is far more persistent than visual sensory information, it is predicted that long auditory SRIs will not incur as large an SRI cost as a long visual SRI because the Long Auditory Store provides a robust recovery mechanism. Long SRIs occur when a response to a given stimulus occurs after the other stimulus. A long visual SRI cost is set to 3 whereas a long auditory SRI is set to 1. A short visual SRI, which is expected to result in a much smaller cost, is set to .5 and a short auditory SRI is set to 0.

In Experiments 2, presenting visual information prior to auditory information was shown to reduce the cost of a long visual SRI. A possible explanation for this somewhat paradoxical finding is that a visual-then-auditory presentation affords a cognitive flow that promotes chunking of visual information prior to chunking of auditory information. If visual chunking is completed prior to the start of auditory chunking, the quality of the visual chunk is likely greater due to less auditory chunking interference. As a result, a long SRI is expected to be less susceptible to information loss over time. In cases when visual information is presented prior to auditory information and there is a long visual SRI, the long visual SRI cost is reduced to 2 (from 3).

-Working Memory SRI Cost Logic-

If Visual is presented together or after Auditory

If Auditory response is immediately after Visual presentation

SRI_Cost_Visual = 3

SRI_Cost_Auditory = 0

Else

SRI_Cost_Visual = .5

SRI_Cost_Auditory = 1

Else

If Auditory response is immediately after Visual presentation

SRI_Cost_Visual = 2

SRI_Cost_Auditory = 0

Else

SRI_Cost_Visual = .5

SRI_Cost_Auditory = 1

Predicting Experiment 1 – 4 Data

To illustrate how the model value is generated, consider Experiment 1 where a participant is presented with visual and auditory information simultaneously. The participant's task is to recall the five dimensions of each information source (e.g., sound/image quality, pitch/image direction, pitch/image range). The manipulation is the order in which the information is recalled. Figure 48 is a depiction of a participant responding to auditory information prior to visual information. Figure 49 is a depiction of a participant responding to visual information before audio. The black boxes are the stimuli presented. The red boxes to the right are the responses to the stimulus. The green boxes are the visual and auditory perception from which long term memory is used (orange line) to chunk the dimensions into working memory (blue box). The Long Auditory Store acts as an extender to auditory perception which allows robust access to auditory information after the auditory presentation.

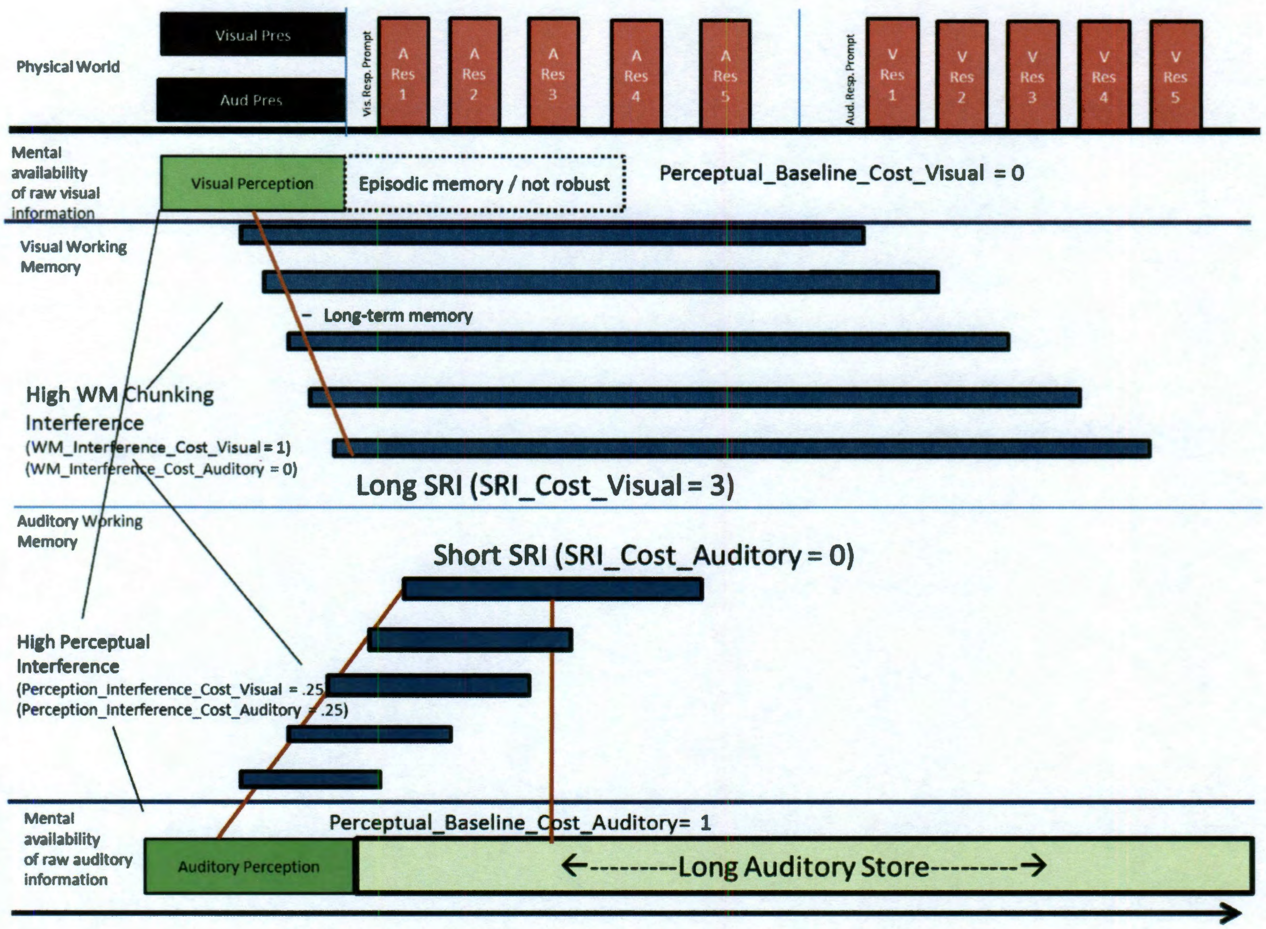

Figure 48. A cognitive flow model of a task with Auditory and Visual information presented together with as auditory response prior to a visual response.

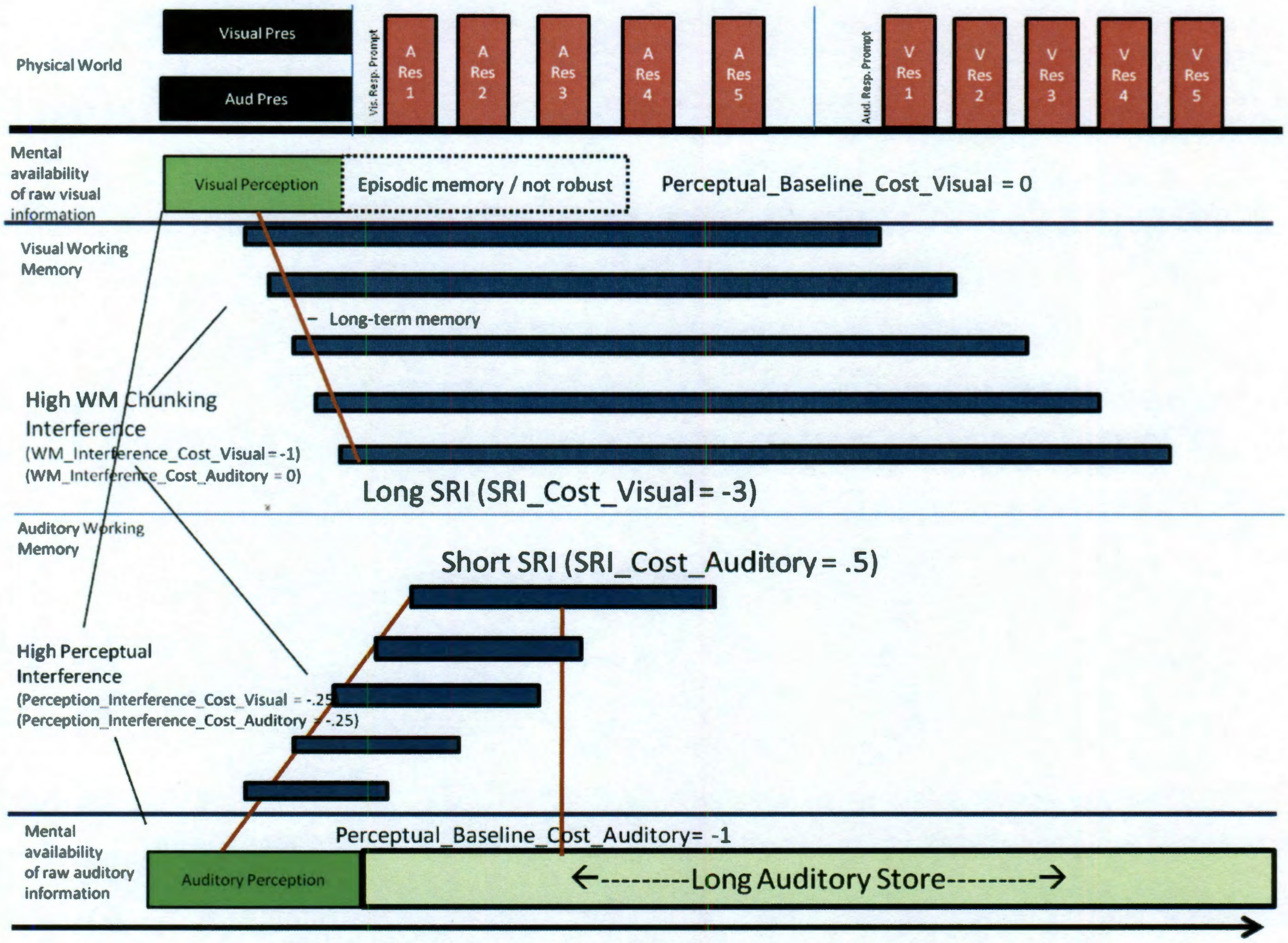

Figure 49. A cognitive flow model of a task with Auditory and Visual information presented together with a visual response prior to an auditory response.

In each figure the costs are illustrated. In the visual then auditory response condition, both visual and auditory stimuli are presented. The perceptual baseline cost is 0 for visual and 1 for auditory. Visual and Auditory perception occur simultaneously incurring a perceptual interference cost of .25 for both the visual and auditory domains. Since visual and auditory information are presented simultaneously, working memory processes are expected to occur in close temporal proximity. This is expected to be particularly disruptive to the storing of visual chunks because there is no robust recovery mechanism. As a result the visual working memory interference cost is 1. Since similar working memory processes do have a robust recovery mechanism, the auditory working memory interference cost is set to 0. Since visual information is recalled first the visual SRI cost is short and set to.5. The auditory SRI is long it receives an SRI cost of 1. In the auditory then visual response condition the only cost difference compared to the visual then auditory response condition is the SRI cost. Since visual information is responded to second the visual SRI is long and has a cost of 3 (instead of 1) and since the auditory SRI is short the cost is 0 (instead of 1). Table 6 outlines the two framework predictions. Since the Visual then Audio condition yielded the highest value, this condition is predicted to result in the most accurate responses.

Table 6. The predicted response cost and accuracy for each condition in Experiment 1.

	Perceptual Cost				Working Memory Cost				Response Cost (rank)	Actual Accuracy (rank)
	Baseline Cost		Interference Cost		SRI Cost		Interference Cost			
	V	A	V	A	V	A	V	A		
Visual Baseline	0	-	0	-	.5	-	0	-	.5 (1)	.88 (1)
Auditory Baseline	-	1	-	0	-	0	-	0	1 (2)	.76 (2)
Visual then Auditory Response	0	1	.25	.25	.5	1	1	0	4 (3)	.55 (1)
Auditory then Visual Response	0	1	.25	.25	3	0	1	0	5.5 (4)	.44 (4)

Using the predicted performance cost value, 97% of the performance variance was accounted for (Figure 50). For the framework to be considered parsimonious, however, a separate performance data set from different participants performing modified tasks should also be explained.

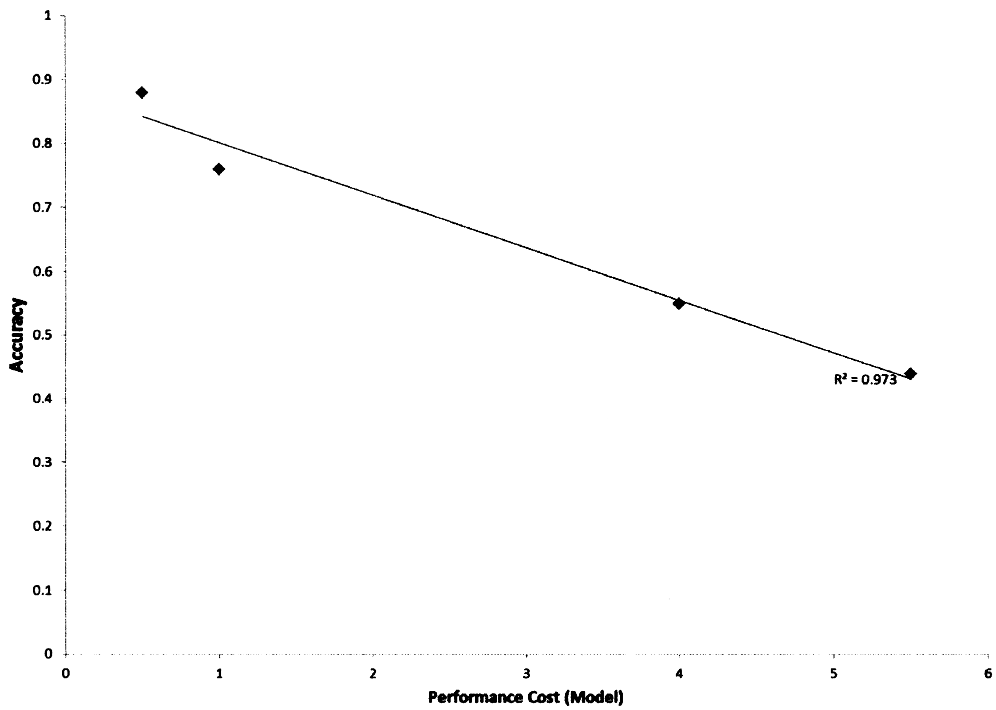

Figure 50. The correlation between performance cost and actual accuracy.

Experiment 2

Figures 51-56 are the proposed timelines of the six dual task conditions. On each figure perceptual and working memory costs are identified. Table 7 outlines the two model predictions for the six conditions in Experiment 2.

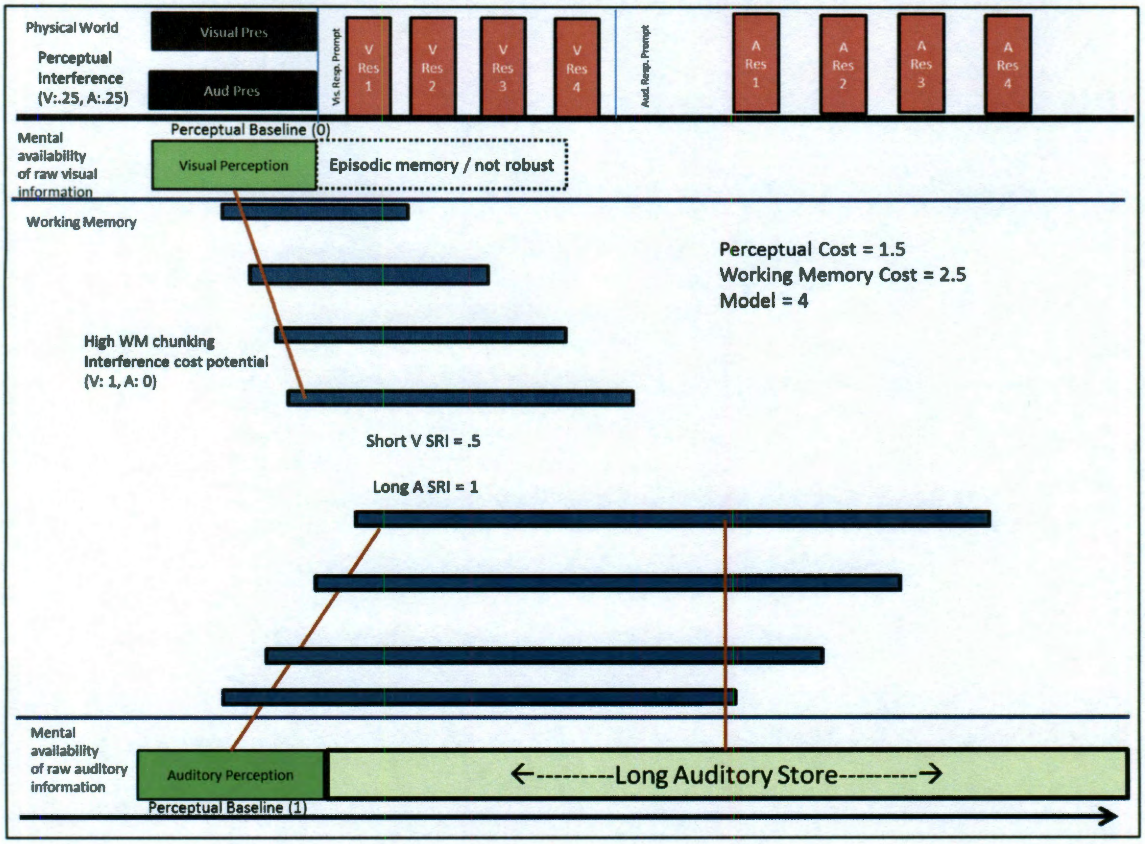

Figure 51. The cognitive flow model of the Visual and Auditory presented together, Visual then Auditory response condition. The model prediction is 4.

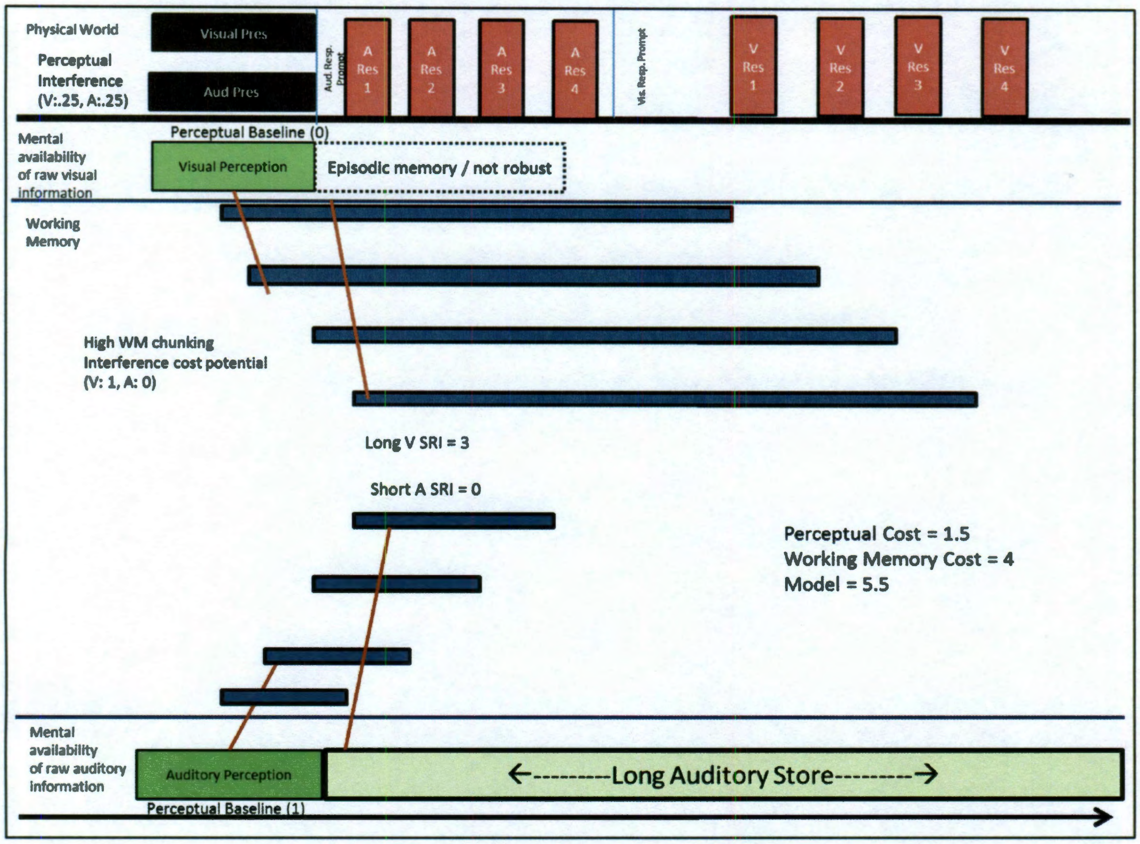

Figure 52. The cognitive flow model of the Visual and Auditory presented together, Auditory then Visual response condition. The model prediction is 5.5.

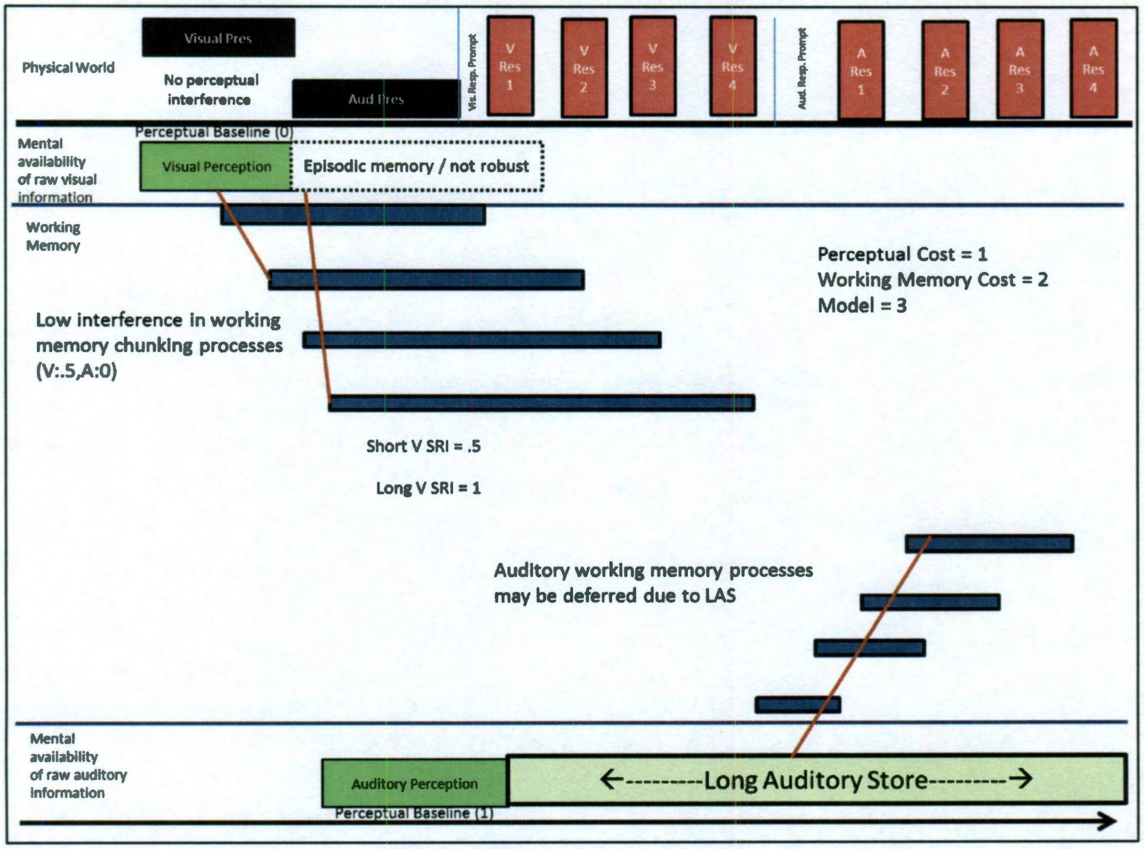

Figure 53. The cognitive flow model of the Visual then Auditory presentation, Visual then Auditory response condition. The model prediction is 3.

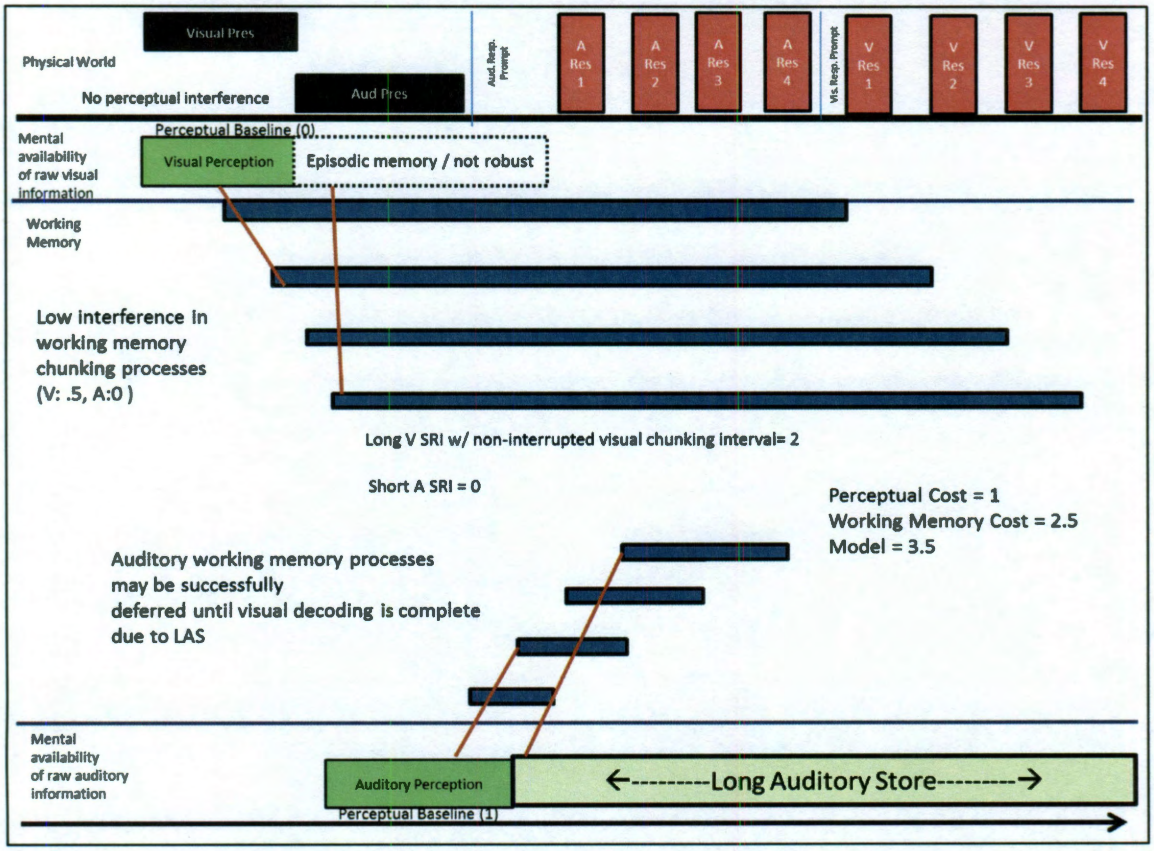

Figure 54. The cognitive flow model of the Visual then Auditory presentation, Auditory then Visual response condition. The model prediction is 3.5.

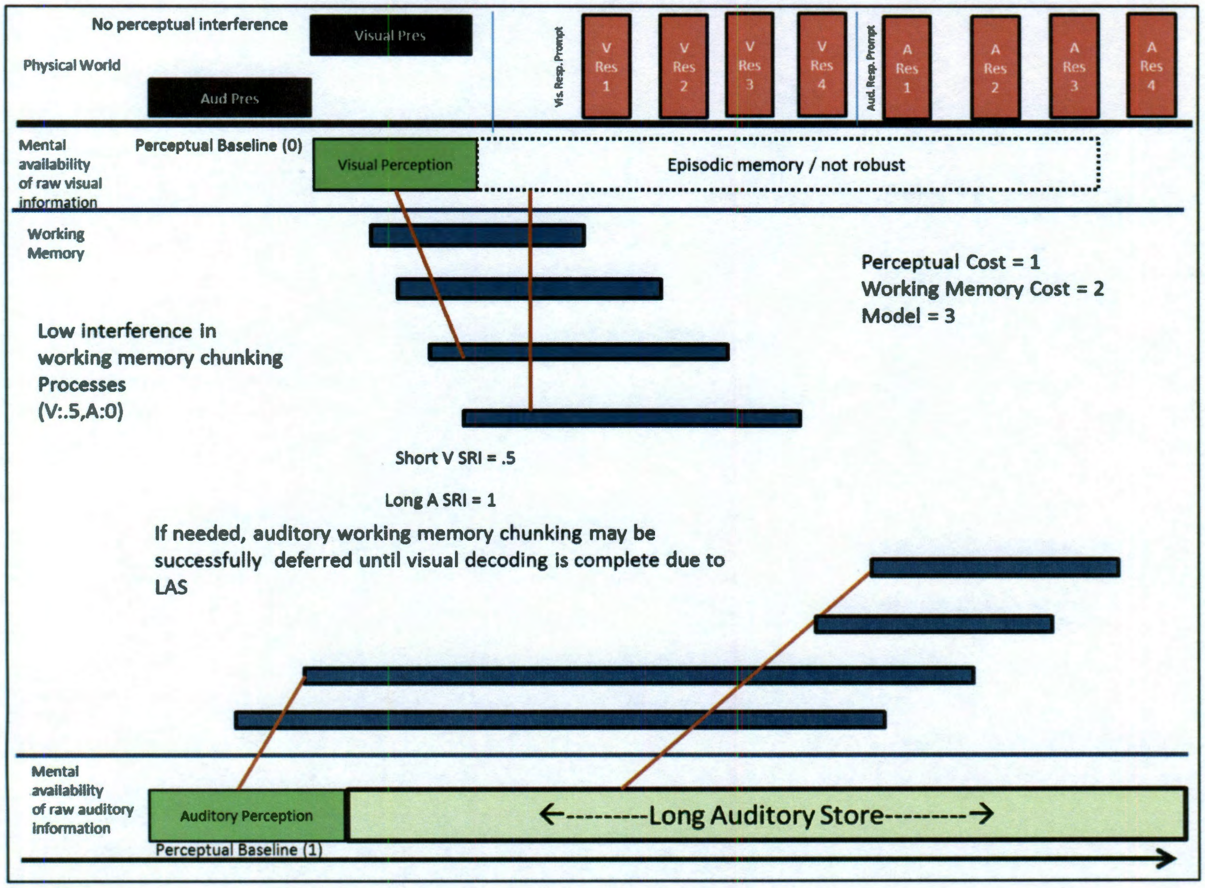

Figure 55. The cognitive flow model of the Auditory then Visual presentation, Visual then Auditory response condition. The model prediction is 3.0.

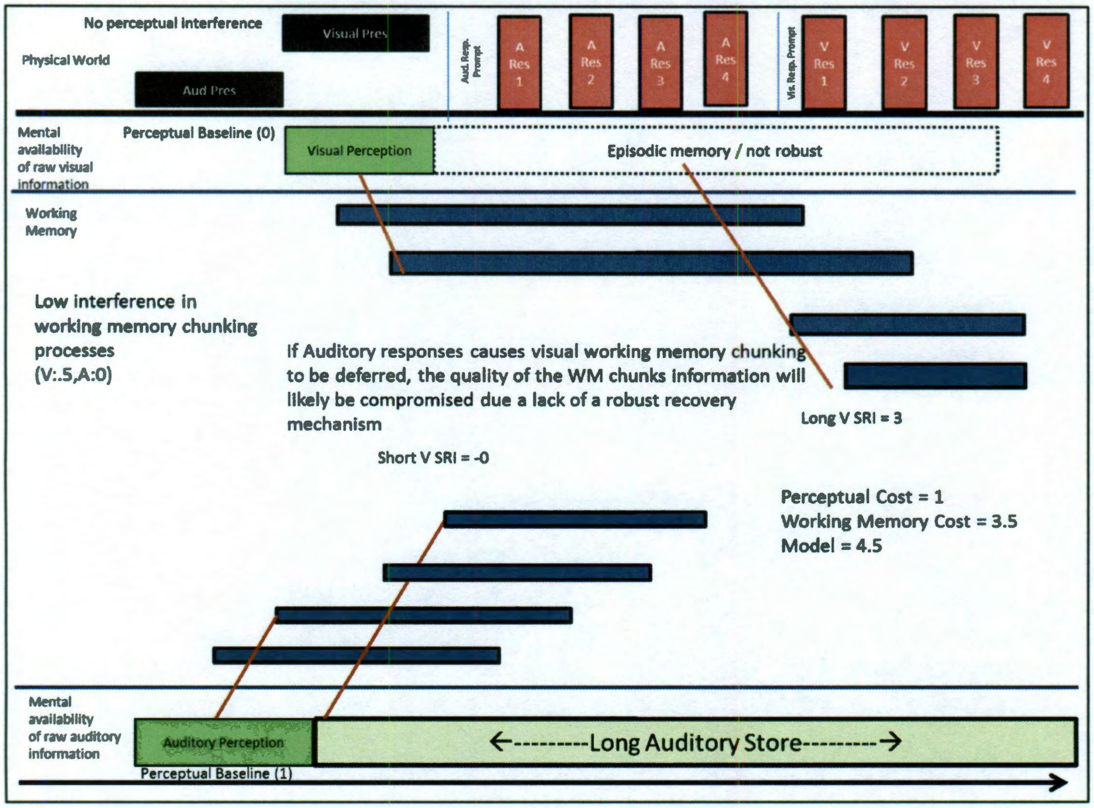

Figure 56. The cognitive flow model of the Auditory then Visual presentation, Auditory then Visual response condition. The model prediction is 4.5.

Table 7. The predicted cost and accuracy for each condition in Experiment 2

	Perceptual Cost				Working Memory Cost				Response	Actual
	Baseline Cost		Interference Cost		SRI Cost		Interference Cost		Cost (rank)	Accuracy (rank)
	V	A	V	A	V	A	V	A		
Visual Baseline	0	-	0	-	.5	-	0	-	.5 (1)	.90(1)
Auditory Baseline	-	1	-	0	-	.5	-	0	1 (2)	.80(2)
(VA) /A->V	0	1	.25	.25	3	0	1	0	-5.5 (8)	.48(8)
(VA)/ V->A	0	1	.25	.25	.5	1	1	0	-4.0(6)	.60(6)
V->A/A->V	0	1	0	0	2	0	.5	0	3.5(5)	.61(5)
V->A/V->A	0	1	0	0	.5	1	.5	0	3.0(3)	.65(3)
A->V/A->V	0	1	0	0	3	0	.5	0	-4.5(7)	.53 (7)
A->V/V->A	0	1	0	0	.5	1	.5	0	3.0(3)	.64 (4)

Relative performance rank between conditions was predicted closely using the proposed framework. Non-significant differences in accuracy such as the difference between (VA)/V->A and V->A/A->V as well as V->A/V->A and A->V/V->A yielded model prediction differences of .5 or less. If the model predicted a difference of greater than .5, significant differences in actual accuracy was found. Most importantly, the two multimodal conditions with the greatest accuracy both received the highest model score. Predicted performance cost accounted for 97% of the performance variance (Figure 57). For Experiments 1& 2 the framework is able to parsimoniously predict relative performance in which different participants, stimulus sets, and conditions were evaluated.

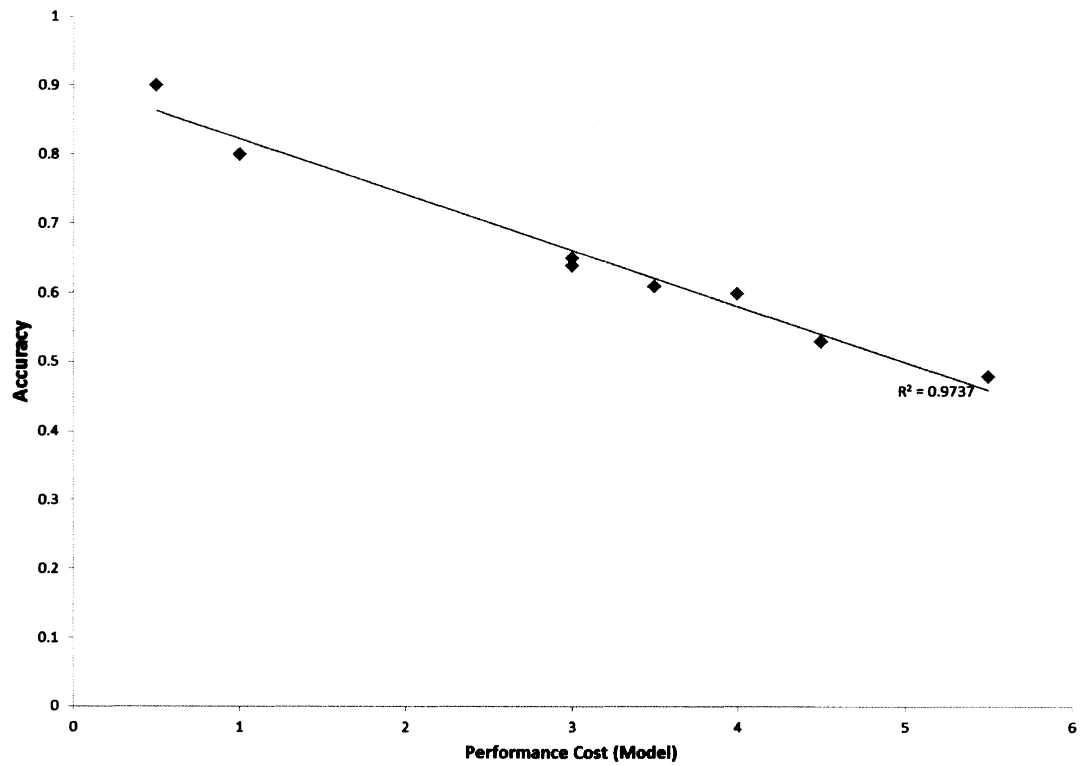

Figure 57. The correlation between performance cost and actual accuracy for Experiment 2.

Experiment 3

Results from Experiment 1-2 in part contributed to the development of the proposed framework while Experiments 3-4 did not. Experiments 3 and 4 differ from the first two experiments in terms of stimuli utilized, modality parings, realism of the scenario, and the response requirements. It is believed that the validity of the framework can be bolstered if results from Experiments 3 and 4 can be predicted by the same framework as Experiments 1 and 2.

In Experiment 3 it was demonstrated that participants tended to respond to the last presented item first. This response pattern will be assumed in the following condition models. Figures 58-61 are illustrations of these four conditions. In Experiment 3 the auditory perceptual baseline cost was reduced to .5. This reduction in perceptual cost was chosen due to the reduction of to-be-remembered auditory dimensions (from Experiments 1 and 2). In addition, for single modality trials, visual working memory interference cost for both kinds of visual working memory chunks was increased by .25 due to the near doubling of chunking activity in visual working memory. Since there are two kinds of visual information chunks in the single modality trials, the total cost increase to visual working memory items is .5.

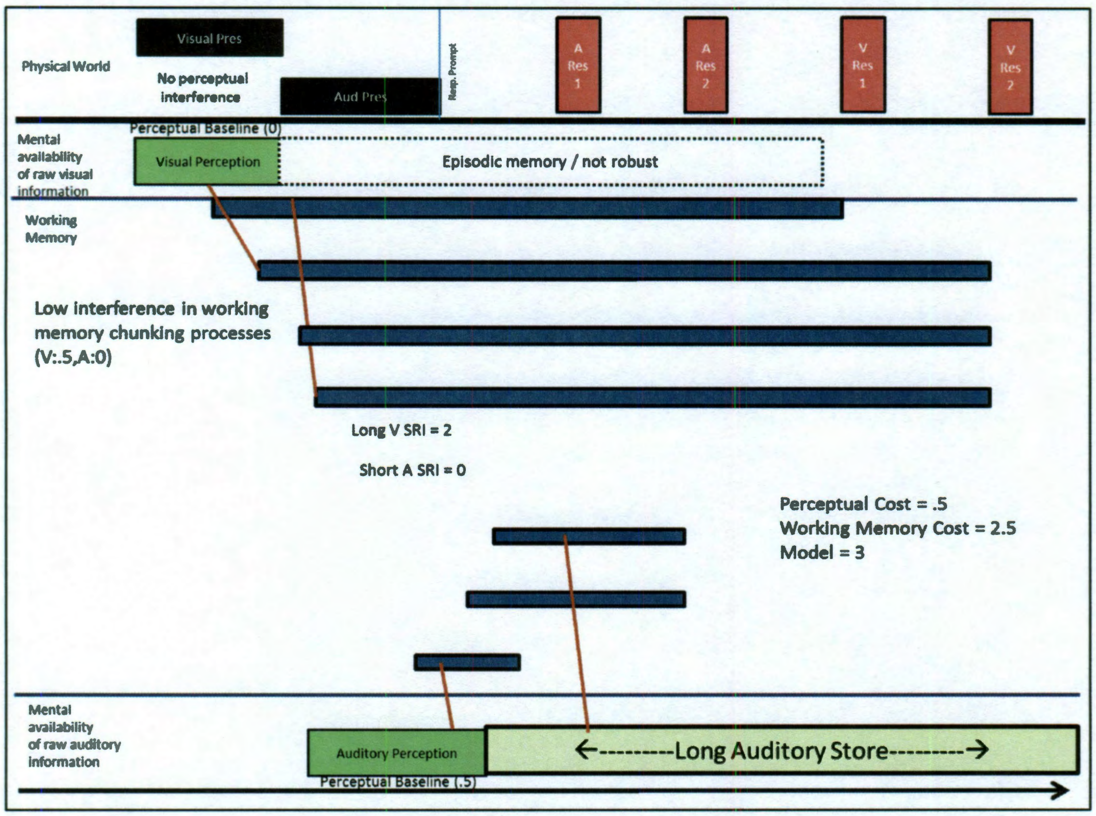

Figure 58. Tthe cognitive flow model of the Visual then Auditory presentation (auditory planes). The model prediction is 2.5.

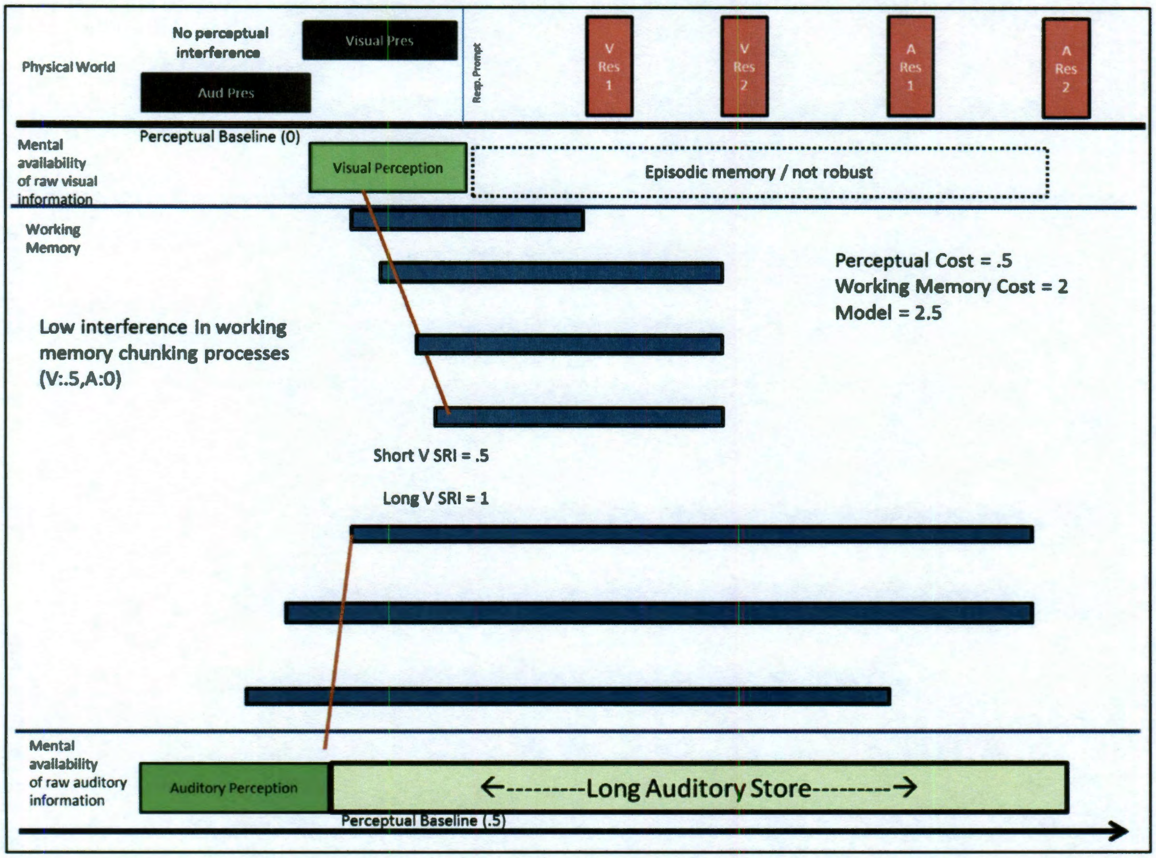

Figure 59. The cognitive flow model of the Auditory then Visual presentation (auditory planes). The model prediction is 2.5.

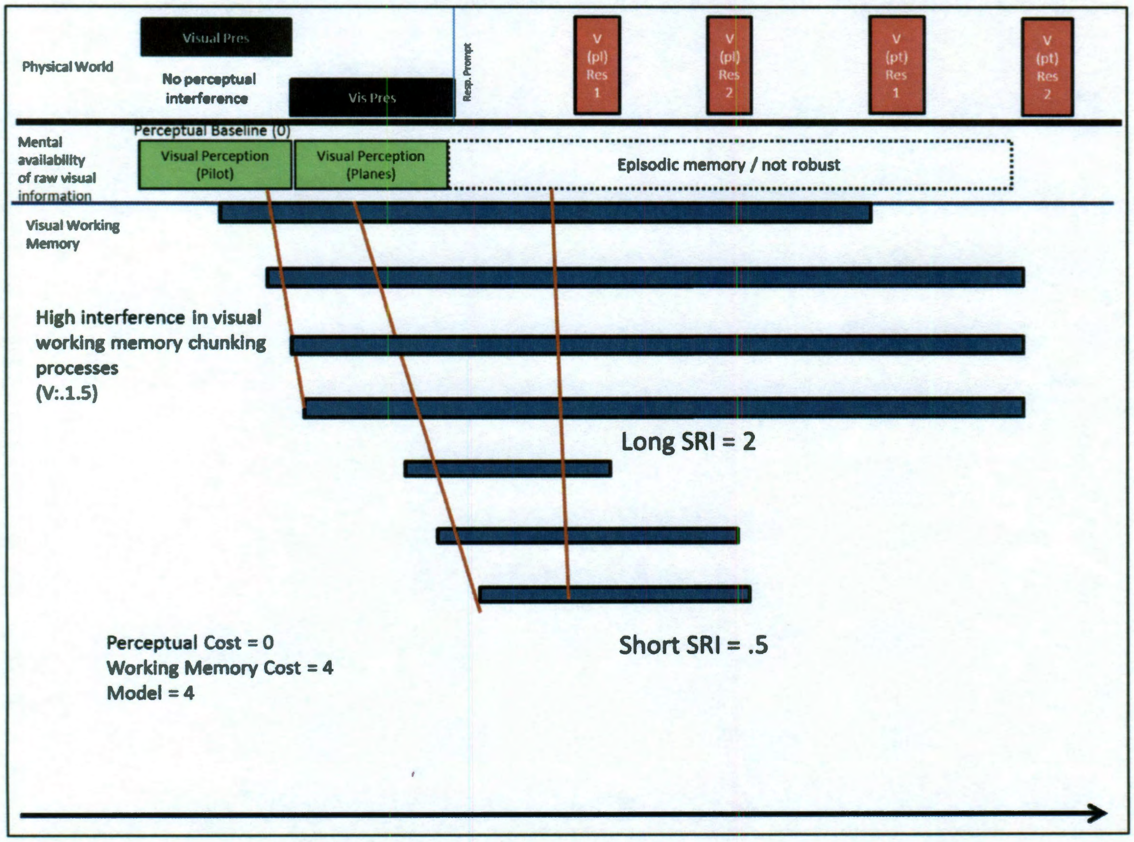

Figure 60. The cognitive flow model of the Auditory then Visual presentation (auditory planes). The model prediction is 4.

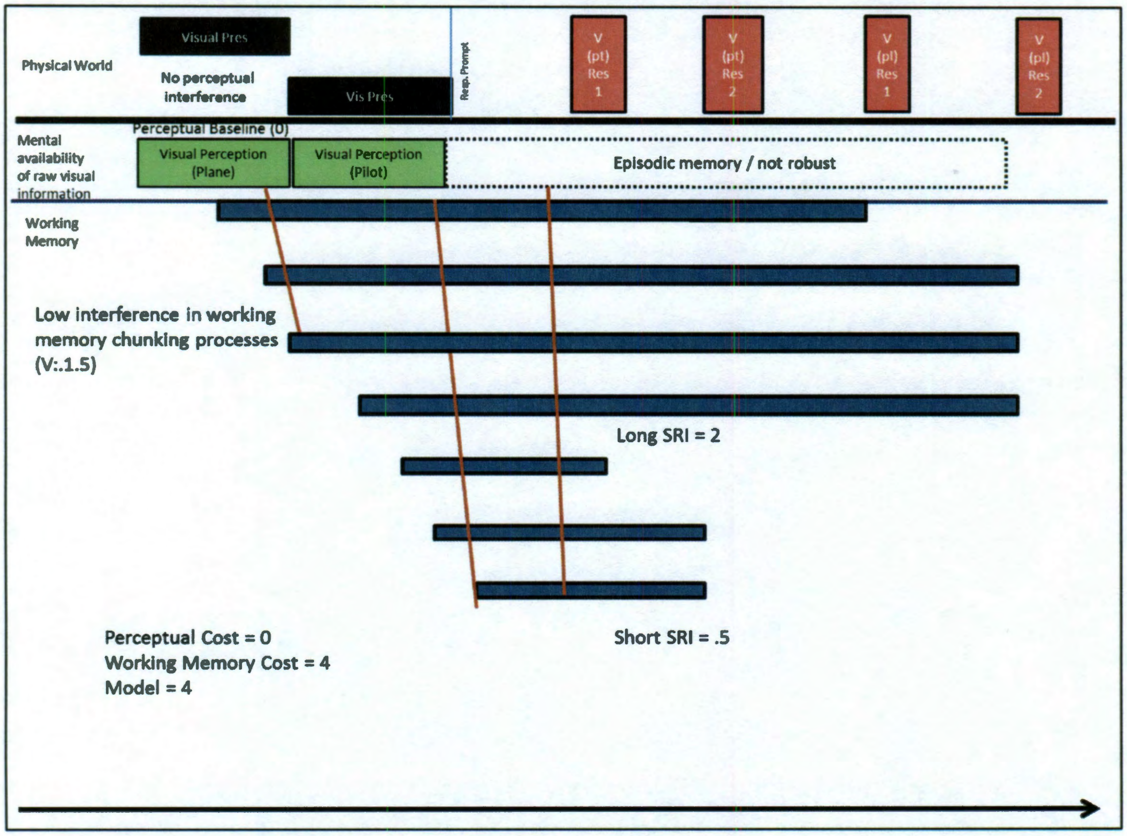

Figure 61. The cognitive flow model of the Visual pilot then Visual planes. The model prediction is 4.

As in Experiments 1 and 2, the predicted cost acted as a strong surrogate accounting for 92% of the variance (Figure 62). It should be noted that a simplifying assumption was made that in all cases the secondly presented information was responded to second. More variance may have been accounted for if this assumption was not made. Table 8 outlines the framework predictions for the six conditions in Experiment 3.

Table 8. The predicted response cost and accuracy for each condition in Experiment 3.

	Perceptual Cost				Working Memory Cost				Predicted	Actual
	Baseline Cost		Interference Cost		SRI Cost		Interference Cost		Cost (rank)	Accuracy (rank)
Dual Modality (audio planes)	V	A	V	A	V	A	V	A		
V->A/A->V	0	.5	0	0	2	0	.5	0	3.0 (2)	.60(2)
A->V/V->A	0	.5	0	0	.5	1	.5	0	2.5(1)	.65(1)
Single Modality (visual planes)	V(isual) pl(ane)	V(isual) p(ilo)t	Vpl	Vpt	Vpl	Vpt	Vpl	Vpl		
Vpl->Vpt/ Vpt->Vpl	0	0	0	0	2	.5	.75	.75	-4.0(3)	.56(4)
Vpt->Vpl/ Vpl->Vpl	0	0	0	0	.5	2	.75	.75	-4.0(3)	.57(3)

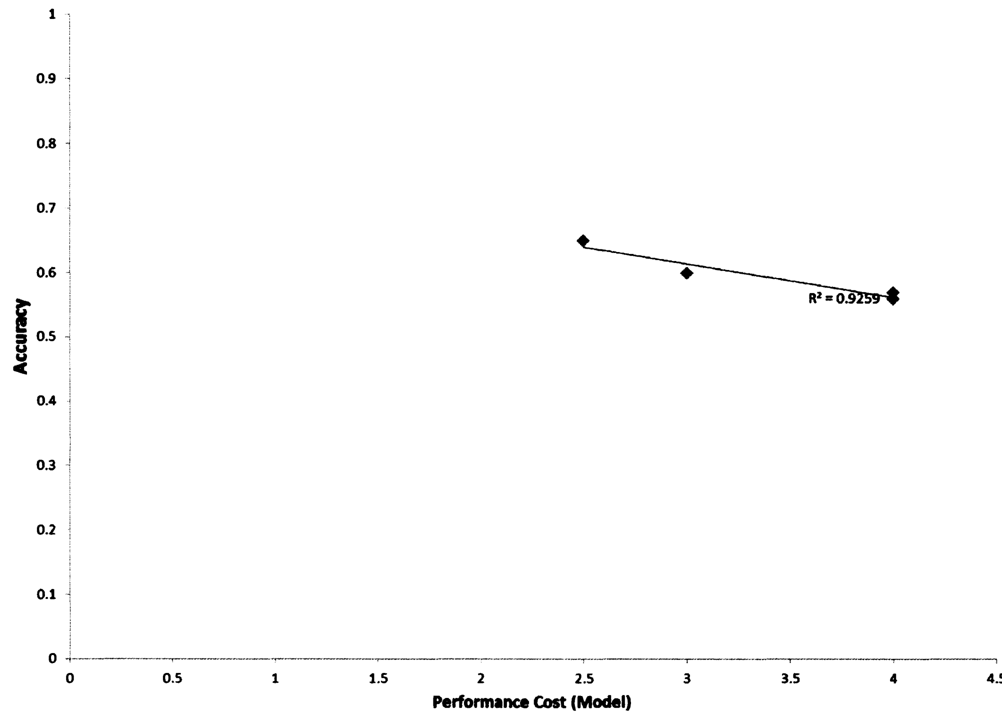

Figure 62. Representation of correlation between performance cost and actual accuracy for Experiment 3

Experiment 4

Since decision making is a different cognitive task than a memory task the expected cognitive flow the decision making task is described in detail. Figures 61-63 represent the physical and mental world in temporal progression from left to right. Physical world actions include the presentation of the visual and auditory sequence and the manual execution of the decision. Below the thick black line is the mental world. In this space perception, working memory, long-term memory, and decision choosing actions are all conducted. Black boxes represent the physical presentation of information. Green boxes represent the perception of the presentation. A white solid box represents echoic memory and a white box with a dashed border represents iconic memory. While perception or sensory memory persists, raw (presentation-style) information is considered to be available.

Blue and purple boxes represent items stored in short-term memory. A blue box stores an attribute of an important dimension (e.g., the number of planes on the radar screen). Purple boxes represent a reduction of the stored attributes into a decision making chunk (e.g., the visual dimension attributes together reflect, "no danger"). Red boxes reflect decision actions. Lines connecting boxes represent long-term memory queries. These queries reflect an operator's effortful accessing of learned rules for the rule-based task. These rules are known because the operator has had prior training. Green lines represent a long-term memory query used to identify which dimensions of a presentation are important and need to be stored for later use. Blue lines represent a long-term memory query used to reduce the dimension attributes stored in short-term memory into a meaningful decision-making unit. Purple lines are a long-term memory queries used to access the rules of the task and deduce a decision using the stored decision making chunks. Dotted lines are used to reflect a compromised or unreliable operation or state due to a high working memory load or lack of information. These compromised states may be the result of earlier compromised states.

To elucidate the flow of this generalized decision making task, consider three components: Visual processing, auditory processing, and decision deduction. An assumption of how this task is processed is that visual and auditory processing will complete before decision processes begin (and vice versa).

To help visualize the model use Figure 61 as a guide. The task begins with visual and auditory information being presented at the same time. Immediately after the start of the visual presentation a visual perception is formed. The start of the visual presentation is represented by the black box in the upper-left portion of the figure. The formation of a visual perception is represented by the green box below the second black box. Based on the task instructions stored in long-term memory information concerning the task contained in the perceptual image (e.g., numerosity, range, and type) are stored in short-term memory. Long-term memory information is represented with green lines. Short-term stores for dimension attributes are represented by the blue boxes. Next, long-term memory is used again to reduce the short-term memory items into a visual decision making chunk stored in short-term memory. This decision making chunk is a summary of the dimensions. Blue lines extending from the blue boxes represent long-term memories used in the reduction of the stored dimension attributes. The purple box represents the decision making chunk. An example of a decision making chunk may be “visual information as a whole communicates danger.” The same information processing steps are completed for auditory information. Once decision chunks are formed for each modality, long-term memory is used to reference the chunks with the known rules to deduce the correct decision. Purple lines represent long-term memory knowledge used to deduce the correct decision. The decision is then executed manually.

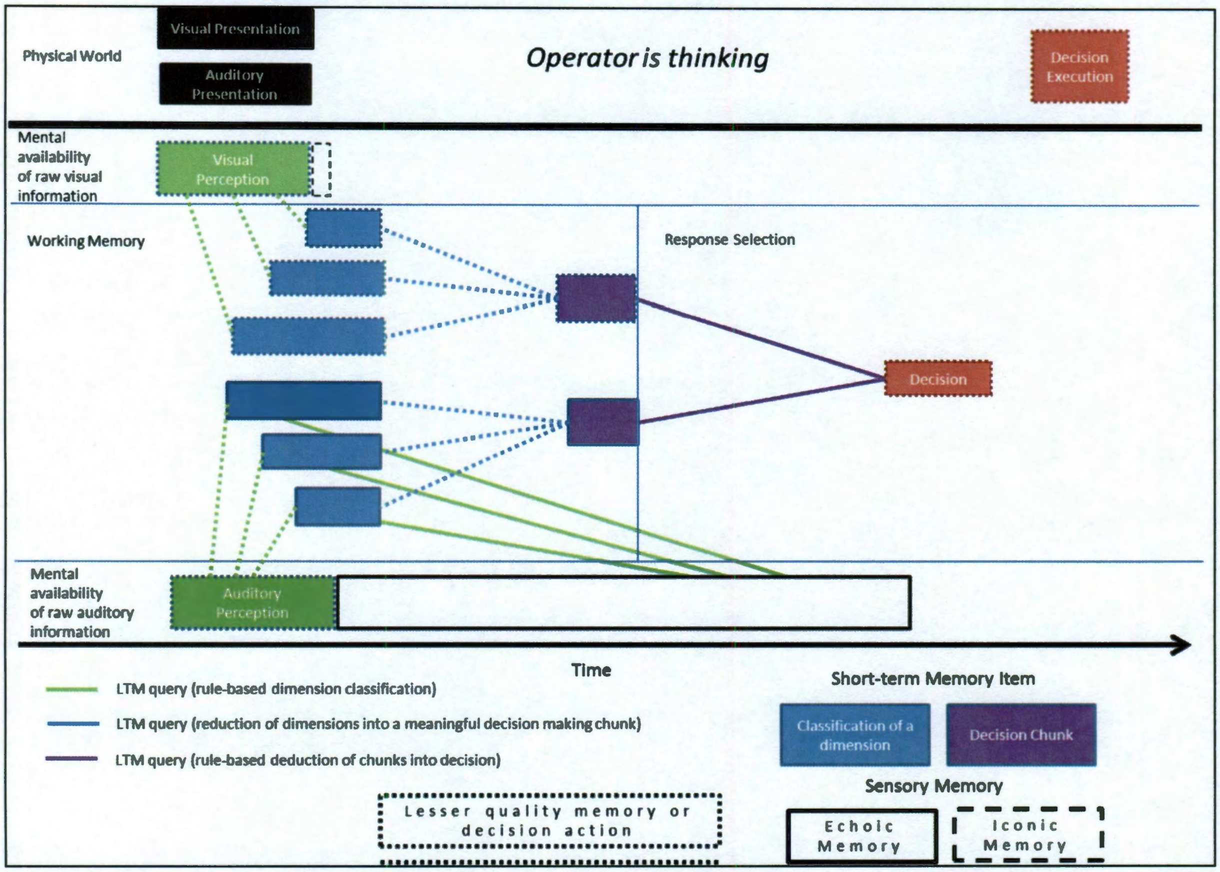

Figure 61. The physical and cognitive flow of a rule-based decision making task when both visual and auditory information are presented together.

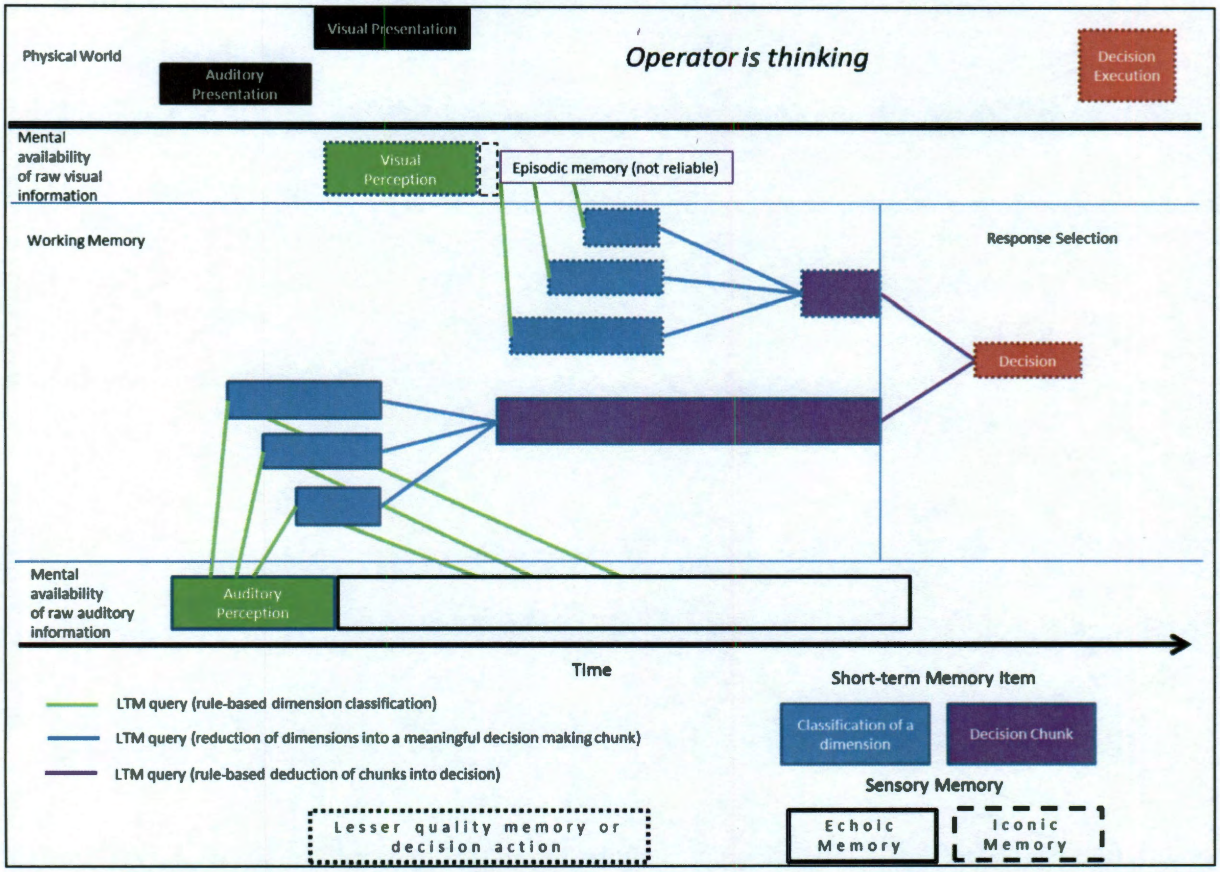

Figure 62. The physical and cognitive flow of a rule-based decision making task when auditory information is presented before visual information.

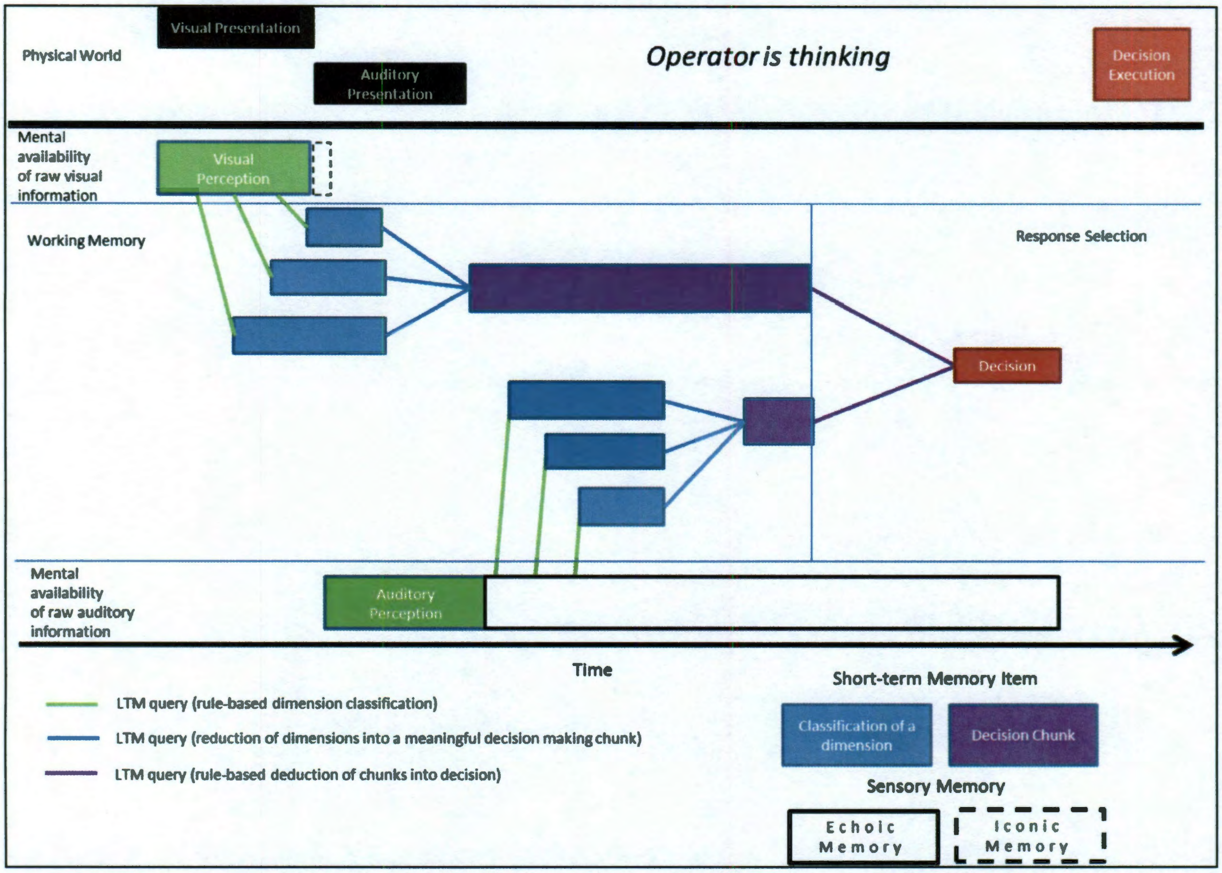

Figure 63. The physical and cognitive flow of a rule-based decision making task when visual information is presented before audio information.

The decision making task in Experiment 4 varied from previous experiments in that only a single response was required, making the average SRI shorter than in previous experiments. However, unlike Experiments 1 - 3, additional long-term memory and chunking processes were required prior to a response/decision which acted to elongate the SRI compared to Experiment 1-3 tasks in which such additional working memory processes were not required. In response to these task changes, the SRI cost for each condition was half the long SRI cost of Experiments 1-3. A long visual SRI is given a cost of 1.5 (half of the normal 3) and a long auditory SRI cost is .5 (half of the normal 1). As in Experiments 1 & 2 when visual information is presented first, the long visual SRI cost is reduced by a third due to an increased chance that visual chunking processes are less likely to be delayed or disturbed by ongoing auditory chunking processes. An undisturbed recoding interval is expected to generally lead to a more reliable visually related working memory chunks. Table 9 outlines the framework predictions for the six conditions in Experiment 4.

Table 9. Predicted Response cost and accuracy for each condition in Experiment 4.

	Perceptual Cost				Working Memory Cost				Predicted	Actual
	Baseline Cost		Interference Cost		SRI Cost		Interference Cost		Cost (rank)	Accuracy (rank)
Dual Modality	V	A	V	A	V	A	V	A		
(VA)	0	.5	.25	.25	1.5	.5	1	0	-4.0 (3)	.80(3)
V->A	0	.5	0	0	1	.5	.5	0	2.5 (1)	.87 (1)
A->V	0	.5	0	0	1.5	.5	.5	0	3.0(2)	.83(2)
Single Modality	Vpl	Vpt	Vpl	Vpt	Vpl	Vpt	Vpl	Vpl		
(Vpl Vpt)	0	0	1	1	1.5	1.5	.75	.75	-6.5(6)	.77(6)
Vpt->Vpl	0	0	0	0	1.5	1.0	.75	.75	-4.0(3)	.80(3)
Vpl -> Vpt	0	0	0	0	1.5	1.0	.75	.75	-4.0(3)	.80(3)

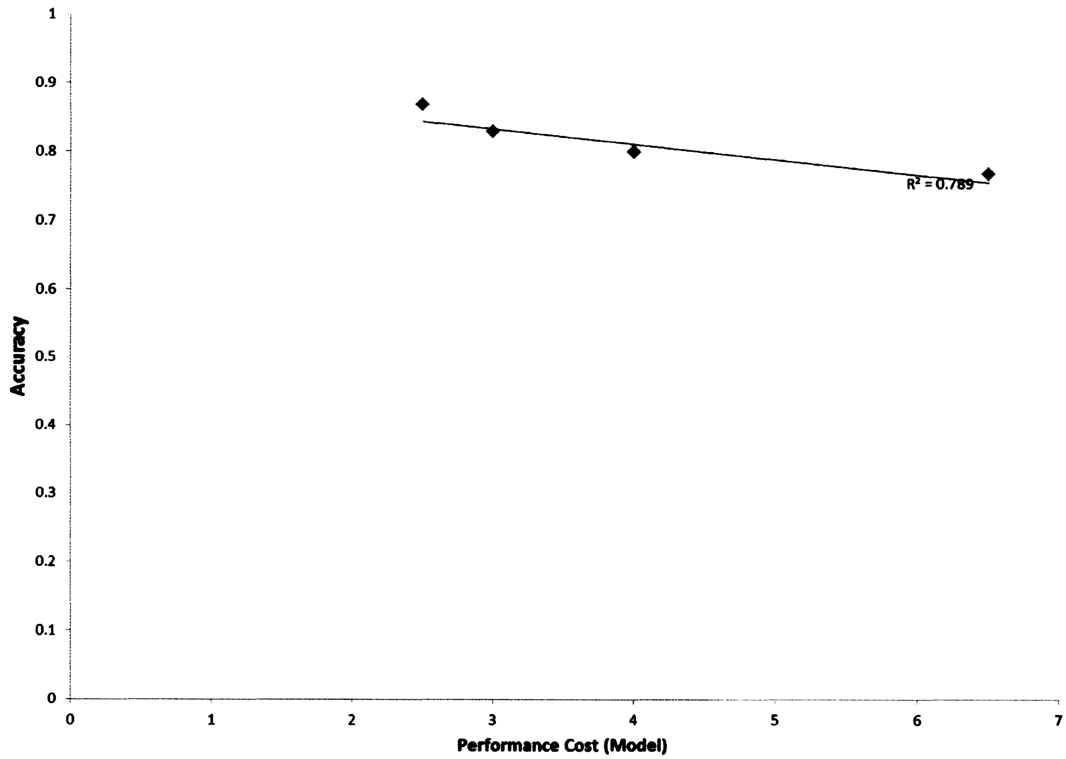

Figure 64. Correlation between performance cost and actual accuracy for Experiment 3

Using the predicted performance cost value, 79% of the performance variance was accounted for (Figure 64). Though less variance was accounted for than in the previous three experiments the strength of the correlation suggests that perceptual and working memory costs observed in Experiments 1-3 were present in the Experiment 4. Success of the framework in predicting Experiment 4 data suggests that it may have utility beyond simple memory tasks.

Conclusion

The fundamental thesis of this research is based on the well-documented finding that the duration of items in echoic memory is longer than the duration of item in iconic memory. One implication of this hypothesis is that visual performance is more time sensitive than auditory performance. Thus, extending the length of time between presentation and response will impact visual performance more severely than auditory performance. As such it was expected that overall recall accuracy and decisions based on the visual and auditory stimulus may be improved if the visual retention interval is shortened.

In all four experiments participants were presented with both auditory and visual information and responded to the information using only their memory of the event. Experiments 1-3 assessed the memory of the multimodal event directly by asking participants to recall details of both the visual and auditory information. Experiment 4 assessed knowledge-based decision making in response to the visual and auditory information.

To a large extent Experiments 1-4 confirmed this hypothesis. In Experiments 1-3, consistent evidence was gathered demonstrating that auditory accuracy was not as affected by presentation and response manipulations as visual accuracy. In Experiments 1-3 shortening the retention interval after the display phase improved overall accuracy. Seemingly counter to the primary hypothesis, however, were the observations where lengthening the visual retention interval by presenting visual information first improved memory and decision making. Though the primary hypothesis of this research was confirmed, other perceptual and working memory factors should be considered when designing a multimodal display.

The effect of manipulating the presentation order: In Experiments 2 through 4 presentation order was manipulated. These experiments are thought to provide cues for designing a more memorable multimodal display. In Experiments 2 and 4, the two experiments where audio-visual information was presented both together and separately, it was observed that presenting the visual and auditory stimulus separately improved overall performance. When proposing the prediction framework I presented an argument that the effect of the staggering of the visual and auditory information would make recall easier and an argument that it would make recall more difficult. The rationale for why staggering the sequences would make recall more difficult is that the first-presented item would have a longer retention interval than if the sequences were presented together. Specifically, presenting the visual and auditory sequence separately adds the second sequence duration to the first presented item's SRI. The rationale for why staggering the sequences would make recall easier is that there would be a reduction in perceptual load. In each of these three experiments it was observed that staggering the sequences lead to greater performance overall. Thus, this benefit was observed across laboratory and real-world inspired memory tasks (Experiments 2 and 3 respectively) as well as a real-world inspired decision making tasks (Experiment 4).

Counter to the core thesis of this research was the finding that performance was enhanced by presenting visual information before auditory information. In Experiment 2 lengthening of the visual retention interval at the display phase was beneficial and this appears to be because it increases the likelihood of an uninterrupted recoding period in which the visual information can be processed and recoded to a less time sensitive format (working memory chunk). Though such recoding is attentionally intensive, it should not interfere with the memory of the auditory presentation because maintaining information in the Long Auditory Store is not thought to require cognitive resources and is less likely to interfere with other cognitive components of the task. In Experiment 4, presenting visual information first was beneficial for decision making because it allowed for an uninterrupted interval to generate a visual decision chunk. Though it is believed that the majority of auditory information processing occurred after the audio presentation, echoic memory provided a reliable way to re-access auditory information and successfully form an audio decision chunk.

In Experiment 3 it was observed that presenting auditory information (audio-planes) prior to visual information (pilot images) led to greater performance overall. Upon further analysis, however, it was discovered that participants tended to respond first to the secondly presented sequence. This behavior is similar to how participants respond to memory items in a free recall test. Since presenting visual information first tended to cause participants to respond to the visual information second, the retention interval for visual items in the recall phase was extended.

The rationale for why overall performance was hurt when and auditory response followed visual presentation is considered a psychologically plausible explanation for the performance effects resulting from presentation order manipulations. The hypothesis presented, however, could be falsified by slightly elongating the delay between the visual and auditory presentation or the delay between a visual presentation and auditory response. If elongating these delays does not lead to greater overall performance, the proposed explanation for why elongating the visual retention interval in the display phase is likely oversimplified or false. The fact that these manipulations were not instituted in the four experiments is a weakness of this research.

The effect of manipulating the response order: Experiments 1 through 3 each assessed how response order is associated with overall performance. Since the response order manipulation assesses performance after the display has completed its role, the first three experiments are thought to provide clues on how train operators to respond to a multimodal display.

In Experiments 1 and 2 participants were forced to respond to either the visual sequence or the audio sequence first. In both these experiments it was found that forcing participants to respond fist to the visual information led to a robust advantage in overall performance. In Experiment 3 participants were allowed to choose their own response order strategy. As in

Experiment 1 and 2 it was observed that shortening the visual response interval was associated the improved performance. Though this relationship cannot be deemed as causal, the correlational direction provides additional supporting evidence that, whether forced or chosen, a visual-then-audio response strategy leads to greater accuracy. It should be noted, however, that the benefit of a visual-then-audio response strategy was much greater when no encoding interval was provided in the presentation phase. This interaction suggests that a substantial portion of the benefit of a visual-then-audio response strategy is due to the fact that it prohibits a visual encoding interval from being disrupted by an auditory response.

The effect of using dual modalities In Experiments 3 and 4, investigations in which single modality performance was compared to multiple modality performance, both recall and decision making was improved over single modality performance by using dual modalities. This result is consistent with many previous observations that using multiple modalities can improve task performance (Navon & Gopher, 1979; Wickens, 1984). The comparison of single and multiple modalities also strengthened the hypothesis that presentation ordering advantages are due to differences in iconic and echoic memory rather than the content of the information. In Experiments 3 and 4 both pilot and plane information were presented. When both types of information were presented visually, no presentation ordering effects were observed. In contrast, when the two types of information were presented using different modalities, ordering presentation effects were observed supporting the hypothesis that the relative performances observed were due to differences in sensory modalities.

Framework Remarks: A parsimonious account of data from Experiment 1 through 4 was achieved using the proposed framework. The framework accounted for performance by summing the expected perceptual and working memory costs. In each of the 4 experiments the predicted response cost was predictive of response quality and was able to predict the condition with the greatest accuracy. The framework remained highly predictive even when different stimuli and responses were required. In each of the experiments the condition(s) with the lowest predicted cost was associated with the highest actual accuracy. The framework has an underlying assumption that stimuli will be held constant. Thus, the framework is not well suited to predict performance if there are large changes in stimulus and response requirements. This limitation is apparent by comparing the accuracy associated with the highest cost prediction between Experiments 1 and 4. In Experiment 1 the highest cost value was 5.5 which were associated with an accuracy of .44. In contrast the highest predicted cost in Experiment 4 was 6.5 which are associated with an accuracy of .77. Clearly if all conditions from each experiment were evaluated simultaneously the framework prediction would not have been as highly correlated. For all data to be predicted together, the modeler would need to fit relative cost values based on task difficulty. This would be similar to how a modeler must use discretion to identify define a task demand scalar using an MRT model.

An additional limitation of the proposed framework is a built-on mechanism to predict absolute performance. One reason for this limitation is the model accounts for relative rather than absolute complexity of the stimuli. In all four Experiments participants had difficulty distinguishing between gradations of tones. While task stimuli was purposefully designed to be difficult to decode as to ensure a performance ceiling is not met, this human auditory perceptual deficit should limit the extent to which auditory information can be used to present abstract sonified information. Most auditory display designers are aware of this limitation, and as a result, the majority of auditory displays are relatively simple beeps and tones.

A further limitation of the proposed framework is that, unlike UTCs, this framework does not address learning across trials. Each trial does not affect procedural memory processes or, more generally, the probability of a successful response in a future trial. Thus this model would not be of use in predicting learning. While UTCs, most extensively ACT-R, are capable of predicting changes in performance order strategies (e.g., respond to auditory information first, delay deep processing of auditory information while processing auditory information) in response to improved performance results (trial learning), neither the timeline nor UTCs are particularly well equipped to explain or predict how an operator would be able to learn how to effectively identify patterns in the auditory stimulus.

An additional weakness of the presented framework is that the potential negative effects of masking are not accounted for. As mentioned, research has shown that echoic memory will be less reliable if language is used and it the auditory stimulus is very long. Three minutes of auditory stimulus cannot be held in echoic memory as effectively as the two seconds of auditory stimuli. This model is best equipped to predict relative performance when information is presented in a brief time interval.

It is predicted that ACT-R is well suited to make accurate predictions for performance observed in Experiments 1-4. As is ACT-R is well suited for modeling cognitive flow with production rules. Perhaps the biggest ACT-R limitation is how it handles the auditory buffer (Long Auditory Store analogue). While the modeler is able to set the length of time auditory information is stored in the buffer, there is not a clear mechanism for adjusting the probability of successful extraction from the auditory buffer. Though the ability to recall auditory information was not as affected by the SRI length as visual information, an effect was still present. If the Long Auditory Store was a perfect store, an auditory performance decrement would not be predicted. Data collected in this investigation suggest that the information in the Long Auditory Store should not be imagined as a binary (available or unavailable) but rather in a state that can continuously degrade over a time interval.

Closing Remarks: In each of the four experiments participants were presented once with a visual and auditory stimulus and were asked to respond using memory. In a real-life situation it is unlikely that a display would be restricted in such a way. In real life if an operator needs to continue perceiving information from a display to make a response, the information should be provided. I argue, however, that in cases when repeated viewings of a display to receive information updates interrupts task performance, a measure of the auxiliary display's effectiveness is its ability to present information such that additional presentations are most often not needed. As an example the armed forces are interested in how multi-modal displays may help soldiers on a battlefield make decisions. It is often the case that for a soldier to be successful he or she must continually have intelligence updates (Glumm et al., 2007). When an update is presented it would be best if the soldier has only to reference it once such that attention can quickly be switched to other important tasks such as staying visually vigilant and directing his or her fellow troops. The more time the soldier has to direct his attention to the auxiliary display, instead of negotiating the environment, the more these other tasks may suffer. Another context where this research may apply is in the vehicle where it is increasingly common for the driver to receive information both auditorily and via text/symbol to assist in a driving decision. As is clear from naturalistic driving studies, driving risk increases most when eyes glances are directed away from the driving scene (and to a far lesser extent when the "mind" is not on the road). The less a driver needs to review display information, the more the driver may remain focused on safe driving habits. A weakness of this research is that the task demands were greater than one might expect from a display. Tasks and stimuli were purposefully designed to be awkward to avoid potential ceiling effects that may obscure potential performance benefits from manipulating the presentation and response orders of a multimodal display. However, in a finely tuned multimodal display it is expected that such manipulations would have a far lesser impact than that observed in Experiments 1-4.

The methodology employed in Experiments 1-4 in turn addresses how effective a display is with just a single presentation. The core thesis of this research was supported in that the longevity of auditory sensory memory at times will have implications on performance in response to the display. Experiment 2 highlighted a somewhat paradoxical finding that elongating the visual retention interval may improve overall performance even though visual sensory memory is not persistent. Hence, though the core thesis of this research did play a critical role in predicting performance, it was clearly not the sole factor. In response, a prediction framework was introduced (which pulled lessons from MRT) that demonstrated an ability to parsimoniously explain results from the four experiments. While the utility proposed framework is likely limited to predicting relative performance in tasks similar to those in Experiments 1-4, I posit that lessons learned may assist those attempting to predict performance in response to a multimodal display. Though strides were made to assess the multimodal presentation and response ordering advantages established in Experiment 1 and 2 in a more real-life task scenario, additional investigations should be made using real-world tasks with real operators for additional confirmation that ordering findings of these laboratory-based experiments lead to tangible real world benefits.

References

Amedi, A., Malach, R., Hendler, T., Peled, S., & Zohary, E. (2001). Visuo-haptic object-related activation in the ventral visual pathway. *Nature Neuroscience, 4,* 324–330.

National Highway Traffic Safety Administration. (2006). *The Driver Workload Metrics Project.* Washington, DC: Angell, L., Auflick, J., Austria, P. A. Kochhar, D., Tijerina, L, Biever, W., . . . Kiger, S.

Baddeley, A.D., Grant, S., Wight, E., & Thompson, N. (1975), Imagery and visual working memory. In: P.M. Rabbitt and S. Dornic (Eds.), *Attention and Performance* (Vols. 5). London: Academic Press.

Baddeley A., Della Sala S. (1996). Working memory and executive control. *Philosophical Transactions of Royal Society B: Biological Science, 351*, 1397–403.

Bolia, R. S., D'Angelo, W. R., & McKinley, R. L. (1999). Aurally aided visual search in three-dimensional space. *Human Factors, 41,* 664-669.

Boltz, M. G. (1998). Tempo discrimination of musical patterns: Effects due to pitch and rhythmic structure. *Perception and Psychophysics, 60,* 1357–1373.

Bonebright, T. L., Nees, M. A., Connerley, T. T., & McCain, G. R. (2001, July). *Testing the effectiveness of sonified graphs for education: A programmatic research project.* Paper presented at the International Conference on Auditory Display, Helsinki, Finland.

Broadbent, D.E. (1958). *Perception and communication.* London: Pergamon Press.

Brooks, L. R. (1968). Spatial and verbal components of the act of recall. *Canadian Journal of Psychology, 22,* 349-368.

Chase, W. G., & Simon, H. A. (1973). The mind's eye in chess. In W. G. Chase (Ed.), *Visual information processing* (pp. 215–281). New York, NY: Academic Press.

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

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

Deese, J., & R. A. Kaufman. (1957). Serial effects in recall of unorganized and sequentially organized verbal material. *Journal of Experimental Psychology, 54*, 180-187.

De Groot, A. D. (1978). *Thought and choice in chess (2nd English Edition).* The Hague: Mouton Publishers.

Efron, R. (1970a). Effect of stimulus duration on perceptual onset and offset latencies. *Attention, Perception, & Psychophysics*, *8,* 231-234.

Efron, R. (1970b). The relationship between the duration of a stimulus and the duration of a perception. *Neuropsychologia, 8,* 37-55.

Fitch, W. T., and G. Kramer. "Sonifying the Body Electric: Superiority of Auditory Over A Visual Display in a Complex, Multivariate System." in Auditory Display: Sonification, Audification, and Auditory Interfaces Proceedings of the proceedings of the 1992 International Conference on Auditory Display, held at the Santa Fe Institute, New Mexico, October 1992. Edited by Gregory Kramer. Mass: Addison-Wesley, 1994.

Flowers, J. H. (2005). In A. Afonso, B. Katz, A. Blum, C. Jacquemin, & M. Denis (Eds.). Thirteen years of reflection on auditory graphing: Promises, pitfalls and potential new directions. *Proceedings of the International Conference on Auditory Display* (pp. 513-518). International Community for Auditory Display.

Frauenberger, C., Stockman, T., & Bourguet, M. L. (2007). In L. M. Mauney & B. N. Walker (Eds.) Pattern design in the context space: A methodological framework for auditory display design. *Proceedings of the International Conference on Auditory Display* (pp. 513-518). International Community for Auditory Display.

Glaser, D. G. (2008). Foraging Memory: Producing Words from One and From Two Semantic Categories (Unpublished thesis). Rice University, Houston, TX.

Glumm, M. M.; Kehring, K. L.; White, T. L. Effects of Multimodal Displays About Threat Location on Target Acquisition and Attention to Visual and Auditory Communications; ARL-TR-4074; U.S. Army Research Laboratory: Aberdeen Proving Ground, MD, 2007.

Haber, R. N., & Standing, L. G. (1969). Direct measures of short-term visual storage. *Quarterly Journal of Experimental Psychology*, *21*, 43-54.

Hartmann, W. M. (1997). *Sounds, Signals and Sensation: Modern Acoustics and Signal Processing*. New York, NY: Springer-Verlag.

Holyoak, K.J. (2005). *The Cambridge handbook of thinking and reasoning*. Cambridge University Press.

Horrey, W. J., Wickens, C. D., & Consalus, K. P. (2006). Modeling drivers' visual attention allocation while interacting with in-vehicle technologies. *Journal of Experimental Psychology: Applied*, 12, 67-86.

Jeon J.Y. & Fricke F.R. (1997), Duration of Perceived and Performed Sounds, *Psychology of Music*, *25*, 70-83.

Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice Hall.

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

Kramer, G., Walker, B. N., Bonebright, T., Cook, P., Flowers, J., Miner, N., & Neuhoff, J. (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, Santa Fe, NM.

Lacherez, P., Seah, E. L., & Sanderson, P. M. (2007). Overlapping melodic alarms are almost indiscriminable. *Human Factors, 49,* 637-645.

Liu, Y. (2001) Comparative study of the effects of auditory, visual and multimodality displays on drivers' performance in advanced traveler information systems. *Ergonomics, 44,* 425-442.

MacKay, D. G. (1973). Aspects of the theory of comprehension, memory, and attention. *Quarterly Journal of Experimental Psychology, 85*, 225-228

Meyer, D. E., & Kieras, D. E. (1997). A computational theory of executive cognitive processes and multiple-task performance: Part 1 Basic mechanisms. *Psychological Review*, *104,* 3-65.

Meyer, D. E., & Kieras, D. (1999). Precise to a practical unified theory of cognition and action: some lessons from EPIC computational models of human multiple-task performance. In D. Gopher and A. Koriat (Eds), *Attention and Performance XVII* (pp. 17-88). Cambridge, MA: MIT Press.

Merabet L. B., Battelli L., Obretenova S., Maguire S., Meijer P., Pascual-Leone A. (2009). Functional recruitment of visual cortex for sound encoded object identification in the blind. *Neuroreport, 20,* 132–138.

Meredith, C.S., & Edworthy, J. (1991). Sources of confusion in intensive therapy unit alarms. In N. Stanton (Ed.) *Human Factors in Alarm Design* (pp. 207-219). Taylor and Francis: London.

Momtahan, K, Hetu, R., & Tansley, B. (1993). Audibility and identification of auditory alarms in the operating room and intensive care unit. *Ergomonics, 36*, 1159-1176.

Morton, J., Crowder, R. G., & Prussin. H. (1971). Experiments with the stimulus suffix effect. *Journal of Experimental Psychology Monograph, 91,* 169-190.

Moray, N. (1967). Where is capacity limited? A survey and a model. *Acta Psychologica*, *27*, 84-92.

Mowbray, G. H. (1964). Perception and retention of verbal information presented during shadowing. *Journal of the Acoustical Society of America*, *36,* 1459-1465.

Murdock, B. B. Jr. (1962). The serial position effect in free recall. *Journal of Experimental Psychology*, *64*, 482-488.

Navon, D., & Gopher, D. (1979). On the economy of the human information processing system. *Psychological Review*, *86*, 214-255.

Nairne, J. S., & Crowder, R. G. (1982). On the locus of the stimulus suffix effect. *Memory and Cognition, 10,* 350-357.

Nikolic, M.I., & Sarter, N.B. (2001). Peripheral visual feedback: A powerful means of supporting effective attention allocation in event-driven, data-rich environments. *Human Factors, 43,* 30-38.

Nees, M.A., & Walker, B.N. (in press). Auditory interfaces and sonification. In Stephanidis, C. (Ed.), *The Universal Access Handbook* (pp. 507-522). New York: CRC Press.

Nees, M. A., & Walker, B. N. (2007, June). *Listener, task and auditory graph: Toward a conceptual model of auditory graph comprehension.* International Conference on Auditory Display, Montreal, Canada.

Neuhoff , J. G., Kramer, G., & Wayand, J. (2002). Pitch and loudness interact in auditory displays: Can the data get lost in the map? *Journal of Experimental Psychology: Applied, 8,* 17–25.

Newell, A. (1973). You can't play 20 questions with nature and win: Projective comments on the papers of this symposium. In W. G. Chase (Ed.), *Visual Information Processing* (pp. 283–308). New York, NY: Academic Press.

Newell, A., & Simon, H. A. (1972). *Human problem solving.* Englewood Cliff, NJ: Prentice-Hall.

Norman, D., & Bobrow, D. (1975). On data-limited and resource-limited processing. *Journal of Cognitive Psychology*, *7*, 44-60.

Parkes, A. M. (1991). Drivers' business decision-making ability whilst using car phones. In E. Lovessey (Ed.), *Contemporary ergonomics: Proceedings of the Ergonomic Society annual conference* (pp. 427–432). London: Taylor & Francis.

Pashler, H. (1998). The psychology of attention. Cambridge, MA: MIT Press.

Patterson, R.D., & Milroy, R. (1980). *Auditory warnings on civil aircraft: the learning and retention of warnings.* London: Civil Aviation Authority.

Peres, S. C., & Lane, D. M. (2003). In D. Skantze, & N. Dahlback (Eds.) Sonification of statistical graphs. *Proceedings of the International Conference on Auditory Display* (pp. 157-160). International Community for Auditory Display.

Peres, S. C., & Lane, D. M. (2005). In J. Kildal, & S. A. Brewster (Eds.) Auditory graphs: The effects of redundant dimensions and divided attention. *Proceedings of the International Conference on Auditory Display* (pp. 169-174). International Community for Auditory Display.

Plomp, R. (1964). Rate of decay of auditory sensation. *Journal of the Acoustical Society of America, 36,* 277-282.

Ramloll, R., Brewster, S., Yu, W., & Riedel, B. (2001). Using non-speech audio sounds to improve access to 2D tabular numerical information for visually impaired uses, *Proceedings of the IHM-HCI* (pp. 515-530).

Rollins, R.A., & Hendricks, R. (1980). Processing for words presented simultaneously to eye and ear. *Journal of Experimental Psychology: Human Perception and Performance, 6,* 99-109.

Seagull, F. J., & Sanderson, P. M. (2001). Anaesthesia alarms in surgical context: An observational study. *Human Factors*, *43*, 66-77.

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs*, *74*, 129.

Sanderson, M. P., Watson, O. M., Russell, J. W., Jenkins, S., Liu, D., Green, N., . . . Krupenia S. S. (2008). Advanced auditory displays and head-mounted displays: advantages and disadvantages for monitoring by the distracted anesthesiologist. *Anesth* Anal, 106, 1787-1797

Sarno, K. J., & Wickens, C. D. (1995). The role of multiple resources in predicting timesharing efficiency. *International Journal of Aviation Psychology*, *5,* 107-130.

Treisman, A., & Davies, A. (1973). Divided attention to ear and eye. In S. Kornblum (Ed.) *Attention and Performance IV* (pp. 101-117). Academic Press: New York.

Treisman, A., & Geffen, G. (1967). Selective attention: Perception or response? *Quarterly Journal of Experimental Psychology*, *19*, 118.

Upson R. (2001). In G. Lorho, J. Marila, & J. Hiipakka (Eds.) Sonification as Mathematics Teaching Tools. *Proceedings of the 2001 International Conference on Auditory Displ*ay (pp. 217-221). International Community for Auditory Display.

Walker, B. N. & Lane, D. M. (2001). In G. Lorho, J. Marila, & J. Hiipakka (Eds.) Psychophysical scaling of sonification mappings: A comparison of visually impaired and sighted listeners. *Proceedings of the International Conference on Auditory Display* (pp. 90–94). International Community for Auditory Display.

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., Kim, J., & Pendse, A. (2007). Musical sound scopes for an accessible aquarium: Bringing dynamic exhibits to the visually impaired. *Proceedings of the International Computer Music Conference* (pp. 268-275). Copenhagen, Denmark.

Walker, B. N., & Kramer, G. (2004). Ecological psychoacoustics and auditory displays: Hearing, grouping, and meaning making. In J. Neuhoff (Ed.), *Ecological psychoacoustics* (pp.150-175). New York: Academic Press.

Walker, B. N. & Mauney, L. M. (2004). In G. Leplaitre, & I. McGregor (Eds.) Individual differences, cognitive abilities and the interpretation of auditory graphs. *Proceedings of the International Conference on Auditory Display* (pp. 6–9). International Community for Auditory Display.

Watson, C. S. & Kidd, G. R. (1994). In D. H. Jameson (Eds.) Factors in the design of effective auditory displays. *Proceedings of the International Conference on Auditory Display* (pp.7-9). International Community for Auditory Display.

Weinstein, L.F., & Wickens, C.D. (1992). Use of nontraditional flight displays for the reduction of central visual overload in the cockpit. *International Journal of Aviation Psychology*, *2,* 121-142.

Watkins, M. J., & Todres, A. K. (1980). Suffix effects manifest and concealed: Further evidence of a 20-second echo. *Journal of Verbal Learning and Verbal Behavior, 19,* 46-53.

Wickens, C. D. (1980). The structure of attentional resources. In R. Nickerson (Ed.), *Attention and performance VIII* (pp. 239–257). Hillsdale, NJ: Erlbaum.

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

Wickens, C.D. (1991). Processing resources and attention. In D.L. Damos (Ed.), *Multiple Task Performance* (pp. 3-34). Washington, D.C: Taylor & Francis.

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

Wickens, C.D. (2008). Multiple resources and mental workload. *Human Factors, 50,* 449-455.

Wickens, C.D., & Liu, Y. (1988). Codes and modalities in multiple resources: A success and qualification. *Human Factors, 45,* 360-380.

Wickens, C. D., Sandry, D., & Vidulch, M. (1983). Compatibility and resource competition between modalities of input, output, and central processing. *Human Factors*, *25*, 227-248.

Wickens, C. D., & Hollands, J. G. (2000). *Engineering psychology and human performance (Third Edition).* Upper Saddle River, New Jersey: Prentice-Hall Inc.

Winkler, I., Reinkainen, K., Naatanen R. (1993). Event-related brain potentials reflect traces of echoic memory in humans. *Perception & Psychophysics,* 53, 443-449.