Internal Representations of Auditory Frequency: Behavioral Studies of Format and Malleability by Instructions

A Dissertation Submitted to
The Academic Faculty

by

Michael A. Nees

In Partial Fulfillment
of Requirements for the Degree
Doctor of Philosophy in the
School of Psychology

Georgia Institute of Technology
December 2009

Internal Representations of Auditory Frequency: Behavioral Studies of Format and Malleability by Instructions

Approved by:

Dr. Bruce N. Walker, Advisor School of Psychology *Georgia Institute of Technology*	Dr. Wendy Rogers School of Psychology *Georgia Institute of Technology*
Dr. Richard Catrambone School of Psychology *Georgia Institute of Technology*	Dr. Terri Bonebright Department of Psychology *DePauw University*
Dr. Gregory Corso School of Psychology *Georgia Institute of Technology*	

Date Approved: October 28, 2009

ACKNOWLEDGEMENTS

I thank my parents, Danny and Susan Nees; their contributions to my successful completion of this project were immense and began long ago when I was very young. I owe this accomplishment to the sacrifices my parents made for my education, and with that knowledge I have appreciated this experience all the more. My wife, Anne Tucker Nees, will always be my biggest supporter. I thank her for all of the encouragement and help, especially during the hectic months during which this project was completed. This dissertation, and indeed much of my career trajectory in psychology, reflects the culmination of over 10 years of mentorship from Terri Bonebright, and it was an honor to have Terri on my committee for this dissertation. My advisor, Bruce Walker, provided a supportive, rigorous, and constructive learning environment throughout my years in the Georgia Tech Sonification Lab. I have have very much appreciated his generous approach to student mentorship. Through every step of the graduate education process, I have been grateful to be working with Dr. Walker. Richard Catrambone, Wendy Rogers and Greg Corso have given me guidance on my scholarship and challenged me to improve upon my ideas and approaches, and I thank them for their help with this dissertation. I am convinced that I ended up with the best possible final product as a result of my committee's helpful advice. I thank my research assistants, Amy Gress and Courtney Lynch, for their hours of work during the data collection phases of these studies. Finally, I would like to thank the members of the Sonification Lab for their friendship and comradery.

TABLE OF CONTENTS

Page

CHAPTER

LIST OF TABLES

LIST OF FIGURES

SUMMARY

Research has suggested that representational and perceptual systems draw upon some of the same processing structures, and evidence also has accumulated to suggest that representational formats are malleable by instructions. Very little research, however, has considered how nonspeech sounds are internally represented, and the use of audio in systems will often proceed under the assumption that separation of information by modality is sufficient for eliminating information processing conflicts. Three studies examined the representation of nonspeech sounds in working memory. In Experiment 1, a mental scanning paradigm suggested that nonspeech sounds can be flexibly represented in working memory, but also that a universal per-item scanning cost persisted across encoding strategies. Experiment 2 modified the sentence-picture verification task to include nonspeech sounds (i.e., a sound-sentence-picture verification task) and found evidence generally supporting three distinct formats of representation as well as a lingering effect of auditory stimuli for verification times across representational formats. Experiment 3 manipulated three formats of internal representation (verbal, visuospatial imagery, and auditory imagery) for a point estimation sonification task in the presence of three types of interference tasks (verbal, visuospatial, and auditory) in an effort to induce selective processing code (i.e., domain-specific working memory) interference. Results showed no selective interference but instead suggested a general performance decline (i.e., a general representational resource) for the sonification task in the presence of an interference task, regardless of the sonification encoding strategy or the qualitative interference task demands. Results suggested a distinct role of internal representations for nonspeech sounds with respect to cognitive theory. The predictions of the processing codes dimension of the multiple resources construct were not confirmed; possible explanations are explored. The practical implications for the use of nonspeech sounds in applications include a possible response time advantage when an external stimulus and the format of internal representation match.

1 INTRODUCTION

Speculations about and empirical examinations of the format of active thought have been prevalent throughout the history of psychology (see, e.g., Galton, 1880; James, 1890; Miller, 1956). In particular, influential accounts of cognitive processes have emphasized a dichotomy between verbal and nonverbal (often synonymous with visuospatial) processing (Baddeley, 1992, 2002, 2003; Mayer & Moreno, 1998; Mayer & Sims, 1994; Paivio, 1991, 2007; Wickens, 1984, 2002, 2008; Wickens & Hollands, 2000; Wickens & Liu, 1988; Wickens, Sandry, & Vidulich, 1983b). These perspectives have converged on a common theme—namely that verbal and visuospatial information are handled by relatively independent processes or conceptual structures that work in parallel during the stage of active information processing that is commonly referred to as "working memory," defined as "the system or systems involved in the temporary maintenance and manipulation of information" (Baddeley, 2002 p. 85). The behavioral consequences of independent verbal and visuospatial processes and their theoretical implications have been examined in considerable detail, and the premise of independence appears to have been corroborated with evidence from neuroscience. Dissociable neural systems for visuospatial and verbal processes have been identified (Anderson, Yulin, Jung, & Carter, 2007; Gruber, 2001; Gruber & Gotschke, 2004; Paulesu, Frith, & Frackowiak, 1993).

Multiple resources theory (Wickens, 1984, 1991, 2002, 2008; Wickens & Hollands, 2000; Wickens & Liu, 1988; Wickens et al., 1983b) perhaps most clearly predicted the potential impact of internal representations on human performance. The multiple resources approach suggested that the processing code—or internal format of stimulus representation—would figure prominently in a person's success or failure during multitasking (see Wickens & Liu, 1988). Specifically, the multiple resources approach posited that simultaneous tasks interfere to the extent that they tax the same member of each of several pairs of resource pools. Resource dichotomies were identified by sensory-perceptual modalities (auditory versus visual), processing codes (verbal and visuospatial), and response modalities (verbal versus manual responses) as depicted in Figure 1. With respect to internal representations or processing codes, then, concurrent tasks where stimuli assume the same internal representation (i.e., two tasks both requiring verbal processing or both requiring visuospatial processing) will interfere with each other more than concurrent tasks that use distinct formats of internal representation (e.g., a verbal task paired with a visuospatial task).

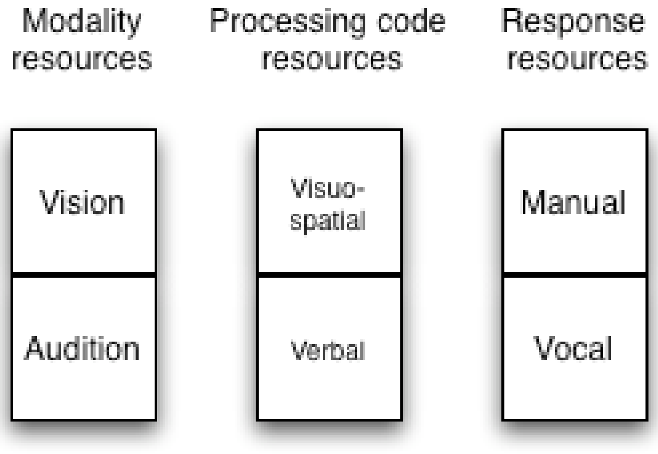

Figure 1: Simplified schematic depiction of the multiple resources approach.

An emerging body of evidence, however, has suggested that the verbal-visuospatial dichotomy of internal representations may be inadequate for at least two reasons. First, this dichotomy may omit a limited set of other plausible formats of representation, and second, existing theory may unnecessarily link (sometimes implicitly and sometimes explicitly) external stimulus representations with determinate internal representational formats.

Representational Formats in Working Memory

Psychology has overwhelmingly embraced the premise of separate representational processing systems for verbal and visuospatial information (Baddeley, 1992, 2002, 2003; Mayer & Moreno, 1998; Mayer & Sims, 1994; Paivio, 1991, 2007; Wickens, 1984, 2002, 2008; Wickens & Hollands, 2000; Wickens & Liu, 1988; Wickens et al., 1983b). Verbal systems "have a linguistic and symbolic base," whereas visuospatial systems "have a spatial analog base" (Wickens & Liu, 1988, p. 601). Verbal processing has a long history of evidence for functional localization and biological bases in the brain, dating back to at least Broca (Broca, 1861/2000) and continuing with modern neuroscience (e.g., Gruber, 2001). The empirical phenomena surrounding verbal representations have been examined intensively in the context of Baddeley's phonological loop—a conceptual working memory structure for processing verbal information (for a review, see Baddeley, 1992). The notion of visuospatial representations with unique sets of properties, however, was controversial for some time (see, e.g., Pylyshyn, 1981).

Whereas the notion of a veridical "picture in the brain" was dismissed (e.g., Pylyshyn, 2003), Shepard (1975) posited a more plausible mechanism of *second order isomorphism*, whereby some pattern of neural activation associated with an external pictorial stimulus is reinstated during the active internal representation of the same stimulus in the absence of a bottom-up percept. Kosslyn and colleagues (1973, 1975, 1976, 1981, Kosslyn, Ball, & Reiser, 1978) and others (e.g., Shepard & Metzler, 1971) supported arguments for a visuospatial format of internal representation with behavioral studies that implied the existence of internal mental images that mimicked the pictorial properties of an external visual stimulus. The hypothesis of second order isomorphism was further supported when neuroscience research (e.g., Farah, Peronnet, Gonon, & Giard, 1988; Kosslyn & Thompson, 2002) showed similar patterns of neural activity in visual processing brain regions during both mental imagery and actual visual perception. Most recent accounts of human information processing have posited a visuospatial format of internal representation with properties that are unique and behaviorally distinguishable from verbal internal representations (Anderson et al., 2007; Baddeley, 2002, 2003; Byrne & Anderson, 2001; Kieras & Meyer, 1997; Mayer & Moreno, 1998; Mayer & Sims, 1994; Meyer & Kieras, 1997; Wickens, 2002, 2008), and even critics of mental imagery research have acknowledged that the processes of representing visuospatial stimuli are distinct in meaningful ways from verbal processes (Pylyshyn, 2002).

Comparatively little attention has been paid to the possibility of an equivalent pseudo-isomorphic auditory representation system in working memory. Current instantiations of dual-process theories (Baddeley, 2002, 2003; Mayer & Moreno, 1998; Mayer & Sims, 1994; Wickens, 2002, 2008) allow for two possible encoding formats for nonspeech sounds: verbal and visuospatial. With regard to verbal encoding of sounds, research has suggested that the small portion of the population who possess absolute pitch can categorically associate a verbal label with the pitch of tones (Levitin & Rogers, 2005). Even listeners without absolute pitch might spontaneously ascribe less sophisticated or less accurate verbal labels to nonspeech sounds. Anecdotal reports (Zatorre & Beckett, 1989), a survey (Mikumo, 1997), and a qualitative analysis (Nees &

Walker, 2008a) have also suggested that people sometimes spontaneously form visuospatial images that depict changes in auditory frequency as pictorial representations of pitch contour. The mechanism for this phenomenon appears to be "'metaphorical' mapping" (Wagner, Winner, Cicchetti, & Garnder, 1981) or "weak synethesia" (Martino & Marks, 2001), whereby auditory frequency generally bears a systematic crossmodal relationship with visual space such that sounds of higher frequency are associated with higher spatial position or "up" (Ben-Artzi & Marks, 1995; Kubovy, 1981; Mikumo, 1997; Szlichcinski, 1979; Wagner et al., 1981; Walker, 2002, 2007). Verbal and visuospatial internal representations for nonspeech sounds are both plausible, yet another pseudo-isomorphic auditory representational format (like a persevering version of echoic memory, see Neisser, 1967) may also be possible, despite its absence in dual-process theories.

Mikumo (1997) reported a taxonomy of encoding formats for melodies that emerged from survey research with both musicians and nonmusicians. One format was "an auditory strategy, in which pitch information was retained in an auditory modality (e.g., by singing, whistling, humming, mental rehearsal of pitches)" (p. 300). Research has confirmed that nonspeech sounds can be rehearsed (Keller, Cowan, & Saults, 1995), and this format of internal representation has been widely referred to as "auditory imagery" (Baddeley & Logie, 1992; Brodsky, Henik, Rubinstein, & Zorman, 2003; Farah & Smith, 1983; Halpern, 1988, 1992; J. D. Smith, Reisberg, & Wilson, 1992). Available evidence has suggested that this auditory pseudo-isomorphic format of representation is not unique to musicians or particular training (Schellenberg & Trehub, 2003), and behavioral data have suggested that these representations indeed preserve the analogical characteristics of a heard stimulus (Halpern, 1989; Levitin, 1994; Levitin & Cook, 1996). Biological evidence has provided corroborating evidence in favor of second order isomorphic auditory imagery, as brain areas associated with auditory perception are also recruited during imagery of sounds (Halpern & Zatorre, 1999; Halpern, Zatorre, Bouffard, & Johnson, 2004; Kraemer, Macrae, Green, & Kelley, 1995; Zatorre & Halpern, 2005).

What remains unclear is the extent to which isomorphic auditory information representations and verbal representations draw upon the same processing structures or mental resources. Some research and theory has suggested that internal representations of speech essentially drop the acoustic properties of the original stimulus (see, e.g., Samuel, 1988) and assume an amodal, verbal format of representation that is shared by visual text (Schumacher et al., 1996); (also see Mowbray, 1953) or an articulatory motor format of representation which likewise drops isomorphic acoustic properties in favor of underlying articulatory motor representations (Liberman & Mattingly, 1985; Liberman & Whalen, 2000). Whereas Baddeley (Baddeley & Logie, 1992) has argued for the phonological loop’s storage component as the explanatory mechanism of auditory imagery, a number of the direct implications of this hypothesis remain unsupported or unresolved in the literature. For example, from this perspective the concurrent processing of internal representations of speech and nonspeech audio should interfere, yet data to this effect are equivocal (for a review, see J. D. Smith et al., 1992) and findings have suggested that pitch (and other nonspeech sounds) and auditory verbal information may be processed independently (or at least without interference from concurrently verbal information) in some circumstances (Bonebright & Nees, 2009; Bonnel, Faita, Peretz, & Besson, 2001; Deutsch, 1970) .

An analogical auditory format of representation in working memory likely preceded the development of language (and therefore articulation) in humans, and it seems likely that the cognitive mechanisms for producing oral language would have piggy-backed upon existing mechanisms for analogical auditory imagery during the co-evolution of language and the perceptual decoders for spoken language (for similar arguments, see Barsalou, 1999; Gruber & Gotschke, 2004). What remains unclear, however, is the extent to which auditory imagery does or does not require articulation (a verbal, domain-specific processing structure) in human cognitive processing.

Malleability of Encoding Formats

In some theoretical approaches to information processing, auditory stimuli have been inextricably linked with the cognitive mechanisms for processing speech. Baddeley (2000) said “the visuospatial sketchpad is assumed to maintain and manipulate visual information…whereas the phonological loop performs a similar function for auditory and verbal material” (p. 127). The Baddeley working memory model, then, explicitly limited acoustic stimuli to processing by the phonological loop, whereas non-text visual stimuli assumed visuospatial representations. Multimedia learning theory (e.g., Mayer & Moreno, 1998; Mayer & Sims, 1994) made approximately the same assumptions regarding the linkage between external and internal representations, and production system theories (e.g., Kieras & Meyer, 1997; Meyer & Kieras, 1997) have similarly advocated modular approaches that link internal processes to a specific modality of input. Whereas multiple resources approaches have separated modalities from internal representation or “processing codes,” the implied (if not imperative) link between the auditory modality and verbal processing has been reflected in representative research supporting the theory (e.g., Wickens & Liu, 1988; Wickens et al., 1983b).

A number of studies have supported the idea that the internal representation of information is not dictated by the external format of the perceived stimulus. Kosslyn’s (1973, 1975, Kosslyn et al., 1978) work on visual imagery, for example, generally used simple instructions to invoke visuospatial representations in lieu of other (e.g., verbal) representational formats. Mathews, Hunt, and MacLeod (1980) showed that people can shift representational strategies at will based on instructions. Research has suggested that visuospatial internal representations (i.e., visual images) can emerge from verbal descriptions (Denis, 2007; Denis, Concalves, & Memmi, 1995; Denis & Zimmer, 1992), and also that visual percepts can be translated into verbal/propositional representations (Clark & Chase, 1972). Perhaps less well-known is research that has suggested that visuospatial representational formats can emerge from nonspeech audio (Mikumo, 1997; Nees & Walker, 2008a; Zatorre & Beckett, 1989), or that auditory imagery can emerge from visual notational representations in music (Brodsky et al., 2003) and perhaps even rhythmic visual patterns (Guttman, Gilroy, & Blake, 2005). Neuroscience studies have shown that the task requirements that invoke different listening strategies recruit different hemispheres for the processing of the same bottom-up stimulus (Brechmann & Scheich, 2005; also see Zatorre, 2003), and encoding strategies can be biologically differentiated at post-attentive (but not pre-attentive) stages of processing (Seppänen, Brattico, & Tervaniemi, 2007). Top-down strategies, then, can dictate the brain areas that process a bottom-up stimulus, and these findings offer support in favor of the arguments for malleable internal representations. The determinants of the format of internal representation seem to be not only the format of the external representation (e.g., pictures, text, speech, or sounds), but also the person's strategy for encoding the information as well as other task dependencies.

Bertolo (2005) and Lopes da Silva (2003) reviewed related studies that suggested congenitally blind people experience visuospatial representations during dreams. Evidence included EEG recordings during sleep that showed activation in the visual cortex, as well as subjective reports of visuospatial imagery during dreams. Congenitally blind people were able to produce drawings of the apparent visual images they had experienced. Bertolo suggested that congenital blindness may leave visual representational areas intact, thereby allowing for the construction of visuospatial-like representations via other modalities without a person ever having experienced visual perception.

Motivations for the Current Research

The incomplete treatment of nonspeech sound in theoretical perspectives can be attributed to a lack of interest in the topic rather than oversight, but the potential contributions of a greater understanding of nonspeech audio encoding for psychology are not trivial. The emerging field of sonification , or nonspeech auditory display (for overviews, see Kramer et al., 1999; Nees & Walker, 2009; Walker & Kramer, 2004), has sought to improve system design by harnessing the advanced sound-production capabilities of modern technology for nonspeech auditory information displays.

The most fundamental arguments for the use of nonspeech sounds in systems have been twofold. First, auditory display researchers (e.g., Stokes, Wickens, & Kite,

1990) have used a “separation-by-modality” argument that usually (directly or inadvertently) invokes the modality dichotomy of multiple resources theory (Wickens, 1984, 1991, 2002, 2008; Wickens & Hollands, 2000; Wickens & Liu, 1988; Wickens et al., 1983b). The exclusive use of visual displays in multitasking, it has been argued, can overtax the mental resources available for visual processing, but diverting some information to the auditory modality can alleviate visual overload. Whereas a number of studies have confirmed the usefulness of auditory displays when vision would otherwise be overtaxed (Brewster, 1997; Brock, Stroup, & Ballas, 2002; Brown, Newsome, & Glinert, 1989) or inappropriate (Brewster & Murray, 2000), it is important to note that this argument alone does not consider the multiple resources approach in its entirety. In particular, issues surrounding internal representations and working memory are not addressed. Second, nonspeech audio has been advocated as an appropriate auditory alternative where speech displays could interfere with actual speech communication, as speech displays could mask similar acoustic stimuli (like conversation) at peripheral sensory stages of processing (see Rossing, 1982) and disrupt selective attention to simultaneous speech (see Broadbent, 1952/1992; Cherry, 1953; Moray, 1969). This argument may hold at the level of within-modality peripheral acoustic masking and selective attention when nonspeech displays are well-designed, as nonspeech auditory displays can be designed such that acoustic masking of speech is avoided (Walker & Kramer, 2004; Watson & Kidd, 1994). The lack of evidence regarding the internal encoding format for nonspeech sounds, however, makes it unclear whether nonspeech sounds and other auditory (i.e., verbal) or visual stimuli can interfere at representational levels in working memory.

Auditory display researchers have yet to thoroughly understand internal representations and examine the possibilities for leveraging representational malleability to an operator's advantage in systems using multimodal displays. If, for example, as Mathews et al. (1980) suggested, "subjects can adopt either strategy [visual or verbal] at will" with comparable patterns of task performance for participants (p. 532), the consequences for ameliorating interference in working memory from multimodal information presentations would be considerable. Another important consideration, however, is the workload involved in translating external representational formats to a different internal format (e.g, forming a visual image from a sentence or sound). In general, research has suggested that transformations of the external representation to a different internal format take time to accomplish (Coney, 1988; Kosslyn, 1976; Tversky, 1975); also see (De Beni, Pazzaglia, & Gardini, 2007), and this may indicate that the transformation process is effortful and demanding. Cognitive load theory (see Chandler & Sweller, 1991), for example, would predict that effortful transformations (of the external stimulus to a different internal format) would have deleterious effects on activities such as learning, as cognitive resources would be invested in the encoding process. The malleability of internal representations may allow for instructions to manipulate the format of internal representations (e.g., by encoding a nonspeech sound as either a verbal or visuospatial representation--the "multiple" aspect of multiple resources approaches), but the potential benefits of averting multitasking working memory conflicts with this approach may be attenuated by the effort required to make such transformations (the "resources" aspect of multiple resources approaches, see Wickens, 2002). Mental workload has been defined as "a hypothetical construct that represents the cost incurred by a human operator to achieve a particular level of performance" (Hart & Staveland, 1988). Wickens and Hollands (Wickens, 2000, also see Wickens, 1984, 2002, 2008) linked the workload construct to the "relationship between resource supply and task demand" (p. 459). From this perspective, performance declines and workload increases when the demands of a particular task exceed the available resources for accomplishing the task.

The bulk of research on internal representations has paid little attention to workload with respect to representational transformations. If the transformation of an external representation to a different internal format demands mental resources, then the extent of the mental workload imposed warrants further investigation. The potential for elevated workload could translate to meaningful performance deficits in scenarios where a system operator is instructed to use a particular encoding strategy. This potential drawback must be weighed against any potential advantages of specifying encoding strategies to avoid interference in working memory in multitasking situations.

Summary

The background and motivations for the current studies can be summarized as follows:

1) Behavioral paradigms that have been developed for studying internal representations include mental scanning (e.g., Halpern, 1988; Kosslyn, 1973; Sternberg, 1966, 1969/2004), sentence-picture verification (e.g., Clark & Chase, 1972; Mathews et al., 1980), and dual-task methodologies (e.g., Bonnel et al., 2001; Navon & Gopher, 1979), which have diagnostic value for determining the mental resources required of a task (Ogden, Levine, & Eisner, 1979; Tsang & Wilson, 1997). Despite the successes of these paradigms at dissociating the properties of verbal and visuospatial representations,

few researchers have attempted to modify any of these paradigms to study nonspeech sounds (for exceptions, see Deutsch, 1970; Halpern, 1988, 1992).

2) To date, it remains unclear how people manipulate, rehearse, and retain information about nonspeech sounds in working memory, although these activities are clearly pervasive in human cognition (Halpern, 1989; Keller et al., 1995; Levitin, 1994; Schellenberg & Trehub, 2003). A better understanding of auditory cognition will clarify gaps in existing theory and also will inform the best practices for implementing nonspeech auditory displays.

3) The current studies applied the major behavioral paradigms of previous research on internal representations to the study of sonifications (nonspeech sounds), pictorial stimuli, and verbal representations. The first two studies demonstrated the malleability of processing codes by using mental scanning and stimulus verification to show dissociable patterns of reaction times across stimulus manipulations for each of three representational encoding strategies: visuospatial imagery, auditory imagery, and verbal representation. A final study examined these three representational strategies for a point estimation sonification task in the presence of an interference task. The third experiment attempted to offer insight into the utility of instructing specific encoding strategies to avoid working memory interference in multitasking situations.

2 EXPERIMENT 1: MENTAL SCANNING OF ENCODED SONIFICATIONS

Sternberg (Sternberg, 1966, 1969/2004) used the mental scanning procedure to make inferences about the properties of internal mental representations of verbal lists from behavioral outcomes (reaction times). Kosslyn (1973) adapted the mental scanning paradigm to study visuospatial internal representations. In the typical mental scanning trial, an internal representation is "viewed" or rehearsed in the absence of an external percept. Participants in Kosslyn's study viewed simple line drawings of objects (e.g., flowers, boats, etc.). After the visual stimulus was removed, participants were instructed to form a mental image of the previously viewed drawing. They were told to focus at a specific spatial location in the image and then scan to a different location to confirm or disconfirm the presence of a particular property. Reaction times to confirm the presence of properties varied as a linear function of the metric distance between the focus point and the property to be verified, which suggested that people were indeed scanning an internal representation that retained the analogical visuospatial properties of the external representation. Another study (Kosslyn et al., 1978) required participants to memorize a map of an island with landmarks, then to image the map and mentally scan between locations. Reaction times for the scanning task increased linearly with increasing metric distance between locations on the map, and the correlation between reaction time and actual distance on the map was $r = .97$.

Whereas Kosslyn's (1973, 1975, Kosslyn et al., 1978) early work on mental scanning most often involved an initial visuospatial stimulus that was removed and later imaged, more recent work has shown that images constructed from verbal descriptions retain metric visuospatial properties. In a series of studies, Denis and Zimmer (1992) used a variety of methods to converge on the finding that the internal representations of maps generated from texts are functionally equivalent to the analog mental images formed from viewing a picture of the described map. Of particular interest, they found that mental scanning times for traversing points in the maps generated from text increased as a linear function of the distance between points on the map, and this finding was successfully replicated (Denis, 2007). Recent neurogimaging research (Mellet et al., 2002) has confirmed that mental scanning of visuospatial representations constructed from verbal descriptions indeed recruit areas of the brain that typically are associated with visual perception—a finding that lends credence to the claim that these representations are in fact visuospatial in nature.

Halpern (1988) modified the mental scanning paradigm to demonstrate a temporal mental scanning effect for songs, which suggested an auditory analog to visual isomorphic images. In the absence of a real auditory percept, participants were asked to make two-choice judgments about the lyrical or musical content of well-known songs, and reactions times increased systematically as participants were asked to make comparisons across increasing spans of time in the songs. This result was taken as evidence that auditory imagery for songs preserved temporal relationships—an auditory parallel to the finding of preserved spatial relationships in visual imagery. Halpern's results were consistent with the possibility of an isomorphic format of internal representation for sounds, yet the examination of other formats of encoding for sound have been mostly overlooked.

If auditory frequency, through cross-modal metaphor (i.e, “weak synesthesia” see Martino & Marks, 2001) with visual spatial position (Kubovy, 1981), can be encoded in a visuospatial, domain-specific, representational module in working memory, it follows that the internal representation of a sonification encoded as a visual image should possess the same demonstrable behavioral properties as a visual image. Past research has shown that visuospatial images generated from either a verbal description (Denis, 2007; Denis & Zimmer, 1992) or a visual percept (Kosslyn, 1973, 1975; Kosslyn et al., 1978) produce patterns of reaction times during mental scanning that suggest metric spatial information is preserved in the internal representation. A visuospatial representation constructed from auditory tones should exhibit these same behavioral properties in a mental scanning task if the internal representation is indeed visuospatial in nature as anecdotal (Zatorre & Beckett, 1989) and qualitative (Mikumo, 1997; Nees & Walker, 2008a) evidence have suggested. Further, this format of representation should be behaviorally distinct from verbal representations or auditory imagery of sonifications.

Participants in Experiment 1 listened to sonifications of temperatures featuring two, three, or four data points (i.e., discrete tones); the distance between data points was varied systematically such that some sonifications featured more pronounced frequency changes (i.e., greater changes in represented value) over time. Within a block of trials, participants were instructed to encode the sounds as either a verbal list, a visuospatial image, or an auditory image. During the verbal condition, participants encoded the data points as a list of values. During the visuospatial imagery condition, participants encoded the sounds as a pictorial image of the mercury in a thermometer. The auditory imagery group was instructed to encode the sonification as they heard it, without any recoding.

Following encoding, they were given a cue to begin to “scan” their respective mental representations. The study used a 3 (encoding strategy) x 3 (number of tones) x 3 (frequency change) within-subjects design, with scanning times from the onset of the cue were recorded as the primary dependent variable.

Hypotheses

Although participants heard exactly the same sound stimuli across each block, different patterns of results were predicted based on the encoding strategy manipulations, which were expected to influence representation of the stimuli in working memory.

Hypothesis 1a

Mental scanning times for the verbal strategy were expected to be unaffected by the overall frequency change in the sonification, but were predicted to increase as a function of the number of data points—corresponding to the number of items in the set to be exhaustively scanned—in the stimulus.

Hypothesis 1b

Mental scanning times for the visuospatial imagery strategy were expected to increase as the overall frequency change increased in the sonification for a given trial, but not as a function of the number of data points. If participants made a pictorial internal representation of a thermometer from the sonifications, then the distance traversed in mentally scanning the image would be affected by the overall amount of change in frequency.

Hypothesis 1c

Sonification durations were held constant across the manipulations of frequency change and the number of tones, thus mental scanning times for the auditory imagery condition were not predicted to be affected by either the frequency change or the number of tones presented in sonifications. Previous research has suggested the auditory representations preserve the absolute temporal aspects of the perceived stimulus (Levitin & Cook, 1996), and the hypothesized flat scanning time across stimulus manipulations would differentiate this encoding strategy from verbal and visuospatial internal representations.

Method

Participants

Participants (*N* = 44, 21 females, *M* age = 19.6 years, *SD* = 1.6) were recruited from undergraduate psychology courses at the Georgia Institute of Technology and received course credit for their participation in the study. All reported normal or corrected to normal vision and hearing. A number of subject-level variables were measured as described below. The restriction of range of individual difference variables in the current sample of undergraduates was expected to preclude any strong conclusions about subject variables and strategy implementation (and individual differences were not the primary focus of the current studies), but these data were collected to look for potential explanations for encoding strategy noncompliance.

Musical Experience Questions

The influence of musical experience on performance with auditory displays has not been firmly established (for a discussion, see Nees & Walker, 2007; Watson & Kidd, 1994), but one study (Neuhoff, Knight, & Wayand, 2002) has suggested a potential influence of individual differences in musical ability on perception of frequency. A brief questionnaire queried participants regarding: 1) the number of years they have played a musical instrument; 2) their number of years of formal musical training (i.e., individual or class instruction in music); and 3) their number of years of experience reading musical notation. Participants reported a mean of 4.55 (*SD* = 3.90) years of formal musical training (i.e., private or class instruction), 4.00 (*SD* = 3.72) years of experience playing a musical instrument, and 4.41 (*SD* = 4.02) years of experience reading musical notation.

Self-reported SAT Scores

Participants who self-reported SAT verbal scores (*N* = 32) had a mean score of 637.81 (*SD* = 78.52). Participants who self-reported SAT math scores (*N* = 35) had a mean score of 708.57 (*SD* = 66.56). Participants who self-reported SAT writing scores (*N* = 25) had a mean score of 600.08 (*SD* = 136.54).

Self-reported Verbal and Spatial Ability Ratings

Mayer and Massa (2003) reported that a brief, two-item self-report rating of verbal and spatial ability—the Verbal-Spatial Ability Rating (VSAR)—captured a significant proportion of the variance associated with longer, multiple-item rating assessments, thus participants' self-report ratings for verbal and spatial abilities were collected. These ratings were expected to offer insight in the event that a participant was unable to implement a visuospatial encoding strategy, as past research has shown that some people with low spatial abilities are unable to use visual imagery effectively (Coney, 1988; Mathews et al., 1980). The mean self-estimated verbal ability rating for the sample was 3.80 (*SD* =0.63), whereas the mean self-estimated spatial ability rating was 3.89 (*SD* = 0.66). Both ability ratings were on a scale of 1 ("very low") to 5 ("very high").

Visuospatial Imagery Ability Scores

A modified version of Paivio's (1978) comparison of mental clocks test were used as an a priori indicator of imagery ability. During this brief test, participants were given pairs of times in a digitial format (e.g., "3:30" and "6:00") and were asked to indicate as quickly as possible which of the pair of times formed a smaller angle on an analog watch face. The task required mental imagery to accomplish, and response times in the original study tended to be inversely related to the angular difference on an analog clock face. Paivio's initial work on the test showed that participants categorized as "high imagers" (based on other spatial abilities measures) were reliably faster to respond across manipulations of angular difference between the two times. This test provided a brief measure that was examined for diagnostic purposes when participants could not follow instructions for the visual imagery encoding manipulation. Research on invoking imagery strategies via instructions has consistently shown that a small percentage of participants are unable to implement the visuospatial strategy as instructed (Kosslyn, 1973; Kosslyn et al., 1978; Mathews et al., 1980), and often these participants have exhibited relatively lower spatial abilities scores on psychometric tests (Coney, 1988; Mathews et al., 1980). In the current study, mean response time for correct responses on the modified mental clocks task (Paivio, 1978) was 6693.64 (*SD* = 2007.70) ms.

Cognitive Style

Research (Mayer & Massa, 2003) has examined information processing with respect to both a) verbal and spatial cognitive *abilities*, operationalized as some quantitative measure of competency, and b) verbal and spatial or visual cognitive *styles* (e.g., Kirby, Moore, & Schofield, 1988; Peterson, Deary, & Austin, 2005), operationalized as a general tendency for using one format of internal representation over another (e.g., favoring words over visual images, etc.). A one-item, self-report rating of cognitive style (the Visual-Verbal Learning Style Rating, VVLSR, from Mayer & Massa, 2003) was collected. Participants' mean self-reported cognitive style score was 2.68 (*SD* = 1.36), with a score of 1 representing a rating of "strongly more visual than verbal," a score of 4 representing a rating of "equally verbal and visual," and a score of 7 representing a rating of "strongly more verbal than visual." (Mayer & Massa, 2003) .

Auditory Imagery Ability Ratings

Given that no validated auditory imagery ability measures exist, a modified version of Seashore's (Seashore, 1919) proposed auditory imagery questionnaire was administered (see Appendix A). Participants reported a mean rating of 3.83 (*SD* = 1.03) across the eight auditory imagery questionnaire items, where a rating of 0 indicated "no image at all," a score of 3 indicated a "fairly vivid" auditory image, and a score of 6 indicated an auditory image "as vivid as actually hearing."

Apparatus

Data collection was administered with a program written with the Macromedia Director 2004 software package. Visual presentations of instructions and responses were made on a 17 in (43.2 cm) Dell LCD computer monitor. Auditory presentations were delivered via Sennheiser HD 202 headphones.

Stimuli

Sonification stimuli depicted the temperature at a weather station on a fictional planet, over the course of one day. Increasing temperatures were represented with increasing frequencies of auditory tones (Walker, 2002, 2007). The change in frequency (and its referent temperature) over the course of the day was manipulated at three levels (small, medium, and large). Small frequency changes were operationally defined as changes in one octave (from musical note C4 to C5) on the equal-tempered musical scale over the course of the day, whereas medium and large stimuli changed two (from C4 to C6) and three (from C4 to C7) octaves, respectively. Each sonification used the same note (C4) as the lower-bound anchor while systematically varying the upper bound anchor for frequency (i.e., temperature) attained during the day. Participants were told that the lower bound of the day corresponded to a starting temperature of 20 degrees Fahrenheit, and that temperature on the planet always increased, albeit to greater or lesser extents, over the course of a day. The maximum temperature value that was possible in the sonification stimuli (C7) corresponded to a temperature of 120 degrees, but participants were told that the maximum temperature was not necessarily achieved each day. Table 1 shows the values that were used as the upper and lower anchors for sonifications for the manipulation of the absolute change in frequency in sonifications.

Table 1: Operational definitions of small, medium, and large changes in frequency for sonifications of temperature on the fictional planet

	Small Δf	Medium Δf	Large Δf
Low f anchor	C4 (262 Hz)	C4 (262 Hz)	C4 (262 Hz)
High f anchor	C5 (523 Hz)	C6 (1047 Hz)	C7 (2093 Hz)

Sonifications also featured two, three, or four discrete tones. For two-tone stimuli, the tones were the anchors dictated by the change in frequency manipulation, as described above. For three-tone stimuli, a random data value between the given anchors was represented with one additional note from the equal-tempered scale. Four-tone stimuli had two notes (i.e., temperature values) in between the anchors. Participants were told that on some days, measures of temperature were sampled more frequently (i.e., three or four times), but each sonification represented the rise in temperatures over the course of only one single day. Four variations on each factorial combination of number of tones and frequency change were created to provide a variety of stimuli.

Each sonification was 800 ms in duration. Discrete tones for sonifications with two tones were 400 ms in length, and three- and four-tone stimuli used tones that were 266 and 200 ms in length, respectively. All discrete tones had 10 ms onset and offset ramps and used the MIDI piano timbre. Sonifications were designed to maintain a constant overall duration to allow for hypothesized patterns of reaction times that could differentiate auditory imagery and verbal encoding strategies.

Procedure

Participants completed the informed consent procedure and demographic questionnaires, then received a brief orientation to the overall task. The computer program explained the relationship between the notes and the temperature changes in the sonifications, and also provided a brief description of each of the three possible encoding strategies (described below) and the scanning task. Participants then experienced the verbal, visuospatial imagery, and auditory imagery encoding conditions in three separate blocks of trials. The order of encoding conditions was counterbalanced across participants. Participants knew the purpose of encoding was for a subsequent memory scanning task.

Verbal List Encoding Condition

Participants received instructions to encode the sounds as a verbal list of words—specifically a list of values, one for each tone—that named the temperatures from the beginning to the end of the day. During instructions, participants saw an example audiovisual animation that depicted a verbal list populating as a sound stimulus was heard. The instructions encouraged participants to forget about sounds and images and focus only on the list of values that they believed the sounds represented.

Visuospatial Imagery Encoding

Participants received instructions to encode the sounds as a visuospatial image—specifically a picture of a thermometer that represented temperature with a vertical line—in their minds. During instructions, participants saw an example audiovisual animation that depicted a visuospatial representation (i.e, a thermometer) forming as the sound stimulus was played. The instructions emphasized that participants were to forget about words and sounds and focus only on the image of the thermometer when encoding and remembering the temperatures for that day (i.e., that trial).

Auditory Imagery Encoding Condition

Participants received instructions to encode the sounds as a pseudo-isomorphic auditory representation by remembering and rehearsing the sonification stimulus as it was perceived. Participants were told to use pitch memory to retain the sounds exactly as they were heard—like a tape recorder in their minds. The instructions encouraged participants to focus only on the sounds.

Task and Instructions

Kosslyn (1973) cautioned that “pilot work had indicated that considerable instructional overkill was necessary to insure [sic] S’s compliance” (p. 92), and Kosslyn, et al. (Experiment 3 1978) found that even if subjects were instructed to make a visual image, they sometimes used an alternate strategy to accomplish the task. In other words, subjects must be explicitly told to consult their internal representations to accomplish the task (e.g., rather than attempting to make another representational transformation to accomplish the task). Following instructions for each block, the experimenter consulted briefly with each participant and emphasized the importance of following the encoding instructions for the block. The experimenter also confirmed through verbal self-report that participants understood the assigned encoding strategy and the scanning task. The computer program reminded participants of their assigned encoding strategy at the beginning of every trial.

On a given trial, participants listened to a sonification of the temperatures for one day on the fictional planet and encoded the stimulus according to the assigned strategy. They listened to the stimulus as many times as they wished, and this number was recorded as the dependent variable *number of times listened*. Participants indicated that they had successfully encoded the stimulus by pressing the spacebar and then saw a brief (3000 ms) blank grey screen immediately followed by a "+" centered on the screen. Participants were encouraged to rehearse their internal representations using the prescribed encoding strategy during the blank screen. The "+" cued participants to begin mental scanning of their respective representations of the stimuli. For the verbal encoding strategy, participants silently read the encoded list of values upon appearance of the "+" cue from the first value in the list to the last value in order at a fast, unchanging rate, and pressed the space bar as soon as their mental scan of the list was complete. For the visuospatial imagery condition, participants, upon seeing the "+" cue, scanned the mercury level in their thermometer visual image as if the mercury were rising at a fast, constant speed from the initial temperature value of the day without stopping until the mercury reached the height of the final temperature of the day. Participants pressed the space bar when the mercury reached the location of the ending temperature for the day in their thermometer image. Finally, in the auditory imagery condition, participants replayed the sonification in their mind (like pressing play on a tape recorder) upon seeing the "+" cue and pressed the space bar as soon as the mental recording was complete. For all conditions, the computer program recorded the time from the onset of the "+" cue until the space bar was pressed as the dependent variable *scanning time*.

Following every trial, participants identified the strategy they had used to encode and remember the sonification during the trial. Participants' choices were limited to "sound [auditory imagery] strategy," "word [verbal] strategy," "picture [visuospatial imagery] strategy," or "not sure" (see Appendix B). Participants selected at least one strategy, and they could choose more than one strategy. Marquer and Pereira (Marquer & Pereira, 1990) advocated for the self-reported corroborations of strategy use as well as an examination of patterns of reaction times in studies of internal representations. Kosslyn's mental imagery experiments (e.g., Kosslyn et al., 1978), for example, used retrospective reports on strategy compliance across a study and eliminated all data from participants who reported strategy compliance below a particular threshold (e.g., 75%), which resulted in the removal of data from 7.6%, 15.4%, 12%, and 6.3% of participants in his Experiments 1, 2, 3, and 4, respectively. Other studies that have manipulated encoding strategies reported eliminating (Reichle, Carpenter, & Just, 2000) or empirically identifying (Mathews et al., 1980) similar proportions of participants who were unable to implement visuospatial imagery encoding strategies, in particular. Dunloksy and Hertzog (1998) reviewed potential flaws in retrospective estimates of strategy implementation (e.g., forgetting) across a study or block of trials and suggested that participants should be queried about strategy use on a trial-by-trial basis. A later study (Dunlosky & Hertzog, 2001) further questioned the validity of retrospective strategy use reports, and found that trial-by-trial reports were preferable, particularly in instances where spontaneous production of strategies was not of interest. In the current study, the strategy compliance question following each trial served as a manipulation check for the encoding strategy independent variable. The trial-by-trial check of the strategy manipulation (Dunlosky & Hertzog, 1998, 2001) was used instead of a retrospective report. Since the current study *assigned* encoding strategies rather than examining spontaneously produced encoding strategies, the trial-by-trial strategy check was chosen to allow for the most precise check of the encoding strategy manipulation.

At the beginning of each of the three blocks, participants completed nine practice trials (one from each of the factorial combinations of the sonification stimulus manipulations). During the testing phase, four repetitions of each of the nine factorial combinations of frequency change and number of data points were randomly interleaved for a total of 36 experimental trials in each of the three encoding strategy blocks. At the end of each block, participants also completed the NASA-Task Load Index (NASA-TLX) (Hart & Staveland, 1988) as a measure of the subjective workload experienced in each encoding condition.

Results

When participants did not indicate use of the appropriate encoding strategy on the post-trial report screen, the participant's scanning time datum for that trial was removed from further analyses. This procedure resulted in the removal of data for 4.9% of all trials (<0.01%, 8.38%, and 5.75% of trials in the auditory imagery, verbal, and visuospatial imagery encoding conditions, respectively) . Statistical outliers—operationally defined as any datum where a participant gave a response that was 3 *SD* beyond her or his own mean scanning time for that factorial cell in the study—resulted in the removal of an additional 0.6% of trials. Thirty-nine of the 44 participants gave complete data across all conditions of the study. Participants whose data sets had empty cells following the removal of data for strategy noncompliance and statistical outliers were included in follow-up analyses for which usable (i.e., strategy compliant and statistically tenable) data were available.

Scanning Time Analyses

A 3 (encoding strategy: auditory imagery, verbal, or visuospatial imagery) x 3 (number of tones: 2, 3, or 4) x 3 (frequency change: small, medium or large—1, 2, or 3 octaves, respectively) repeated measures ANOVA was performed on the scanning time dependent variable. Greenhouse-Geisser corrections were used in all analyses where sphericity assumptions were violated. Results (see Figure 2) showed significant main effects of strategy, F (1.50,57.02) = 20.01, $p < .001$, partial $\eta^2 = .35$, number of tones, F (1.47,55.83) = 64.20, $p < .001$, partial $\eta^2 = .63$, and frequency change, F (1.35,51.24) = 6.69, $p = .007$, partial $\eta^2 = .15$, as well as significant interactions of strategy with number of tones, F (2.31,87.59) = 4.50, $p = .01$, partial $\eta^2 = .11$, and strategy with frequency change, F (1.67,63.31) = 9.49, $p = .001$, partial $\eta^2 = .20$. Nonsignificant effects included the interaction of number of tones with frequency, F (3.24,123.11) = 0.61, $p = .62$, and the three-way interaction, F (4.68,177.99) = 1.28, $p = .28$.

Follow-up pairwise comparisons showed that, collapsed across the number of tones and frequency change manipulations, the auditory imagery strategy ($M = 1432.21$, $SE = 78.09$) resulted in faster scanning times than the verbal strategy ($M = 1748.66$, $SE = 121.70$, $p = .01$) or the visuospatial imagery strategy ($M = 2362.67$, $SE = 187.43$, $p < .001$). The verbal strategy scanning times were also significantly faster than the visuospatial imagery scanning times ($p = .005$). Overall main effects should be interpreted cautiously in light of the significant interactions. The omnibus three-way analysis showed a number of effects warranting follow-up, thus analyses continued with a series of two-way ANOVAs, one at each level of the encoding strategy manipulation, to test the primary hypotheses of the study.

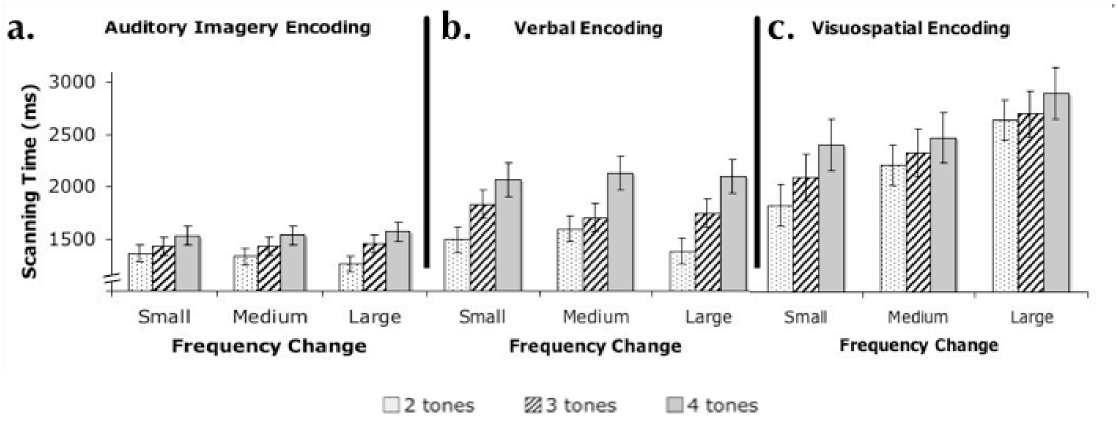

Figure 2: Mean mental scanning times as a function encoding strategy, frequency change, and the number of tones. Error bars represent standard error of the mean.

For the verbal encoding strategy, 41 participants provided full data after trials with statistical outliers or incorrect strategies were removed, so only data from these 41 participants are reported for these analyses. Hypothesis 1a predicted that scanning times under this encoding condition would be sensitive *only* to the effect of the number of tones (which corresponded to the number of temperature values to be serially—that is phonologically—scanned in participants' verbal representations). Results (see Figure 2, panel b) showed a significant main effect of number of tones, $F\,(1.38,55.15) = 54.43$, $p < .001$, partial $\eta^2 = .58$. The main effect of frequency change was not significant, $F\,(1.38,55.09) = 0.59$, $p = .50$, nor was the interaction of number of tones with frequency change, $F\,(3.35,133.85) = 1.93$, $p = .12$. For the main effect of tones, a significant linear increasing trend described the pattern of scanning times as the number of tones increased, $F\,(1,40) = 64.87$, $p < .001$, partial $\eta^2 = .62$.

For the visuospatial encoding strategy, 42 participants provided full data after trials with statistical outliers or incorrect strategies were removed, so only data from these 42 participants are reported for these analyses. Hypothesis 1b predicted that scanning times under this encoding condition would be sensitive *only* to the effect of frequency change (which corresponded to the metric distance to be scanned in participants' visuospatial images). Results (see Figure 2, panel c) showed significant main effects of number of tones, $F\,(1.46,59.95) = 9.93$, $p = .001$, partial $\eta^2 = .20$, and frequency change, $F\,(1.29,52.99) = 10.34$, $p = .001$, partial $\eta^2 = .20$. The interaction of number of tones with frequency change was not significant, $F\,(3.23,133.63) = 0.79$, $p = .51$. For the main effect of tones, a significant linear increasing trend described the pattern of scanning times as the number of tones increased, $F\,(1,41) = 12.24$, $p = .001$, partial $\eta^2 = .23$. For the main effect of frequency change, a significant linear increasing trend described the pattern of scanning times as the number of tones increased, F (1,41) = 11.60, p = .001, partial η^2 = .22.

For the auditory imagery strategy, all 44 participants provided full data after trials with statistical outliers or incorrect strategies were removed. Hypothesis 1c predicted that scanning times under this encoding condition would be faster than the other conditions and sensitive to neither the effect of frequency change nor to the effect of the number of tones, as scanning times were expected affected only by the duration of the stimuli, which was held constant across the independent variables. Results (see Figure 2, panel a) showed a significant main effect of number of tones, F (1.44,61.96) = 13.30, $p < .001$, partial η^2 = .24. The main effect of frequency change was not significant, F (1.80,77.57) = 0.10, p = .87, nor was the interaction of number of tones with frequency change, F (3.14,134.92) = 1.10, p = .35. For the main effect of tones, a significant linear increasing trend described the pattern of scanning times as the number of tones increased, F (1,43) = 17.03, $p < .001$, partial η^2 = .28.

Exploratory Analyses for the Number of Times Listened

The number of times participants listened to a stimulus during the study portion of a trial was of secondary interest with respect to the hypotheses of the current study, but these data were examined with a 3 (encoding strategy) x 3 (number of tones) x 3 (frequency change) repeated measures ANOVA. In particular, I was interested to test whether participants needed fewer stimulus presentations to encode the stimulus using the auditory imagery strategy, as this encoding format required analogical representation of the stimulus as it was heard (as opposed to the recoding of the stimulus that was required for the other encoding strategies). Greenhouse-Geisser corrections were used in all analyses where sphericity assumptions were violated. Results (see Figure 3) showed significant main effects of the number of tones, *F* (1.49,56.95) = 30.63, *p* < .001, partial η^2 = .45, and nonsignificant main effects for strategy, *F* (1.50,56.95) = 0.24, *p* = .724, and frequency change, *F* (2,76) = 1.51, *p* = .227**.** The analysis showed significant interactions of strategy with number of tones, *F* (3.05,115.90) = 3.54, *p* = .009, partial η^2 = .09, and of the number of tones with frequency change, *F* (2.53,95.97) = 5.98, *p* = .002, partial η^2 = .14. Nonsignificant effects included the interaction of strategy with frequency change, *F* (3.24,123.117) = 1.20, *p* = .312, and the three-way interaction, *F* (5.22,198.29) = 1.02, *p* = .42. Follow-up analyses continued with a series of two-way ANOVAs, one at each level of the encoding strategy manipulation.

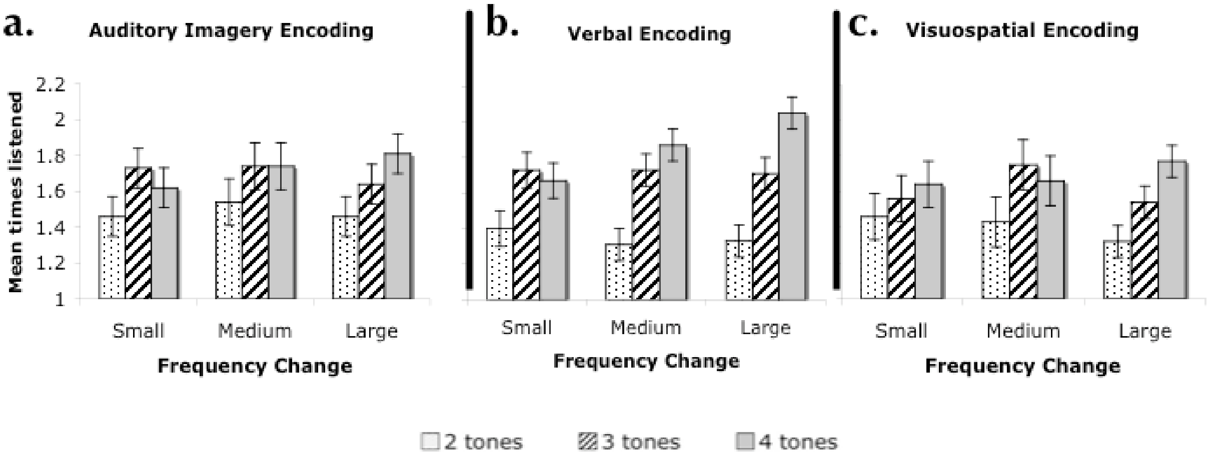

Figure 3: Mean numbers of times each sonification was listened to during encoding as a function of encoding strategy, frequency change, and the number of tones. Error bars represent standard error of the mean.

For the auditory imagery strategy, results (see Figure 3, panel a) showed a significant main effect of number of tones, $F(2,86) = 13.25$, $p < .001$, partial $\eta^2 = .24$. The main effect of frequency change was not significant, $F(1.65,70.79) = 1.53$, $p = .23$, nor was the interaction of number of tones with frequency change, $F(3.16,136.07) = 2.15$, $p = .08$. For the main effect of tones, a significant linear increasing trend described the the number of times listened as the number of tones increased, $F(1,43) = 20.21$, $p < .001$, partial $\eta^2 = .32$.

For the verbal encoding strategy, results (see Figure 3, panel b) showed a significant main effect of number of tones, $F(1.31,52.57) = 22.99$, $p < .001$, partial $\eta^2 = .37$, and a significant interaction of number of tones with frequency change, $F(3.09,123.59) = 3.48$, $p = .02$, partial $\eta^2 = .08$. The main effect of frequency change was not significant, $F(1.59,63.75) = 1.42$, $p = .25$. For the main effect of the number of tones, a significant linear increasing trend described the number of times listened as the number of tones increased, $F(1,40) = 27.08$, $p < .001$, partial $\eta^2 = .40$. The significant interaction is visible in Figure 3, panel b, where the linear increasing trend did not hold for the verbal encoding strategy when sonifications had small frequency changes.

For the visuospatial encoding strategy, results (see Figure 3, panel c) showed significant main effects of number of tones, $F(2,82) = 15.76$, $p = .001$, partial $\eta^2 = .28$. The main effect of frequency change, $F(1.97,80.90) = 1.05$, $p = .35$, and the interaction of number of tones with frequency change, $F(2.46,100.82) = 1.80$, $p = .16$, were not significant. For the main effect of tones, a significant linear increasing trend described the number of times listened as the number of tones increased, $F(1,41) = 31.25$, $p < .001$, partial $\eta^2 = .43$.

TLX Analyses

A one-way repeated measures ANOVA was performed on the NASA-TLX subjective workload scores across strategy conditions. A significant effect of strategy was found, F (2,86) = 16.96, $p < .001$, partial $\eta^2 = .28$. Paired comparisons showed that the auditory imagery encoding strategy ($M = 9.14$, $SE = 0.50$) resulted in significantly lower workload than both the verbal encoding strategy ($M = 11.61$, $SE = 0.37$, $p < .001$) and the visuospatial encoding strategy ($M = 11.25$, $SE = 0.45$, $p < .001$). The verbal and visuospatial encoding strategies were not significantly different from one another ($p = .99$).

Exploratory Correlations of Scanning Time with Subject Variables

Exploratory analyses examined the relationship between the subject level variables (musical experience data, self-reported SAT verbal scores, self-reported ratings of verbal and spatial ability and cognitive style, and the comparison of mental clocks test score) and the mental scanning time variable. These analyses are included in Appendix C. Most of the correlations were nonsignificant, likely due to restriction of range on individual difference variables in the sample, and no strong conclusions can be drawn about the relationship between individual difference variables and performance for the study tasks.

Discussion

In Experiment 1, I manipulated the encoding strategy for the sonification stimuli and also varied the stimulus properties (frequency change and number of tones) in a configuration that allow for hypothesized patterns of reaction times that would differentiate each encoding strategy.

The primary dependent variable was the mental scanning time, and dependent variables of secondary interest included the number of times listened to each sonification and the NASA-TLX measure of subjective workload for each encoding strategy.

Mental Scanning Times

Hypothesis 1 predicted that the verbal encoding strategy would show effects of the number of tones but not the frequency change, and this was confirmed. The pattern of results showed that participants seemed to be able to recode the nonspeech auditory stimuli into verbal representations in working memory. As predicted, mental scanning times increased as the number of tones (i.e., the number of items in participants' verbal representations) increased, yet the manipulation of frequency change did not affect scanning times under conditions of verbal encoding. The considerable effect of the number of tones on scanning times was consistent with past research on exhaustive scanning of verbal lists (e.g., Sternberg, 1969/2004).

Hypothesis 2 predicted that the visuospatial encoding strategy would show effects of frequency but not the number of tones, and this was partially confirmed. Unlike the verbal encoding condition, scanning times under the visuospatial encoding strategy increased as the frequency change in the sonifications increased. This effect was consistent with my predictions for participants who used auditory frequency as a cross-modal referent for visual spatial position in constructing visual images from the sonifications. Interestingly, however, mental scanning times also increased as the number of tones increased. A similar effect was found by Kosslyn et al. (1978, Experiment 1) in a mental scanning study of visual images; their study showed that mental scanning times were affected by both the distance traversed in a visual image and the number of items traversed in the mental image. Kosslyn et al. described this finding as a per-item inspection cost, whereby intervening items contributed a main effect to scanning times due to brief, compulsory pauses at each item during scanning. These results replicated this finding, albeit with a paradigm that used nonspeech auditory stimuli to inform participants' internal visuospatial representations. These findings also suggested that the effects of the number of items scanned and the distance traversed (corresponding to frequency change) were approximately equal.

Hypothesis 3 predicted that the scanning times for the auditory imagery encoding condition would match the veridical times of the sonification stimuli and show effects of neither the number of tones nor the frequency change, and this was also partially confirmed. Overall scanning times for the auditory imagery encoding condition were faster, as predicted, than the scanning times for the verbal and visuospatial encoding conditions, although the scanning times for auditory images were considerably longer than the veridical length of the sound stimuli perceived. Past research has suggested that pseudo-isomorphic internal representations of sounds maintain veridical characteristics of the stimulus (Halpern, 1989; Levitin, 1994; Levitin & Cook, 1996), including temporal aspects of the sound stimulus (Levitin & Cook, 1996). These studies, however, required the mental reproduction of longer stimuli than the 800 ms sonifications of the current study. Future research should examine mental scanning times while systematically varying the duration of stimuli. The additional time that was consistently required to scan auditory images (beyond the 800 ms length of the actual stimuli) in the current study may represent an artifact of the mental scanning task (e.g., reflective of the additional processing requirement of planning and executing a response in addition to replaying the sound in one's mind). Alternatively, the internal representation of more brief sounds may be systematically biased toward overestimating the duration of the sound that was heard.

Like the verbal and visuospatial encoding conditions, the auditory imagery encoding condition also exhibited increasing mental scanning times as the number of tones increased. Whereas the effect of the number of tones was smaller for the visuospatial and auditory imagery conditions as compared to the verbal encoding condition, the results suggested a universal, per-item scanning cost that persevered across cognitive representational formats for the mental scanning task.

Number of Times Listened

On average, participants tended to listen to sonifications with more tones more times during encoding. This finding persisted across encoding strategies and suggested that participants needed to hear the sonification stimulus more times to encode and/or recode it into an internal representation as the number of tones increased, regardless of encoding strategy. This main effect should be interpreted in light of the interaction for the verbal encoding strategy, whereby small frequency changes did not demonstrate the same linear increasing trend as the number of tones increased. Figure 3, panel b showed that participants required fewer times listening to the four-tone sonification stimuli, but only when the frequency change was small. A plausible interpretation of this finding is that the small frequency change facilitated verbal labeling, as the values (assigned to the tones) in the verbal list would be closer together and be differentiated over a smaller range. Perhaps the most obvious expected outcome for the number of times listened variable—that participants using the auditory imagery encoding strategy would need fewer times listened to encode the sonification stimuli as an auditory image—was not supported by the analyses.

Subjective Mental Workload

The NASA-TLX subjective workload scores showed lower perceived workload for the auditory imagery condition. This finding is consistent with the notion that the recoding of a stimulus from its external format requires mental effort. The auditory imagery condition required no recoding of the percept, and thus perceived workload was lower for this strategy as compared to the verbal or visuospatial encoding strategies.

Conclusions

The overall patterns of mental scanning times in Experiment 1 differentiated verbal, visuospatial, and auditory imagery encoding, and this suggested that nonspeech auditory stimuli can indeed be encoded flexibly in a variety of representational formats in working memory. An unexpected but interesting outcome of the current study was the finding of a per-item scanning cost that was universal across the encoding strategies examined here, but was most pronounced with the verbal encoding strategy. This finding replicates and expands upon the previous finding of Kosslyn et al. (Kosslyn et al., 1978) and suggests that mental scanning paradigms are sensitive (to varying degrees of effect) to the number of items present in the representation, regardless of the domain-specific format of the representation in working memory. There are a number of possible explanations for this effect, but two prominent possibilities are suggested by the literature and the pattern of results obtained here: 1) yet to be determined visuospatial and auditory rehearsal mechanisms in working memory are involved in mental scanning and operate to reinstate representations in a serial manner (i.e., with a per-item access cost), much like the articulatory mechanism of verbal working memory; and/or 2) a lingering effect of the auditory stimulus was present in addition to the effects of the encoding strategy manipulation.

Regarding the first possibility, the articulatory/phonological loop component of verbal working memory has been studied extensively, and the active rehearsal and maintenance of verbal material has been shown to occur in a serial fashion such that the time to review items in verbal memory increases as a function of the number of items to be reviewed (e.g., Baddeley, 1992, 2002; Sternberg, 1966, 1969/2004). No corresponding rehearsal mechanism for visuospatial representations has been identified. Candidate processes for the maintenance of visuospatial representations have been proposed (e.g., selective visual attention, see Awh, Jonides, & Reuter-Lorenz, 1998), (or eye movements, see Tremblay, Saint-Aubin, & Jalbert, 2006), yet other accounts have claimed that no visuospatial rehearsal mechanism exists (Washburn & Astur, 1998). Similarly, the only candidate rehearsal process that has been suggested for auditory imagery has been the same phonological loop that is involved in verbal rehearsal (Baddeley & Logie, 1992), but this proposal seems flawed given that many sounds that cannot necessarily be articulated can nonetheless be remembered (e.g., Crowder, 1993; Deutsch, 1970 ; Keller et al., 1995; Schellenberg & Trehub, 2003 also see). Until researchers better understand how analogical representations (both visual and auditory) are rehearsed and maintained in working memory, it will be difficult to rule out the possibility that these processes proceed serially (much like the rehearsal of verbal information) and are therefore subject to increasing time effects as the number of items (or perhaps the amount of visual or auditory information contained in the representation) increases. Kosslyn (1981), for example, elaborated a theoretical account of visuospatial imagery whereby a series of functions generated an analogical visual image from information stored in long-term memory, and he further theorized that a “find” function was required for both mental image retrieval/generation and mental image scanning. Such a function might reference the number of items in the original stimulus, and this type of mechanism (and a parallel operation for auditory imagery) would explain the universal per-item scanning time cost found in the current study.

A complementary interpretation, as mentioned above, involves the possibility that some trace of the initial auditory stimulus persisted into the domain-specific internal representation. Biological evidence that complements this explanation has been found in a PET study that showed that domain-specific internal representations constructed from different modalities of input showed similar patterns of neural activation (i.e., domain-specificity of the internal representation regardless of the modality of input), yet maintained distinct biological markers for the modality of input (Mellet et al., 2002). In the current study, results showed evidence of domain-specific internal representation as function of encoding strategy, yet the universal per-item cost could plausibly result from a lingering, stimulus-specific effect of the input stimulus. This interpretation is discussed further in the General Discussion.

Summary of Experiment 1

The primary findings of Experiment 1 can be summarized as follows:

1) Results generally confirmed the hypotheses that mental scanning times would differentiate distinct encoding formats. Mental scanning times under visuospatial encoding were sensitive to the frequency change manipulation, which was a metaphorical

indicator of metric space in participants' visual images. Mental scanning times under verbal encoding were sensitive to the number of tones manipulation, which corresponded to the number of items in verbal lists. Mental scanning times under auditory imagery encoding were fastest overall (compared to the other encoding strategies) and closest to the veridical length of the external stimulus, as would be expected with little or no recoding.

2) Across all 3 encoding strategies, participants's mental scanning times and the number of times they listened to the initial stimulus were sensitive to the number of tones manipulation. This unexpected but interesting finding suggests a universal per-item scanning cost that persists across encoding strategies. One plausible explanation of this finding is that rehearsal/scanning mechanisms across all 3 types of representation require some process that serially generates or scans the representation. Another possible explanation is that the initial external stimulus had a lingering effect in addition to the observed effects of encoding strategy.

3) Subjective workload was lowest under conditions of auditory imagery encoding, as the external stimulus did not have to be recoded into a different internal format.

3 EXPERIMENT 2: A SOUND-SENTENCE-PICTURE VERIFICATION TASK

The sentence-picture verification task was originally used by Clark and Chase (1972) to derive a predictive model of sentence-picture comparisons. Clark and Chase presented a simple sentence (e.g., “star is above plus”) *simultaneously* with a pictorial representation that was either consistent or inconsistent with the sentence (see Figure 4, panels A and B). Participants performed a two-choice reaction time task, whereby they either confirmed or denied that the sentence described the picture. The researchers varied the linguistic complexity of the sentence stimulus by including negations (e.g., “the star is not above the plus”). Clark and Chase arrived at a model of sentence-picture verification that seemed to accurately fit their data; their model presupposed a common, immutable, propositional format of internal representation for both sentences and pictures, whereby participants converted pictures to propositional representations to perform comparisons.

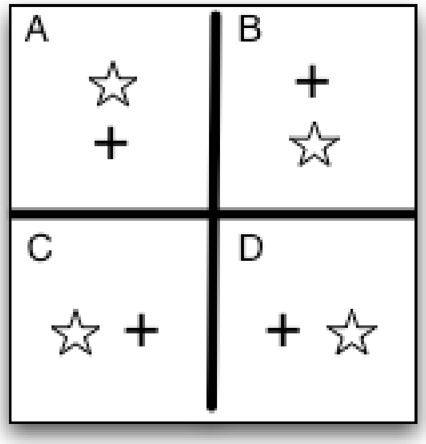

Figure 4: Representative picture stimuli used in various studies using the sentence-picture verification task.

Clark and Chase's (1972) propositional model seemed to fit their data well for the simultaneous-presentation version of the sentence-picture verification task, but often the comparison of two stimuli in different formats occurs after a delay, whereby the first stimulus must be temporarily held in working memory. Tversky (1975) was one of the first researchers to empirically demonstrate the use of a visuospatial imagery (also called "pictorial" or "iconic") strategy in sentence-picture verification. Tversky's data suggested that the imagery strategy required a successive (rather than simultaneous) presentation of the sentence followed by the picture. With a 5 s delay between sentence and picture, linguistic complexities (e.g., verbal negations) in the sentence no longer impacted reaction times for the picture, yet linguistic manipulations showed effects when the picture appeared at the same time as the sentence. Additionally, subjects' response times were much faster in the successive presentation condition (808 ms as compared to 2168 ms). In a study that complemented Tversky's work, Glushko and Cooper (1978) manipulated the time allowed during the comprehension (also referred to as "preparation" or "study") phase of the task (when the sentence was being studied and internalized) and found that, consistent with a visuospatial imagery stategy, subsequent picture verification reaction times decreased as participants were allowed more comprehension time (also see Coney, 1988).

Macleod, Hunt, and Mathews (1978) further qualified the conditions under which participants used a visual imagery strategy in sentence-picture verification. They allowed preparation or comprehension times for the initial stimulus to vary as a subject-controlled variable, which was measured along with verification time. Individuals' data patterns were examined against the predicted pattern of reaction time results from a propositional model, specifically the constituent comparison model proposed by Carpenter and Just (1975). Whereas the overall group averages fit the propositional model fairly well, individual data patterns revealed a subgroup of participants whose data patterns did not match the predicted patterns of a propositional strategy. This subgroup showed much longer comprehension (study) times, much shorter verification times, and no effects of linguistic complexity on verification times, all of which suggested that this subgroup were spontaneously forming a mental image of the sentence to compare to the picture during verification. This inference was corroborated by the additional finding that the pictorial strategy subgroup exhibited significantly higher psychometric test scores for spatial ability as compared to the subgroup using the propositional strategy.

Whereas the findings of Macleod et al. (1978) might have been interpreted to suggest pictorial representation as a skill only available to those with high spatial ability, a follow-up study suggested otherwise. Mathews, Hunt, and MacLeod (1980) used a sentence-picture verification task to examine internal representations and their flexibility. On the first day of their three-day study, participants performed the sentence verification task under spontaneous (i.e., non-prescribed) strategy conditions. Again, whereas most participants exhibited patterns of comprehension and verification reaction times that were consistent with the predictions of the propositional constituent comparison model, a subset of participants showed the longer comprehension times and shorter verification times that would be expected with a visual imagery strategy. On days two and three of their study, participants were instructed to use either a linguistic (propositional) or a visual imagery strategy to accomplish the sentence-picture verification task. The study used a within-subjects design and counterbalanced the instructional manipulation on the second and third days. The data suggested that participants were in fact able to control and adjust their representational strategies as prescribed. Participants, when given a propositional strategy, produced patterns of comprehension and verification reaction times that were extremely consistent with the predictions of the constituent comparison model (Carpenter & Just, 1975). Mean data under conditions of pictorial instruction conformed nearly perfectly to the predicted reaction time patterns of an imagery strategy, whereas the mean fit from the spontaneous condition of day one fell in between but tended to resemble the propositional strategy—exactly as would be predicted with heterogeneous strategies where a majority of participants using a propositional approach.

Remarkably, then, Mathews et al.'s (1980) data suggested that a simple instructional manipulation could induce shifts in the format of internal representation of information, regardless of participants' spontaneously preferred encoding strategy. A few caveats to this result, however, should be mentioned. Under instructions to use the imagery strategy, participants with high spatial ability spent significantly longer (817 ms, on average) studying the sentence during the comprehension phase, and were also faster (by 134 ms on average) at responding during the verification phase of the task. Also, some participants (6 out of 32) seemed to be altogether unable to effectively implement an imagery strategy. No such parallel was found for the linguistic strategy, which was readily acquired by all participants.

Kroll and Corrigan (1981) twice replicated (Experiment 1 and Experiment 2, Block 1) the pattern of results observed by MacLeod et al. (1978) and further qualified the conditions under which a visuospatial representation strategy may be spontaneously adopted. Specifically, the finite, two-item stimulus set typically employed in the traditional sentence-picture verification task allowed participants to infer that a negative sentence (e.g., “the star is not above the plus”) implied that the mental image of “plus above star” allowed for the quickest comparison with the verification (picture) stimulus. For some blocks in their studies, Kroll and Corrigan introduced two unexpected alternative images to the stimulus set: the horizontal variations depicted in Figure 3, panels C and D. With four possible stimuli, a given sentence could not be recoded into a single image (during comprehension) that captured the range of possible verification picture stimuli. Results showed that when the task demands and stimuli were such that an imagery strategy became maladaptive, most subjects who used the imagery strategy would spontaneously switch to a propositional strategy. Further, some who preferred an imagery strategy produced results that suggested they could make shifts in strategy on a block-to-block basis based on the size of the verification stimulus set; they would use an imagery strategy when the verification set had two members and a propositional strategy when the verification set had four members. Related work has shown that when participants are not given a specific encoding strategy for the comprehension phase of a sentence-picture verification task, they seem to actively adapt their strategy based on tasked demands, such as the expected format of the verification stimulus (Noordzij, van der Lubbe, & Postma, 2005).

Another variation of the task has compared performance across sentence-picture and picture-picture verification tasks (Noordzij, van der Lubbe, Neggers, & Postma, 2004). Two distinct patterns of reaction times emerged. Participants who were thought to be using a visual imagery strategy showed the same pattern of verification reaction times for sentence-picture and picture-picture verification, whereas those using a propositional strategy were markedly faster at picture-picture as compared to sentence-picture verification, presumably due to the time needed to recode the second picture into a proposition when it was preceded by a sentence.

The finding of flexibility in representational format via instructional manipulations has been corroborated with biological evidence. Reichle, Carpenter, and Just (2000) found that, for an overwhelming majority of participants, fMRI techniques could reliably differentiate the strategy (verbal or visuospatial) that had been prescribed and was being used by participants across blocks based on a within-subjects instructional manipulation. Verbal strategies were characterized by relatively higher levels of activity in Broca's area, and visuospatial strategies showed relatively higher activation in parietal areas associated with visuospatial processing. Interestingly, their work also confirmed biological distinctions based on psychometric tests scores—decreased fMRI activation in verbal and visuospatial processing areas was respectively associated with higher verbal or visuospatial abilities.

Taken together, the collective results of three decades of sentence-picture verification tasks have suggested that: 1) development of a mental image requires time for the image to form or be instated in working memory (Coney, 1988; Tversky, 1975); 2) when no time is allowed for an image of the sentence to form, participants will spontaneously use a propositional strategy as evidenced by the persistence of effects associated with linguistic complexities (Clark & Chase, 1972); 3) some subjects will spontaneously employ a visual imagery strategy if the time between comprehension and verification allows for the formation of an image (Tversky, 1975), and this may be especially true of subjects with high spatial ability (Coney, 1988; MacLeod et al., 1978);

and 4) verification reaction time is faster when the external representation (e.g., a picture) of the verification stimulus matched participants' internal representation (e.g., a mental image) of the remembered stimulus (Glushko & Cooper, 1978; MacLeod et al., 1978).

Despite the convergent and highly consistent pattern of results that have emerged from using sentence-picture tasks to study internal representations, surprisingly few modifications to the original stimuli and format of the task have been attempted. The fundamental reasoning of stimulus verification paradigms could be adapted to study the encoding and comparison of a great variety of stimuli, yet studies to date have generally made only minor departures from the task as it was presented by Clark and Chase (1972) to study sentence comprehension processes.

Experiment 2 modified the sentence-picture verification procedure to examine internal representations for sounds, sentences, and pictures in a within-subjects design. Stimuli depicted fictitious stock prices, and the encoding strategy was manipulated at three levels (auditory imagery, visuospatial imagery, and verbal encoding) in different blocks of the study via instructions. The external format of the stimulus was also varied at three levels (sentences, pictures, and sounds) within each block for both participant-controlled preparation time (i.e., study or encoding) and verification time (i.e., deciding whether the second "verification" stimulus matched the encoded stimulus). Trials consisted of: 1) a preparation or study period for a stimulus (either a sentence, a picture, or a sound), during which participants were instructed to encode the study stimulus as an auditory image, a visuospatial image, or a verbal representation, regardless of the study stimulus format; 2) a brief delay; and 3) the presentation of a verification stimulus (either a sentence, a picture, or a sound) to which participants made a speeded comparison of their respective internal representations to confirm or disconfirm a match with the verification stimulus's depicted state for the stock (i.e., either "price increased" or "price decreased").

Hypotheses

Hypothesis 2a

Across all three formats of internal representation, study times for the initial stimulus were predicted to be relatively faster when the external format of the study stimulus matched the instructed internal format (i.e., when no recoding was required). Participants who were instructed to encode the stimulus as an auditory image, for example, were expected to need less study time when the external stimulus was already a sound as compared to when the external stimulus was verbal (i.e., a sentence) or visuospatial (i.e., a simple graph) and had to be recoded into an auditory image.

Hypothesis 2b

Regardless of the external format of the stimulus during preparation, verification times of the stock as increasing or decreasing were predicted to be fastest when the external verification stimulus format matched the internal representation format used in working memory during the study period. If the external verification stimulus was a sentence, participants were expected to respond faster under the verbal encoding condition (as compared to the visuospatial or auditory imagery encoding conditions). If the external verification stimulus was a picture, participants were expected to respond faster under the visuospatial imagery encoding condition (as compared to the verbal or auditory imagery encoding conditions). If the external verification stimulus was a sound, participants were expected to respond faster under the auditory imagery encoding condition (as compared to the verbal or visuospatial imagery encoding conditions).

Method

Participants

Participants (*N* = 39, 13 females, *M* age = 20.0 years, *SD* = 1.7) were recruited from undergraduate psychology courses at the Georgia Institute of Technology and received course credit for their participation in the study. All reported normal or corrected to normal vision and hearing.

Demographic and subject-level variables were the same as those in Experiment 1. Participants reported a mean of 3.92 (*SD* = 4.38) years of formal musical training (i.e., private or class instruction), 3.26 (*SD* = 4.00) years of experience playing a musical instrument, and 3.72 (*SD* = 4.15) years of experience reading musical notation. Participants who self-reported SAT verbal scores (*N* = 31) had a mean score of 619.1 (*SD* = 81.7). Participants who self-reported SAT math scores (*N* = 33) had a mean score of 692.1 (*SD* = 62.7). Participants who self-reported SAT writing scores (*N* = 22) had a mean score of 671.4 (*SD* = 114.0). The mean self-estimated verbal ability rating for the sample was 3.74 (*SD* =0.82), and the mean self-estimated spatial ability rating was 3.72 (*SD* = 0.86). Both ability ratings were on a scale of 0 to 5. Participants' self-reported cognitive style score was 2.46 (*SD* = 1.14), with a score of 1 representing a rating of "strongly more visual than verbal," a score of 4 representing a rating of "equally verbal and visual," and a score of 7 representing a rating of "strongly more verbal than visual." (Mayer & Massa, 2003) Participants reported a mean rating of 3.82 (*SD* = .81) across the auditory imagery questionnaire items, where a rating of 0 indicated "no image at all," a score of 3 indicated a “fairly vivid” auditory image, and a score of 6 indicated an auditory image “as vivid as actually hearing.” Mean response time for correct responses on the modified mental clocks task (Paivio, 1978) was 6670.81 (*SD* = 2146.77) ms.

Apparatus

The apparatus were the same as described in Experiment 1.

Stimuli

Sound Stimuli

Sound stimuli consisted of two-tone sonifications that represented the opening and closing price of a stock over the course of a trading day. Like in past research using sentence-picture verification (Carpenter & Just, 1975; Clark & Chase, 1972; Coney, 1988; Glushko & Cooper, 1978; Kroll & Corrigan, 1981; MacLeod et al., 1978; Mathews et al., 1980), only a limited stimulus set was required. Sound stimuli for Experiment 2 used only two discrete tones—C4 (262 Hz) and C5 (523 Hz)—for each sonification. Each tone was 100 ms in length (with 10 ms onset and offset ramps) and synthesized with the MIDI piano instrument. For a given trial with a sonification stimulus, the sonification represented either an increasing stock price (C4 followed by C5) or a decreasing stock price (C5 followed by C4).

Pictorial Stimuli

Pictorial stimuli featured unlabeled line graphs that depicted the trend of the price of the stock over the course of the trading day. Two line graphs were used—one that showed the stock price increasing and one that showed the stock price decreasing (see Figure 5).

Figure 5: Examples of pictorial stimuli depicting the stock price increasing (left panel) and the stock price decreasing (right panel).

Verbal Stimuli

Verbal stimuli described the trend of the stock over the trading day with a two-word phrase (i.e., "price increased" or "price decreased"). These stimuli were presented in large (approximately 40 point) font at the center of the screen.

Procedure

After the informed consent procedure, demographic data were collected, and participants completed the battery of subject-level measures. Participants received instructions about the format of the task; participants were told that they would need to remember the state of the stock in the first (study) stimulus for later comparison with the second (verification) stimulus. All participants were given 36 instructional trials without a prescribed encoding strategy (i.e., under spontaneous strategy conditions) as an introduction to the sound-sentence-picture verification task. Participants then experienced 72 trials of each encoding strategy, as described below, and the order of the encoding strategies was counterbalanced across participants.

Visuospatial Imagery Encoding Block

During this block, participants were instructed to encode the study stimulus as a simple visuospatial image, like a graph that represented greater stock price as higher up on the Y-axis (see Figure 5). During instructions, participants were shown example audiovisual animations that depicted the visual graph representations as sound stimuli were heard or as sentences were read. Instructions emphasized that an image was to be formed during the study portion of a trial for every trial within the block, regardless of whether the given external representation was a sound, picture, or sentence.

Verbal (Sentence) Encoding Block

During this block, participants were instructed to encode the study stimulus as a sentence that stated either "price increased" or "price decreased." During instructions, participants were shown an example audiovisual animation that depicted a sentence as a sound stimulus was heard or as a picture stimulus was viewed. Instructions emphasized that a sentence was to be formed during the study portion of a trial for every trial within the block, regardless of whether the given external representation was a sound, picture, or sentence.

Auditory Imagery Encoding Block

During this block, participants were instructed to encode the study stimulus like the two note sonification stimuli, with pitch increasing (note C4 followed by C5) if the stock price increased or pitch descending (note C5 followed by C4) if the stock price decreased. During instructions, participants were shown an example audiovisual animation that played a sonification as a sentence or a picture stimulus was viewed. Instructions emphasized that an auditory image was to be formed during the study portion of a trial for every trial within the block, regardless of whether the given external representation was a sound, picture, or sentence.

Instructions and Task

At the beginning of each of the three blocks of the encoding strategy manipulation, all participants gave verbal confirmation to the experimenter (following the instructions and before the test trials began) that they understood the encoding strategy. The trial format was modeled after Mathews et al. (1980). Participants were instructed to keep their left index finger on the “Z” key and their right index finger on the “?” key, and they were told to press either key to begin a trial. A “Z” or “?” keypress initiated the study stimulus (either a sonification, picture, or sentence), and participants encoded the study stimulus in their assigned representational formats. Participants kept their left index finger on the “Z” key and their right index finger on the “?” key throughout the trial, and they pressed either key whenever they had encoded the stimulus in the prescribed format. Following the keypress, the verification stimulus appeared after a 3000 ms delay that showed a blank grey screen. Participants’ task was to press the “Z” key if the state depicted in the verification stimulus (i.e., stock price increased or stock price decreased) matched the state of their encoded representations, or to press the “?” key for mismatches. The mapping of keys (i.e., left and right index finger responses) to confirmation responses was counterbalanced across participants. Participants were instructed to achieve both encoding and verification as quickly as possible while following encoding instructions and avoiding errors, and participants received feedback about their reaction time following each trial.

For each factorial combination of the three encoding stimulus formats, three verification stimulus formats, and two possible states for each format (stock price increased or decreased)[1], a total of 36 possible pairings of stimuli were possible. The first 36 spontaneous trials were considered practice trials, and these practice trials sampled the full spectrum of factorial stimulus combinations. Following the spontaneous block of trials, participants experienced 72 trials (2 repetitions of the 36 possible stimulus combinations) using each of the three encoding strategies. Participants underwent a total of 252 trials. Presentation of the 36 possible stimulus combinations within a block were randomly interleaved.

The first dependent variable of Experiment 2 was *study time*, operationally defined as the time from the beginning of the presentation of the encoding stimulus on a given trial (initiated by the first “Z” or “?” keypress) until the participant presses the “Z” or “?” key to indicate that they had formed their internal representation of the stimulus. The second dependent variable was the *verification time*, operationally defined as the time from when the verification stimulus appeared until the participant pressed a response key to confirm or disconfirm that the encoding and verification stimuli matched. In addition to the encoding and verification dependent variables, participants completed the NASA-TLX as a measure of subjective workload for their respective encoding conditions collapsed across the format of study and verification stimuli.

---

[1] It is well-established in sentence-picture verification that verification times are faster when the study stimulus matches the verification stimulus than when the stimuli do not match (Reichle et al., 2000; Tversky, 1975). While this effect was expected, it was not of interest to the current study. As such, the number of matching and mismatching trials were balanced across the manipulations of interest here, and all analyses were collapsed across this variable.

Results

Missing Data and Data Excluded for Strategy Noncompliance

One participant was unable to implement the prescribed encoding strategy across all conditions of the study and was excluded case-wise from further analyses. Appendix D shows the participant's subject-level data. Although the subject-level data of one participant is inadequate to draw any strong conclusions about the role of the subject variables in predicting strategy compliance, it is interesting to note that the participant reported no musical experience and "very low" self-ratings of verbal and spatial ability, which may explain the participant's inability to adhere to any of the strategies across the study. Four additional participants were excluded from one or more sets of analyses for strategy noncompliance, but the data from these participants were included in all analyses where strategy-compliant data were available.

Excluding the participant who was dropped case-wise, strategy noncompliance data elimination proceeded on a trial-by-trial basis for the remaining 38 participants. Data for 3.1% of trials across the study were excluded because a participant indicated she or he did not use the instructed encoding strategy. Data from individual trials were considered outliers and excluded from intracondition averages if the participant's verification response was incorrect or if the response time was greater than 3 standard deviations from their individual means for a given factorial condition (see Mathews et al., 1980), and this led to the elimination of an additional 7.6% of data across all trials.

Results for Study Time

A 3 (encoding strategy: verbal, visuospatial imagery, or auditory imagery) x 3 (study stimulus format: sentence, picture, or sound) repeated measures ANOVA was performed on the study time dependent variable. Greenhouse-Geisser corrections were used in all analyses where sphericity assumptions were violated. Results showed significant main effects of strategy, *F* (2,66) = 4.27, $p = .018$, partial $\eta^2 = .12$, study stimulus format, *F* (1.40,46.19) = 18.72, $p < .001$, partial $\eta^2 = .36$, and the interaction, *F* (2.86,95.19) = 3.03, $p = .035$, partial $\eta^2 = .04$. Follow-up analyses proceeded to examine the hypotheses in the simplest, most direct, and most powerful method with a series of single degree of freedom contrasts.

Hypothesis 2a predicted relatively faster study times when the format of the external study stimulus matched the prescribed encoding strategy format. Study times were expected to be longer when the external study stimulus had to be recoded into a different format. A series of single degree-of-freedom contrasts compared study times as a function of strategy for each type of study stimulus. Planned comparisons tested the hypotheses that the format of the study stimulus would show significantly faster study times for encoding strategy that matched the study stimulus format as compared to the other two encoding strategies. Results of the planned comparisons for the study time dependent variable are shown in Table 2 and depicted in Figure 6.

Table 2: Contrasts for study times as a function of study stimulus

Encoding strategy comparison	Mean study time difference in ms (*SE*)	df	F	p	partial η^2	N
	Sentence study stimulus					
Verbal vs visuospatial*	-373.22(139.9)	(1, 35)	7.12	.01	.17	36
Verbal vs auditory*	-278.34(129.4)	(1, 35)	4.63	.04	.12	36
	Picture study stimulus					
Visuospatial vs verbal	221.16(128.2)	(1, 34)	2.98	.09	--	35
Visuospatial vs auditory	14.35(131.3)	(1, 34)	0.01	.91	--	35
	Sound study stimulus					
Auditory vs verbal	-1.29(104.9)	(1, 35)	<0.01	.99	--	36
Auditory vs visuospatial*	-397.62(165.2)	(1, 35)	6.03	.02	.15	36

* = $p < .05$

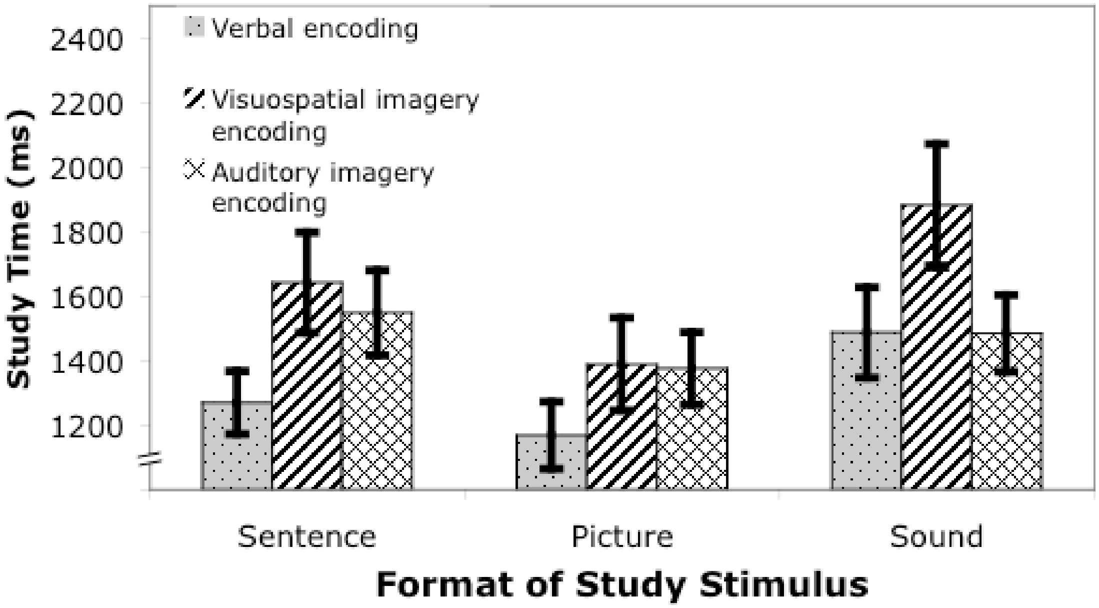

Figure 6: Results for study time as a function of the study stimulus and encoding strategy in Experiment 2. Error bars represent standard error.

Planned contrasts confirmed Hypothesis 2a only for the sentence study stimulus, as the verbal encoding strategy resulted in significantly faster study times than either visuospatial or auditory imagery encoding strategies. When the study stimulus was a picture, the visuospatial encoding strategy was not significantly faster than the verbal or auditory imagery encoding strategies, and in fact the verbal encoding strategy resulted in significantly faster study times for pictures than the visuospatial encoding strategy at a marginal (p = .09) level of significance. Finally, when the study stimulus was a sound, the auditory imagery encoding strategy was significantly faster than the visuospatial encoding strategy, but the auditory imagery encoding strategy was not significantly faster than the verbal encoding for sounds.

Omnibus Results for Verification Time

A 3 (encoding strategy: verbal, visuospatial imagery, or auditory imagery) x 3 (study stimulus format: sentence, picture, or sound) x 3 (verification stimulus format: sentence, picture, or sound) repeated measures ANOVA was performed on the scanning time dependent variable. Greenhouse-Geisser corrections were used in all analyses where sphericity assumptions were violated. Results showed significant main effects of study stimulus format, F (2,64) = 4.65, p = .015, partial η^2 = .13, verification stimulus format, F (1.54,49.26) = 23.89, p < .001, partial η^2 = .43, the interaction of strategy with verification stimulus format, F (2.97,95.00) = 24.41, p < .001, partial η^2 = .43, and the interaction of study stimulus format with verification stimulus format, F (3.01,96.37) = 8.09, p < .001, partial η^2 = .20. Nonsignificant effects included the main effect of strategy, F (1.69,54.04) = 2.98, p = .068, the interaction of strategy with study stimulus format, F (3.27,104.71) = 0.40, p = .77, and the three-way interaction, F (5.66,181.13) = 1.28, p = .27. Follow-up analyses again proceeded with a series of contrasts that examined the specific hypotheses regarding verification time.

Results for Verification Time as a Function of Encoding Strategy and Verification Stimulus

Hypothesis 2b predicted that the verification times would be fastest for a given verification stimulus when it matched the format of participants' encoding strategies. This pattern of results was predicted regardless of (i.e., collapsed across) the format of the study stimulus. Consequently, for each verification stimulus format, the verification times for the two verification stimuli that did not match the encoding strategy were compared to the format that matched the encoding strategy. Results are presented in Table 3 and Figure 7.

Table 3: Contrasts for verification times as a function of encoding strategy and verification stimulus

Encoding strategy comparison	Mean verification time difference in ms (*SE*)	df	F	p	partial η^2	N
	Sentence verification stimulus					
Verbal vs. visuospatial*	-139.82(47.4)	(1, 34)	8.70	.006	.20	35
Verbal vs auditory*	-180.84(40.4)	(1, 34)	20.00	<.001	.37	35
	Picture verification stimulus					
Visuospatial vs verbal	- 70.66(43.6)	(1, 35)	2.63	.114	--	36
Visuospatial vs auditory*	-239.02(50.7)	(1, 35)	22.23	<.001	.39	36
	Sound verification stimulus					
Auditory vs verbal	- 64.84(36.7)	(1, 35)	3.12	.086	--	36
Auditory vs visuospatial*	-200.21(65.3)	(1, 35)	9.40	.004	.21	36

* = $p < .05$

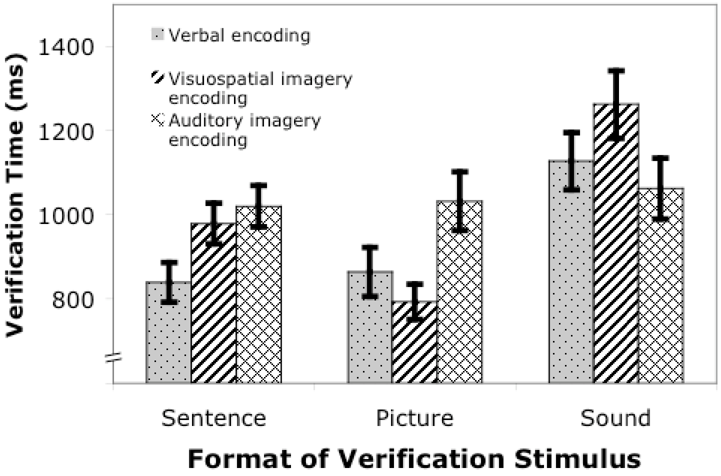

Figure 7: Results for verification time as a function of encoding strategy and the format of the verification stimulus in Experiment 2. Error bars represent standard error of the mean.

The overall pattern of results for the analysis of verification times as a function of encoding strategy and verification stimulus (collapsed across the format of the study stimulus) generally confirmed the predictions of Hypothesis 2b, with the exception that two of the planned comparisons were not significant. When the verification stimulus was a sentence, participants verified the sentence stimulus faster using the verbal encoding strategy (which matched their internal representation) than when they used the visuospatial or auditory imagery encoding strategies. When the verification stimulus was a picture, participants verified the picture stimulus significantly faster using the visuospatial imagery strategy than when they used the auditory imagery encoding strategy, but their verification times were not statistically faster (*p* = .114) for the visuospatial as compared to the verbal encoding strategy. When the verification stimulus was a sound, participants verified the sound stimulus significantly faster using an auditory imagery encoding strategy than a visuospatial encoding strategy, but their verification times were not statistically faster (*p* = .086) as compared to the verbal encoding strategy.

Results for Verification Time as a Function of Study Stimulus

Hypothesis 2b also predicted that verification times would not differ when the format of the verification stimulus (collapsed across strategies) matched the format of the study stimulus, as the participant was expected to have recoded the original stimulus. These analyses are reported in Table 4 and depicted in Figure 8.

Table 4: Contrasts for verification times as a function of study stimulus and verification stimulus

Study stimulus	Mean verification time difference in ms (*SE*)	df	F	p	partial η^2	N
	Sentence verification stimulus					
Sentence vs picture	-9.66(17.97)	(1,34)	0.62	.44	--	35
Sentence vs sound	-22.61(28.79)	(1,34)	0.29	.59	--	35
	Picture verification stimulus					
Picture vs sentence	-30.77(26.0)	(1,35)	1.40	.24	--	36
Picture vs sound	-42.09(25.1)	(1,35)	2.81	.10	--	36
	Sound verification stimulus					
Sound vs sentence*	-144.91(40.7)	(1,35)	12.68	.001	.26	36
Sound vs picture*	-184.20(44.1)	(1,35)	17.45	<.001	.33	36

* = $p < .05$

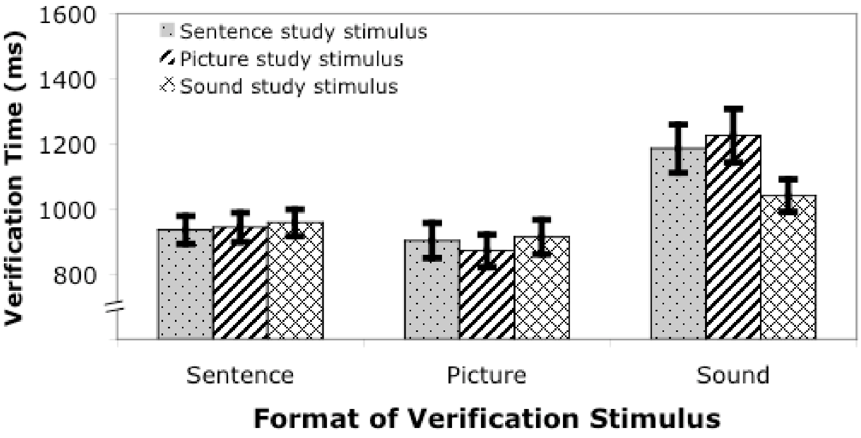

Figure 8: Results for verification time as a function of the format of the study stimulus and the format of the verification stimulus in Experiment 2. Error bars represent standard error of the mean.

The predicted pattern of results held for both sentence and picture verification stimuli, as neither showed faster verification times when the format of the verification and study stimuli matched. Interestingly, however, when the verification stimulus was a sound, there was a significant tendency to respond faster when the study stimulus was also a sound, regardless of the encoding strategy that was used.

Exploratory Analysis of the Spontaneous Strategy Practice Block

The first practice block of the sound-sentence-picture verification task featured 36 trials that allowed the participants to practice the task under spontaneous conditions without a prescribed encoding strategy or explicit instructions regarding the possible strategies. A set of contrast analyses looked at the study times during the practice block as a function of the format of the study stimulus. Results of these analyses are listed in Table 5.

Table 5: Study times during the practice block as a function of the study stimulus

Study stimulus comparison	Mean study time difference in ms (S.E)	df	F	p	partial η^2
Sentence vs picture	211.15(122.9)	(1,37)	2.95	.094	--
Sentence vs sound*	- 555.5(172.1)	(1,37)	10.42	.003	.22
Picture vs sound*	-766.7(156.2)	(1,37)	24.10	<.001	.39

* = $p < .05$

The analysis showed that encoding time was significantly slower when the study stimulus was a sound. Spontaneous strategy implementation most likely resulted in heterogeneous strategy use across participants, as the results follow the same pattern as the results of Figure 6 (study times during test trials a function of encoding strategy) if the Figure 6 results were to be collapsed across encoding strategies.

The verification times for the practice block were analyzed as a function of the study stimulus. Contrasts that parallel the same analyses for the test blocks (collapsed across study strategy—see Table 4 and Figure 8) are shown in Table 6.

Table 6: Contrasts for verification times as a function of study stimulus and verification stimulus during the practice block

Study stimulus comparison	Mean verification time difference in ms (*SE*)	df	F	p	partial η^2	N
	Sentence verification stimulus					
Sentence vs picture	-35.37(79.1)	(1, 38)	0.62	.20	--	39
Sentence vs sound	62.36(68.8)	(1, 38)	0.21	.82	--	39
	Picture verification stimulus					
Picture vs sentence	-86.58(66.6)	(1, 38)	1.69	.20	--	39
Picture vs sound	-5.43(46.5)	(1, 38)	1.39	.25	--	39
	Sound verification stimulus					
Sound vs sentence*	-429.12(137.7)	(1, 38)	9.71	<.01	.21	39
Sound vs picture	-192.71(137.4)	(1, 38)	1.97	.17	--	39

*$p < .05$

This analysis for the practice block showed that only one significant difference, as sound verification stimuli were responded to reliably faster when the study stimulus had been a sound as compared to a sentence. Those results are pictured in Figure 9.

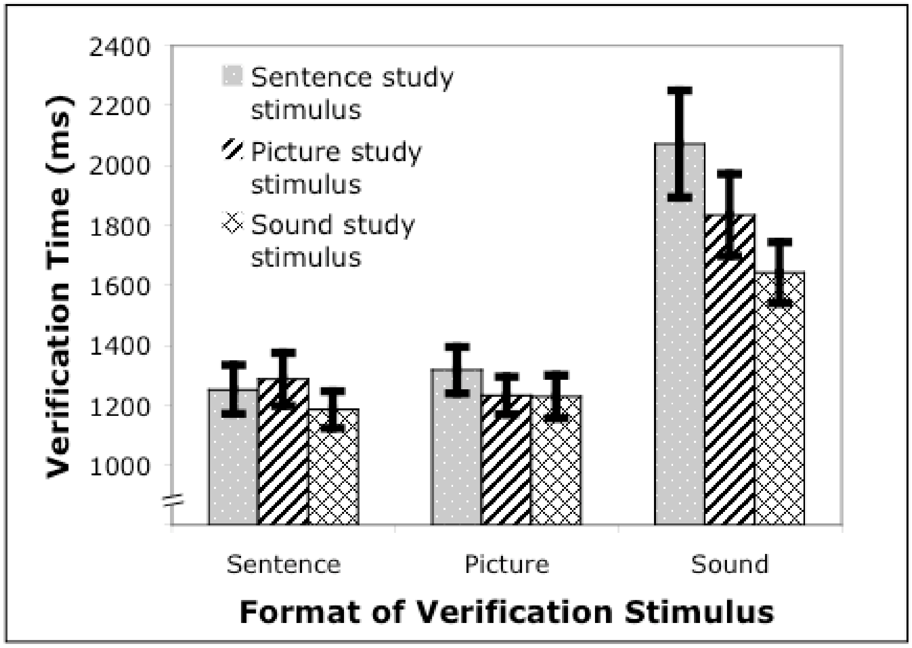

Figure 9: Results for verification time as a function of the format of the study stimulus and the format of the verification stimulus in the practice block of Experiment 2. Error bars represent standard error.

NASA-TLX Workload Analyses

The NASA-TLX data were collected at the end of each encoding strategy block, thus the TLX composite scores for the verbal (*M* =10.91, *S.D.* = 2.35), visuospatial (*M* =11.85, *S.D.* = 2.06), and auditory encoding strategies(*M* =10.94, *S.D.* = 2.09) were compared with planned contrasts. Results are shown in Table 7. TLX planned comparisons showed that the visuospatial encoding strategy induced higher subjective workload than both the verbal and the auditory encoding strategies.

Table 7: Results for the NASA-TLX assessment of composite subjective workload as a function of encoding strategy

Encoding strategy comparison	Mean TLX difference (S.E)	df	F	p	partial η^2
Verbal vs visuospatial*	-0.94(.41)	(1,32)	5.30	.028	.14
Verbal vs auditory	-0.03(.44)	(1,32)	0.01	.945	--
Auditory vs visuospatial*	-0.91(.39)	(1,32)	5.57	.025	.15

* = $p < .05$

Further Exploratory Analyses

Exploratory analyses examined the relationship between the subject level variables (musical experience data, self-reported SAT verbal scores, self-reported ratings of verbal and spatial ability and cognitive style, and the comparison of mental clocks test score) and performance variables. These analyses are included in Appendices F, G, and H. Most of the correlations were nonsignificant, likely due to restriction of range on individual difference variables in the sample, and no strong conclusions can be drawn about the relationship between individual difference variables and performance for the study tasks.

Discussion

Experiment 2 extended the widely-used sentence-picture verification task to include nonspeech auditory stimuli. The sentence-picture paradigm has shown that people can internally represent simple sentences in verbal/propositional or visuospatial formats (Coney, 1988; Glushko & Cooper, 1978; MacLeod et al., 1978; Mathews et al., 1980; Tversky, 1975). Past research with the paradigm has also confirmed that the internal format of representation can be altered via instructional manipulations (Mathews et al., 1980; Reichle et al., 2000).

In general, the results for verification times in Experiment 2 showed that nonspeech auditory, verbal, and pictorial stimuli can each be represented internally as verbal representations, visuospatial images, or auditory images. This finding was evident (with some exceptions) in the patterns of verification times as a function of encoding strategy for each of the three respective types of verification stimuli. In general, a given format of verification stimulus (sentence, picture, or sound) was verified fastest when participants were using the encoding strategy that matched the verification stimulus (Figure 7).

The results for study time were not fully consistent with the hypothesis that study time should decrease when the study stimulus format and encoding strategy matched, as the predicted pattern of results was only found in full for the sentence study stimulus, where the verbal encoding strategy resulted in faster study times than the other encoding strategies. Study times for pictures under the visuospatial encoding strategy were not faster than study times for pictures using verbal or auditory imagery encoding. For a sound study stimulus, encoding times were significantly lower for the auditory imagery as compared to the visuospatial imagery encoding strategy, but the study times did not differ from the verbal encoding strategy. This suggested that participants could assign a verbal label to a sound as quickly as they could retain an auditory image, but the generation of a visuospatial image from a sound took longer. The discrepancies in this findings could possibly be attributable to the relatively familiarity of transformational encoding processes. It may be possible, for example that the process of verbal encoding (i.e, labeling) is fast for a stimulus in any format, as the process of verbally encoding (i.e., assigning labels to) stimuli in different modalities (e.g., vision) is familiar (e.g., Carpenter

& Just, 1975; Colegate & Eriksen, 1970). The sonification stimuli were assumed to be relatively novel, yet were still associated with a phrase relatively quickly in the verbal encoding condition with a sonification study stimulus. While it is not especially surprising that a visual stimulus was quickly encoded verbally (nearly faster than it was encoded as an image), interestingly, the picture stimulus was also encoded as an auditory image as quickly it was encoded as a visuospatial imagery, which may be explained in part by the crossmodal relationship between visual spatial and auditory frequency (Ben-Artzi & Marks, 1995; Kubovy, 1981; Mikumo, 1997; Szlichcinski, 1979; Wagner et al., 1981; Walker, 2002, 2007).

The TLX analysis suggested that overall (composite) subjective workload was highest under conditions of visuospatial imagery encoding. As mentioned above, verbal encoding is a familiar process (e.g., Colegate & Eriksen, 1970), thus the relative difference between visuospatial imagery and verbal encoding was not surprising. Interestingly, however, the auditory imagery strategy induced the same subjective workload as the verbal encoding strategy. Whereas past research has strongly suggested that visuospatial encoding is effortful (or at least requires time, see Coney, 1988; MacLeod et al., 1978; Tversky, 1975), this is the first study to explicitly compare the workload across encoding strategies with mixed formats of external stimuli. Whereas Experiment 1 showed that subjective workload was lower with an auditory encoding strategy for nonspeech auditory stimuli, Experiment 2 suggested that auditory imagery was less subjectively demanding than visuospatial encoding and on par with verbal encoding across different formats of the external stimulus to be encoded.

An unexpected but interesting finding from Experiment 3 was the lingering effect of an auditory study stimulus that interacted with format of the verification stimulus such that verification times were faster for sounds when sounds had been studied, regardless of the encoding strategy. This finding is discussed in more detail in the General Discussion.

Summary of Experiment 2

The primary findings of Experiment 2 can be summarized as follows:

1) The results for study time only partially confirmed the predictions of Hypothesis 2a that participants would require less time to encode a stimulus when their encoding strategy matched the format of the external stimulus. The strategy compliance manipulation check and the general pattern of findings for verification times (discussed next) suggested that encoding instructions were followed across encoding strategies, so the study time results may reflect effects of relative familiarity or efficacy with a given encoding strategy and/or a given study stimulus.

2) The pattern of results for verification time generally confirmed Hypothesis 2b and dissociated the three encoding strategies as distinct formats of internal representation. For a given verification stimulus (sentence, picture, or sound), the mean verification times for that stimulus format were fastest when the encoding strategy (verbal encoding, visuospatial imagery, or auditory imagery) matched the verification stimulus format.

3) For verification times, I did not expect an interaction between the format of the external study stimulus and the format of the external verification stimulus, because the study stimulus was to be recoded according to the encoding strategy. In other words, recoding was expected to drop or inhibit the properties of the original external representation in favor of the format dictated by the encoding strategy. This pattern of

results held with one interesting exception: a lingering effect of the sound study stimulus was found. Across encoding strategies, participants responded faster to a sound verification stimulus when then the study stimulus had also been a sound. This finding may suggest a relatively protracted effect of auditory sensory memory and is examined in the context of all three Experiments in the General Discussion.

4 EXPERIMENT 3: INTERFERENCE TASKS FOR DIAGNOSING CONFLICTS IN REPRESENTATIONAL PROCESSING

Interference task paradigms have been used extensively in psychology, often to diagnose the locus of information processing conflicts (Ogden et al., 1979; Tsang & Wilson, 1997). These paradigms usually examine performance of the task of primary interest both alone and in the presence of another task or set of tasks. If performance of the primary task declines in the presence of a secondary task, then the two tasks are inferred to draw upon the same mental resources at some stage of information processing.

Patterns of interference, then, have been used to understand the relative dependence or independence of processing for concurrent or temporally proximal stimuli. Much of the evidence supporting the Baddeley model of working memory was derived from studying patterns of task interference (for a review, see Baddeley, 1992), and interference studies have been proffered as evidence of the processing codes dimension, in particular, of multiple resources theory (Wickens & Liu, 1988). This behavioral paradigm seems especially suitable for studying the inherently unobservable conceptual mechanisms or mental resources associated with internal representations, as we can hold constant the demands of resource pools associated with modalities and responses (see Figure 1 and Wickens, 2002, 2008) such that any observed conflicts can be attributed to internal processes. Of note, interference tasks can be constructed such that all aspects of the tasks are time-shared (e.g., Wickens, Kramer, Vanasse, & Donchin, 1983a; Wickens & Liu, 1988), or the interference task can be arranged such that the tasks are sequential with overlapping aspects. With sequential but overlapping paradigms, one stimulus must be retained in working memory during an interference task, and a delayed response to or comparison with the first stimulus is required (e.g., Bonebright & Nees, 2009; Deutsch, 1970; Fausset, Rogers, & Fisk, 2008).

The sequential but overlapping interference task paradigm should be able to discern if the perceptual-representational demands of the interference task draw upon the same representational resources that are required for maintaining the primary task stimulus in working memory. For example, visual perception and visuospatial internal representation share some of the same processes or mechanisms (Farah, 1985; Finke, 1980; Finke, Freyd, & Shyi, 1986; Podgorny & Shepard, 1978) and have at least some common biological correlates (Farah et al., 1988; Kosslyn & Thompson, 2002). Parallel findings support shared representational and biological processes for auditory perception and imagery (Farah & Smith, 1983; Halpern & Zatorre, 1999; Halpern et al., 2004; Kraemer et al., 1995; Zatorre & Halpern, 2005). Obviously internal representation is not perception per se, and imagery can usually be readily distinguished from a veridical perceptual experience in the normal population. Existing research has described neural activities that are distinct for respective representational and perceptual processes (Behrmann, Moscovitch, & Winocur, 1994), yet neuroscience has also discovered a considerable degree of overlapping activity that is unique to neither imagery nor perception (Kosslyn & Thompson, 2002; Mechelli, Price, Friston, & Ishai, 2004). The hard boundaries between imagery and perception have yet to be established (see Kosslyn & Thompson, 2003), but available data suggest the two processes draw upon many of the same mental resources within a given representational format (e.g., visuospatial, verbal, or auditory).

An unresolved question involves the extent to which this overlap in perceptual and representational processing leads to meaningful conflicts during multitasking. The resolution of this dilemma is especially critical for understanding the appropriate use of sound in systems, as auditory displays have been deployed in multitasking scenarios where other (often visual) stimuli must concurrently be processed. Experiments 1 and 2 sought to establish that the internal representation of nonspeech sounds have distinct behavioral properties (vis-à-vis the well-established verbal and visuospatial representational dichotomy) and that representations in general can be manipulated by instructions. The general confirmation of these phenomena were an important step toward understanding internal representations of nonspeech sounds, but a greater concern for the use of sounds in systems is the extent to which these representational conflicts result in meaningful task interference. Wickens (Wickens, 2007) offered a set of three criteria for operationalizing the mental resources construct: 1) separable resource pools are biologically dissociable; 2) resources predict behavioral outcomes and information processing conflicts; and 3) descriptions of resources can be translated into design heuristics. The biological plausibility of at least three distinct formats of representation generally has been supported (see, e.g., Baddeley, 2003; Gruber, 2001; Zatorre, 2003). Until theoretical representational conflicts can be demonstrated to predict interference and conflicts among tasks, however, the distinction between representational formats will not be translated into design heuristics. If, however, interference can be either induced or avoided by straightforward encoding instructions, then this finding will provide a concrete heuristic that will aid in the design of systems and tasks that use auditory displays.

The approach tested herein was approximately commensurate with the multiple resources perspective. This approach, though imperfect (see Navon, 1984), has offered concise and parsimonious explanations of a number of empirical phenomena (e.g., timesharing and stimulus-response compatibility) while also capturing a number of the broader themes of cognitive psychology (e.g., cognition as stages-of-processing and working memory as visuospatial and verbal representational systems). These advantages coupled with an explicit concern for heuristics for practical applications make the multiple resources construct useful for the purposes of the current discussion.

As mentioned above, the construct of workload—a critical consideration in human factors task and system design—has been overlooked in much of the literature on internal representations. Whereas most people may be able to encode a particular stimulus as instructed (Mathews et al., 1980), this transformation may be effortful. Further, in transformation, the accuracy of the representation of the external stimulus may be compromised (Brandimonte, Hitch, & Bishop, 1992), and tasks that require consultation of the transformed internal representation may become more difficult than tasks accomplished using an internal representation that matches the external stimulus (e.g. Chandler & Sweller, 1991).

Experiment 3 examined performance for understanding the information contained in sonifications in the presence of three different types of interference tasks. Participants in Experiment 3 heard brief, four-note sonifications that represented the price of a stock at the opening, mid-morning, mid-afternoon, and closing of a trading day. In a between-subjects manipulation, they were instructed to use a verbal strategy, a visuospatial imagery strategy, or an auditory imagery strategy to encode the sonifications. During the single task portion of the study, participants were asked, after encoding the stimulus according to their prescribed instructions, to estimate the stock price in dollars at a queried time of day (a point estimation task). Participants were then introduced to three additional blocks of trials in a within-subjects manipulation that required performance of the point estimation sonification task using their respective encoding strategies in the presence of each of three interference tasks: a verbal interference task, a visuospatial interference task, and an auditory interference task.

Hypotheses

Participants were predicted to experience greater disruption to performance of the sonification task when the interference task drew upon the same theoretical representational resources as their prescribed encoding strategy for the sonification.

Hypothesis 3a

Participants in the visuospatial imagery encoding group were predicted to experience greater disruption of the sonification task from the visuospatial interference task relative to baseline single-task performance as well as compared to interference task conditions with a verbal or auditory task.

Hypothesis 3b

Participants in the verbal encoding group were predicted to experience greater disruption of the sonification task from the verbal interference task relative to baseline single-task performance as well as compared to interference task conditions with a visuospatial or auditory task.

Hypothesis 3c

Participants in the auditory imagery encoding group were predicted to experience greater disruption of the sonification task from the auditory interference task relative to baseline single-task performance as well as compared to interference task conditions with a verbal or visuospatial task.

Hypothesis 3d

Workload measurements using the NASA-TLX were also hypothesized to vary with the locus of predicted interference effects. Specifically, it was predicted that the additional demand of an interference task that taxed the same representational resources (as the encoding format of the sonification stimulus) should be reflected in an increase in subjective workload relative to single task baseline conditions. Likewise, it was predicted that the addition of an interference task that taxed a different set of representational resources (from the encoding format) would have significantly less (and perhaps even a negligible) impact on perceived workload (see Wickens & Hollands, 2000). If representational systems are to be viewed as pools of resources (Wickens, 2002, 2008; Wickens & Hollands, 2000; Wickens & Liu, 1988), then perceived workload relates to the interplay of task demands with available resources (Wickens & Hollands, 2000).

Method

Participants

Participants, *N* = 55, 21 females, *M* = 19.9 (*SD* = 1.48) years of age, were recruited from undergraduate psychology course at the Georgia Institute of Technology and received either course credit or nominal monetary compensation ($10.00 per hour) for their participation in the study.

Demographic and subject variables were the same as those in Experiments 1 and 2. Participants reported a mean of 3.26 (*SD* = 3.49) years of formal musical training (i.e., private or class instruction), 3.98 (*SD* = 4.3) years of experience playing a musical instrument, and 4.02 (*SD* = 4.46) years of experience reading musical notation. Participants who self-reported SAT verbal scores (*N* = 42) had a mean score of 619.1 (*SD* = 78.6). Participants who self-reported SAT math scores (*N* = 43) had a mean score of 710.0 (*SD* = 74.7). Participants who self-reported SAT writing scores (*N* = 28) had a mean score of 628.6 (*SD* = 75.1). The mean self-estimated verbal ability rating for the sample was 3.75 (*SD* =0.87), and the mean self-estimated spatial ability rating was 3.76 (*SD* = 0.72). Both ability ratings were on a scale of 0 to 5. Participants' self-reported cognitive style score was 2.64 (*SD* = 1.30), with a score of 1 representing a rating of "strongly more visual than verbal," a score of 4 representing a rating of "equally verbal and visual," and a score of 7 representing a rating of "strongly more verbal than visual." (Mayer & Massa, 2003) Participants reported a mean rating of 4.04 (*SD* = .76) across the auditory imagery questionnaire items, where a rating of 0 indicated "no image at all," a score of 3 indicated a "fairly vivid" auditory image, and a score of 6 indicated an auditory image "as vivid as actually hearing." Mean response time for correct responses on the modified mental clocks task (Paivio, 1978) was 7514 (*SD* = 2976.73) ms.

Apparatus

The apparatus were the same as described in Experiment 1.

Primary Sonification Task Stimuli

The stimuli for Experiment 3 featured sonifications that depicted the price of a fictional stock over the course of a trading day at opening, mid-morning, mid-afternoon, and closing, in that order. The four prices of the stock throughout the day were represented with 200 ms notes in the MIDI piano timbre, and 300 ms separated each note. Each sonification was 1700 ms in length. Participants were given the overall scaling for mapping frequency to dollars in the opening instructions. Stock prices ranged between a low value of 6 dollars (MIDI C4) to a high value of 106 dollars (MIDI C7). To give participants a scaling context, the notes representing the opening and closing prices were available during the study period for a sonification stimulus. The four data points presented in each sonification stimulus were randomly chosen from values within this range. A set of sonification stimuli were constructed within these stimulus parameters, and each stimulus was used in 4 different point estimation trials (each querying the value of the stock price at one of the 4 times of day represented in the sonification).

Procedure

After the informed consent procedure, demographic data and subject variables were collected. Participants were then randomly assigned to one of three encoding conditions for the primary sonification task. Participants were aware that the purpose of encoding was to remember the sonification to perform the point estimation task after a brief delay (in the single task block) or after an interference task.

Visuospatial Imagery Encoding Condition

Participants in this condition were instructed to encode and rehearse the sounds as a visuospatial image, like a visual graph that represented higher stock prices as higher up on the visual Y-axis and time of day on the visual X-axis. During instruction, participants were shown an example audiovisual animation that depicted a visuospatial representation forming as a sound stimulus was heard.

Verbal List Encoding Condition

Participants in this condition were instructed to encode and rehearse the sounds as a verbal list, like a list of values, one for each tone, that named the stock prices from the beginning to the end of the day. During instruction, participants were shown an example audiovisual animation that depicted a verbal list populating as a sound stimulus was heard.

Auditory Imagery Encoding Condition

Participants in this condition were instructed to encode and rehearse the sounds in the pseudo-isomorphic format of the sonifications, whereby the sensory experience of hearing the tones was to be retained in working memory (see, e.g., Keller et al., 1995). During instruction, participants were encouraged to silently rehearse the notes during encoding and to employ covert strategies for retaining pitch information in the original format of the stimulus.

Sonification Task and Instructions

Following instructions for their respective encoding conditions, participants underwent a block of 30 single-task point estimation sonification trials. Each trial began with a screen that allowed the participant to click one of three buttons that played the lower bound ($6) tone, the upper bound tone ($106), or the actual sonification stimulus for the trial, respectively. Upon clicking the button for the sonification stimulus, the sonification played. Participants were permitted to listen to the stimulus and the reference tones as many times as needed before proceeding. When ready, participants pressed either the “Z” or the “?” key to continue[2]. Upon the conclusions of a 10 s delay that showed a blank grey screen, participants saw the query for the point estimation task.

The point estimation task asked participants to estimate the price of the stock at a given time of the trading day (e.g., “What was the price of the stock at mid-afternoon?”). Performance data and task analyses of point estimation tasks with sonified displays of quantitative data have been reported in previous research (e.g., D. R. Smith & Walker, 2005; Walker & Nees, 2005). The dependent variable for the point estimation task with sonifications was the root mean squared error (*RMSE*) of responses in dollars. For sonification trials across all blocks, the particular sonification stimulus and the queried time of day were randomly chosen, and trials varying the sonification stimulus and queried time of day were randomly interleaved. Participants received feedback that gave the correct answer following every point estimation trial with the sonification task throughout the study, including after interference trials.

Interference Tasks

Following the first block of single-task point estimation trials, participants experienced three additional blocks of the sonification task in the presence of three different interference tasks. Researchers (Kahneman, 1973; Wickens, 1984) have suggested that a variety of interference tasks should be required to dissociate representational systems. The respective interference tasks were inserted during the time featured the 10 s delay in the single task condition—after the encoding of a sonification stimulus, but before the query regarding the price of the stock at a given time of day (see

---

[2] Whereas using these particular keys to advance from the study/encoding screen to the point estimation question was not crucial for the single task trials, the nature of responses in later interference trials required participants to position their index fingers on these keys to advance to the interference portion of a trial.

Figure 10). Participants heard and encoded the sonified stock prices according to their assigned strategies, and they indicated with either the “Z” or “?” key that encoding was complete. Immediately upon the keypress, the interference task began. The order of presentation of the interference task blocks was counterbalanced across participants.

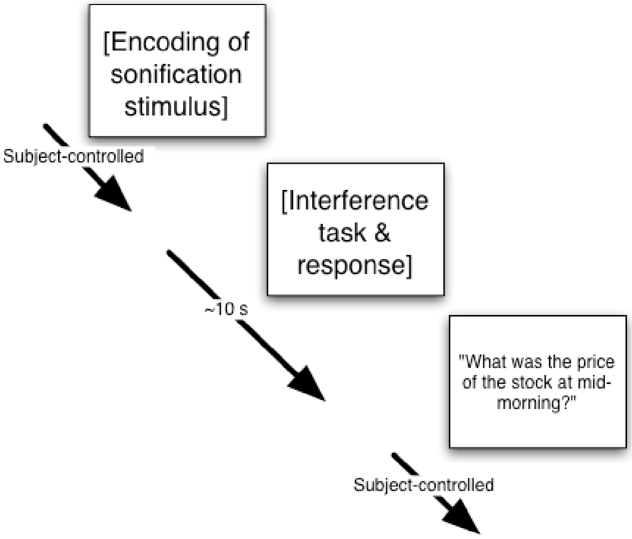

Figure 10: Flow of an interference task trial.

Participants were instructed to divide their mental resources equally between both tasks. While the allocation policy between the primary sonification task and the respective interference tasks was held constant and equal, the difficulty of the each secondary task was manipulated at two levels. When allocation policy remained unchanged, the primary sonification task should have been sensitive to increases in the difficulty of the interference task, provided the combination of tasks sufficiently taxes the same pool of representational resources (Fisk, Derrick, & Schneider, 1986; Gopher, Brickner, & Navon, 1982; Wickens, 1984; Wickens & Liu, 1988). Insensitivity of the primary task to changes in the task difficulty would suggest that the two tasks draw upon different pools of resources, but sensitivity to task difficulty would offer another important diagnostic tool for assessing whether the sonification task conflicted with interference tasks at the level of internal representations in working memory.

Visuospatial Interference Task

For this task, participants made two-choice judgments about Shepard-Metzler type block stimuli (Shepard & Metzler, 1971). A visual block figure was presented as a standard stimulus. The standard was followed by a 5000 ms delay that featured a blank grey screen, and then a comparison visual block figure appeared. The comparison stimulus was either a rotated version of the standard stimulus, or a rotated mirror image of the standard. Participants' task was to judge whether the comparison stimulus was a rotated depiction of the standard or a rotated mirror image of the standard, and responses were made using the "Z" and "?" keys on the computer keyboard. Mappings of responses ("standard" and "mirror") to keys were counterbalanced across participants. Stimuli consisted of images of three-dimensional block figures that were first used by Shepard and Metzler and recently offered as a stimulus library by Peters and Battista (Peters & Battista, 2008). Difficulty of the task was manipulated by varying the angle of rotation of the comparison stimulus (compared to the standard) at two levels: 15 degrees (easy) and 45 degrees (difficult).

Verbal Interference Task

Participants viewed a brief (1200) ms presentation of a set of upper case consonants in a modified version of the Sternberg memory scanning task (Sternberg, 1966). After a 5000 ms delay (i.e., a blank screen), participants saw a single lower case consonant. Their task was to determine whether or not the lower case consonant was a member of the original consonant set. Responses were made using the "Z" and "?" keys on the computer keyboard, and mappings of responses (“yes” and “no”) to keys were counterbalanced across participants. Difficulty of the verbal interference task was manipulated by varying the number of consonants in the initial set at two levels: 4 consonants (easy) and 8 consonants (difficult).

Auditory Interference Task

The stimuli and task here were modeled after the interference task described by Deutsch (1970). Participants heard a standard tone from the equal-tempered musical scale in the octave above middle C, followed by a series of to be ignored tones. A 1000 ms delay followed the interference tones, and a final comparison tone was heard. Participants’ task was to judge whether the comparison tone was the same as or different from the standard. For “different” trials, the comparison tones were one semi-tone different from the standard. All tones were 200 ms in duration with 10 ms onset and offset ramps, and stimuli used the MIDI saxophone instrument. The initial standard tone and each interference tone were all separated by 300 ms of silence. Difficulty of the auditory interference task was manipulated by varying the number of intervening tones between the standard and comparison at two levels: 4 tones (easy) and 8 tones (difficult) (see Deutsch, 1970; Massaro, 1970; Ries & DiGiovanni, 2007).

Implementation of the Interference Tasks

The time between the beginning of the interference task and the query for the stock price was not permitted to exceed 10 s across interference tasks (see Figure 7). Participants were instructed to respond as quickly and as accurately as possible to the interference tasks, and 10 s allowed ample time for a response to be recorded for the interference tasks used here. Participants who did not log a response within the allotted

10 s time window had the trial terminated and were given feedback that encouraged them to respond as quickly as possible to the interference task. Terminated interference trials (n = 7 across all 1,080 verbal interference trials; n = 40 across all 1,080 visuospatial interference task trials; n = 14 across all 1,080 auditory interference task trials) were repeated from the beginning. Upon logging a response to the interference task, participants were queried about the price of the stock at a given time of day. Participants completed the strategy compliance question (Appendix B) following each trial as a check on the encoding strategy implementation for the sonification task.

Participants were given the opportunity for a brief break at the conclusion of each block of trials. The beginning of each block using interference tasks introduced each task with detailed instructions. A set of 10 practice trials for each interference task alone were used to familiarize participants with the requirements for the interference task. Damos (1991) reviewed research suggesting that the amount of single task practice does not seem to affect later dual-task performance; the primary purpose of all practice trials herein was to familiarize participants with the tasks and task combinations. After the practice trials had been completed, participants began the block of 20 experimental trials that required performance of the sonification task in the presence of the interference task.

The primary dependent variable of Experiment 3 was the RMSE of participants' responses to the point estimation task across the 4 blocks of the study. Accuracy data were also collected for each interference task. Participants completed the NASA-TLX as a measure of subjective workload following each of the four blocks (single task, verbal interference, visuospatial interference, and auditory interference) of the study.

Results

Missing Data and Data Excluded for Strategy Noncompliance

Appendix F lists the encoding strategy, specific missing data, and reasons for missing data for all participants who gave incomplete data. Four participants (2 in the visuospatial encoding condition and one each in the auditory and verbal encoding conditions) reported complete strategy noncompliance for one or more blocks of the block study and were excluded case-wise from further analyses. Appendix G shows the data for the subject-level variables for these participants. In general, the subject-level variables do not allow for any strong conclusions to be drawn about the reasons the four participants were noncompliant with the instructed encoding strategy. The additional four cases of missing data were the result of a computer crash, a drop-out, and two instances where participants did not complete all study tasks in the time allotted and had to leave the study. Where possible, participants who gave partial data for reasons unrelated to strategy compliance were included in analyses.

Excluding the four participants who were dropped case-wise, strategy noncompliance data elimination proceeded on a trial-by-trial basis for the remaining 51 participants. Data for 11.6% of trials across the study were excluded because a participant indicated she or he did not use the instructed encoding strategy for the sonification trial. Strategy compliance across blocks of the study was analyzed in greater detail below. Data from individual sonification trials would have been considered outliers and excluded from intracondition averages if the participants' RMSE had been greater than 3 standard deviations from their individual means for a given factorial condition, but no datum in the study met this elimination criteria.

Primary Analyses

Hypotheses 3a -3c predicted that participants would experience greater disruption to performance of the sonification task when the interference task drew upon the same theoretical representational resources as their prescribed encoding strategy for the sonification. This hypothesis was tested with a 3 (encoding strategy) by 4 (block: single task, verbal interference, visuospatial interference, or auditory interference) mixed ANOVA on the RMSE dependent variable for the sonification task. Forty-eight participants gave complete data across all conditions of the study. Sphericity assumptions were upheld for the main effects of the within-subjects block manipulation, *W* = .90, $\chi2(5) = 4.42$, *p* = .49. Results showed a significant main effect of interference block, *F* (3,135) = 6.25, *p* = .001, partial $\eta^2 = .12$ (see Figure 11). Nonsignificant effects included the main effect of strategy , *F* (2,45) <0.01, *p* = .99, and the interaction of interference block with strategy, *F* (6,135) = 1.29, *p* = .27.

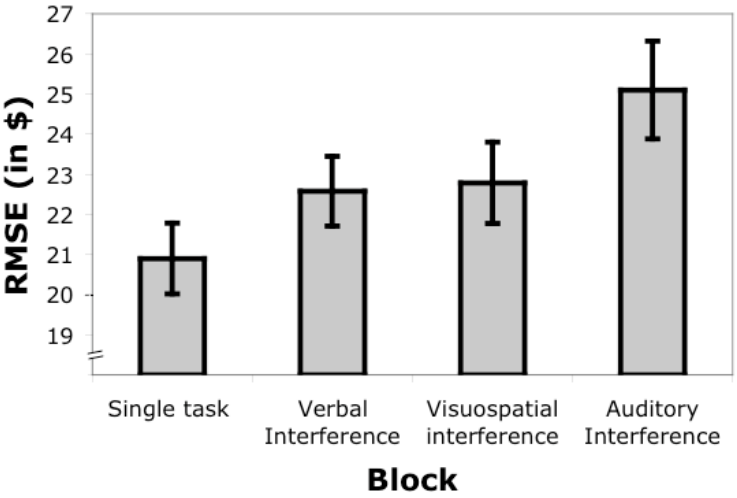

Figure 11: RMSE on the point estimation task as a function of block (interference task) of the study. Error bars represent standard error of the mean. Also see Appendix H.

Follow-up pairwise comparisons for the block manipulation (see Table 8) revealed that performance of the sonification task with the visuospatial and auditory interference tasks resulted in significantly worse performance than performance of the sonification task with no interference task (i.e., the single task block). Performance of the sonification task with the verbal interference task was not significantly different from the single task block. Performance of the sonification task with the auditory interference task was significantly worse as compared to both the visuospatial and verbal interference blocks.

Table 8: Pairwise comparisons of performance of the sonification task between blocks of the study (i.e., intererence tasks)

Interference block comparison	Mean RMSE difference (SE)	df	F	p	partial η^2
Single task vs verbal	-1.59(.91)	(1,45)	3.01	.090	--
Single task vs visuospatial*	-1.81(.89)	(1,45)	4.17	.047	.09
Single task vs auditory*	-4.21(1.12)	(1,45)	14.07	.001	.24
Verbal vs visuospatial	-0.22(.88)	(1,45)	0.06	.802	--
Verbal vs auditory*	-2.63(1.02)	(1,45)	6.52	.014	.13
Visuospatial vs auditory*	-2.40(1.04)	(1,45)	5.31	.026	.11

* = $p < .05$

A corollary of hypotheses 3a-3c was that performance on the sonification task should be sensitive to the difficulty of the interference task, but only when the encoding strategy matched the interference task and thus taxed the same hypothetical pool of representational resources. Given that selective interference as a function of encoding strategy—the strictest prediction of multiple resources—was not confirmed, there was little reason to expect that the corollary predictions of multiple resources theory regarding the difficulty manipulation would be confirmed. A series of interaction contrasts examined RMSE for the sonification task at each type of interference task for both difficulty levels. No significant interaction of difficulty with strategy was obtained for the verbal interference task, *F* (2,48) = 1.61, *p* = .21, the visuospatial interference task, *F* (2,48) = 1.11, *p* = .34, or the auditory interference task, *F* (2,49) = 1.07, *p* = .35. A parallel analysis examined performance on the interference task dependent variable (operationalized as the percent of correct responses on the interference task for the block) as a function of strategy and the difficulty of the interference task. This analysis examined the possibility that sonification task performance was maintained across strategies at the expense of interference task performance when the encoding strategy used the same format of representation as the interference task. Again, a series of interaction contrasts examined the interference task percent correct across each type of interference task and difficulty level. No significant interaction of difficulty with strategy was obtained for the verbal interference task, *F* (2,48) = 1.34, *p* = .27, the visuospatial interference task, *F* (2,48) = 0.77, *p* = .47, or the auditory interference task, *F* (2,49) = 2.42, *p* = .10.

Strategy Compliance Analyses

One alternate possibility that might reveal an effect of encoding strategy would be if strategy compliance had decreased selectively as a function of encoding strategy and interference task. In this case, for example, a person using a verbal encoding strategy would have been unable to maintain that encoding strategy in the presence of a verbal interference task. This outcome would be reflected in an interaction between block (interference task) and strategy for the percent of trials in a given block that participants were in compliance with their assigned strategy. A 3 (strategy) by 4 (block) ANOVA on the percent of strategy compliance (derived from the trial-by-trial strategy compliance question) examined this possibility for the 48 participants who gave usable data across all 4 blocks of the study. Sphericity assumptions were violated for the main effects of the within-subjects block manipulation, $W = .28$, $\chi 2(5) = 55.83$, $p < .01$, thus a Greenhouse-Geisser correction was used. Results showed a significant main effect of interference block, $F\,(2.10,94.60) = 5.38$, $p = .005$, partial $\eta^2 = .11$; nonsignificant effects included the effect of strategy, $F\,(2,45) = 1.24$, $p = .299$, and the interaction of interference block with strategy, $F\,(4.20,94.60) = 0.52$, $p = .73$. Strategy compliance did not vary as a function of the encoding strategy. Follow-up pairwise comparisons for the block manipulation revealed that strategy compliance for the single task block was significantly lower ($M = .84$, $SD = .25$) than strategy compliance for the verbal ($M = .92$, $SD = .19$) or the visuospatial blocks ($M = .91$, $SD = .19$) but not the auditory imagery block ($M = .87$, $SD = .26$). The verbal encoding strategy block also showed significantly higher strategy compliance than the auditory imagery block.

Table 9: Comparison of blocks for the percent of trials in compliance

Interference block comparison	Mean % compliance difference (SE)	df	F	p	partial η^2
Single task vs verbal*	-.07(.02)	(1,45)	9.09	.004	.17
Single task vs visuospatial*	-.07(.02)	(1,45)	11.67	.001	.21
Single task vs auditory	-.03(.03)	(1,45)	1.10	.299	--
Verbal vs visuospatial	.003(.01)	(1,45)	0.13	.722	--
Verbal vs auditory*	.04(.02)	(1,45)	4.42	.041	.09
Visuospatial vs auditory	.04(.02)	(1,45)	3.78	.058	--

* = $p < .05$

Exploratory Test of a General Resource Hypothesis

Given that the manipulation of encoding strategy failed to show selective release from working memory interference as a function of the resource demands of the interference task for any of the measures used here, the next most plausible model of interference effects was the general resource model (Kahneman, 1973). If the combination of the sonification task and the interference tasks behaved as a general (i.e., non-selective) resource model, then the increase in difficulty of the interference task should have had a general effect on performance of the sonification task, regardless of the format of the interference task. This possibility was examined with a one-tailed t-test that compared RMSE for the sonification task in the presence of easy interference tasks (collapsed across format and encoding strategy) to RMSE for performance of the sonification task in the presence of difficult interference tasks (collapsed across format and encoding strategy). The t-test showed that performance of the sonification task was significantly better (i.e., showed less error) in the presence of an easy versus a difficult interference task, $t(47) = -1.90$, $p = .032$; mean RMSE during easy interference tasks = 22.14, *SD* =6.5; mean RMSE during hard interference tasks = 23.88, *SD* = 7.38. A parallel analysis confirmed that the mean percent correct across interference tasks was higher for easy (*M* = 71.6, *SD* =9.5) as compared to difficult (*M*= 67.0, *SD* = 11.4) interference tasks (collapsed across format), $t(47) = 2.72$, $p = .005$.

TLX Analyses

A 3 (encoding strategy) by 4 (block: single task, verbal interference, visuospatial interference, or auditory interference) mixed ANOVA on the TLX composite scores dependent variable examined the possibility that the encoding strategy selectively affected perceived workload for interference tasks (see Figure 12). Forty-seven participants gave complete TLX data across all conditions of the study[3]. Sphericity assumptions were upheld for the main effect the within-subjects block manipulation, $W = .80$, $\chi 2(5) = 9.48$, $p = .09$. Results showed a significant main effect of interference block, $F\ (3,132) = 8.86$, $p < .001$, partial $\eta^2 = .17$. Nonsignificant effects included the main effect of strategy , $F\ (2,44) = 1.00$, $p = .38$, and the interaction of interference block with strategy, $F\ (6,132) = 1.26$, $p = .30$. Follow-up pairwise comparisons for the block manipulation are shown in Table 10.

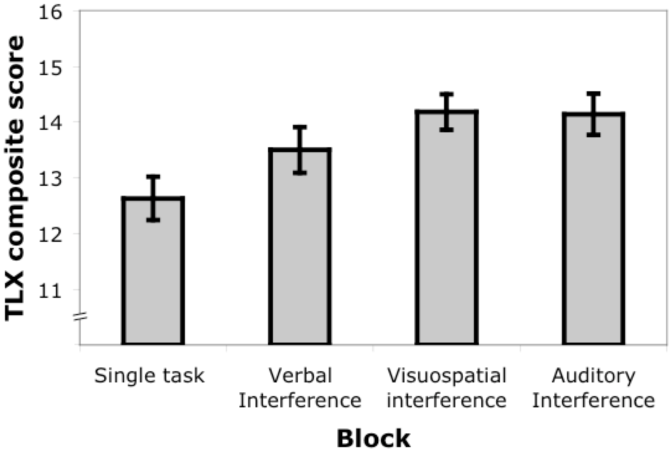

Figure 12: TLX composite score as a function of study block. Error bars represent standard error of the mean. Also see Appendix H.

---

[3] One participant in the visuospatial encoding condition (who was not excluded for strategy noncompliance) provided all study data except for the visuospatial interference block TLX, because the participant ran over the expected duration of the study and had to leave the study early. That participant was excluded from the TLX analyses.

Table 10: Comparison of subjective workload between blocks of the study

Interference block comparison	Mean TLX difference (S.E)	df	F	p	partial η^2
Single task vs verbal*	-0.87(.37)	(1,44)	5.56	.023	.11
Single task vs visuospatial*	-1.54(.37)	(1,44)	17.87	<.001	.29
Single task vs auditory*	-1.51(.41)	(1,44)	13.91	.001	.24
Verbal vs visuospatial*	-0.67(.33)	(1,44)	4.17	.047	.09
Verbal vs auditory*	-0.64(.29)	(1,44)	4.91	.032	.10
Visuospatial vs auditory	-0.03(.29)	(1,44)	0.14	.906	--

* = $p < .05$

Results showed that perceived workload was significantly higher for all interference task conditions as compared to the single task condition. Workload during the visuospatial and auditory interference task blocks were both significantly higher than during the verbal block, but were not significantly different from each other.

Exploratory Analysis of the Number of Times Listened

A 4 (block) x 3 (encoding strategy) mixed ANOVA examined the average number of times participants listened to each sonification stimulus before advancing in the trial. Greenhouse-Geisser corrections were used in all analyses where sphericity assumptions were violated. Results showed a significant effect of block, F (2.04,93.90) = 22.67, $p <$ .001, partial $\eta^2 = .33$. The effect of encoding strategy was not significant, F (2,46) = 1.95, p =.15, and the interaction of block with encoding strategy was not significant, F (6,138) = 1.49, p = .21. Follow-up contrasts on the significant effect of block are shown in Table 11. Participants listened to the sonification significantly more times in the initial single task block (M = 7.01 times listened, SD = 4.36) than the verbal interference (M = 4.57, SD = 2.94), visuospatial interference (M = 4.16, SD = 2.88), or auditory interference (M = 4.17, SD = 2.74) blocks. None of the interference blocks were significantly different from each other. This suggests that participants listened to the sonification more times as they were learning the task in the initial single task block, but participants were not selectively influenced to listen to the sonification more times as a function of either encoding strategy or the demands of a particular interference task.

Table 11: Contrasts across blocks for the number of times listened to each sonification stimulus

Interference block comparison	df	F	p	partial η^2
Single task vs verbal*	(1,46)	21.71	<.001	.32
Single task vs visuospatial*	(1,46)	52.55	<.001	.53
Single task vs auditory*	(1,46)	30.62	<.001	.40
Verbal vs visuospatial	(1,46)	1.68	.200	--
Verbal vs auditory	(1,46)	1.67	.203	--
Visuospatial vs auditory	(1,46)	<0.01	.967	--

* = $p < .05$

Exploratory Correlations of Performance Variables with Subject Variables

Exploratory analyses examined the relationship between the subject level variables (musical experience data, self-reported SAT verbal scores, self-reported ratings of verbal and spatial ability and cognitive style, and the comparison of mental clocks test score) and performance variables. These analyses are included in Appendix I. Most of the correlations were nonsignificant, likely due to restriction of range on individual difference variables in the sample, and no strong conclusions can be drawn about the relationship between individual difference variables and performance for the study tasks.

Discussion

Experiment 3 attempted to offer behavioral evidence of task interference as a function of the format of internal representation and the qualitative aspects of interference tasks. Participants were predicted to be able to encode the sonification stimuli flexibly as verbal representations, visuospatial images, or auditory images, depending on the instructions given. When these respective representations were temporarily maintained in working memory, patterns of interference were predicted to dissociate encoding strategies based on the perceptual-representational demands of the interference task.

Potential Limitations and Methodological Shortcomings of Experiment 3

The results of Experiment 3 failed to find any evidence of selective interference of domain-specific tasks as a function of the sonification task encoding strategy in working memory. The attribution of the source of interference (or a lack thereof) in resource models can be difficult. Perhaps the most obvious explanation for the lack of an effect of encoding strategy is that participants failed to follow the encoding strategy instructions. In this case, encoding strategies would have presumably been heterogeneous across the interference task blocks, and no selective interference as a function of processing code would be observable. The trial-by-trial manipulation check on encoding strategy (which allowed for the elimination of data from strategy non-compliant trials) and the prior results that affirmed the effective implementation of the same encoding strategy manipulation in Experiments 1 and 2, taken together, suggested that this finding is unlikely to stem from encoding strategy noncompliance on the sonification task. Nonetheless, given the inherently unobservable nature of encoding strategies (which were implemented to influence processing codes), the possibility of strategy noncompliance is difficult to rule out entirely.

If participants did effectively implement encoding strategies, the lack of a significant interaction of encoding strategies with the interference task type on the sonification task *RMSE* dependent variable might indicate that performance on the sonification task was preserved across interference tasks to the detriment of performance for the specific interference task that matched the encoding strategy. Results also suggested, however, that the encoding strategy had no interaction with performance for the interference tasks as would be alternately predicted (given that sonification task performance showed no strategy effects) under multiple resources models if participants had shown selective decline in interference task performance as a function of strategy. Since the selective interference predictions of multiple resources were not confirmed, it was not surprising to find that the predicted interaction of task difficulty with strategy was also absent, as the selective effects of task difficulty as a function of processing code was a corollary prediction from the multiple resources approach.

Another plausible explanation of the lack of encoding strategy effects would be the case where participants initially encoded the sonification stimulus according to the prescribed strategy, but spontaneously recoded the stimulus (to a different processing code format) based on the demands of the interference task. The data suggested that participants did not avert selective interference by spontaneously switching encoding strategies (i.e., recoding their representations of the sonification stimuli) as a function of the interference task demands, as the encoding strategy and interference task block manipulations did not interact for the percent of strategy compliance measurement. Again, it remains possible that the trial-by-trial self-reported manipulation check on encoding strategies was not effective (though it generally appeared to be effective in Experiments 1 and 2) for the interference task paradigm.

A remaining scenario that would uphold the multiple resources approach would be the case where participants were able to selectively adjust their encoding strategies for the respective interference tasks as a function of their encoding strategies for the sonification task. Strategy use for the encoding tasks was not measured in the current study, as interference tasks were selected under the assumption that the interference tasks chosen demanded specific representational resources. This assumption seems warranted, however, as the literature seems to rule out the possibility that: 1) the mental rotation task could be accomplished with any strategy other than visuospatial imagery; 2) the verbal interference task could be accomplished with any strategy other than a verbal strategy; 3) the auditory interference pitch comparison task could accomplished with any strategy other than an auditory or pitch memory strategy. Still, without a firm indicator (beyond surface task characteristics) of the cognitive strategies participants used for the interference tasks, it remains possible that the processing code resources taxed by the interference tasks did not meet the experimental assumptions of domain specificity.

Theoretical implications of Experiment 3

The information processing model that was supported by the current data was a general resource model (e.g., Kahneman, 1973) of representational processing codes in working memory. The model was "general" in that there was no selective interference, and it was a "resource" (rather than a serial or indivisible process) in that performance of the sonification task did not collapse altogether[4] in the presence of an interference task. Three of the findings from the current study support this interpretation. First, collapsed

---

[4] A pilot simulation that generated random responses within the range of possible responses (i.e., chance performance) for each trial in the sonification task revealed that RMS error on the sonification task would average about 42.9 (*SD* =5.0) dollars at the chance level of performance.

across encoding strategy (which had no effect on performance), performance of the sonification task in the presence of each of the interference tasks was worse than performance of the sonification task alone. Although the difference in sonification task RMSE between single task and verbal interference tasks was not statistically significant, the NASA-TLX analyses suggest that the verbal interference task was simply easier (and thus not as disruptive to sonification task performance) than the other interference tasks. The second piece of evidence in support of a general resource model was the finding that, although interference task difficulty did not interact with strategy, interference task difficulty had a significant general effect on performance of the sonification task such that difficult interference tasks disrupted performance of the sonification task more than easy interference tasks regardless of the encoding strategy or the type of interference task. Finally, the TLX analyses showed no effect of encoding strategy, but revealed that, collapsed across strategy, all interference task conditions showed significantly higher perceived workload than the single task sonification condition.

A careful re-examination of the literature on the processing codes dimension of multiple resources showed that it is perhaps the least studied dimension of the model. In one of the few published accounts that explicitly tested the predictions of the processing codes component of multiple resources, Wickens and Liu (1988) found qualified support for the verbal-visuospatial processing codes dichotomy using a paradigm where all aspects of both tasks were time-shared (as opposed to the sequential-but-overlapping paradigm of the current study). A concurrent visuospatial task (involving mental rotation) disrupted another visuospatial task (involving tracking) more than a concurrent verbal task. The study did not test the (equally theoretically valid) co-prediction that a verbal secondary should interfere more than the spatial secondary task with a verbal primary task. Furthermore, when secondary task difficulty was varied, only a few of the predicted effects were observed. The authors found a paradoxical dual-task increase in performance of the verbal task under difficult (as compared to easy) conditions that was explained by calling upon the general activation or arousal component of the Kahneman (1973) model while simultaneously rejecting the model's general resource in favor of differentiated resources. Essentially, the authors claimed that difficult dual-task conditions increased arousal, which in turn recruited/motivated the deployment of more resources and increased performance *only* for the secondary task that required different resources than the primary task. A strict test of this explanation would require the same pattern of results for a difficult versus easy spatial task in the presence of a verbal primary task.

The empirical support for the inclusion of the processing codes resource dimension (reviewed in Wickens, 1984, 2002) came in part from the literature surrounding Baddeley's (e.g., Baddeley, 1992, 2002) multi-component working memory model, which often showed patterns of selective interference as a function of the format of information to be remembered . Interestingly, however, the precise nature of the type of interference to be expected *within* a given slave system (the phonological loop or the visuospatial sketchpad) was often not indicated in research supporting the multicomponent model. These studies often showed interference for the sake of interference without specifying the model of interference that was to be expected, so it remains unclear whether the slave system "components" of multi-component working memory should behave as multiple resources or multiple serial processors. This lack of specificity leaves open the criticism that the selectiveness of interference in the empirical work supporting multi-component working memory (and perhaps the processing codes dimension of multiple resources by proxy) was the result of selective changes in the quantitative difficulty across interference tasks (rather than selective interference due to the qualitatively similar demands of the tasks). For example, Logie and Edworthy (1986) found that interference tasks involving either articulatory suppression or comparison of homophones disrupted memory for melodies, whereas an interference task involving comparisons of visual symbols did not. The surface interpretation of these results, “shared mechanisms in the processing of verbal and musical material,” sounds compelling upon first examination, but without some data (e.g., a difficulty manipulation, a measure of workload, etc.) regarding the relative difficulty of each task, the possibility remains that the visual matching task presented the only combination of tasks in the study that was difficult enough to overtax a general, undifferentiated processor.

Baddeley’s model, however, did specify that the central executive, a domain-general manager of the domain-specific slave systems, behaved as “limited capacity pool of general resources” (Baddeley, 2002 p. 89). To distinguish the extent to which a particular task engages domain-specific versus domain-general processes remains untenable with behavioral methods alone. Studies (Morris, Phillips, Thibault, & Chaparro, 2008) have claimed that mere rehearsal mostly engages domain-specific processes, whereas the manipulation of information in working memory (e.g., reordering verbal information, etc.) engages executive processes. This account seems plausible, but neurobiological corroboration is perhaps the only way to truly distinguish whether a particular task requires specific or general processes (Schneider, 2008, Feb 13).

Multiple resources has no central executive or domain-general component, although Wickens has acknowledged such processes with assertions that the multiple resources model predicts improved but not perfect time-sharing when different resource pools are engaged (Wickens, 2008). This general dual-task cost was quantified as a general dual-task deficit in a computational iteration of the multiple resources model (Wickens, 2002).

The multiple resources account distinguishes working memory from perception but the processing code dimension of the theory has been collapsed across these dimensions in its predictions about task interference. Further, studies affirming the predicted patterns of results under the assumptions of multiple resources (e.g., Brill, Mouloua, Gilson, Rinalducci, & Kennedy, 2008; Ferris, Hameed, Penfold, & Rao, 2007; Jeon, Davison, Nees, Wilson, & Walker, 2009; Wickens & Liu, 1988) have used truly time-shared dual-task paradigms where all aspects of both tasks overlapped, whereas Experiment 3 of the current study used the sequential-but-overlapping paradigm. The results of the current study taken together with recent literature (e.g., other studies that have not confirmed the predictions of multiple resources, with various proffered explanations, when tasks were not fully time-shared, see Bonebright & Nees, 2009; Fausset et al., 2008; Nees & Walker, 2008b) suggest that representation in working memory in the absence of perception, such as in sequential-but-overlapping interference task paradigms, may operate as a general resource rather than as dichotomous (verbal and visual) or trichotomous (verbal, visual, and auditory) pools of separable resources. This qualification implies that the usefulness of the distinction between verbal and spatial processing codes is derived from its heuristic power when information from the concurrent tasks must be perceived simultaneously, as studies (Brill et al., 2008; Ferris et al., 2007; Wickens & Liu, 1988) have shown selective release from interference as a function of the processing code of the perceived information.

In cases where information must first be encoded and then maintained/rehearsed in the presence of an intervening task, the results of Experiment 3 suggested that the qualitative format of the internal representation (i.e., the encoding strategy) offers no heuristic value in predicting interference. General interference is to be expected regardless of the qualitative combinations of encoding strategies (i.e., processing codes) and tasks, and this interference appears to be a function of the combined difficulty of both tasks. This interpretation is entirely consistent with the notion of a central executive in working memory that behaves as a general resource, but (as described above) it remains difficult, without corroborating evidence, to make any strong claims about the relative involvement of a domain-general central executive in behavioral tasks like those used in Experiment 3. One can reasonably ask whether the "timesharing" (Wickens, 2008) predictions of multiple resources theory were even intended to apply to sequential but overlapping task paradigms. Was multiple resources theory intended to be used to predict interference for paradigms where only working memory representational resources were time-shared (i.e, without an element of concurrent perception)? Wickens and Liu (1988) used language that strongly suggested such task paradigms fall under the purview of multiple resources when they ascribed observed effects regarding the paradoxical compatibility of manual (over vocal) responses for a verbal task as a function of the "*concurrency* between processing stages" (p. 607, italics retained from the original), which meant "a substantial requirement to maintain working memory load".

It is important to note that this qualification leaves the majority of the predictions of multiple resources theory intact. Further, the multiple resources approach represents one perspective in an unresolved debate about the nature of fundamental constraints on human information processing. Arguments continue to made in favor of each of four persepectives: a) the central, serial bottleneck perspective (e.g., Pashler, 2000); b) the multiple, serial bottlenecks perspective (e.g., Byrne & Anderson, 2001); c) the central, allocatable capacity perspective (e.g., Kahneman, 1973; Tombu & Jolicœur, 2002); and d) the multiple allocatable capacities, or multiple resources, perspective (e.g., Gopher et al., 1982; Wickens, 2002). Debate continues in part as a result of the difficulties in distinguishing these models empirically (Navon & Miller, 2002). The most plausible resolution will likely acknowledge a multi-layered system with a variety of psychological mechanisms and operations, some of which behave serially and others of which are divisible and allocatable like a capacity or resource (Kahneman, 1973). An information processing system that operates at all levels (e.g., input modalities, perception, working memory, response modalities) as dichotomous resources would offer a parsimonious explanation if it was correct, but there is no inherent reason to expect homogeneous constraints on human information processing at each of these levels.

Other models of human performance may also offer insight regarding the results obtained in Experiment 3. For example, Ratwani, Andrews, Sousk, and Trafton (2008) recently examined the effects of the modality of an *interruption*—a term which perhaps most accurately captures the true configuration of the sequential-but-overlapping dual-task paradigm used here—on performance of a primary task. In their study, a visual primary task (completing a series of commands on a visual computer interface in the correct order) was interrupted by a digit-adding task that was presented either: 1) in the visual modality, where the primary task screen was replaced by the interruption task presentation until the interruption task was complete; 2) in the auditory modality, where the primary task screen disappeared until performance of the interruption task was complete; or 3) in the auditory modality, where participants were allowed to view the primary task screen during the interruption but could not resume until the interruption task was complete. Results showed that the visual interference task and the auditory interference task with an invisible screen were not significantly different from each other, and both interfered with resumption of the primary task more than the auditory interference condition where the screen remained visible. The study compared the predictions of multiple resources models (which have already been explored in detail here) to the predictions of a memory for goals model, which has suggested that interference tasks are more or less disruptive to a primary task to the extent that cues during interference do or do not allow resumption of the primary task. The multiple resources approach easily explains the disruption of a visual task by a visual interference task, but it does not readily explain the finding that an auditory-verbal task could also interfere with a visual task. Instead, the researchers suggested that the availability of cues for resuming the primary task during the interference task (as opposed to the specific or selective interference properties of the interference task) predicted the extent to which interference was observed. The current study's finding of a general interference effect, then, fit the predictions of the memory for goals model in that the single task conditions allowed for the participants to focus exclusively on resuming the primary sonification task after the delay period, whereas each of the interference tasks generally disrupted the maintenance of cues and interfered with the resumption of the primary task.

Summary of Experiment 3

The primary findings of Experiment 3 can be summarized as follows:

1) The pattern of interference with the sonification task was not selective as a function of the qualitative combination of the encoding strategy and the demands of the interference task. Instead, the pattern of interference showed a general effect whereby performance of the sonification task decreased in the presence of any interference task, regardless of its qualitative aspects.

2) The hypotheses regarding auditory representation as a distinct pool of processing code resources were not confirmed. Since no evidence of distinct verbal or visuospatial representational processes was found either, this is not necessarily evidence against the distinctness of auditory representation. Rather, the results of Experiment 3, which were predicated on the predictions of a multiple resources approach, seemed to suggest that representational interference in working memory (when a stimulus has been encoded and is not actively being perceived) does not show the selectivity predicted by multiple resources models. This may be attributable to the sequential-but-overlapping interference task design, and may further suggest that the processing codes dimension of multiple resources theory is most relevant for internal representation during actual perception (i.e., true time-sharing of tasks as opposed to maintenance and rehearsal of a representation in the absence of perception).

3) Again, as in Experiments 1 and 2, the data are consistent with the interpretation that the auditory sensory representation of the initial external sonification

stimulus may have had a lingering, protracted influence across all encoding strategies. This was evident from the finding that the auditory interference task was most disruptive to the point estimation sonification task. An alternative explanation—that the auditory interference task was simply quantitatively harder than the other interference tasks—was not entirely supported by the TLX subjective workload measurements, as the visuospatial interference block equivalent to the auditory interference block with respect to workload yet interfered with the sonification task significantly less.

5 GENERAL DISCUSSION

The pattern of results obtained in Experiments 1 and 2 generally confirmed the hypothesis that nonspeech sounds can be flexibly encoded in working memory as visuospatial images, verbal representations, or auditory images. In general, patterns of reaction times for both the mental scanning and sound-sentence-picture verification tasks differentiated verbal, visuospatial, and auditory encoding of stimuli. The results of Experiment 3 failed to differentiate internal representations with domain-specific interference tasks. This finding is perhaps most plausibly interpreted as a shortcoming in the model of predicted task interference (multiple resources) than it is a refutation of the existence of domain-specific working memory representations, as ample data and theory have supported the uniqueness of visuospatial, verbal, and now auditory working memory processes. Instead the results of Experiment 3 suggest that, in some circumstances, general interference in working memory may occur regardless of the domain-specificity of representations.

The pattern of results across studies suggested the possibility of a lingering effect of the auditory stimulus that might be attributable to a persevering form of echoic or auditory sensory memory. This effect was most directly observed in Experiment 2 when a sound verification stimulus was verified faster if the study stimulus had also been a sound, regardless of the encoding strategy manipulation (see Figure 8). Sentence and picture verification stimuli, on the other hand, showed no congruency effects with the format of the study stimulus. While the other studies did not directly examine the possibility of protracted auditory sensory memory, the results from Experiment 1 and Experiment 3 are also consistent with this interpretation. In Experiment 1, each encoding strategy showed sensitivity to the number of tones presented in the initial sonification stimulus, which is consistent with the interpretation that the effect of the initial sound stimulus lingered in addition to and in spite of the encoding strategy manipulation, which was also effective. In Experiment 3, the auditory interference task was significantly more disruptive to performance of the sonification task than the other interference tasks. While this could possibly result from strictly quantitative interference if the auditory interference task was simply more difficult than the other interference tasks, the TLX data suggested that the visuospatial interference task block was as difficult as the auditory interference task block, yet it was not as disruptive to the sonification task.

The duration of the “echoic” or auditory sensory store has never been established, and differently researchers have varied widely in their estimates of its time course (ranging from a few hundred milliseconds up to a few seconds, see Cowan, 1984; Hicks, 1974; Neisser, 1967). Cowan proposed a dual-process acoustic sensory memory system that featured a short store (of up to 350 ms) and a long store (of up to 20 s). The short store essentially represented the fleeting perceptual present, whereas the long store was post-perceptual. The specific properties of the long memory store have yet to be firmly established, but the current results suggest that memory for the perceptual qualities of a nonspeech auditory stimulus may linger for several seconds following the presentation of the stimulus, even when the initial percept has been successfully recoded into a different domain (e.g., when the sound is remembered as picture or image). Modifications of the sound-sentence-picture verification task may prove useful for establishing a more determinate duration of this effect, as discussed below.

Contributions of These Studies to the Human Factors of Sonification

The use of sonifications in systems will require an understanding of how people rehearse and maintain nonspeech sound stimuli. To date the bulk of auditory display research has focused on the sensory-perceptual aspects of information processing for nonspeech sounds; very little attention has been paid to auditory cognition. The results of these studies suggest that nonspeech sounds can be encoded flexibly as verbal representations, visuospatial images, or auditory images.

Experiment 1 suggested that the time required for the mental rehearsal and examination of an internal representation in the absence of the external percept will be affected by the number of items present in the original sound percept, regardless of the temporal characteristics of the sound. A task that requires the internal re-examination of a sonification stimulus, then, will require more time as the number of tones presented in the sonification increases. This effect was not attributable to the duration of the stimulus, which was held constant, and it was found across encoding strategies.

Experiment 2 offered perhaps a more general rule across external formats of stimuli. Specifically, for a task that requires a person to encode an initial stimulus and then to compare the initial stimulus to a different stimulus following a delay, the person should generally be able to respond faster if they use an encoding strategy for the initial stimulus that matches the format of the post-delay comparison stimulus.

Regarding Experiment 3, despite the high likelihood that auditory displays will be implemented in the presence of other competing (often visual) tasks, only a few studies (e.g., Bonebright & Nees, 2009; Brock et al., 2002; Jeon et al., 2009; Nees & Walker, 2008b; Peres & Lane, 2005) to date have examined performance with sonifications in the presence of another task. To this author's knowledge, no other studies have examined sonification performance while systematically varying the qualitative aspects of potential interference tasks. With respect to the implementation of nonspeech sounds in systems, the results of Experiment 3 suggested the following about scenarios where a person must encode a sonification stimulus, attend to a different task, and then return to answer a question about the sonification based on their internal representation:

1) The encoding strategy for the sonification stimulus will have no effect upon performance and will not interact with the qualitative aspects of the interference task. Because the current study did not test a condition with a spontaneous or uninstructed encoding strategy, the remote possibility exists that a prescribed strategy could have a general effect on performance as compared to conditions where participants selected their own encoding strategies. For example, it is possible that participants would experiment with a variety of encoding strategies under spontaneous conditions, but it is unclear whether this would have a positive or negative impact on encoding vis-à-vis a consistent, prescribed encoding strategy across trials. Although more research is needed, the current results suggest that a system designer need not instruct a particular encoding strategy for sonification tasks, as all strategies resulted in extremely similar performance across conditions.

2) Regarding the effect of specific interference tasks, the results suggested that any inteference task of sufficient difficulty will result in a decrease in performance (as compared to performance with sonifications alone) when the interference task overlaps a retention period for the information encoded from the sonification. NASA-TLX results

likewise indicated that participants experienced higher subjective workload while performing the point estimation sonification task in the presence of an interference task.

3) The results here also indicated, however, that performance of the sonification task will proceed (albeit with increased error) in the presence of an interference task. The specific constraints of the tasks and systems for which sonifications are deployed will determine whether the increase in error due to an interference task is tolerable.

4) The auditory interference task showed the most pronounced interference with performance of the sonification task as measured by RMSE. The perceived workload of the combined sonification and auditory interference tasks was also higher than all other conditions except the combined visuospatial intererence and sonification tasks. This finding regarding the auditory interference task is discussed in the greater context of all three experiments above, but this suggests that a system designer should be careful when implementing sonifications in tasks where other auditory (i.e., nonverbal and non-visual) stimuli must be processed, even when the stimuli are separated in time and not perceived simultaneously.

5) Finally, the mere "separation-by-modality" justification for implementing auditory displays in the presence of other visually taxing tasks was not supported in the current study, as the visuospatial mental rotation task significantly disrupted the auditory sonification task across encoding strategies.

Overall, these studies suggested that the internal representation of nonspeech sounds in working memory may be either verbal, visuospatial, or auditory in format, depending on the listener's conscious encoding strategy. The encoding strategy may have meaningful effects on the duration of mental examination of nonspeech sounds in the absence of a percept. The encoding strategy may also affect the time it takes a listener to compare a sound to another subsequent stimulus. No support was found for the hypothesis that encoding strategy manipulations can offer a release from selective interference in domain-specific working memory, but more research is needed to examine this finding with respect to task requirements.

Future directions

A number of potentially informative follow-up investigations were suggested by the findings of the current studies. Regarding the duration of auditory sensory memory, the delay between encoding and mental scanning in Experiment 1 and the delay between encoding and stimulus verification in Experiment 2 were both held constant at 3000 ms. A straightforward extension of each of these studies would manipulate the time between encoding and scanning or verification in an effort to determine the temporal boundaries of what appears to be a lingering effect of the initial auditory stimulus. For example, if after a longer delay between encoding and scanning, mental scanning times for the auditory imagery and/or visuospatial imagery encoding conditions no longer showed an effect of the numbers of tones, the temporal crossover point for this phenomenon would be a candidate for the duration of Cowan’s (1984) long auditory store. Perhaps a more compelling demonstration would use the same design as Experiment 2, but would manipulate the time delay between the end of the encoding process and the appearance of the verification stimulus. If the facilitative effect of the sound study stimulus paired with the sound verification stimulus (collapsed across encoding strategies) shown in Figure 8 were to disappear with a longer interval between encoding and verification, then the temporal crossover boundary of the effect would again be a candidate estimate for the duration of auditory sensory memory.

Future research should examine and corroborate the instructional manipulations of the current study with biological evidence from neuroimaging. Research continues to accumulate (e.g., Belardinelli et al., in press; Farah et al., 1988; Halpern & Zatorre, 1999; Halpern et al., 2004; Kosslyn & Thompson, 2002; Kraemer et al., 1995; Mellet et al., 2002; Zatorre, 2003) to suggest that domain-specific representation (e.g., visual imagery and auditory imagery) correlates with activation of the brain areas responsible for processing those same domain-specific stimuli during actual perception (e.g., the visual and auditory cortices, respectively). The instructional manipulations used in all three studies here lend themselves to very straightforward and precise hypotheses regarding domain-specific activation as a function of encoding strategy, as one would expect to observe domain-specific biological activity during the successful implementation of a given encoding strategy (e.g., verbal encoding, visuospatial imagery, or auditory imagery), regardless of the format of the external stimulus to be encoded.

Neuroimaging may also eventually confirm or refute the hypothesis that interference in working memory in sequential but overlapping paradigms is domain general. The role of executive processes, for example, should be reflected in biological processes that are active across tasks and encoding strategies if the interference is indeed general. Clearly more behavioral research is also needed to identify the precise nature of interference in tasks that overlap. Inconclusive research to date, for example, seems to suggest that manipulations of verbal and spatial processing codes generally conform to the task interference predictions of multiple resources under timesharing where the perceptual aspects of both tasks overlap (e.g., Brill et al., 2008; Ferris et al., 2007; Jeon et al., 2009; Wickens & Liu, 1988), but Experiment 3 suggested a general representational resource for the sequential but overlapping task design when only representational requirements overlap. Studies should use a battery of domain-specific interference tasks to look for patterns of selective interference for a sonification task during the perception of sonification stimuli (rather than after the stimuli had been encoded in working memory). While perception and representation have been shown to share representational processing (Kosslyn & Thompson, 2002; Mechelli et al., 2004), the biological boundaries of the two processes have yet to be defined (Behrmann et al., 1994; Kosslyn & Thompson, 2003). Experiment 3 suggested that the same holds true for the behavioral boundaries of task interference under conditions of overlapping perception and representation.

Conclusions

Although more research is needed, the current studies began to place pseudo-isomorphic auditory representations alongside the traditional verbal and visuospatial representational domains of working memory. These findings have important consequences for theoretical accounts of working memory and information processing. The present findings showed that the differentiation of verbal, visuospatial, and nonspeech auditory information processing in working memory warrants further examination. Similarly, the present paradigm suggests that individual encoding strategies for nonspeech stimuli are malleable, and also that the external format of the stimulus does not necessarily dictate the domain-specific working memory representation of the stimulus. Results suggested that representational interference is general and not selective, which failed to confirm the predictions of the multiple resources approach. Future work should expand the paradigms used here to examine the temporal properties of auditory sensory memory, and also to qualify the precise circumstances in which the representational demands of concurrent tasks will create general or selective information processing conflicts.

APPENDIX A: MODIFIED AUDITORY IMAGERY QUESTIONNAIRE (MODIFIED FROM SEASHORE, 1919, P. 216)

For the next few questions, you'll be asked to rate how well you feel you can imagine some specific auditory experiences in your pitch memory. Pitch memory is like when a song gets stuck in your head. Some people can vividly imagine hearing a song or a certain sound in their mind--so vivid that it's almost like actually hearing the song or sound. Other people find it more difficult to imagine sounds. The ability to imagine sounds is called "auditory imagery," and you'll be asked to form some auditory images and rate how vivid these images are.

1. Can you imagine the sound of a gun shot? How vivid is your auditory image?
2. Can you imagine the sound of clinking water glasses? How vivid is your auditory image?
3. Can you imagine the sound of the ringing of bells? How vivid is your auditory image?
4. Can you imagine the sound of the hum of bees? How vivid is your auditory image?
5. Can you imagine the characteristic tone quality of a piano? How vivid is your auditory image?
6. Can you imagine the characteristic tone quality of a flute? How vivid is your auditory image?
7. Can you imagine the sound of your favorite song? How vivid is your auditory image?
8. Can you imagine the sound of your psychology instructor’s voice? How vivid is your auditory image?

[Participants answered each of the questions above by choosing a rating from the following scale:]

0 – No image at all
1 – Very faint
2 – Faint
3 – Fairly vivid
4 – Vivid
5 – Very vivid
6 – As vivid as actually hearing

APPENDIX B: POST-TRIAL STRATEGY COMPLIANCE QUESTION

Please check the box of the strategy you used during memory scanning on the last trial. Check all that apply. Please be honest, even if you used the wrong strategy—it is very important that we know which strategy you used.

--picture strategy

--word strategy

--sound strategy

--not sure

Please briefly describe any problems you had with using the strategy you were given for this block on the last trial. (Leave blank if you had no problems).

[text box for answer]

APPENDIX C: CORRELATIONS OF SUBJECT VARIABLES WITH SCANNING TIMES IN EXPERIMENT 1

Table 12: Correlation coefficients for subject variables and average scanning time across pairings of the frequency change and numbers of tones manipulations for the picture strategy

Number of tones	two			three			four		
Frequency change	small	medium	large	small	medium	large	small	medium	large
Music 1[1]	-.20	-.25	-.24	-.22	-.18	-.25	-.08	-.14	-.23
Music 2[2]	-.11	-.14	-.09	-.10	-.06	-.12	.03	-.01	-.08
Music 3[3]	-.19	-.19	-.15	-.27	-.07	-.16	-.05	-.14	-.12
SAT V[4]	-.23	-.18	.08	-.13	-.06	-.06	-.08	-.09	.09
SAT M[5]	-.13	-.21	.04	-.04	-.07	-.09	-.06	-.01	.14
SAT W[6]	-.01	-.09	.13	.14	.06	-.02	.09	.13	.18
V abil[7]	-.25	-.23	-.18	-.13	-.27	-.19	-.21	-.19	-.20
Sp abil[8]	-.09	-.06	.09	-.11	.11	-.04	-.04	-.03	.13
Cog style[9]	-.17	-.17	-.20	-.09	-.18	-.17	-.06	-.16	-.20
Clock[10]	.17	.15	-.07	.21	.08	.07	.22	.11	-.02
Aud imag[11]	-.26	-.30*	-.18	-.29	-.23	-.20	-.25	-.23	-.10

**p* < .05; ***p* < .01

Table 13: Correlation coefficients for subject variables and average scanning time across pairings of the frequency change and numbers of tones manipulations for the sentence strategy

Number of tones	two			three			four		
Frequency change	small	medium	large	small	medium	large	small	medium	large
Music 1[1]	-.39**	-.39**	-.36*	-.37*	-.26	-.24	-.38*	-.31*	-.37*
Music 2[2]	-.35*	-.32*	-.33*	-.33*	-.24	-.19	-.33*	-.26	-.31
Music 3[3]	-.39**	-.41**	-.35*	-.37*	-.29	-.29	-.37*	-.35*	-.39**
SAT V[4]	-.17	-.23	-.37*	-.20	-.44**	-.34*	-.35*	-.30*	-.26
SAT M[5]	-.13	-.21	-.38*	-.16	-.43**	-.31*	-.30	-.24	-.17
SAT W[6]	-.03	-.12	-.27	-.07	-.20	-.13	-.17	-.10	>-.01
V abil[7]	-.20	-.33*	-.41**	-.19	-.40**	-.29	-.40**	-.29	-.31*
Sp abil[8]	-.34*	-.30	-.31*	-.36*	-.38*	-.31*	-.40*	-.29	-.25
Cog style[9]	.12	-.09	-.01	-.01	.18	-.25	-.14	-.17	-.22
Clock[10]	.04	.10	-.12	-.07	-.11	-.09	-.05	-.21	-.11
Aud imag[11]	-.20	-.37*	-.25	-.16	-.28	-.07	-.18	-.19	-.26

**p* < .05; ***p* < .01

Table 14: Correlation coefficients for subject variables and average scanning time across pairings of the frequency change and numbers of tones manipulations for the auditory strategy

Number of tones	two			three			four		
Frequency change	small	medium	large	small	medium	large	small	medium	large
Music 1[1]	-.11	-.49**	-.29	-.41**	-.34*	-.17	-.39**	-.34*	-.27
Music 2[2]	-.04	-.41**	-.19	-.33*	-.27	-.11	-.32*	-.28	-.22
Music 3[3]	-.25	-.51**	-.35*	-.41**	-.41**	-.31*	-.33*	-.34*	-.31*
SAT V[4]	-.22	-.20	-.14	-.07	-.17	-.17	-.10	-.14	-.10
SAT M[5]	.02	-.11	.05	.01	<.01	.10	.07	-.01	.01
SAT W[6]	-.02	-.10	-.02	-.09	-.07	.08	.01	.04	.02
V abil[7]	-.06	-.20	-.08	-.11	-.05	-.02	-.21	-.14	-.05
Sp abil[8]	.03	-.14	-.03	-.01	-.09	.04	.10	.08	-.03
Cog style[9]	-.36*	-.13	-.23	-.22	-.13	-.24	-.18	-.19	-.07
Clock[10]	.12	<.01	.04	.09	.03	.13	.12	.09	.08
Aud imag[11]	-.14	-.27	-.15	-.15	-.16	-.09	.04	-.18	-.07

*p < .05; **p < .01

[1] M1: Years having played a musical instrument
[2] M2: Years of formal musical training
[3] M3: Years experience reading musical notation
[4] SATV: Self-reported SAT verbal score
[5] SATM: Self-reported SAT math score
[6] SATW: Self-reported SAT writing score
[7] V abil: Self-reported verbal ability
[8] S abil: Self-reported spatial ability
[9] Cog style: Self-reported cognitive style (lower scores indicate more verbal; higher scores indicate more spatial)
[10] Clock: Mental clock average response time for correct answers
[11] Aud imag: Total score on auditory imagery questionnaire (higher scores indicate more vivid self-reported auditory imagery)

APPENDIX D: SUBJECT-LEVEL DATA FOR THE PARTICIPANT EXCLUDED CASE-WISE IN EXPERIMENT 2

The table below shows the subject-level data collected the participant who was excluded case-wise from Experiment 2 based on strategy compliance. The participant reported no musical experience and self-reported "very low" verbal and spatial abilities, which perhaps offers some explanatory value with respect to the participant's inability to implement prescribed encoding strategies across the entire study.

Table 15: Demographic and subject level variables for participant who gave incomplete data due to encoding strategy noncompliance in Experiment 2

Subj[1]	M1[3]	M2[3]	M3[4]	SATV[5]	SATM[6]	SATW[7]	Vab[8]	Sab[9]	Cog[10]	Aud[11]	Clock[12]
024	0.0	0.0	0.0	--	760	--	1	1	1	2.1	4028.27

---

[1] Subj: Participant number
[2] M1: Years having played a musical instrument
[3] M2: Years of formal musical training
[4] M3: Years experience reading musical notation
[5] SATV: Self-reported SAT verbal score
[6] SATM: Self-reported SAT math score
[7] SATW: Self-reported SAT writing score
[8] Vab: Self-reported verbal ability
[9] Sab: Self-reported spatial ability
[10] Cog: Self-reported cognitive style (lower scores indicate more verbal; higher scores indicate more spatial)
[11] Aud: Average score on auditory imagery questionnaire (higher scores indicate more vivid self-reported auditory imagery)
[12] Clock: Mental clock average response time for correct answers

APPENDIX E: CORRELATIONS OF SUBJECT VARIABLES WITH VERIFICATION PERFORMANCE IN EXPERIMENT 2

Table 16: Correlation coefficients for subject variables and average verification times across pairings of study with verification stimuli for the verbal strategy

Study stim	sentence			picture			sound		
Verification stim	sentence	picture	sound	sentence	picture	sound	sentence	picture	sound
Music 1[1]	-.19	-.04	-.16	-.11	-.08	-.19	-.14	<-.01	-.01
Music 2[2]	-.11	.02	-.11	-.02	.01	-.16	-.05	.07	.08
Music 3[3]	-.21	-.03	-.18	-.08	-.08	-.24	-.13	.01	.03
SAT V[4]	-.33*	-.17	-.09	-.34*	-.34*	-.01	-.29	-.07	-.14
SAT M[5]	-.21	-.18	-.01	-.08	-.14	-.02	-.15	-.02	.02
SAT W[6]	-.29	-.09	-.14	-.31	-.29	-.02	-.23	-.01	-.18
V abil[7]	.17	.24	.05	-.13	.10	.02	-.02	.08	-.11
Sp abil[8]	.03	.12	.09	-.18	-.04	-.03	-.09	-.01	-.15
Cog style[9]	-.08	.04	-.15	-.07	-.04	-.15	-.06	.11	.05
Clock[10]	.20	.10	.14	.13	.22	.19	.20	-.07	.25
Aud imag[11]	.12	.14	-.06	-.07	.17	.07	.01	-.08	.06

*$p < .05$; **$p < .01$

Table 17: Correlation coefficients for subject variables and average verification time across pairings of study with verification stimuli for the picture strategy

Study stim	sentence			picture			sound		
Verification stim	sentence	picture	sound	sentence	picture	sound	sentence	picture	sound
Music 1[1]	-.30	-.06	-.30	-.25	-.13	-.36*	-.18	-.22	-.21
Music 2[2]	-.14	.07	-.17	-.19	-.03	-.23	-.10	-.09	-.03
Music 3[3]	-.21	-.02	-.22	-.27	-.09	-.29	-.09	-.14	-.08
SAT V[4]	-.29	-.29	-.04	-.49**	-.34*	-.28	-.18	-.39*	-.07
SAT M[5]	-.23	-.19	-.07	-.17	-.22	-.08	-.11	-.16	-.10
SAT W[6]	-.39*	-.42**	-.32	-.38*	-.29	-.32	-.12	-.33*	-.25
V abil[7]	.15	.21	.20	-.28	.26	.06	.14	.10	.08
Sp abil[8]	-.20	-.11	-.23	-.43**	-.15	-.26	-.10	-.14	-.07
Cog style[9]	.19	.15	.10	-.15	.08	.03	.04	.01	.23
Clock[10]	.41*	.29	.34	.19	.55**	.34*	.18	.41*	.16
Aud imag[11]	.18	.21	.09	-.05	.32*	.04	.21	.21	.10

*$p < .05$; **$p < .01$

Table 18: Correlation coefficients for subject variables and average verification time across pairings of study with verification stimuli for the auditory strategy

Study stim	sentence			picture			sound		
Verification stim	sentence	picture	sound	sentence	picture	sound	sentence	picture	sound
Music 1[1]	-.10	-.19	-.25	-.19	-.21	-.16	-.14	-.02	-.18
Music 2[2]	-.06	-.17	-.17	-.11	-.15	-.03	-.06	.03	-.14
Music 3[3]	-.13	-.22	-.25	-.20	-.24	-.11	-.13	-.06	-.20
SAT V[4]	-.32	-.27	-.23	-.19	-.25	-.07	-.38*	-.14	-.26
SAT M[5]	-.31	-.10	.01	-.31	-.12	-.09	-.24	-.16	-.21
SAT W[6]	-.17	-.26	-.38	-.31	-.24	-.14	-.43**	-.20	-.19
V abil[7]	.11	.08	.06	.24	-.02	.14	-.14	.19	.09
Sp abil[8]	.06	.12	.03	.22	-.09	.16	-.15	.02	.11
Cog style[9]	-.06	-.08	-.06	-.05	-.19	.05	-.04	.04	-.15
Clock[10]	-.01	.07	.09	.13	.06	.11	.09	.26	.27
Aud imag[11]	.04	.01	-.05	.13	.03	.07	-.08	.24	.14

**p* < .05; ***p* < .01

[1] M1: Years having played a musical instrument
[2] M2: Years of formal musical training
[3] M3: Years experience reading musical notation
[4] SAT V: Self-reported SAT verbal score
[5] SAT M: Self-reported SAT math score
[6] SAT W: Self-reported SAT writing score
[7] V abil: Self-reported verbal ability
[8] Sp abil: Self-reported spatial ability
[9] Cog style: Self-reported cognitive style (lower scores indicate more verbal; higher scores indicate more spatial)
[10] Clock: Average reaction time for correct answers for the mental clock test.
[11] Aud imag: Total score on auditory imagery questionnaire (higher scores indicate more vivid self-reported auditory imagery)

APPENDIX F: EXPLANATIONS OF INCOMPLETE PARTICIPANT DATA IN EXPERIMENT 3

Table 19: Explanations of incomplete participant data in Experiment 3

Participant number	Encoding condition	Blocks missing	Reason for incomplete data
52005	Verbal	Verbal interference Spatial interference	Computer crash during data collection
52015	Auditory	Auditory interference Verbal interference Spatial interference	Strategy compliance
52020	Verbal	Auditory interference Spatial interference	Quit study
52022	Verbal	Spatial interference	Quit study (didn’t finish in time)
52023	Verbal	Spatial interference TLX	Quit study (didn’t finish in time)
52042	Verbal	Single task	Strategy compliance
52044	Visual	Spatial interference	Strategy compliance
52047	Visual	Auditory interference Spatial interference	Strategy compliance

APPENDIX G: SUBJECT-LEVEL DATA FOR PARTICIPANTS EXCLUDED CASE-WISE IN EXPERIMENT 3

The table below shows the subject-level data collected for each of the participants who were excluded case-wise based on strategy compliance.

Table 20: Demographic and subject level variables for participants who gave incomplete data due to encoding strategy noncompliance compliance

Subj[1]	Strat[2]	M1[3]	M2[4]	M3[5]	SATV[6]	SATM[7]	SATW[8]	Vab[9]	Sab[10]	Cog[11]	Aud[12]	Clock[13]
52013	A	9.0	9.0	9.0	690	710	--	3	3	6	4.0	6887.0
52042	Ve	7.0	7.0	7.0	580	700	680	3	3	2	4.9	7589.2
52044	Vi	0.0	1.0	1.0	600	800	600	3	3	1	3.8	7262.3
52047	Vi	0.0	0.0	0.0	540	790	640	3	4	1	4.1	8424.1

---

[1] Subj: Participant number

[2] Strat: Encoding strategy; A = Auditory, Ve = Verbal, Vi = Visuospatial

[3] M1: Years having played a musical instrument

[4] M2: Years of formal musical training

[5] M3: Years experience reading musical notation

[6] SATV: Self-reported SAT verbal score

[7] SATM: Self-reported SAT math score

[8] SATW: Self-reported SAT writing score

[9] Vab: Self-reported verbal ability

[10] Sab: Self-reported spatial ability

[11] Cog: Self-reported cognitive style (lower scores indicate more verbal; higher scores indicate more spatial)

[12] Aud: Average score on auditory imagery questionnaire (higher scores indicate more vivid self-reported auditory imagery)

[13] Clock: Mental clock average response time for correct answers

APPENDIX H: ADDITIONAL FIGURES FROM EXPERIMENT 3

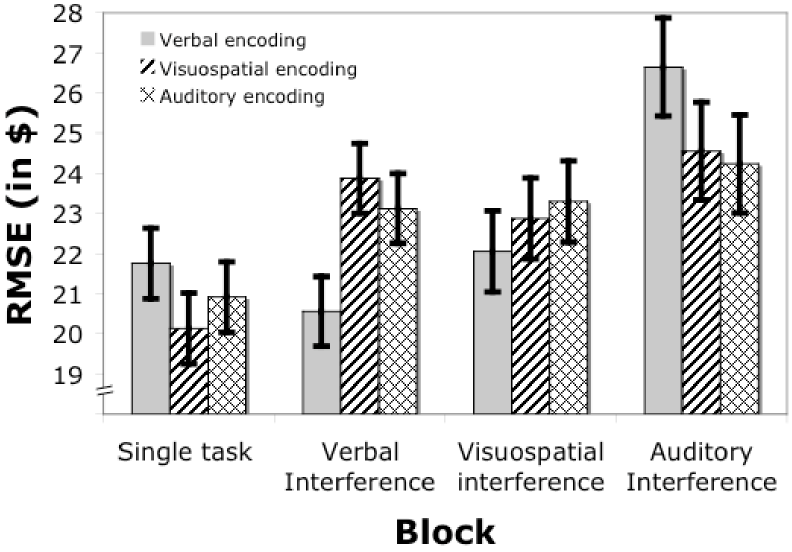

Figure 13: Alternate version of Figure 11 with the nonsignificant effect of encoding strategy depicted.

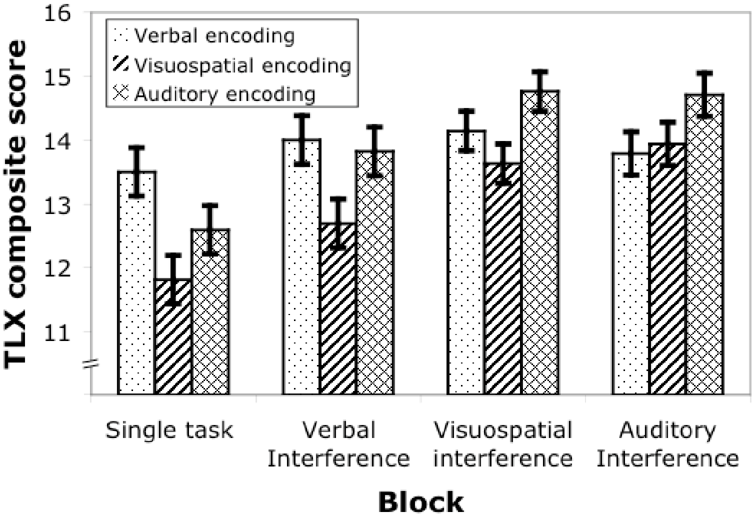

Figure 14: Alternate version of Figure 12 with the nonsignificant effect of encoding strategy depicted.

APPENDIX I: CORRELATIONS OF SUBJECT-LEVEL VARIABLES WITH PERFORMANCE VARIABLES IN EXPERIMENT 3

Table 21: Correlation coefficients for subject variables with average RMSE, average interference task percent correct, and TLX composite scores across study blocks

	RMSE				Percent correct			TLX			
	single	verbal	spatial	auditory	verbal	spatial	auditory	single	verbal	spatial	auditory
Music 1[1]	-.07	-.25	-.12	-.26	.12	.02	.39**	-.02	-.10	.05	.09
Music 2[2]	-.11	-.24	-.27	-.28	.13	.01	.39**	-.06	-.08	.04	.09
Music 3[3]	-.6	-.14	-.16	-.29*	.10	.03	.34*	-.14	-.10	.00	.10
SAT V[4]	-.24	-.34*	-.30*	-.18	-.10	.07	.17	-.12	-.12	-.19	-.03
SAT M[5]	-.26	-.37**	-.34*	-.23	-.07	.13	.07	-.18	-.17	-.12	-.09
SAT W[6]	-.09	-.29*	-.10	-.17	-.01	.23	.12	.01	-.10	-.05	-.03
V abil[7]	-.01	.15	.07	.12	.04	-.26	.38**	.09	.12	.26	.26
Sp abil[8]	-.01	-.13	.02	.15	.02	.18	-.11	-.13	-.03	-.10	-.01
Cog style[9]	-.05	.24	.10	.16	-.04	-.16	.20	-.03	.03	.11	.26
Clock[10]	-.04	.00	.00	.11	.07	-.19	.15	.04	.03	-.04	-.07
Aud imag[11]	-.01	.08	.09	-.07	.21	.02	.27	-.10	.12	.13	.28

*$p < .05$; **$p < .01$; *N*'s contributing to these correlations range from 47 to 51

[1] M1: Years having played a musical instrument
[2] M2: Years of formal musical training
[3] M3: Years experience reading musical notation
[4] SAT V: Self-reported SAT verbal score
[5] SAT M: Self-reported SAT math score
[6] SAT W: Self-reported SAT writing score
[7] V abil: Self-reported verbal ability
[8] Sp abil: Self-reported spatial ability
[9] Cog style: Self-reported cognitive style (lower scores indicate more verbal; higher scores indicate more spatial)
[10] Clock: Average reaction time for correct answers for the mental clock test.
[11] Aud imag: Total score on auditory imagery questionnaire (higher scores indicate more vivid self-reported auditory imagery)

REFERENCES

Anderson, J. R., Yulin, Q., Jung, K., & Carter, C. S. (2007). Information processing modules and their relative modality specificity. *Cognitive Psychology*, 54, 185-217.

Awh, E., Jonides, J., & Reuter-Lorenz, P. A. (1998). Rehearsal in spatial working memory. *Journal of Experimental Psychology: Human Perception and Performance*, 24(3), 780-790.

Baddeley, A. D. (1992). Is working memory working? The fifteenth Bartlett lecture. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, 44(1), 1-31.

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

Baddeley, A. D. (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience*, 4(10), 829-839.

Baddeley, A. D., & Andrade, J. (2000). Working memory and the vividness of imagery. *Journal of Experimental Psychology: General*, 129(1), 126-145.

Baddeley, A. D., & Logie, R. H. (1992). Auditory imagery and working memory. In D. Reisberg (Ed.), *Auditory Imagery* (pp. 179-197). Hillsdale, NJ: Lawrence Erlbaum Associates.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, 22(04), 577-660.

Behrmann, M., Moscovitch, M., & Winocur, G. (1994). Intact visual imagery and impaired visual perception in a patient with visual agnosia. *Journal of*

Experimental Psychology: Human Perception and Performance, 20(5), 1068-1087.

Belardinelli, M. O., Palmeiro, M., Sestieri, C., Nardo, D., Di Matteo, R., Londei, A., et al. (in press). An fMRI investigation on image generation in different sensory modalities: The influence of vividness. *Acta Psychologica*, 132(TBD), TBD.

Ben-Artzi, E., & Marks, L. E. (1995). Visual-auditory interaction in speeded classification: Role of stimulus difference. *Perception & Psychophysics*, 57(8), 1151-1162.

Bertolo, H. (2005). Visual imagery without visual perception? *Psicológica*, 26, 173-188.

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

Bonnel, A.-M., Faita, F., Peretz, I., & Besson, M. (2001). Divided attention between lyrics and tunes of operatic songs: Evidence for independent processing. *Perception & Psychophysics*, 63(7), 1201-1212.

Brandimonte, M. A., Hitch, G. J., & Bishop, D. V. (1992). Verbal recoding of visual stimuli impairs mental image transformations. *Memory & Cognition*, 20(4), 449-455.

Brechmann, A., & Scheich, H. (2005). Hemispheric shifts of sound representation in auditory cortex with conceptual listening. *Cerebral Cortex*, 15(5), 578-587.

Brewster, S. (1997). Using non-speech sound to overcome information overload. *Displays*, 17, 179-189.

Brewster, S., & Murray, R. (2000). Presenting dynamic information on mobile computers. *Personal Technologies*, 4(4), 209-212.

Brill, J. C., Mouloua, M., Gilson, R. D., Rinalducci, E. J., & Kennedy, R. S. (2008). Effects of secondary loading task modality on reserve attentional capacity. *Proceedings of the Human Factors and Ergonomics Society 52nd Annual Meeting* (pp. 1219-1223), New York, NY.

Broadbent, D. E. (1952/1992). Listening to one of two synchronous messages. *Journal of Experimental Psychology: General*, 121(2), 125-127.

Broca, P. (1861/2000). Remarks on the seat of the faculty of articulated language, following an observation of aphemia (loss of speech), Bulletin de la Societe Anatomique (Vol. 6, pp. 330-357).

Brock, D., Stroup, J. L., & Ballas, J. A. (2002). Using an auditory display to manage attention in a dual task, multiscreen environment. *Proceedings of the International Conference on Auditory Display (ICAD2002)*, Kyoto, Japan.

Brodsky, W., Henik, A., Rubinstein, B. S., & Zorman, M. (2003). Auditory imagery from musical notation in expert musicians. *Perception & Psychophysics*, 65(4), 602-612.

Brown, M. L., Newsome, S. L., & Glinert, E. P. (1989). An experiment into the use of auditory cues to reduce visual workload. *Proceedings of the ACM CHI 89 Human Factors in Computing Systems Conference (CHI 89)* (pp. 339-346).

Byrne, M. D., & Anderson, J. R. (2001). Serial modules in parallel: The psychological refractory period and perfect time-sharing. *Psychological Review*, 108(4), 847–869.

Carpenter, P. A., & Just, M. A. (1975). Sentence comprehension: A psycholinguistic processing model of verification. *Psychological Review*, 82(1), 45-73.

Chandler, P., & Sweller, J. (1991). Cognitive Load Theory and the Format of Instruction. *Cognition and Instruction*, 8(4), 293-332.

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *The Journal of the Acoustical Society of America*, 25, 975.

Clark, H. H., & Chase, W. G. (1972). On the process of comparing sentences against pictures. *Cognitive Psychology*, 3(3), 472-517.

Colegate, R. L., & Eriksen, C. W. (1970). Implicit speech as an encoding mechanism in visual perception. *The American Journal of Psychology*, 83(2), 208-215.

Coney, J. (1988). Individual differences and task format in sentence verification. *Current Psychology*, 7(2), 122-135.

Cowan, N. (1984). On short and long auditory stores. *Psychological Bulletin* 96(2), 341-370.

Crowder, R. G. (1993). Auditory memory. In S. McAdams & E. Bigand (Eds.), *Thinking in Sound: The Cognitive Psychology of Human Audition* (pp. 113-145). Oxford: Clarendon Press.

Damos, D. (1991). Dual-task methodology: Some common problems. In D. Damos (Ed.), *Multiple-task Performance* (pp. 101-119). Washington, D.C.: Taylor & Francis.

De Beni, R., Pazzaglia, F., & Gardini, S. (2007). The generation and maintenance of visual mental images: Evidence from image type and aging. *Brain and Cognition*, 63, 271-278.

Denis, M. (2007). Assessing the symbolic distance effect in mental images constructed from verbal descriptions: A study of individual differences in the mental comparison of distances. *Acta Psychologica*.

Denis, M., Concalves, M. R., & Memmi, D. (1995). Mental scanning of visual images generated from verbal descriptions: Toward a model of image accuracy. *Neuropsychologia*, 33(11), 1511-1530.

Denis, M., & Zimmer, H. D. (1992). Analog properties of cognitive maps constructed from verbal descriptions. *Psychological Research*, 54(4), 286-298.

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

Dunlosky, J., & Hertzog, C. (1998). Aging and deficits in associative memory: What is the role of strategy production? *Psycholoy and Aging*, 13(4), 597-607.

Dunlosky, J., & Hertzog, C. (2001). Measuring strategy production during associative learning: The relative utility of concurrent versus retrospective reports. *Memory & Cognition*, 29(2), 247-253.

Farah, M. J. (1985). Psychophysical evidence for a shared representational medium for mental images and percepts. *Journal of Experimental Psychology: General*, 114(1), 91–103.

Farah, M. J., Peronnet, F., Gonon, M. A., & Giard, M. H. (1988). Electrophysiological evidence for a shared representational medium for visual images and visual percepts. *Journal of Experimental Psychology: General*, 117(3), 248–257.

Farah, M. J., & Smith, A. F. (1983). Perceptual interference and facilitation with auditory imagery. *Perception & Psychophysics*, 33(5), 475-478.

Fausset, C. B., Rogers, W. A., & Fisk, A. D. (2008). Understanding the required resources in line graph comprehension. *Proceedings of the Human Factors and*

Ergonomics Society 52nd Annual Meeting (pp. 1830-1834), New York, New York.

Ferris, T., Hameed, S., Penfold, R., & Rao, N. (2007). Tactons in multitask environments: The interaction of presentation modality and processing code. *Proceedings of the Human Factors and Ergonomics Society 51st Annual Meeting* (pp. 1481-1485), Baltimore, MD.

Finke, R. A. (1980). Levels of equivalence in imagery and perception. *Psychological Review*, 87(2), 113-132.

Finke, R. A., Freyd, J. J., & Shyi, G. C. (1986). Implied velocity and acceleration induce transformations of visual memory. *Journal of Experimental Psychology: General*, 115(2), 175-188.

Fisk, A. D., Derrick, W. L., & Schneider, W. (1986). A methodological assessment and evaluation of dual-task paradigms. *Current Psychology*, 5(4), 315-327.

Galton, F. (1880). Statistics of mental imagery. *Mind*, 5, 301-318 [Electronic version]. Retrieved September 312, 2007, from http://psychclassics.yorku.ca/Galton/imagery.htm.

Glushko, R. J., & Cooper, L. A. (1978). Spatial comprehension and comparison processes in verification tasks. *Cognitive Psychology*, 10, 391-421.

Gopher, D., Brickner, M., & Navon, D. (1982). Different difficulty manipulations interact differently with task emphasis: Evidence for multiple resources. *Journal of Experimental Psychology: Human Perception and Performance*, 8(1), 146-157.

Gruber, O. (2001). Effects of domain-specific interference on brain activation associated with verbal working memory task performance. *Cerebral Cortex*, 11(11), 1047-1055.

Gruber, O., & Gotschke, T. (2004). Executive control emerging from dynamic interactions between brain systems mediating language, working memory and attentional processes. *Acta Psychologica*, 115, 105-121.

Guttman, S. E., Gilroy, L. A., & Blake, R. (2005). Hearing what the eyes see: Auditory encoding of visual temporal sequences. *Psychological Science*, 16(3), 228-235.

Halpern, A. R. (1988). Mental scanning in auditory imagery for songs. *Journal of experimental psychology. Learning, memory, and cognition*, 14(3), 434-443.

Halpern, A. R. (1989). Memory for the absolute pitch of familiar songs. *Memory & Cognition*, 17(5), 572-581.

Halpern, A. R. (1992). Musical aspects of auditory imagery. In D. Reisberg (Ed.), *Auditory Imagery* (pp. 1-27). Mahwah, NJ: Lawrence Erlbaum Associates.

Halpern, A. R., & Zatorre, R. J. (1999). When that tune runs through your head: A PET investigation of auditory imagery for familiar melodies. *Cerebral Cortex*, 9(7), 697-704.

Halpern, A. R., Zatorre, R. J., Bouffard, M., & Johnson, J. A. (2004). Behavioral and neural correlates of perceived and imagined musical timbre. *Neuropsychologia*, 42(9), 1281-1292.

Hart, S. G., & Staveland, L. E. (1988). Development of the NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock & N.

Meshkati (Eds.), *Human Mental Workload* (pp. 239-250). Amsterdam: North Holland Press.

Hicks, R. G. (1974). Duration of echoic memory. *Journal of Auditory Research*, 14, 125-132.

James, W. (1890). The Principles of Psychology [Electronic version]. Retrieved September 12, 2007, from http://psychclassics.yorku.ca/James/Principles/index.htm.

Jeon, M., Davison, B. K., Nees, M. A., Wilson, J., & Walker, B. N. (2009). Enhanced auditory menu cues improve dual-task performance and are preferred wth in-vehicle technologies. *Proceedings of the First International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI 2009)* (pp. TBD), Essen, Germany.

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

Keller, T. A., Cowan, N., & Saults, J. S. (1995). Can auditory memory for tone pitch be rehearsed? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(3), 635-635.

Kieras, D. E., & Meyer, D. E. (1997). An overview of the EPIC architecture for cognition and performance with application to human-computer interaction. *Human-Computer Interaction*, 12, 391-438.

Kirby, J. R., Moore, P. J., & Schofield, N. J. (1988). Verbal and visual learning styles. *Contemporary Educational Psychology*, 13(2), 169-184.

Kosslyn, S. M. (1973). Scanning visual images: Some structural implications. *Perception and Psychophysics*, 14(1), 90-94.

Kosslyn, S. M. (1975). Information representation in visual images. *Cognitive Psychology*, 7(3), 341-370.

Kosslyn, S. M. (1976). Can imagery be distinguished from other forms of internal representation? Evidence from studies of information retrieval time. *Memory & Cognition*, 4(3), 291-297.

Kosslyn, S. M. (1981). The medium and the message in mental imagery: A theory. *Psychological Review*, 88(1), 46-66.

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

Kosslyn, S. M., & Thompson, W. L. (2002). Shared mechanisms in visual imagery and visual perception: Insights from cognitive neuroscience. In M. S. Gazzaniga (Ed.), *The New Cognitive Neurosciences* (2nd ed., pp. 975-985). Cambridge, MA: MIT Press.

Kosslyn, S. M., & Thompson, W. L. (2003). When is early visual cortex activated during visual mental imagery? *Psychological Bulletin* 129(5), 723-746.

Kraemer, D. J. M., Macrae, C. N., Green, A. E., & Kelley, W. M. (1995). Sound of silence activates auditory cortex. *Nature*, 378, 496-498.

Kramer, G., Walker, B. N., Bonebright, T., Cook, P., Flowers, J., Miner, N., et al. (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: International Community for Auditory Display (ICAD).

Kroll, J. F., & Corrigan, A. (1981). Strategies in sentence-picture verification: The effect of an unexpected picture. *Journal of Verbal Learning and Verbal Behavior*, 20(5), 515-531.

Kubovy, M. (1981). Concurrent pitch segregation and the theory of indispensable attributes. In M. Kubovy & J. Pomerantz (Eds.), *Perceptual Organization* (pp. 55-99). Hillsdale, NJ: Lawrence Erlbaum.

Levitin, D. J. (1994). Absolute memory for musical pitch: Evidence from the production of learned melodies. *Perception & Psychophysics*, 56, 414-414.

Levitin, D. J., & Cook, P. R. (1996). Memory for musical tempo: Additional evidence that auditory memory is absolute. *Perception & Psychophysics*, 58, 927-935.

Levitin, D. J., & Rogers, S. E. (2005). Absolute pitch: Perception, coding, and controversies. *Trends in Cognitive Sciences*, 9(1), 26-33.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, 21(1), 1-36.

Liberman, A. M., & Whalen, D. H. (2000). On the relation of speech to language. *Trends in Cognitive Sciences*, 4(5), 187-196.

Logie, R. H., & Edworthy, J. (1986). Shared mechanisms in the processing of verbal and musical material. In D. G. Russell, D. Marks & J. Richardson (Eds.), *Imagery 2*. Dunedin, New Zealand: Human Performance Associates.

Lopes da Silva, F. H. (2003). Visual dreams in the congenitally blind? *Trends in Cognitive Sciences*, 7(8), 328-330.

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

Marquer, J., & Pereira, M. (1990). Reaction times in the study of strategies in sentence–picture verification: A reconsideration. *The Quarterly Journal of Experimental Psychology Section A*, 42(1), 147-168.

Martino, G., & Marks, L. E. (2001). Synthesia: Strong and weak. *Current Directions in Psychological Science*, 10(2), 61-65.

Massaro, D. W. (1970). Retroactive interference in short-term recognition memory for pitch. *Journal of Experimental Psychology* 83(1), 32-39.

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

Mayer, R. E., & Massa, L. J. (2003). Three facets of visual and verbal learners: Cognitive ability, cognitive style, and learning preference. *Journal of Educational Psychology*, 95(4), 833-846.

Mayer, R. E., & Moreno, R. (1998). A split-attention effect in multimedia learning: Evidence for dual processing systems in working memory. *Journal of Educational Psychology*, 90(2), 312-320.

Mayer, R. E., & Sims, V. K. (1994). For whom is a picture worth a thousand words? Extensions of a dual-coding theory of multimedia learning. *Journal of Educational Psychology*, 86(3), 389-401.

Mechelli, A., Price, C. J., Friston, K. J., & Ishai, A. (2004). Where bottom-up meets top-down: Neuronal interactions during perception and imagery. *Cerebral Cortex*, 14(11), 1256-1265.

Mellet, E., Bricogne, S., Crivello, F., Mazoyer, B., Denis, M., & Tzourio-Mazoyer, N. (2002). Neural basis of mental scanning of a topographic representation built from a text. *Cerebral Cortex*, 12(12), 1322-1330.

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(1), 3-65.

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

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

Moray, N. (1969). *Listening and Attention*. Baltimore, MD: Penguin Books.

Morris, N., Phillips, C., Thibault, K., & Chaparro, A. (2008). Sources of secondary task interference with driving: Executive processes or verbal and visuo-spatial rehearsal processes. *Proceedings of the Human Factors and Ergonomics Society 52nd Annual Meeting* (pp. 1556-1559), New York, New York.

Mowbray, G. H. (1953). Simultaneous vision and audition: The comprehension of prose passages with varying levels of difficulty. *Journal of Experimental Psychology*, 46, 365-372.

Navon, D. (1984). Resources: A theoretical soup stone? *Psychological review*, 91(2), 216-234.

Navon, D., & Gopher, D. (1979). On the economy of the human-processing system. *Psychological Review*, 86(3), 214-255.

Navon, D., & Miller, J. (2002). Queuing or sharing? A critical evaluation of the single-bottleneck notion. *Cognitive Psychology*, 44, 193-251.

Nees, M. A., & Walker, B. N. (2007). Listener, task, and auditory graph: Toward a conceptual model of auditory graph comprehension. *Proceedings of the International Conference on Auditory Display (ICAD2007)* (pp. 266-273), Montreal, Canada.

Nees, M. A., & Walker, B. N. (2008a). Encoding and representation of information in auditory graphs: Descriptive reports of listener strategies for understanding data. *Proceedings of the International Conference on Auditory Display (ICAD 08)*, Paris, FR (24-27 June).

Nees, M. A., & Walker, B. N. (2008b). Encoding of information in auditory displays: Initial research on flexibility and processing codes in dual-task scenarios. *Proceedings of the 52nd Meeting of the Human Factors and Ergonomics Society (HFES 2008)* (pp. 1820-1824), New York, NY.

Nees, M. A., & Walker, B. N. (2009). Auditory interfaces and sonification. In C. Stephanidis (Ed.), *The Universal Access Handbook* (pp. 507-521). New York: Lawrence Erlbaum Associates.

Neisser, U. (1967). *Cognitive Psychology*. New York: Meredith Publishing Company.

Neuhoff, J. G., Knight, R., & Wayand, J. (2002). Pitch change, sonification, and musical expertise: Which way is up? *Proceedings of the International Conference on Auditory Display*, 351–356.

Noordzij, M. L., van der Lubbe, R. H., Neggers, S. F., & Postma, A. (2004). Spatial tapping interferes with the processing of linguistic spatial relations. *Can J Exp Psychol*, 58(4), 259-271.

Noordzij, M. L., van der Lubbe, R. H. J., & Postma, A. (2005). Strategic and automatic components in the processing of linguistic spatial relations. *Acta Psychologica*, 119(1), 1-20.

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

Paivio, A. (1978). Comparisons of mental clocks. *J Exp Psychol Hum Percept Perform*, 4(1), 61-71.

Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology Revue Canadienne de Psychologie*, 45(3), 255-287.

Paivio, A. (2007). *Mind and Its Evoluation: A Dual Coding Theoretical Approach*. Mahwah, NJ: Lawrence Erlbaum Associates.

Pashler, H. (2000). Task switching and multitask performance. *Attention and performance XVIII: Control of cognitive processes*, 277–309.

Paulesu, E., Frith, C. D., & Frackowiak, R. S. J. (1993). The neural correlates of the verbal component of working memory. *Nature*, 362(6418), 342-345.

Peres, S. C., & Lane, D. M. (2005). Auditory graphs: The effects of redundant dimensions and divided attention. *Proceedings of the International Conference on Auditory Display (ICAD2005)* (pp. 169-174), Limerick, Ireland.

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

Peterson, E. R., Deary, I. J., & Austin, E. J. (2005). A new measure of Verbal–Imagery Cognitive Style: VICS. *Personality and Individual Differences*, 38(6), 1269-1281.

Podgorny, P., & Shepard, R. N. (1978). Functional representations common to visual perception and imagination. *Journal of Experimental Psychology: Human Perception and Performance*, 4(1), 21-35.

Pylyshyn, Z. (1981). The imagery debate: Analogue media versus tacit knowledge. *Psychological Review*, 88(1), 16-45.

Pylyshyn, Z. (2002). Mental imagery: In search of a theory. *Behavioral and Brain Sciences*, 25, 157-238.

Pylyshyn, Z. (2003). Return of the mental image: are there really pictures in the brain? *Trends in Cognitive Sciences*, 7(3), 113-118.

Ratwani, R., Andrews, A., Sousk, J., & Trafton, J. G. (2008). The effect of interruption modality on primary task resumption. *Proceedings of the Human Factors and Ergonomics Society 52nd Annual Meeting* (pp. 393-397), New York, NY.

Reichle, E. D., Carpenter, P. A., & Just, M. A. (2000). The neural bases of strategy and skill in sentence–picture verification. *Cognitive Psychology*, 40(4), 261-295.

Ries, D. T., & DiGiovanni, J. J. (2007). Release from interference in auditory working memory for pitch. *Hearing Research*, 230(1-2), 64-72.

Rossing, T. D. (1982). *The Science of Sound*. Reading, MA: Addison-Wesley.

Samuel, A. G. (1988). Central and peripheral representation of whispered and voiced speech. *Journal of Experimental Psychology: Human Perception and Performance*, 14(3), 379-388.

Schellenberg, E. G., & Trehub, S. E. (2003). Good pitch memory is widespread. *Psychological Science*, 14(3), 262-266.

Schneider, W. (2008, Feb 13). The synergy of the control and representational function of the human brain. Professional lecture given at the Georgia Institute of Technology, Atlanta, GA.

Schumacher, E. H., Lauber, E., Awh, E., Jonides, J., Smith, E. E., & Koeppe, R. A. (1996). PET Evidence for an amodal verbal working memory system. *Neuroimage*, 3(2), 79-88.

Seashore, C. E. (1919). *The Psychology of Musical Talent*. New York: Silver Burdett & Co.

Seppänen, M., Brattico, E., & Tervaniemi, M. (2007). Practice strategies of musicians modulate neural processing and the learning of sound-patterns. *Neurobiology of Learning and Memory*, 87(2), 236-247.

Shepard, R. N. (1975). Form, formation, and transformation of internal representations. *Information Processing and Cognition: The Loyola Symposium*.

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

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

Smith, J. D., Reisberg, D., & Wilson, M. (1992). Subvocalization and auditory imagery: Interactions between the inner ear and inner voice. In D. Reisberg (Ed.), *Auditory Imagery* (pp. 95-119). Hillsdale, NJ: Lawrence Erlbaum Associates.

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

Sternberg, S. (1969/2004). Memory-scanning: Mental processes revealed by reaction-time experiments. In D. A. Balota & E. J. Marsh (Eds.), *Cognitive Psychology: Key Readings* (pp. 48-74). New York: Psychology Press.

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

Szlichcinski, K. P. (1979). The art of describing sounds. *Applied Ergonomics*, 10(3), 131-138.

Tombu, M., & Jolicœur, P. (2002). All-or-none bottleneck versus capacity sharing accounts of the psychological refractory period phenomenon. *Psychological Research*, 66(4), 274-286.

Tremblay, S., Saint-Aubin, J., & Jalbert, A. (2006). Rehearsal in serial memory for visual-spatial information: Evidence from eye movements. *Psychonomic Bulletin & Review*, 13(3), 452-457.

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

Tversky, B. (1975). Pictorial encoding of sentences in sentence-picture comparison. *The Quarterly Journal of Experimental Psychology*, 27(3), 405-410.

Wagner, S., Winner, E., Cicchetti, D., & Garnder, H. (1981). "Metaphorical" mapping in human infants. *Child Development*, 52, 728-731.

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

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

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., & Nees, M. A. (2005). Conceptual versus perceptual training for auditory graphs. *Proceedings of the Human Factors and Ergonomics Society 49th Annual Meeting* (pp. 1598-1601), Orlando, FL.

Washburn, D. A., & Astur, R. S. (1998). Nonverbal working memory of humans and monkeys: Rehearsal in the sketchpad? *Memory & Cognition*, 26(2), 277-286.

Watson, C. S., & Kidd, G. R. (1994). Factors in the design of effective auditory displays. *Proceedings of the International Conference on Auditory Display (ICAD1994)*, Sante Fe, NM.

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

Wickens, C. D. (1991). Processing resources and attention. In D. 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(2), 159-177.

Wickens, C. D. (2007). How many resources and how to identify them? Commentary on Boles et al. and Vidulich and Tsang. *Human Factors*, 49(1), 53-56.

Wickens, C. D. (2008). Multiple resources and mental workload. *Human Factors*, 50(3), 449-455.

Wickens, C. D., & Hollands, J. G. (2000). Attention, time-sharing, and workload. In C. D. Wickens & J. G. Hollands (Eds.), *Engineering Psychology and Human Performance* (3rd ed., pp. 439-479). Upper Saddle River, NJ: Prentice Hall.

Wickens, C. D., Kramer, A., Vanasse, L., & Donchin, E. (1983a). Performance of concurrent tasks: A psychophysiological analysis of the reciprocity of information-processing resources. *Science*, 221(4615), 1080-1082.

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

Wickens, C. D., Sandry, D. L., & Vidulich, M. (1983b). Compatibility and resource competition between modalities of input, central processing, and output. *Human Factors*, 25(2), 227-248.

Zatorre, R. J. (2003). Hemispheric asymetries in the processing of tonal stimuli. In *The Asymmetrical Brain*. Cambridge, MA: MIT Press.

Zatorre, R. J., & Beckett, C. (1989). Multiple coding strategies in the retention of musical tones by possessors of absolute pitch *Memory & Cognition*, 17(5), 582-589.

Zatorre, R. J., & Halpern, A. R. (2005). Mental Concerts: Musical Imagery and Auditory Cortex. *Neuron*, 47(1), 9-12.