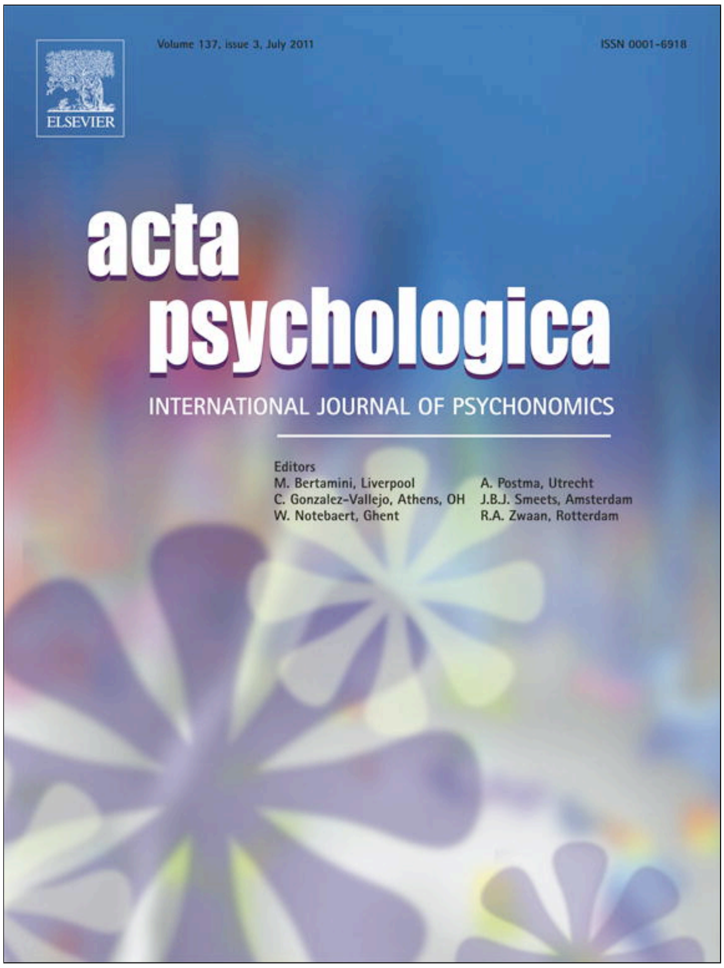
Volume 137, issue 3, July 2011
ISSN 0001-6918
ELSEVIER
acta psychologica
INTERNATIONAL JOURNAL OF PSYCHONOMICS
Editors
M. Bertamini, Liverpool
C. Gonzalez-Vallejo, Athens, OH
W. Notebaert, Ghent
A. Postma, Utrecht
J.B.J. Smeets, Amsterdam
R.A. Zwaan, Rotterdam

Acta Psychologica 137 (2011) 309–317

Contents lists available at ScienceDirect

Acta Psychologica

journal homepage: www.elsevier.com/ locate/actpsy

Mental scanning of sonifications reveals flexible encoding of nonspeech sounds and a universal per-item scanning cost

Michael A. Nees *, Bruce N. Walker **

Georgia Institute of Technology, School of Psychology, 654 Cherry Street, Atlanta, GA, 30332, USA

ARTICLE INFO

Article history:
Received 8 June 2010
Received in revised form 9 March 2011
Accepted 22 March 2011
Available online 19 April 2011

PsycINFO classification:
2300
2326
2340
2343

Keywords:
Mental imagery
Sonification
Encoding
Nonspeech sounds
Working memory
Auditory imagery

ABSTRACT

A mental scanning paradigm was used to examine the representation of nonspeech sounds in working memory. Participants encoded sonifications – nonspeech auditory representations of quantitative data – as either verbal lists, visuospatial images, or auditory images. The number of tones and overall frequency changes in the sonifications were also manipulated to allow for different hypothesized patterns of reaction times across encoding strategies. Mental scanning times revealed different patterns of reaction times across encoding strategies, despite the fact that all internal representations were constructed from the same nonspeech sound stimuli. Scanning times for the verbal encoding strategy increased linearly as the number of items in the verbal representation increased. Scanning times for the visuospatial encoding strategy were generally slower and increased as the metric distance (derived metaphorically from frequency change) in the mental image increased. Scanning times for the auditory imagery strategy were faster and closest to the veridical durations of the original stimuli. Interestingly, the number of items traversed in scanning a representation significantly affected scanning times across all encoding strategies. Results suggested that nonspeech sounds can be flexibly represented, and that a universal per-item scanning cost persisted across encoding strategies. Implications for cognitive theory, the mental scanning paradigm, and practical applications are discussed.

1. Introduction

Humans are constantly perceiving and processing all manner of sights, sounds, and other environmental stimuli. Cognitive psychologists, in particular, have attempted to understand how people mentally represent or think about these diverse varieties of stimulation. Influential theoretical accounts have emphasized dual-process models of working memory that distinguish verbal (i.e., symbolic speech) from visuospatial (i.e., analogical imagery) representational processes (Baddeley, 1992; Mayer & Moreno, 1998; Wickens, 2002). To date, relatively little attention has been paid to the internal coding of *nonspeech* sounds, such as music, environmental sounds, and sonifications, including nonspeech auditory information displays like those found increasingly in digital devices. These commonly encountered stimuli are not readily classifiable as verbal or visuospatial, yet the processing of nonspeech sounds is clearly an important component of human cognition (for an overview, see McAdams & Bigand, 1993) that has been given little consideration with respect to encoding and representation. Given that nonspeech sounds pervade many environments and their presence may affect other concurrent information processing tasks (Salame & Baddeley, 1989; Shelton, Elliot, Eaves, & Exner, 2009), research to understand the encoding of nonspeech sounds in working memory is also important for practical reasons.

Evidence has suggested that nonspeech sounds might assume a non-symbolic (i.e., nonverbal) format of cognitive representation that also is not visuospatial, but instead is best described as *auditory imagery*—a pseudo-isomorphic auditory analog to visual imagery (e.g., Crowder, 1989; Farah & Smith, 1983; Keller, Cowan, & Saults, 1995; Reisberg, 1992). The phenomenon of auditory imagery, however, has not been fully explored in the context of the well-established dichotomy of verbal and visuospatial processing in working memory. Though Baddeley and Logie (1992) identified the phonological loop – the verbal mechanism of working memory – as "the seat of auditory imagery" (pp. 180), that conclusion implies little or no functional distinction between the auditory representation of nonspeech sounds and the representation of verbal information in working memory. Empirical evidence (e.g., Deutsch, 1970) and some theories of speech perception (Liberman & Mattingly, 1985) challenge this perhaps overly simplistic interpretation.

Another theoretical difficulty is that descriptions of working memory (Baddeley, 1992; Mayer & Moreno, 1998; Wickens, 2002) have often implicitly linked (and sometimes explicitly wed) the internal coding of a stimulus to its external format (e.g., verbal, pictorial, etc.), although a number of studies have suggested that the format of the

* Corresponding author: Tel.: +1 317 258 0507; fax: +1 404 894 8905.
** Corresponding author:Tel.: +1 404 894 8265; fax: +1 404 894 8905.
E-mail addresses: mnees1984@yahoo.com (M.A. Nees), bruce.walker@psych.gatech.edu (B.N. Walker).

0001-6918/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.actpsy.2011.03.012

cognitive representation of a stimulus is malleable by strategy, individual differences, and/or task demands during encoding (e.g., Kosslyn, Ball, & Reiser, 1978; Mathews, Hunt, & MacLeod, 1980). In other words, the domain-specific working memory module engaged by a given stimulus may be manipulated. There is little doubt that many nonspeech sounds can be recoded with a verbal label such as a musical note name, and this labeling ability has been identified as the primary (but not the only) encoding mechanism in absolute pitch (Zatorre & Beckett, 1989). Other research, however, has suggested that nonspeech sounds may also be encoded as visuospatial images. For example, pitch has exhibited systematic crossmodal relationships with visual space such that sounds of higher pitch were congruent with higher spatial position or "up" (Ben-Artzi & Marks, 1995; also see Kubovy, 1981), and this and related phenomena have been dubbed "weak synesthesia" (Martino & Marks, 2001). Both qualitative (Mikumo, 1997) and anecdotal (Zatorre & Beckett, 1989) reports have suggested that musicians may form images from sounds that are visuospatial in nature (e.g., a visual image of notes on a musical staff to encode a melody), and a recent examination of descriptive reports of encoding strategies for auditory stimuli showed that participants engaged in auditory imagery, verbal labeling, and visuospatial imagery when they performed tasks with nonspeech sounds (Nees & Walker, 2008).

Introspective and anecdotal reports alone have not provided enough evidence to draw strong conclusions regarding the independence of verbal, visuospatial, and nonverbal auditory representations in working memory, as these data can be subjective and problematic (for a discussion, see Evans, 1990). Whereas qualitative introspections have provided some insights regarding the internal representation of stimuli, a more rigorous approach should corroborate and confirm these reports with more objective measures. The current experiment used response times derived from the mental scanning paradigm to try to differentiate three formats of internal representation in working memory: verbal representation, visuospatial imagery, and (nonverbal) auditory imagery.

1.1. Mental scanning

In the typical mental scanning trial, an internal representation is "viewed" or mentally examined in the absence of an external percept. Sternberg (1966, 1969/2004) first used the mental scanning procedure to make inferences from behavioral outcomes (reaction times) about the properties of memory for verbal materials. Of relevance to the current experiment, his work showed that the time to perform an exhaustive internal scan of a list of digits increased as a function of the number of items in the list (also see, for example, Klatzky, Juola, & Atkinson, 1971). Kosslyn et al. (1978) showed that, under conditions of visuospatial imagery (as contrasted with Sternberg's digit lists), mental scanning times varied as a function of the metric distance scanned in an image. The experiment 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 distance between locations on the map, and the correlation between reaction time and actual distance on the map was $r=.97$.

Recent work has shown that images constructed from verbal descriptions (rather than percepts of images) also have 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 (also see Noordzij, van der Lubbe, & Postma, 2005). 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, 2008). Neurogimaging research (Mellet et al., 2002) has confirmed that mental scanning of visuospatial representations constructed from verbal descriptions indeed recruits 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 distinct to visuospatial processing (also see Kirchoff & Buckner, 2006; Kosslyn & Thompson, 2003). No research to date, however, has empirically confirmed that metric spatial relations are present in visual images derived via auditory frequency's metaphorical relationship with distances in space (Ben-Artzi & Marks, 1995; also see Kubovy, 1981).

Halpern (1988), however, demonstrated a temporal mental scanning effect for songs. 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. The further examination of this and other formats of encoding for sounds has been mostly overlooked. Halpern suggested, "The ideal control group for this comparison is a logical impossibility—that is a group that is told *not* to use an imagery representation (which would be equivalent to saying 'don't think of an elephant'). Perhaps a control group could be instructed specifically to use a nonimagery strategy." (pp. 441). The ideal control, however, may not be a group that is instructed to use a generic nonimagery strategy, but rather may be groups that are instructed to use plausible alternative encoding strategies to auditory imagery. Halpern's series of studies, therefore, compared a specifically instructed auditory imagery condition to a spontaneous encoding strategy control condition. Most participants in the control condition exhibited patterns of reaction times that suggested that they spontaneously adopted the same auditory imagery strategy as the instructed condition. Halpern suggested the potential for a comparison condition that did not use auditory imagery, but she did not specify what a "non[auditory]imagery strategy" might entail. The current experiment specifically addressed this dilemma by instructing the use of three theoretically plausible encoding strategies for nonspeech sounds: auditory imagery, visuospatial imagery, and verbal encoding.

1.2. Motivations and overview of the current experiment

In the current study, we attempted to demonstrate that listeners are flexible in the encoding and representation of nonspeech sounds in working memory. We did this by differentiating the mental scanning times associated with nonspeech auditory, verbal, and visuospatial encoding strategies, respectively, for the same nonspeech sound stimuli. Halpern's (1988) work on mental scanning appears to be the only research to date to examine mental scanning of sounds, and no research has shown within the same experimental paradigm that nonspeech sounds can be represented via verbal processes, visuospatial imagery, and auditory imagery in working memory. Our experiment used a set of manipulations that were expected to differentiate mental scanning times based on the instructed format of encoding of the stimulus.

Participants listened to sonifications—specifically, nonspeech auditory representations where higher pitched tones represented higher temperatures on a fictional planet. Sonifications featured either two, three, or four data points (i.e., discrete tones) that represented temperature readings; the distance between data points was varied systematically such that some sonifications featured more pronounced frequency changes (i.e., greater changes in represented temperature values) 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. In the Verbal condition, participants encoded the data points as a verbal list of temperatures (e.g., 20 °F, 40 °F, 45 °F, etc.). In the Visuospatial Imagery condition, participants encoded the sounds as a pictorial image of the mercury in a thermometer, where a greater frequency change corresponded to a higher level of mercury in the thermometer image. In the Auditory Imagery condition, participants were instructed to encode the sonification as a sound exactly as they heard it (like a tape recorder), without any recoding. Following encoding, they were given a cue to begin to "scan" their respective mental representations. The experiment used a 3 (encoding strategy) × 3 (number of tones) × 3 (frequency change) within-subjects design, and scanning times from the onset of the cue were recorded as the primary dependent variable. We also measured subjective workload across encoding strategies using the NASA Task Load Index (NASA-TLX) (Hart & Staveland, 1988).

1.3. 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.

1.3.1. Hypothesis 1

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 (see Sternberg, 1966, 1969/2004) – in the stimulus. Consistent with this previous finding on scanning of memory for verbal stimuli, we had good reason to expect that adding more items to a verbal list (dictated by the manipulation of the number of tones in the sounds) would increase mental scanning times, yet the actual numbers in the list (dictated by the frequency change manipulation) would not affect the time to mentally scan the list.

1.3.2. Hypothesis 2

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. When participants made a pictorial internal representation of a thermometer from the sonifications with greater frequency changes – which represented metrically longer visual depictions of the mercury level – the distance traversed in mentally scanning the image should have been affected by the overall amount of change in frequency (see Denis & Zimmer, 1992; Kosslyn et al., 1978). We did not expect the number of tones manipulation to have an effect in this condition, as participants were instructed to scan through intermediate values in the thermometer image without pausing their scanning process.

1.3.3. Hypothesis 3

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 that auditory representations preserve the temporal aspects of the perceived stimulus (Halpern, 1988; Levitin & Cook, 1996), thus the auditory imagery encoding condition was hypothesized to show the fasting mental scanning times (closest to the veridical duration of the auditory stimuli). The hypothesized flat scanning times across stimulus manipulations would differentiate this encoding strategy from verbal and visuospatial internal representations.

1.3.4. Hypothesis 4

Though little research has examined the mental workload associated with transforming percepts into different representational formats, the formation of visual images, for example, has been shown to require time (see, e.g., Tversky, 1975), which may be indicative of increased mental workload. We hypothesized that overall scores on the NASA-TLX (Hart & Staveland, 1988) would show lower perceived workload for the auditory imagery condition as compared to the verbal and visuospatial representational conditions, as the auditory imagery condition did not require participants to recode the initial percept into a different format in working memory.

2. Method

2.1. 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.

2.2. 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.

2.3. Stimuli

The stimuli in the current experiment were designed to allow for hypotheses that would differentiate three distinct formats of encoding for nonspeech sounds. Sonification stimuli depicted the temperature at a weather station on a fictional planet, over the course of one day. Higher temperatures were represented with higher 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 jumps in one octave (from musical note C4 to C5) on the equal-tempered musical scale, whereas medium and large stimuli changed two (from note C4 to C6) and three (from note 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 always corresponded to a starting temperature of 20 °F, and that the 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 was 120 °C (represented by note C7), but participants were told that the maximum temperature was not necessarily achieved every day.

Sonifications also featured two, three, or four discrete tones. For two-tone stimuli, the tones were the lower and upper 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. The precise values of the interpolated tones in three- and four-tone sonifications were not of interest—the significance of the manipulation was the addition of more items (i.e., data points, represented with tones) to the sonification.

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.

2.4. 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 and instructions for each of the three possible encoding strategies (described below) and the scanning task. Participants then experienced the Verbal, Visuospatial Imagery, and Auditory Imagery 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.

2.4.1. Verbal 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 to an audiovisual animation that depicted a verbal list becoming populated as a sound stimulus was heard. The instructions encouraged participants to forget about sounds and images, and participants were told to focus only on the list of values that they believed the sounds represented.

2.4.2. Visuospatial Imagery condition

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 height of the vertical line in the animation (i.e., the mercury in the thermometer) increased as the frequency of the tones increased, as higher frequency represented higher attained temperature in the sonifications. 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). The thermometer image was chosen as a visual metaphor for auditory frequency, because its vertical representation of quantity was amenable to translations from auditory frequency to the single, constrained visual dimension of height.

2.4.3. Auditory Imagery condition

Participants received instructions to encode the sounds as a pseudo-isomorphic auditory representation by remembering and rehearsing the sonification stimulus exactly 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.

2.5. 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. (1978, Experiment 3) found that even if subjects were instructed to make a visual image, they sometimes used an alternate strategy to accomplish the scanning task. In other words, subjects must be explicitly told to consult their internal representations (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. Importantly, the instructions were carefully worded so as to instruct specific encoding and scanning procedures without suggesting that any particular encoding strategy should exhibit any particular shortening or lengthening of mental scanning times as a function of the stimulus manipulations (for a thorough discussion of this concern with the mental scanning paradigm, see Kosslyn, 1980, Chapter 12). 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. Participants indicated that they had successfully encoded the stimulus by pressing the spacebar and then saw a brief (3000 ms) blank gray 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 (i.e., used their “inner voices” to produce from memory) the encoded list of values upon appearance of the “+” cue. Participants were told to progress from the first value in their mental list to the last value in order at a fast, unchanging rate, and to press the space bar as soon as their mental scan of the list was complete. For the visuospatial imagery condition, participants were instructed, upon seeing the “+” cue, to scan 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 were told to scan through any intermediate values in their representation without pausing and to press 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 were instructed to replay the sonification in their mind (like pressing “play” on the tape recorder in their minds) upon seeing the “+” cue, and they pressed the space bar as soon as the mental recording had finished playing. 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”. Participants selected at least one strategy, and they could choose more than one strategy. In an effort to reduce the potential for demand characteristics in the identification of strategy compliance, the strategy compliance questionnaire included these instructions: “Please be honest even if you used the wrong strategy—it is very important that we know which strategy that you used.” The experimenter also explained the importance of answering the strategy compliance questions honestly, regardless of their compliance or noncompliance with the instructed encoding strategy. Marquer and 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) used retrospective reports on strategy compliance across an experiment 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 percentages of participants who were unable to implement visuospatial imagery encoding strategies, in particular. Dunlosky and Hertzog (1998) reviewed potential flaws in retrospective estimates of strategy implementation (e.g., forgetting) across an experiment or block of trials and suggested that participants should be queried about strategy use on a trial-by-trial basis. A later experiment (Dunlosky & Hertzog, 2001) found that trial-by-trial reports were preferable to retrospective strategy use reports, 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. Since the current experiment *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 encoding strategy 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-TLX (Hart & Staveland, 1988) as a measure of the subjective workload experienced in each encoding condition.

3. Results

When a participant did not indicate use of the appropriate encoding strategy on the post-trial report screen, the participant's data for that trial were 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 all follow-up analyses for which usable (i.e., strategy compliant and statistically tenable) data were available.[1]

3.1. Scanning time analyses

A 3 (encoding strategy: auditory imagery, verbal, or visuospatial imagery) × 3 (number of tones: 2, 3, or 4) × 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 Fig. 1) showed significant main effects of strategy, F (1.50,57.02) = 20.01, p<.001, partial η^2 = .35, number of tones, F (1.47,55.83) = 64.20, p<.001, partial η^2 = .63, and frequency change, F (1.35,51.24) = 6.69, p = .007, partial η^2 = .15, as well as significant interactions of strategy with number of tones, F (2.31,87.59) = 4.50, p = .01, partial η^2 = .11, and strategy with frequency change, F (1.67,63.31) = 9.49, p = .001, partial η^2 = .20. The interaction of the number of tones with frequency changes was not significant, F (3.24,123.11) = 0.61, p = .62, nor was 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*) and 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). 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.

For the Auditory Imagery condition, all 44 participants provided full data after trials with statistical outliers or incorrect strategies were removed. Results (see Fig. 1, 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 the 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.

For the Verbal condition, 41 participants provided full data after trials with statistical outliers or incorrect strategies were removed. Results (see Fig. 1, panel b) showed a significant main effect of the number of tones, F (1.38,55.15) = 54.43, p<.001, partial η^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 the 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 η^2 = .62.

For the Visuospatial Imagery condition, 42 participants provided full data after trials with statistical outliers or incorrect strategies were removed. Results (see Fig. 1, panel c) showed significant main effects of the number of tones, F (1.46,59.95) = 9.93, p = .001, partial η^2 = .20, and frequency change, F (1.29,52.99) = 10.34, p = .001, partial η^2 = .20. The interaction of the 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 η^2 = .23. For the main effect of frequency change, a significant linear increasing trend described the pattern of scanning times as change in frequency increased, F (1,41) = 11.60, p = .001, partial η^2 = .22.

3.2. NASA-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 η^2 = .28. Paired comparisons showed that the auditory imagery encoding strategy (M = 9.14, SE = 0.50) resulted in a significantly lower perceived 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).

4. Discussion

We manipulated the instructions for the encoding strategy for sonification stimuli. We varied the stimulus properties (frequency change and number of tones) in a configuration that allowed for hypothesized patterns of mental scanning times – based on the existing literature – that would differentiate each encoding strategy if participants could in fact use a given strategy to rehearse and mentally scan their respective representations in working memory. The primary dependent variable was the mental scanning time, and a dependent variable of secondary interest was the NASA-TLX measure of subjective workload for each encoding strategy. Results generally

[1] This approach was adopted in an effort to analyze all usable data that were collected, given that a participant could have had difficulty with using one particular encoding strategy but not the others. Based on the concerns of a reviewer, we also conducted all follow-up analyses with only the 39 participants that gave full data across all conditions. In these analyses, the patterns and magnitudes of significant and nonsignificant effects were unaffected, so we report the follow-up analyses with all usable data here.

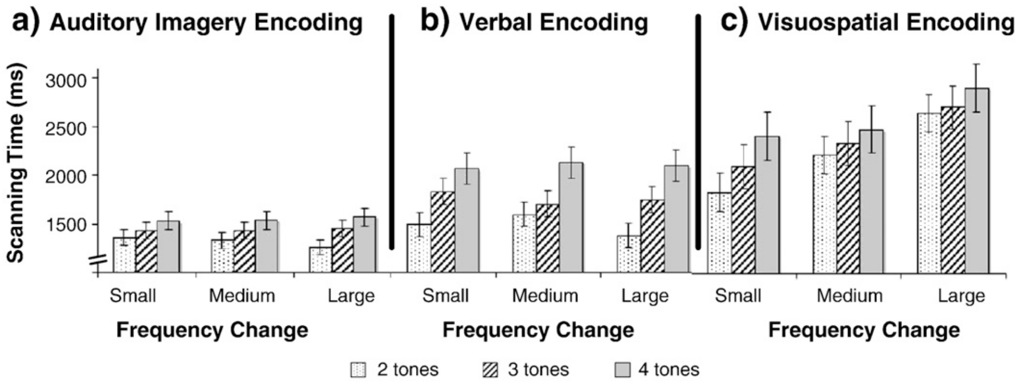

Fig. 1. Mean mental scanning times as a function of encoding strategy, frequency change, and the number of tones. Error bars represent standard error of the mean.

confirmed the hypotheses that mental scanning times would differentiate distinct encoding formats. Mental scanning times under visuospatial encoding were generally longer and were sensitive to the frequency change manipulation, which was a metaphorical indicator of metric space in participants' visual images. To our knowledge, this is the first study to offer quantitative empirical evidence that nonspeech sounds can be encoded as visuospatial images, and that these images seem to have the same properties as visuospatial images formed from visual percepts. Mental scanning times under verbal encoding were sensitive to the number of tone 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 lengths of the external sound stimuli.

Across all three encoding strategies, participants' mental scanning times were sensitive to the number of tones manipulation. This somewhat unexpected but interesting finding suggests a universal per-item scanning cost that persisted across encoding strategies. One plausible explanation for this finding, which is discussed in more detail below, is that rehearsal/scanning mechanisms across all three types of representation require some process that serially generates or scans the representation. Another possible explanation is that the initial external sound stimulus had a lingering effect in addition to the observed effects of encoding strategy.

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; thus perceived workload was lower for this strategy as compared to the verbal or visuospatial encoding strategies.

4.1. Nonspeech sounds and representations in working memory

The results of this experiment offer some tentative evidence for an independent representational format for nonspeech auditory imagery that exists alongside the dual-process descriptions of verbal and visuospatial working memory that are common in current cognitive theory (e.g., Baddeley, 1992; Mayer & Moreno, 1998; Wickens, 2002). Further research is needed to determine the extent to which nonspeech auditory imagery representation actually occurs independent of other forms of representation. It has been suggested that the mechanisms of verbal rehearsal are also responsible for the maintenance of nonspeech sounds in working memory (e.g., Baddeley & Logie, 1992), yet the data on this subject are equivocal. The current results show that the mental scanning paradigm can be used to differentiate patterns of reaction times for verbal rehearsal and scanning processes as compared to nonspeech auditory imagery rehearsal and scanning processes. More research, and likely new research paradigms altogether, are needed to disentangle this complex question regarding whether speech and nonspeech auditory stimuli receive mutually exclusive cognitive representations.

There have effectively been four approaches for dealing with encoding strategies in memory experiments. First, many researchers are not interested in encoding strategies, so individual differences and heterogeneity in such strategies are assumed to be handled by random assignment to conditions and are therefore ignored altogether. A second approach is to choose stimuli and tasks that are assumed to induce homogeneous use of encoding strategies across participants. These approaches may, for example, use unusual symbols that are not amenable to verbal labeling to induce an imagery strategy in memory, or alternatively, take steps to induce verbal encoding by making imagery strategies difficult and inefficient, such as by changing from a capitalized font for a study set to a lower case font for a test probe in a memory task (see, e.g., the Symbol Probe Task and the Letter Probe Task, respectively, in Henson, Burgess, & Frith, 2000). The third approach is to allow participants to use spontaneous encoding strategies during data collection, then to use post hoc methods such as retrospective questionnaires (e.g., Kirchoff & Buckner, 2006) or model-fitting of individual data patterns (e.g., Mathews et al., 1980) to determine which participants used a given encoding strategy. Finally, the fourth approach is to instruct participants to use a particular encoding strategy for a task and thereby explicitly manipulate the encoding strategy with instructions and a careful manipulation check to screen for strategy compliance (e.g., Kosslyn et al., 1978). We chose the fourth approach in our experiment to establish which possible strategies are plausible candidates for encoding of nonspeech sounds, but we believe that experimental paradigms using the second and third approaches will further strengthen and confirm (or disconfirm) our results and interpretations here.

For example, the current study instructed encoding strategies across all conditions and did not allow for a spontaneous encoding strategy to be selected by participants. Research on the sentence–picture verification task has suggested that, while encoding strategies vary according to individual differences or task demands, a verbal encoding may be the default strategy adopted by a majority of participants (Hunt, 1978). During debriefing, our own participants have reported using verbal strategies to encode sonifications about twice as often as they report visuospatial or auditory imagery strategies (Nees & Walker, 2008). A given participant's proclivity toward a particular encoding strategy for nonspeech sounds warrants further research. The small number of participants in the current study whose data were removed as a result of strategy noncompliance may have been prescribed a strategy that contradicted their spontaneous encoding preference. This type of result has been reported in studies where visuospatial imagery is prescribed for a task; a small number of participants have been unable to successfully form images (Kosslyn et al., 1978; Mathews et al., 1980). A related question involves the extent to which visuospatial imagery of nonspeech sounds occurs with sounds less amenable to visual imagery than the stimuli used here. Future research should examine, for example, the conditions under which a person will spontaneously use a given encoding strategy. Halpern's (1988) work suggested that most people will spontaneously adopt an auditory imagery encoding strategy for sounds, and our subjective workload results suggest that this approach may be the least cognitively taxing approach for remembering nonspeech sounds. Other research has shown that most people will adapt their encoding strategy to perform a task most efficiently (e.g., by using imagery when it affords easier task performance, see, e.g., Kroll & Corrigan, 1981), and new research paradigms might be able to demonstrate a similar phenomenon with encoding of nonspeech sounds.

4.2. Per-item scanning costs in mental scanning paradigms

Denis and Kosslyn (1999) noted that, whereas mental scanning was originally formulated as a paradigm for studying internal representation, the literature on mental scanning has branched to include an interest in the process in and of itself. As such, an unexpected but interesting outcome of the current experiment 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. (1978, Experiment 1) and suggests that during mental scanning participants are sensitive (to varying degrees of effect) to the number of items present in the representation, regardless of the specific format of the representation in working memory. The per-item scanning cost was noted in Kosslyn's early work, yet this effect, to our knowledge, has never been fully explained or replicated in the literature in the decades since it was first noticed.

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) visuospatial and auditory rehearsal mechanisms in working memory are involved in mental scanning and operate to reinstate and scan 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 original, external auditory percept was present in addition to the effects of the encoding strategy manipulation.

Regarding the first possibility, the active rehearsal and maintenance of verbal material in the articulatory/phonological loop component of verbal working memory 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). Performance of some visuospatial tasks like the Corsi Block-Tapping Test (see Kessels, van Zandvoort, Postma, Kappelle, & de Haan, 2000) requires serial visuospatial production from memory. A mechanism to account for serial processes with visuospatial representations has not been fully identified in the same fairly unanimous way that the phonological loop of verbal working memory has been identified as the mechanism of serial verbal processing. Candidate processes for the maintenance of visuospatial representations include “online…discrete shifts in spatial attention” (Awh, Jonides, & Reuter-Lorenz, 1998 pp. 788). Other research has similarly suggested that an *inner scribe* actively maintains sequential information about movements and spatial locations (Logie & Pearson, 1997). Pearson (1999) showed that an interference task that disrupted the inner scribe affected mental scanning times, but not *blink transformations*—mentally switching from one distinct image to another rather than scanning within an image. These results suggest that the inner scribe may be closely tied to the process of mentally scanning a visual image.

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 at least partially inadequate 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). Reisberg, Smith, Baxter, and Sonenshine (1989) showed that only covertly articulated auditory images can be reinterpreted ambiguously, whereas unarticulated auditory images are unambiguous. This finding suggested that different properties of representations – or perhaps different representations altogether – may emerge when articulatory processing is engaged. Though articulatory processes can clearly operate on verbal representations and also at least can imitate some nonspeech auditory stimuli, articulation seems to be neither a necessary condition of, nor a sufficient explanation for, auditory imagery for nonspeech sounds. The distinction may be that some working memory tasks require only the articulatory component or the passive store of the phonological loop, yet other tasks require both components (Smith, Reisberg, & Wilson, 1992). A mechanism to account for the maintenance and rehearsal of inarticulable nonspeech sounds in working memory remains to be explained with future research.

Another interpretation, as mentioned above, involves the possibility that some trace of the initial auditory stimulus persisted in the domain-specific internal representation. Biological evidence that complements this explanation has been found in a PET experiment that showed that domain-specific internal representations constructed from different modalities of input showed unique patterns of neural activation (i.e., domain-specificity of the internal representation irrespective of the modality of the stimulus), yet the representations 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 a function of encoding strategy, yet the universal per-item cost could plausibly result from a lingering, stimulus-specific effect of the number of tones present in the auditory stimulus.

4.3. A note on demand characteristics and the mental scanning paradigm

A reviewer of this study was especially concerned with the possibility that unique task demands across the different encoding strategies caused the observed patterns of mental scanning times rather than the encoding strategy itself or the format of internal representation being scanned. It is well-established that simply instructing participants “to scan” their memories leaves ambiguity as to how participants approach the task (see Denis & Kosslyn, 1999, pp. 603) and results in heterogeneous approaches to mental scanning, so careful instructions that were specific to mental scanning for each particular encoding strategy were required. We do not view this criticism as particularly problematic, however, because the instructions *were* the manipulation. We took care with our instructions to minimize the potential introduction of demand characteristics that would imply expectations about hypothesized scanning times. We believe that the differentiation of mental scanning times for different encoding strategies for the same stimuli within the same experiment is a strength rather than a weakness of our approach, because the bulk of mental scanning to date has examined encoding strategies piecemeal with either no control groups or spontaneous (uninstructed) strategy control groups, which can be problematic due to individual differences in strategy use (see, e.g., Halpern, 1988; MacLeod, Hunt, & Mathews, 1978).

Historically, the mental scanning paradigm has been subject to similar methodological criticisms surrounding demand characteristics (e.g., Pylyshyn, 1981). This line of criticism posits that participants produce patterns of reaction times in mental scanning tasks not as a function of the structure or format of their own internal representations of information, but instead as a function of their expectations regarding how they “should” respond. These expectations, it has been suggested, are informed by their assumptions regarding the hypotheses of the experimenter (inferred from the experimental instructions) or their own suppositions about the correct pattern of reaction times that should result from performing the task (inferred from their own perceptual experiences or attention to particular features of the stimuli). Although this criticism of mental scanning studies has already been addressed at length (Denis & Carfantan, 1985; Denis & Kosslyn, 1999; Kosslyn, 1980, Chapter 12; Reed, Hock, & Lockhead, 1983), it seemingly inevitably arouses concern for the critical reader of studies that use mental scanning. If the demand characteristics hypothesis is drawn upon to explain the pattern of results shown here, then we must conclude that participants successfully anticipated some hypotheses regarding the predicted pattern of reaction times that were implicit (at best) in the instructions, while participants failed to respond in accordance with other very explicit experimental instructions—for example, the instructions to scan through intermediary temperatures, where present, without stopping during visuospatial image scanning. Further, in our study, frequency demonstrated a metaphorical relationship with metric distance, but only when the participant used a visuospatial encoding strategy. We know of no perceptual or imagery theory that would predict this symbolic distance effect for frequency representations in working memory per se, unless changes in frequency are converted to a visuospatial image that associates greater distances with greater frequency changes. We see no way in which the appeal to task demands offers a more complete or plausible interpretation of the observed results for mental scanning times for the instructions that were used in this experiment. We believe that the demand characteristics hypothesis is considerably less parsimonious – not to mention less consistent with the existing literature and theory – than our interpretation of the results: namely, that the mental scanning times differentiated different encoding strategies that informed different formats of internal representation with a universal, per-item scanning cost across representational formats.

4.4. Practical relevance

The overall patterns of mental scanning times in this experiment differentiated verbal, visuospatial, and auditory imagery encoding, and this suggest that nonspeech auditory stimuli can indeed be encoded flexibly in a variety of representational formats in working memory.

Past research, and Multiple Resources Theory (see Wickens, 2002) in particular, has suggested that distinct representational processes (i.e., verbal and visuospatial processing codes) in working memory can proceed relatively independently with little interference with concurrent tasks. The question of whether a distinct working memory mechanism or module for nonspeech sound will function independently of verbal or visuospatial process remains unresolved. This question is of particular interest for practical applications, as the implementation of nonspeech sounds in information-rich environments has become more feasible and more common. The appropriateness of nonspeech sounds as information displays will depend upon the extent to which processing and remembering the information contained in nonspeech sounds can be done in concert with other concurrent (often visual or verbal) tasks.

4.5. Conclusions

Using a mental scanning paradigm, we offered evidence that nonspeech sounds can be flexibly coded in working memory. We also found evidence that suggested a per-item scanning cost that persisted in mental scanning across encoding strategies as the number of items to be scanned increased. This experiment offered initial insight as to the nature of the representation of nonspeech sounds in working memory, which is a pervasive mental activity that has been largely overlooked in the literature to date. Our results suggested that nonspeech sounds can be encoded as verbal representations, visual images, or auditory images in working memory, although the extent to which these different formats of representation are truly modular and independent should be examined further. Representations of nonspeech sounds in working memory may have a considerable impact on learning and information processing in scenarios where diverse stimuli (e.g., verbal, visual, and nonspeech auditory stimuli) are present.

Acknowledgments

This work partially fulfilled dissertation requirements for the first author's doctoral degree. We thank Terri Bonebright, Richard Catrambone, Gregory Corso, and Wendy Rogers for insightful comments and feedback on this experiment. We also acknowledge the valuable contributions of Courtney Lynch and Amy Gress during data collection for this project. Portions of this research were supported by NSF Career Award #IIS-0644076 to Bruce N. Walker.

References

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

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

Crowder, R. G. (1989). Imagery for musical timbre. *Journal of Experimental Psychology. Human Perception and Performance*, *15*(3), 472–478.

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.

Denis, M. (2008). 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*, *127*(1), 197–210.

Denis, M., & Carfantan, M. (1985). People's knowledge about images. *Cognition*, *20*, 49–60.

Denis, M., & Kosslyn, S. M. (1999). Does the window really need to be washed? More on the mental scanning paradigm. *Current Psychology of Cognition*, *18*(4), 593–616.

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? *Psychology 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.

Evans, J. S. B. T. (1990). Verbal reports of cognitive strategies: A note on Marquer and Pereira. *Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, *42*(1), 169–170.

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

Halpern, A. R. (1988). Mental scanning in auditory imagery for songs. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *14*(3), 434–443.

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.

Henson, R. N. A., Burgess, N., & Frith, C. D. (2000). Recoding, storage, rehearsal, and grouping in verbal short-term memory: An fMRI study. *Neuropsychologia*, *38*, 426–440.

Hunt, E. (1978). The sentence-verification paradigm: A case study of two conflicting aproaches to individual differences. *Intelligence*, *2*(2), 129–144.

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), 635–645.

Kessels, R. P. C., van Zandvoort, M. J. E., Postma, A., Kappelle, L. J., & de Haan, E. H. F. (2000). The Corsi Block-Tapping Task: Standardization and normative data. *Applied Neurophysiology*, *7*(4), 252–258.

Kirchoff, B. A., & Buckner, R. L. (2006). Functional-anatomic correlates of individual differences in memory. *Neuron*, *51*, 263–274.

Klatzky, R. L., Juola, J. F., & Atkinson, R. C. (1971). Test stimulus representation and experimental context effects in memory scanning. *Journal of Experimental Psychology*, *87*(2), 281–288.

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

Kosslyn, S. M. (1980). *Image and mind.* Cambridge, MA: Harvard University Press.

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

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

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

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

Logie, R. H., & Pearson, D. G. (1997). The inner eye and the inner scribe of visuo-spatial working memory: Evidence from developmental fractionation. *European Journal of Cognitive Psychology*, *9*(3), 241–257.

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.

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., & 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.

McAdams, S., & Bigand, E. (1993). *Thinking in sound: The cognitive psychology of human audition.* Oxford: Clarendon Press.

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.

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

Nees, M. A., & Walker, B. N. (2008). 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)*.

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.

Pearson, D. G. (1999). Mental scanning and spatial processes: A role for the inner scribe? *Current Psychology of Cognition*, *18*(4), 564–573.

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

Reed, S. K., Hock, H. S., & Lockhead, G. R. (1983). Tacit knowledge and the effect of pattern configuration on mental scanning. *Memory & Cognition*, *11*(2), 137–143.

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.

Reisberg, D. (Ed.). (1992). *Auditory Imagery*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Reisberg, D., Smith, J. D., Baxter, D. A., & Sonenshine, M. (1989). "Enacted" auditory images are ambiguous; "Pure" auditory images are not. *The Quarterly Journal of Experimental Psychology*, *41A*(3), 619–641.

Salame, P., & Baddeley, A. D. (1989). Effects of background music on phonological short-term memory. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, *41*(1), 107–122.

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

Shelton, J. T., Elliot, E. M., Eaves, S. D., & Exner, A. L. (2009). The distracting effects of a ringing cell phone: An investigation of the laboratory and classroom setting. *Journal of Environmental Psychology*, *29*, 513–521.

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.

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

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.

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

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.