ENCODING AND REPRESENTATION OF INFORMATION IN AUDITORY GRAPHS: DESCRIPTIVE REPORTS OF LISTENER STRATEGIES FOR UNDERSTANDING DATA

Michael A. Nees & Bruce N. Walker

Sonification Lab, School of Psychology
Georgia Institute of Technology
Atlanta, GA, USA 30332
mnees@gatech.edu; bruce.walker@psych.gatech.edu

ABSTRACT

While a growing wealth of data have offered insights into the best practices for auditory display design and application, little is known about how listeners internally represent and use the information presented in auditory displays. At the conclusion of three separate studies, participants responded to an open-ended question about the strategies they used to perform auditory graphing tasks. We report a descriptive analysis of these qualitative responses. Participants' comments were coded by two raters along a number of dimensions that were chosen to represent a comprehensive set of encoding and task strategy possibilities. These descriptive analyses suggest that auditory graph listeners use a variety of strategies to cognitively represent the data in the display. Furthermore, these qualitative data offer a number of insights and questions for future research on information representation for auditory displays.

1. INTRODUCTION

Over the past two decades, an abundance of research has examined the role of sound as a means of information display in human-machine systems. The accumulated data have offered best-practice design suggestions for auditory displays [e.g., 1, 2], as well as both theoretical accounts [see 3, 4, 5] and practical applications [e.g., 6, 7] of the use sound in a system. To date, however, very few researchers have considered how a listener encodes and represents the information in auditory displays at an internal, cognitive level.

Seminal theoretical accounts of information processing from psychology and human factors have generally been concerned with the encoding and representation of visual (e.g., spatial) information or linguistic (verbal) information [e.g., 8, 9]. Explanations of information processing for sound have generally treated the auditory modality as an agent for processing speech, which accordingly assumes a verbal (i.e., articulatory-phonological) internal representation. Research from related fields (e.g., music perception), however, has suggested that sonifications—nonspeech auditory information displays—may assume a variety of internal formats of representation.

Mikumo [10, 11] and others [e.g., 12] have identified at least four possible formats that listeners might use for the internal representation of the information conveyed by auditory frequency. With *verbal representations,* the information is encoded with verbal labels in an articulatory-phonological manner (such as when musical notes are labeled with their names, e.g., C#). *Visuospatial representations* encode information in a picture-like image [probably with many of the same properties of mental images, see, e.g., 13] that captures contour changes in frequency as a picture in the mind (e.g., like a visual graph). With *motor representations*, the information is encoded as a motor program that preserves rhythmic patterns (e.g., tapping) or perhaps even fingering positions for musical instruments (in well-trained musicians) that correspond to the frequencies of tones. Finally, *sensory-musical representations* encode the information as an isomorphic representation [or, more plausibly, as a 2nd-order isomorphic representation, see, e.g., 14] whereby the listener attempts to preserve the actual sensory experiences associated with the sounds that were heard (e.g., by whistling, humming, or any other attempt to access sensory memory).

The current study examined the roles of these encoding strategies as well as several additional strategies for accomplishing tasks with auditory graphs, a class of auditory displays that use sound to convey quantitative data. Auditory graphs typically represent changes along the visual Y-axis with frequency, while changes along the X-axis are mapped to time [1, 3, 15]. For three separate auditory graph studies [whose primary findings are reported elsewhere, see 16, 17], open-ended responses were recorded during a debriefing protocol. Participants were simply asked to briefly describe the strategy they used to accomplish the study tasks. We present a descriptive analysis of these qualitative responses, which were coded by two independent raters to examine both encoding strategies and other strategies for data analysis with auditory graphs.

2. METHODS

2.1. Participants

The qualitative data of the current study were gathered as part of a debriefing protocol during three separate studies [16 Experiments 1 and 2, 17]. Across all three studies, participants (*N* = 131; 74 males and 57 females) were recruited from psychology classes at the Georgia Institute of Technology and were compensated with course extra credit for their participation.

2.2. Procedure and Data

The procedures for the three studies are reported elsewhere and will be summarized briefly here. The first study was concerned with the effects of data complexity on performance of a point estimation task with auditory graphs. Participants listened to auditory graphs of a stock price over the course of a 10-hour trading day, and they were asked to estimate the price of the stock at a randomly selected hour of the trading day (i.e., "What was the price of the stock at 2 pm?"). Data complexity was manipulated by varying the number of trend reversals in the data (either 0, 1, 2, or 3 trend reversals during the trading day) as well as the data density (either 1, 2, 4, or 8 data points were presented per second). The second study examined the effects of the same manipulations on a local trend identification task, in which participants were asked to identify whether the data were increasing, decreasing, or staying the same over a given 1 hour period in the trading day (i.e., "Was the price of the stock increasing, decreasing, or staying the same between 1 pm and 2 pm?"). The third study for which the qualitative data were collected examined the relative efficacy of different types of contextual enhancements for tracking time in auditory graphs. Participants in this study performed the point estimation task with either 1) no concurrent contextual cues; 2) a snare drum that sounded every hour of the trading day on the hour; 3) an intensity burst in the data that sounded every hour on the hour; or 4) a speech cue that provided the time of day every hour on the hour.

Stimuli for each of these studies were sonified using the Sonification Sandbox [18]. Data changes in the auditory graphs were mapped to changes in auditory frequency with the MIDI piano instrument timbre. Scaling anchors were used for maximum and minimum values in a data set such that the minimum data value ($6) was mapped to MIDI note G2, whose frequency was 98 Hz, and the maximum data value ($106) corresponded to MIDI note B6 at 1979.5 Hz. A positive polarity mapping was employed such that increases in data were represented by increases in auditory frequency, and data values that fell between notes on the chromatic scale were adjusted in frequency (i.e., pitch bent) to correspond to exact MIDI note values.

The data of interest in the current study, which were not reported or examined in previous reports, were qualitative responses to an open-ended query that was presented to participants at the end of each of these three studies: "Please briefly describe the strategy you used to accomplish the study tasks." Participants' responses were recorded at the end of each of these three studies, and they were allowed to write as much or as little as they desired.

The data presented here are the scores of two independent raters' evaluations of participant responses to this question. The raters used the coding scheme below to examine the responses of each participant, which were compiled across the three studies.

2.3. Coding Scheme

Qualitative data were scored according to the presence or absence of a statement indicating that a particular strategy was used. If the strategy was mentioned, a one (i.e., present) was scored for that particular dimension; the participant was given a zero for dimensions that were not mentioned in their responses. Strategies were operationalized and data were coded along the following dimensions:

1. A **verbal encoding strategy** was present if a participant**:** a) used naming or labeling (e.g., any mention of musical note names); b) compared an assigned verbal label to another verbal label (such as using the name of the opening tone as a verbal reference); c) mentioned counting in specifically labeled increments (e.g., in increments of 10 dollars); or d) gave any indication that she or he labeled specific dollar values or times of day with a verbal tag and calculated from the labeled anchor.
2. A **visuospatial encoding strategy** was present if a participant: a) mentioned that she or he mentally drew a picture, built a picture, or created a graph in the mind; or b) gave a response that explicitly mentioned characteristics of a visual graph (e.g., slope, line, top, or bottom)
3. A **motor encoding strategy** was present if a participant: a) mentioned the use of (non-articulatory) movement such as tapping with the hands or feet, etc.; or b) used a strategy that involved counting on the fingers, moving physically with the mouse or fingers, or "drawing" on the desk with finger to remember the shape of the data.
4. A **sensory-musical encoding strategy** was present if the participant: a) mentioned humming, whistling, or vocalizing (either overtly or covertly) any part of the stimulus (e.g., the melody or the pitch of a tone); b) indicated a strategy that involved maintaining some isomorphic representation of the sounds; c) mentioned that he or she quickly replayed the auditory graph to hear the initial sound for comparison with sensory (e.g., isomorphic, as in echoic memory) stores; or d) made any mention of "hearing" an isomorphic representation of the sound in their mind (e.g., "recorded the pitch in my head").

In addition to documenting comments about encoding strategies, we also coded for the use of the following four additional categories of information: uses of auditory context, counting, arithmetic, and different strategies based on stimulus properties. A previous task analysis [19] suggested that contextual judgments about pitch, temporal judgments (which may be aided by counting), and arithmetic calculations may be required to accomplish some auditory graphing tasks, and the use of different strategies for encoding frequency may be possible subject to task and stimulus dependencies, etc. [see, e.g., 12].

5**. Context**: Auditory context generally refers to the intentional addition of tick marks, reference tones, a priori knowledge about the opening stock price, etc. [see 19, 20]. This dimension of the coding scheme was concerned with the use of context as a strategy for accomplishing tasks with auditory graphs. Strategic used of context was coded as present when: the participant used reference tones, opening stock price, auditory tick marks, etc. For the context study [17], context refers to the speech cues, snare drum (ticks), intensity pulses or bursts, highest, lowest, or starting values, etc. Context was also defined as the use of surrounding or nearby tones to determine the sonified stock price.
6. **Counting**: The participant mentioned the use of counting, either explicitly or implicitly (e.g., using number of beats per sound implies counting implies that the number of beats was counted).
7. **Arithmetic**: The participant mentioned the use of mathematical operations. Note that references to "estimation" were not counted as mentions of arithmetic operation, as everyone had to estimate to accomplish the task.
8. **Differential strategy** use based on stimulus was also coded. This included when participants mentioned having used different strategies for different stimuli.

It is important to note that the coding scheme categories were *not* mutually exclusive and any strategy could overlap with any number of other strategies. Research has suggested that multiple encoding strategies may be possible for a given auditory stimulus [10, 12], and it may be possible that these distinct internal representations may be simultaneously activated or selectively inhibited based upon task demands and other contingencies [for a description of possible theoretical mechanisms, see 21]

3. RESULTS

Participant responses to the strategy question had a mean length of 32 (*SD* = 22.7) words, where numbers and symbols (e.g., "$") were counted as one word each. For all ratings the percentage of agreement was calculated as a general index of the stability or consistency of the rating scheme across the two raters. Percentage agreement [see 22] was defined with respect to the total number of possible ratings ($N = 131$) and the total number of discrepant ratings according to the equation:

$$\text{Percentage agreement} = \frac{131 - \text{total number of discrepancies}}{131} X\ 100\% \quad (1)$$

Percentages of agreement for each dimension of the coding scheme are reported in Table 1, as well as the average percentage (between the two raters) of participants who reported using a particular strategy.

Strategy dimension	Percentage agreement between raters	Average percentage of strategy use
Verbal encoding	85	20.6
Visuospatial encoding	94	8.4
Motor encoding	95	10.7
Sensory-musical encoding	89	9.2
Context	75	65.7
Counting	83	55.8
Arithmetic	97	3.0
Differential	92	8.0

Table 1: Percentage agreement between the 2 raters along each dimension of the coding scheme and average percentage of participants who reported using a particular strategy.

Data on the individual strategies are discussed in more detail below.

3.1. Verbal encoding strategy

Between the two raters, the average percentage of participants who reported using a verbal encoding strategy was 20.6%. Responses that were coded as indicating the use of verbal labeling as an encoding strategy included:

> *"I guessed all prices in increments of 10 from the base [opening] value of 6."*

> *"I tried to stay in increments of 25."*

> *"Halftones and slight variations were given a slight difference in estimation against a 06>-18~31~43<-56->68~79~91->106 pitch estimate."*

> *"...compared to [the] low note that sounded maybe like a C#"*

3.2. Visuospatial encoding strategy

The average percentage of participants reporting the use of a visuospatial encoding strategy was 8.4%. Typical responses that were coded as visuospatial in nature were:

> *"I visually saw the graph in my mind."*

> *"I just tried to picture a graph in my head."*

> *"I pictured the shape of the graph."*

> *"[I] created an image of notes on a scale"*

> *"I kind of followed a tune and made the graph in my head."*

> *"First, as the sound was playing I tried to picture the equivilent [sic] line graph."*

> *"[I] drew a visual graph in my head as to how the pitch looked."*

> *"[I] listened carefully drew [a] virtual diagram."*

3.3. Motor encoding strategy

The average percentage of participants reporting the use of a motor encoding strategy was 10.7%. Typical responses that were coded as reflecting a motor strategy were:

> *"I tapped my foot to try to keep the beat."*

> *"I counted the snare beats on my fingers and remembered the finger I lifted at the correct time (thumb at 8 am, index finger at 9, etc.)."*

> *"[I] tried to graph the sounds with my finger."*

> *"At one point I was kinda [sic] trying to draw a graph out on the table in front of me."*

> *"I tapped my foot every second."*

> *"I counted every second on my fingers."*

> *"I also sort of traced a graph on the screen with the cursor."*

3.4. Sensory-musical encoding strategy

The average percentage of participants reporting the use of a sensory-musical encoding strategy was 9.2%. Typical responses that were coded as sensory in nature were:

> *"I recorded the first pitch in my mind."*

> *"[I] tried to keep the initial sound in my head."*

> *"[I] tried to keep the starting pitch in my mind the whole time."*

> *"I used pitch memory."*

> *"[I] estimated the pitch by ear."*

3.5. Use of auditory context

The average percentage of participants reporting the use of context was 65.7%. Typical responses that were coded as the use of auditory context were:

"I just compared the pitch to either the highest, lowest, or starting and did my best."

"[I] used surrounding tones to determine best answer."

"I used the initial value."

"[I] used the max/min pitch as a reference."

"I used the first sound to represent a referance [sic] point."

"[I] compared [the sounds] to highest and lowest pitch of the day, compared to [the] known opening price's pitch."

"I counted the snare drum beats for time position."

3.6. Use of counting

The average percentage of participants reporting the use of a counting strategy was 55.8%. Typical responses that were coded as relying on counting were:

"I kept a mental count of the hours."

"I counted the beats."

"I counted using intervals."

"I counted out the seconds as they passed."

"I counted out the seconds starting at 8 in my head."

"[I] counted to the specific time."

"I tried to count seconds."

"I counted the number of hours by simulating a metronome."

3.7. Use of arithmetic

The average percentage of participants reporting the use of arithmetic was 3.0%. Typical responses that were coded as relying on arithmetic were:

"I then subtracted to estimate the value of the stock from its pitch."

"…then adding and subtracting increments of 10…"

3.8. Use of different strategies for different auditory graphs

The average percentage of participants reporting the use of a different strategies depending on difference in the auditory graph stimuli (or based on a switch in strategy at some point during the study) was 8.0%. Typical responses that were coded as differential were:

"For the one- and two-note/hour graphs, I counted each note by the hour it represented and noted the difference between each note or the hours in question; for the multiple notes/hour graphs, I counted the seconds so that I could focus on the difference during the second (hour) in question."

"When the beats were 1 or 2 per second, I counted beats. When they were 4 or 8 per second, I counted time in my head."

"For time, if it was 1 or 2 beats per second, I would count the beats, otherwise I would count seconds in my head."

4. CONCLUSIONS

These descriptive analyses suggest that listeners use a variety of encoding strategies for accomplishing tasks with auditory graphs. Our data offer evidence that all of the possible encoding strategies for frequency that have been suggested in past research in other fields [e.g., music perception, see10, 11, 12] are indeed employed by people using an auditory graph. Specifically, listeners may: 1) attempt to assign verbal labels to data points; 2) imagine a visuospatial or picture-like representation of the data, probably much like a mental image [see, e.g., 13]; 3) use manual or pedal motor codes to encode the information presented in the sounds; and/or 4) attempt to remember and maintain a veridical or isomorphic representation of the sensory experience of the sound heard. We likewise found evidence that both auditory context (in the form of intentional context, reference tones, and the use of surrounding or initial tones) seems to be used relatively often to perform auditory graph tasks. Finally, counting seems to be a critical strategy used by listeners to orient to temporal aspects of the display.

4.1. Limitations of the current study

There are several limitations to the data reported here. First, the strategy question was part of a debriefing protocol that participants competed after performing the tasks of primary interest. The open-ended nature of the question allowed participants to say as little or as much as they preferred, and the mean response length of 32 words suggests that responses were not especially detailed. While the coding scheme was simply designed to detect the presence or absence of a particular strategy, a participant's failure to mention using a strategy *does not* mean that the participant did not use that particular strategy. Participant may have simply failed to mention the use of any given strategy in those few short sentences that were given at the end of a long study. Similarly, participants may not have been consciously aware of the cognitive strategies they used to accomplish the auditory graphing tasks, or they may have been unable to adequately capture their internal strategies in their verbal descriptions. Furthermore, participants may have been biased to report the use of certain strategies, including the use of auditory context, as a result of training and instruction (in the use of auditory graphs, context, and reference tones) that were a necessary precursor for the primary tasks in these studies.

These limitations notwithstanding, the descriptive analyses presented here confirm the applicability of past research on frequency encoding [10-12] to auditory graph research. Furthermore, the data suggest a wealth of interesting and heretofore unexplored possibilities for research on both theoretical and applied aspects of auditory displays.

4.2. Internal representation of sounds and cognitive theory

From a theoretical perspective, very little research has considered how non-speech sounds are internalized and represented by a human listener. Baddeley [e.g., 9, 23, 24] argued strongly for the phonological loop of working memory as a module for processing *all* sounds, both verbal and nonverbal. The full implications of this suggestion are beyond the scope of the current discussion [for a thorough discussion, see 21], but briefly, Baddeley's contention suggested : 1) that the cognitive rehearsal and reinstatement of non-speech sounds employs the articulatory apparatus (i.e., either overt or covert vocalization), and 2) that the concurrent processing of speech and nonspeech audio will interfere at the level of working memory or cognitive representation, even when peripheral acoustic masking has been adequately addressed. Other theorists [25, 26] have vehemently disagreed, arguing instead that speech sounds are processed separately from nonspeech sounds. While available data are equivocal [for a review, see 27], the resolution of this debate will be critical to understanding best practice use of auditory displays in multi-tasking scenarios where speech and nonspeech audio are present.

As mentioned above, most information processing theories have emphasized the relative independence of verbal and visuospatial cognitive representations [e.g., 8, 9], yet these approaches have paid little attention to the cognitive psychology of nonspeech sounds. The data of the current study reaffirm the findings of Mikumo [10, also see 12], who suggested at least four possible formats of cognitive representation for auditory frequency. More research is needed to fully understand the theoretical and applied implications of multiple formats for the cognitive representation of nonspeech sounds [see 21], and it seems likely that revisions to cognitive theory will be required to accommodate a more nuanced understanding of the cognitive processes for nonspeech sound stimuli. Currently, even the more detailed cognitive architectures [e.g., 28, 29] allow for only rudimentary representation (i.e., of frequency and duration) for nonspeech sounds.

4.3. Current and future research

Perhaps the most important research question for the successful use of sonifications involves whether certain formats of internal representation conflict with concurrent visual tasks at a cognitive level during multimodal information processing scenarios. Research has long suggested, for example, that the concurrent auditory and visual presentation of independent verbal information (i.e., via simultaneous spoken speech and visual text) results in profound impairments in the comprehension of at least one of the two messages [30]. This apparent conflict at the cognitive level of verbal representation (despite presentation to separate modalities) should be further examined, as similar information processing conflicts may occur as a function of each of the four representational formats described here.

A pilot study [31] is underway to examine both the malleability of internal representations for auditory graphs, as well as the potential for multimodal cognitive conflicts based on the format of internal representations of nonspeech audio. Participants are being trained to encode the information presented in auditory graph stimuli as either a) verbal lists of values, or b) visuospatial mental images of a graph. After extensive practice with encoding the auditory graphs in these respective formats, participants are asked to perform auditory graphing tasks in the presence of one of two possible visual distractor tasks. The first distractor task [a modified version of the task used by 32] requires verbal processing, while the other distractor task [a mental rotation task, see 33] requires visuospatial processing. The Sternberg task is predicted to show relatively greater interference for participants who encode the auditory graphs as verbal lists of values, while the mental rotation distractor task is predicted to show relatively greater interference for participants who encode the auditory graphs as visuospatial mental images. Such a dissociation would confirm the diagnostic value of the dual-task methodology for studying conflicts at the level of internal cognitive representations in multimodal display scenarios.

The general lack of knowledge about how sound is encoded and represented presents a significant obstacle to the effective deployment of auditory displays. We expect that more research in this area will have a significant impact on our understanding of the appropriate use of auditory displays in man-machine systems.

5. REFERENCES

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

[2] J. H. Flowers, "Thirteen years of reflection on auditory graphing: Promises, pitfalls, and potential new directions," in *International Conference on Auditory Display (ICAD 2005)*, Limerick, Ireland, 2005.

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

[4] C. Frauenberger, T. Stockman, and M.-L. Bourguet, "Pattern design in the context space: A methodological framework for auditory display design," in *13th International Conference on Auditory Display (ICAD 07)*, Montreal, Canada, 2007, pp. 513-518.

[5] A. de Campo, "Toward a data sonification design space map," in *13th International Conference on Auditory Display (ICAD 07)*, Montreal, Canada, 2007, pp. 342-347.

[6] M. Watson and P. M. Sanderson, "Sonification helps eyes-free respiratory monitoring and task timesharing," *Human Factors,* vol. 46, pp. 497-517, 2004.

[7] D. McKeown and S. Isherwood, "Mapping candidate within-vehicle auditory displays to their referents," *Human Factors,* vol. 49, pp. 417-428, 2007.

[8] C. D. Wickens, "Multiple resources and performance prediction," *Theoretical Issues in Ergonomics Science,* vol. 3, pp. 159-177, 2002.

[9] A. D. Baddeley, "Is working memory still working?," *European Psychologist,* vol. 7, pp. 85-97, 06 2002.

[10] M. Mikumo, "Multi-encoding for pitch information of tone sequences," *Japanese Psychological Research,* vol. 39, pp. 300-311, 1997.

[11] M. Mikumo, "Motor encoding strategy for pitches of melodies," *Music Perception,* vol. 12, pp. 175-197, 1994.

[12] R. J. Zatorre and C. Beckett, "Multiple coding strategies in the retention of musical tones by possessors of absolute pitch " *Memory & Cognition,* vol. 17, pp. 582-589, 1989.

[13] S. M. Kosslyn, "Information representation in visual images," *Cognitive Psychology,* vol. 7, pp. 341-370, 07 1975.

[14] R. N. Shepard, "The mental image," *American Psychologist,* vol. 33, pp. 125-137, 1978.

[15] J. H. Flowers, D. C. Buhman, and K. D. Turnage, "Cross-modal equivalence of visual and auditory scatterplots for exploring bivariate data samples.," *Human Factors,* vol. 39, pp. 341-351, 1997.

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

[17] S. Cantu, "The effect of context on root mean error in point estimation tasks for auditory graphs," in *School of Psychology* Atlanta, GA: Georgia Institute of Technology, 2007.

[18] B. K. Davison and B. N. Walker, "Sonification Sandbox reconstruction: Software standard for auditory graphs," in *ICAD 07 - Thirteenth Annual Conference on Auditory Display*, Montreal, Canada (26-29 June), 2007, p. TBD.

[19] D. R. Smith and B. N. Walker, "Effects of auditory context cues and training on performance of a point estimation sonification task," *Applied Cognitive Psychology,* vol. 19, pp. 1065-1087, 2005.

[20] B. N. Walker and M. A. Nees, "An agenda for research and development of multimodal graphs," in *International Conference on Auditory Display (ICAD2005)*, Limerick, Ireland, 2005, pp. 428-432.

[21] M. A. Nees, "Internal processing codes for auditory information displays: Implications for theory and interface design," Unpublished doctoral preliminary examination paper. Georgia Institute of Technology, 2008.

[22] W. A. Rogers, B. Meyer, N. Walker, and A. D. Fisk, "Functional Limitations to Daily Living Tasks in the Aged: A Focus Group Analysis," *Human Factors,* vol. 40, 1998.

[23] P. Salame and A. D. Baddeley, "Noise, unattended speech, and short-term memory," *Ergonomics,* vol. 30, pp. 1185-1194, 1987.

[24] P. Salame and A. D. Baddeley, "Effects of background music on phonological short-term memory," *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology,* vol. 41, pp. 107-122, 1989.

[25] A. M. Liberman and I. G. Mattingly, "The motor theory of speech perception revised," *Cognition,* vol. 21, pp. 1-36, 1985.

[26] A. M. Liberman and D. H. Whalen, "On the relation of speech to language," *Trends in Cognitive Sciences,* vol. 4, pp. 187-196, 2000.

[27] J. D. Smith, D. Reisberg, and M. Wilson, "Subvocalization and auditory imagery: Interactions between the inner ear and inner voice," in *Auditory Imagery*, D. Reisberg, Ed. Hillsdale, NJ: Lawrence Erlbaum Associates, 1992, pp. 95-119.

[28] D. E. Kieras and D. E. Meyer, "An overview of the EPIC architecture for cognition and performance with application to human-computer interaction," *Human-Computer Interaction,* vol. 12, pp. 391-438, 1997.

[29] D. E. Meyer and D. E. Kieras, "A computational theory of executive cognitive processes and multiple-task performance: Part 1. Basic mechanisms," *Psychol Rev,* vol. 104, pp. 3-65, 1997.

[30] G. H. Mowbray, "Simultaneous vision and audition: The comprehension of prose passages with varying levels of difficulty," *Journal of Experimental Psychology,* vol. 46, pp. 365-372, 1953.

[31] M. A. Nees and B. N. Walker, "Encoding of information in auditory displays: Initial research on flexibility and processing codes in dual-task scenarios.," in *52nd Meeting of the Human Factors and Ergonomics Society (HFES 2008)*, New York, NY, in press.

[32] S. Sternberg, "High-speed scanning in human memory," *Science,* vol. 153, pp. 652-654, 1966.

[33] R. N. Shepard and J. Metzler, "Mental rotation of three-dimensional objects," *Science,* vol. 171, pp. 701-703, 1971.