APPLIED COGNITIVE PSYCHOLOGY
Appl. Cognit. Psychol. (2008)
Published online in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/acp.1457

Most Earcons Do Not Interfere with Spoken Passage Comprehension

TERRI L. BONEBRIGHT[1]* and MICHAEL A. NEES[2]

[1]*Department of Psychology, DePauw University, Greencastle, USA*
[2]*School of Psychology, Georgia Institute of Technology, Atlanta, USA*

SUMMARY

A cross-modal dual attention experiment was completed by 198 undergraduates in three blocks that each consisted of an orientation task and a concurrent listening task. For the orientation task, participants located regions on an LCD that were cued by speech or one of four types of symbolic auditory cues (i.e. earcons); the concurrent task required participants to listen to and answer questions about GRE sample test passages. Results indicated the orientation task had no effect on comprehension of the passages compared to a passage-only control for four of the five auditory cue types. All auditory cues resulted in high performance for the orientation task, with speech and complex sounds exhibiting the highest performance. Implications for auditory display design and for assistive technologies for visually impaired persons are discussed. Copyright © 2008 John Wiley & Sons, Ltd.

Interest in auditory displays as an alternative or accompaniment to visual displays has flourished in recent years. Researchers have extensively investigated the use of non-speech auditory earcons to convey information regarding *processes and actions* in the context of interface design (see Blattner, Sumikawa, & Greenberg, 1989; Kramer, 1994). Earcons have been used in a variety of computer applications, including as an assistive technology for visually and physically impaired people (Brewster, Raty, & Kortekangas, 1996; Hankinson & Edwards, 1999; Stevens, Edwards, & Harling, 1997), in navigation aids for software menus (Brewster, 1998), and to assist people's awareness of one another while working together on distributed projects (Isaacs, Walendowski, & Ranganthan, 2002). Such abstract sounds that use variations in timbre, pitch, and rhythm have been effectively mapped onto actions and objects that have no inherent auditory representation in software packages (Brewster, Wright, & Edwards, 1995). A trade-off exists, however, in that the abstract nature of earcons may require a period of learning or formal training in their use, since they may be more difficult to learn and retain than sounds with at least some degree of ecological relationship to their referent (for a review, see Edworthy & Hellier, 2006).

Two of the primary advantages of auditory information displays involve the potential for sound to serve alerting and orienting functions (Kramer, 1994). Many environments in which an auditory information display might be implemented, however, may be saturated

*Correspondence to: Terri L. Bonebright, Department of Psychology, DePauw University, 7 E. Larabee St., Greencastle, IN 46135, USA. E-mail: tbone@depauw.edu with other sounds including, but not limited to, human speech. Although studies in psychoacoustics have made the prediction of masking effects tenable in highly controlled environments with limited stimulus sets (Watson & Kidd, 1994), the prediction of acoustic masking in more complex, ecologically valid settings is a difficult problem (Edworthy, 1998). Furthermore, recent research on central masking (i.e. non-peripheral masking that occurs beyond the cochlea in the auditory system) suggests that issues of acoustic masking may indeed be a very complex phenomenon that cannot be fully explained by predictions based on models of the acoustic periphery (see Durlach, Mason, Kidd, Arbogast, Colburn, & Shinn-Cunningham, 2003). Despite these warranted concerns, few studies have examined the extent to which auditory displays interfere with concurrent listening tasks and vice versa.

CROSS-MODAL ATTENTION IN CUING SPATIAL LOCATION

Many studies have examined cross-modal cuing (and inhibition) of visual spatial attention using spatialised audio. This research has confirmed that auditory signals can facilitate visual attention to a spatial location (e.g. Eimer, 2001; McDonald, Teder-Salejarvi, & Hillyard, 2000; Mondor & Amirault, 1998, for reviews, also see Schmitt, Postma, & De Haan, 2000; Spence, McDonald, & Driver, 2004), and spatial audio has been shown to decrease the time required to perform a three-dimensional visual search (Bolia, D'Angelo, & McKinley, 1999). Little research to date, however, has examined whether symbolic auditory cues—non-spatialised sounds that convey meaning via auditory dimensions such as pitch and timbre—or verbal cues can be used to effectively capture or orient visual spatial attention.

In one of the few studies examining non-spatialised verbal auditory cues, Ho and Spence (2006) found that such cues do not automatically facilitate visual attention, although they acknowledge that their testing paradigm may have allowed participants to simply ignore the auditory cues and use visual information. Furthermore, a previous study by Ho and Spence (2005) had found that, during a simulated driving task, auditory cues presented behind the participant decreased reaction time for critical events occurring in the rear-view mirror—located physically in front of the participant—which suggests that auditory cues need not necessarily coincide with a target's spatial location to be useful for orienting visual attention. Ben-Artzi and Marks (1995) have demonstrated that the frequency of non-spatialised sounds can facilitate classification of visual stimuli that varied in spatial location, provided that the auditory and visual stimuli were congruent (i.e. higher frequency sounds were coupled with higher spatial location).

AUDITORY DUAL TASK PERFORMANCE

While many questions remain about how well symbolic auditory cues can facilitate visual *spatial* attention, Kramer (1994) further identified the potential for sounds to interfere with concurrent speech communication in real world applications as a significant obstacle to the implementation of auditory displays. Accordingly, multiple resource models (e.g. Wickens, 2002) would predict that performance with an auditory display would interfere with performance on a concurrent listening task to the extent that the multi-task scenario imposed a sufficiently high workload by taxing limited modality resources and also to the extent to which the codes of the concurrent information (defined by a continuum anchored by verbal information at one end and manual/spatial information at the other) interfered with each other.

Research regarding the effects of a secondary task on speech comprehension is somewhat conflicted. Tun, O'Kane, and Wingfield (2002) found that both younger and older adults suffered worse recall of speech under conditions where either meaningful or meaningless distractor speech was concurrently presented. Other research has found that target speech signals can be detected with surprisingly high accuracy in the presence of up to five competing speech signals (Simpson, Brungart, Iyer, Gilkey, & Hamil, 2006), but the intelligibility of the message in such displays decreases dramatically with each added speaker (see Ericson, Brungart, & Simpson, 2003). Browsing a website while listening to irrelevant auditory information has been shown to have no negative impact on either task, while relevant, assistive speech improved performance on the browsing task (Fang, Xu, Brzezinski, & Chan, 2006).

Despite the high likelihood that auditory displays will be deployed and used in the context of dual or multi-tasking scenarios involving other auditory or visual input concurrently, surprisingly few studies have examined the degree to which concurrent tasks interfere with auditory displays and vice versa. Janata and Childs (2004) showed that sonified displays presented as accompaniments to visual displays aided monitoring stock data. Interestingly, the positive contribution of the auditory display was even more pronounced when a secondary number-matching task was added. Peres and Lane (2005) found that the addition of a concurrent visual monitoring task to an auditory monitoring task initially produced deficits in performance of the auditory task; however, performance soon (i.e. after around 25 dual task trials) returned to pre-dual task levels.

AUDITORY DISPLAYS AS ASSISTIVE TECHNOLOGIES IN CLASSROOMS

Although visually impaired students are commonly integrated into mainstream classrooms, these learners face a number of obstacles (see Dimigen, Roy, Horn, & Swan, 2001) that could be addressed by assistive technologies developed within electronic classrooms. For example, many visually impaired persons have a degree of intact visual function (about 124 million people worldwide have 'low vision', see Resnikoff et al., 2004), and for such students, reading from a traditional chalkboard while taking notes can be problematic (see Kalloniatis & Johnston, 1994). In an electronic classroom, however, these students may have access to lecture notes (e.g. from an electronic blackboard) and slides either via a 'talking' digital whiteboard (Berque, 2003) or a personal display that has been adapted (i.e. magnified, etc.) to meet their needs. Given that such students may need to fixate a few inches from the display to view it properly, it seems likely that additional cues could help orient the viewer not only to *when* an activity occurs on the display but also *where* that activity is located (Berque, Bonebright, Kinnett, Nichols, & Peters, 2003). Wickens' multiple resource model (for an overview, see Wickens, 2002) would suggest that an already taxed visual system would be a poor choice for the modality of delivery for such orienting cues, but auditory cues could be used as an alternative.

Accordingly, researchers have attempted to represent a visual grid system with earcons. Edwards' (1989) Soundtrack interface, for example mapped the frequencies of sounds to a grid system within the display, but his use of simple square wave tones proved to be of limited usefulness to visually impaired listeners. Most listeners circumvented the need for

DOI: 10.1002/acp

pitch information by simply counting tones to determine grid coordinates in the Soundtrack interface. Brewster et al. (1996) used a grid system mapped with earcons for assisting physically disabled people to scan a display and choose an action, but they only performed a brief trial with one participant. Given the limited research in this area, the usefulness of sound to orient a listener to a grid location remains unclear.

CURRENT STUDY

The current study was designed to mimic conditions that students with low vision would face in classrooms that provide assistive technology in the form of an LCD display with magnification (Betz, 2002). The LCD in such classrooms presents material that would normally be displayed using either an overhead projector, a whiteboard, or a chalkboard. In such cases, the student needs to attend not only to any visual information that is present but also to any information that is being presented orally by the instructor. Thus, the type of material used for the concurrent listening task in this study needed to have a similar complexity and range of topics to what a student might actually experience in a classroom. In addition, this listening task had to require the participants to remember the material, which would be expected for most course instruction. Our goal for the orientation task was to provide a valid test of a student working with an LCD, such that he or she would need to know both *when* and *where* to look on the screen.

Therefore, we assessed several different types of earcons for their usefulness in orienting visual attention to regions of the visual display, while participants concurrently listened to a GRE test passage for which they answered questions. The earcons included simple tones (a pitch mapping), instruments (a timbre mapping), sweeps (a pitch-change mapping), pitched instruments (a pitch *and* timbre) mapping, and speech (a verbal mapping) to indicate regions in the visual display. These sounds were produced using a rule-based, system following Blattner et al.'s (1989) and Brewster et al.'s (1995) specific guidelines and Hereford and Winn's (1994) general approach to using sound for interfaces to form 'families' of abstract sounds that corresponded to the rows and columns of the visual display grid.

Hypotheses

For the display orientation task, we predicted that the speech and pitched instruments conditions would exhibit the best performance. Speech is a familiar stimulus that should be easy to associate with regions of a visual display, especially since the grid display in the current study matched the mappings used on calculators. Pitched instruments have a redundant coding scheme that provides multiple mappings (i.e. both pitch *and* timbre variations) to represent locations. Both simple tones and instruments have no redundant mapping; thus, we expected that they would not be as helpful as speech or pitched instruments. Finally, sweeps offer a pitch change orientation that is non-redundant and which may be a less salient manipulation than a pitch manipulation (see Walker & Ehrenstein, 2000).

We predicted that the speech stimuli would interfere with the concurrent listening task due to the common verbal code of the information (see Wickens, 2002); thus comprehension of the speech passages would be worse with speech orienting cues compared to all earcons. We further predicted that a control group that only listened to the passages (i.e. the single task condition) would outperform participants who also performed the concurrent orienting task, regardless of the type of auditory cue used.

METHOD

Participants

Participants were 204 primarily Caucasian undergraduate students who were recruited from psychology courses and offered a choice of either extra credit or a payment of $5.00 for their contribution. The data from four participants were not available due to computer malfunction. Two additional participants were identified as outliers (i.e. more than two standard deviations from the mean) on more than three measures, and their data were excluded. The final sample included 198 (60 males and 138 females) participants ranging in age from 17 to 25 years old ($M = 19.75$, $\mathrm{SD} = 1.21$).

Materials and apparatus

Passages

Three passages with their accompanying questions were chosen from a Graduate Record Examination (GRE) general test practice booklet (*GRE: Practicing to take the general test*, 1998) to represent a range of subject areas (evolution of intelligence, comparison of scientific disciplines, and myths about Amazonian society). We required more than one passage with different topics to make sure that participants' previous knowledge would not unduly affect their scores and that they would represent a range of material that might be reasonable to expect in college courses. The passages were read aloud by a male undergraduate student, who also worked as a radio announcer, and recorded onto cassette tapes. The resulting spoken passages ranged in length from 3.03 to 3.08 minutes and were presented to participants using an Onkyo TA-RW490 cassette recorder and two Koss M80 stereo speakers placed 60 cm from each other on either side of a laptop computer 50 cm in front of the participants.

Computer hardware and software

A Dell Latitude laptop with a 28.5 cm × 21.5 cm LCD screen was used as the visual display and to present all sound stimuli for the orientation task and to record screen location responses. A mouse was attached to facilitate movement of the cursor on the LCD while participants made location selections. The software for the experiment was written using Visual Basic 6.0, and Sony open-air headphones (model MDR-15) were used to deliver the sound stimuli from the computer to the participant. All sound files (except for the speech sounds) were synthesized mono-aural sounds produced and edited using Metasynth and SoundEdit, and saved as .WAV files. Spoken number stimuli were recorded by a female speaker using PeakDV and an Audio-technica MB 4000C microphone. A female speaker was chosen to produce these stimuli to provide a non-conflicting auditory stream with the male voice used for the passages (see Ericson et al., 2003).

Experimental conditions for the sound/screen task

The LCD screen was divided into nine rectangles (each 9 cm × 6.5 cm) in a grid arranged in three rows and three columns, and five sets of nine sounds were produced to reference the regions of the screen (see Figure 1). The task whereby sound cues were used to orient participants to a display region is henceforth referred to as the sound/screen task, and the experimental conditions were:

DOI: 10.1002/acp

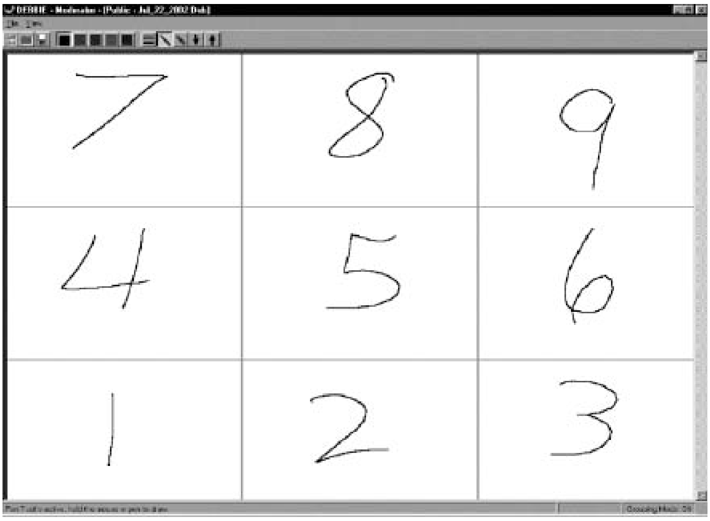

Figure 1. Screen regions designated by number

Speech: The speech condition used spoken number mappings to cue the different regions in the display grid. The numbers were mapped onto the screen positions in the same manner as a calculator. The numbers were recorded by a female speaker saying the numbers 1 through 9 and ranged in duration from 435.5 to 626.9 milliseconds.

Tones: The tones were 240, 540 and 840 Hz modified sine waves that were faded out using offset ramps through the entire 200 millisecond tone. They were mapped with the 240 Hz tone on the bottom row, the 540 Hz tone on the middle row and the 840 Hz tone on the top row, as research has suggested that relatively higher pitch is generally associated with relatively higher spatial location (see Rigas & Alty, 2005; Walker & Ehrenstein, 2000). The columns were indicated from left to right by one, two and three repetitions of each sound without breaks—the same method used for the other three sets of sounds described below—and the resulting durations ranged from 200 to 600 milliseconds.

Sweeps: The sweeps were composed of three different sounds, all with a duration of 230.1 milliseconds. The sound for the middle row of the display was a steady sine wave at 540 Hz. The sound for the upper row was a sine wave that began with a 540 Hz tone with a two-octave ramp up to produce a tone with an upward sweep while the sound for the bottom row also began at the same pitch followed by a two-octave downward ramp in frequency. The duration range for these stimuli was 230.1–690.2 milliseconds.

Instruments: The instruments chosen were a synthesized saxophone, cello and flute all at 540 Hz, which were mapped onto the bottom, middle and top rows, respectively.

Pitched instruments: The pitched instruments were the same as the instruments with the addition of a pitch cue such that the saxophone was at 240 Hz, the cello at 540 Hz, and the flute was at 840 Hz each with a duration of 229.8 milliseconds. The duration range for both the instruments and the pitched instruments was 229.8–689.3 milliseconds.

The final experimental condition was a passage-only control, during which participants only performed the passage comprehension task.

Design and procedure

The study employed a 6 (Sound) × 3 (Block) mixed design. The experimental conditions manipulated the mapping of sounds with the screen positions, and this between groups variable had six levels (speech, tones, sweeps, instruments, pitched instruments and passage-only control) with 33 participants in each condition. Block was a within-subjects variable that had three levels, one for each set of 36 screen responses and one passage; thus, participants completed a total of 108 screen responses and listened to and answered questions for three passages during the 1-hour procedure.

Participants were randomly assigned to one of the six experimental conditions. During training participants were given instructions about the layout of the screen grid and the associated sounds, and they twice heard each of the nine sounds played one at time while the corresponding screen region lit up. Participants then received three practice trials with feedback for the sound/screen task. During both practice and experimental trials, participants had 5 seconds within which to respond to a sound by clicking the mouse on the chosen region. For all trials, feedback was given within the chosen region using a 6 cm × 2 cm white rectangle containing either the word 'Correct' or 'Incorrect' in bold 36 pt black letters. Participants were instructed to perform the sound/screen task as quickly as possible while still maintaining accuracy.

After completing the training and practice trials for the sound/screen task, participants were told that they would also be performing a second task. The experimenter explained that they would hear a short passage, similar to the type found on the SAT or ACT, for which they would answer questions after both the passage and the sound/screen task were finished. They were encouraged to do the best they could on both tasks simultaneously.

For testing, participants were randomly assigned to one of six possible orders for the passages as a control for order effects. They completed three blocks of experimental trials, and within each block, each sound with its corresponding grid position was presented four times in a random order for each participant. Participants, therefore, were required to respond to a sound cue by selecting a visual display region approximately every 5–6 seconds. The passages were played at a normal speaking level (approximately 60 dB SPL). After the story and the sounds and screen matching were completed, participants answered seven questions about the passage that were taken from the GRE practice test. Upon completion of all three blocks, participants filled out a follow-up survey.

For the passage-only condition, the participants performed only the passage comprehension task, which was used to provide a baseline. After the completion of the three passages, these participants also completed a follow-up questionnaire.

Dependent measures

The dependent measures included the number of correct screen responses out of 36 for each block (i.e. participant responses for the sound-to-screen matching), and the mean response time for screen responses (latency to choose the screen position after the sound began). Of note, the response time was recorded from the onset of a given sound, thus the actual reaction time of the participant was potentially confounded with the different durations for the sound stimuli. We address this problem using two methods that will be discussed in Results Section. Passage comprehension scores, which were the number of correct answers out of seven questions for a single passage for a total score of 21 for each participant, were also collected.

The follow-up questions required participants to indicate on a 5-point scale their responses to questions addressing their perceptions of how easy it was to use the sounds and the degree of annoyance and distraction caused by the sounds. All participants could also respond to an open-ended question about any other comments they had about the experiment.

RESULTS

Passage comprehension scores

A within-groups ANOVA to test the effect of the three passage types on mean passage scores revealed a significant difference, $F(2{,}394) = 20.63$, $p < .001$, $\eta_p^2 = .10$. Follow-up analyses showed significant differences among the means for all the passages except between the science and the Amazon passages: science: $M = 3.11$, SD = 1.50; Amazon: $M = 3.24$, SD = 1.69; intelligence: $M = 3.93$, SD = 1.60. This analysis suggested that our goal of having passages and questions that varied in difficulty but did not produce floor or ceiling effects was achieved.

A 6 (Sound) × 3 (Block) mixed factorial ANOVA comparing the mean passage scores among all six experimental conditions for the three blocks showed a non-significant interaction, $F(10{,}384) = 1.39$, $p = .18$, and two significant main effects (sound, $F(5{,}192) = 4.07$, $p = .002$, $\eta_p^2 = .09$; block, $F(2{,}384) = 6.80$, $p = .001$, $\eta_p^2 = .03$).

Follow-up analyses for the main effect for block showed that mean passage scores increased from block 1 ($M = 3.12$, SD = 1.70) to block 2 ($M = 3.58$, SD = 1.61), but that there was no difference between blocks 2 and 3 ($M = 3.57$, SD = 1.57). Contrary to our hypothesis, Tukey's HSD *Post hoc* tests for the main effect for sound condition revealed no differences among the scores for tones, speech, instruments, pitched instruments, and the passage-only control. Sweeps showed the lowest mean passage scores compared to all conditions except for instruments (see Table 1 for total passage score means).

Sound/screen task responses

Number of correct responses

A 5 (Sound without the passage-only control) × 3 (Block) mixed factorial ANOVA was performed using the number of correct screen responses for each block as the dependent variable. The analysis revealed a significant interaction, $F(8{,}320) = 4.97$, $p < .001$, $\eta_p^2 = .13$ and that both main effects were significant (sound, $F(4{,}160) = 14.13$, $p < .001$,

Table 1. Mean total passage comprehension scores (out of 21) for all sound/screen conditions and the passage-only control collapsed across block

Sound condition	Mean	SD	*n*
Passage-only control	11.03^{a}	3.36	33
Speech	11.00^{a}	3.75	33
Tones	11.12^{a}	3.10	33
Sweeps	8.24^{b}	2.46	33
Instruments	9.30^{ab}	3.50	33
Pitched instruments	10.82^{a}	3.46	33

Higher scores indicate better comprehension.
Means indicated by different superscript letters are significantly different from one another.

DOI: 10.1002/acp

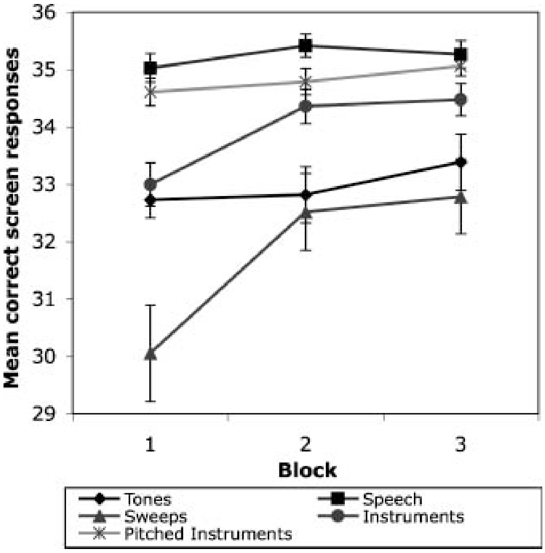

Figure 2. Mean number of correct screen responses (±SE) for sound conditions by block

$\eta_p^2 = .26$; block, $F(2,320) = 24.61, p < .001, \eta_p^2 = .11$). Due to the significant interaction, it is important to note that the main effects may be misleading and should be considered within the context of the interaction.

Follow-up analyses for the main effect for block showed that the correct number of screen responses increased from block 1 ($M = 33.09$, SD = 3.19) to block 2 ($M = 33.98$, SD = 2.62), but there was no difference between blocks 2 and 3 ($M = 34.20$, SD = 2.5), which matches the results for the passage comprehension scores. The means for the screen responses also show that participants performed at almost ceiling level across all blocks.

Tukey's HSD test of the main effect for sound condition revealed that participants listening to speech ($M = 35.24$, SD = 1.05) had a higher number of correct screen responses than participants listening to sweeps ($M = 31.79$, SD = 3.74), tones ($M = 32.98$, SD = 2.25), or instruments ($M = 33.95$, SD = 1.36) but were not different from those participants who heard pitched instruments ($M = 34.82$, SD = .93). Sweeps had the lowest number of correct screen responses compared to all other conditions except tones.

Simple effects were analysed for the sound conditions within each block to investigate their interaction. These analyses showed that in block 1, speech had a higher mean number of correct screen responses than all other conditions except pitched instruments, sweeps had fewer correct responses than any other condition, and tones were equal to instruments but had fewer correct responses than speech and pitched instruments (see Figure 2). In block 2, speech had a higher mean number of correct screen responses than tones and sweeps and no differences with instruments and pitched instruments. For block 3, sweeps were equivalent in correct responses to tones, but had fewer correct responses than all other conditions. For all three blocks, pitched instruments had the same pattern of differences such that the responses were equivalent with speech and instruments but greater than tones and sweeps.

Response times for screen responses

As mentioned previously, the response time was recorded from the onset of each stimulus. The non-speech sound stimuli were constructed in such a way that there was a systematic

DOI: 10.1002/acp

bias in terms of when the information became available for each target column. In contrast, the speech sounds may have been recognized before the entire stimulus was heard by listening only to the initial phoneme of the sound. Thus, we took a conservative approach and performed two analyses: one with the full stimuli durations subtracted from the response times, and another analysis in which the mean recognition times for the speech stimuli[1] were substituted for the full duration for the speech stimuli. The results from the two analyses were identical and are reported in detail below.

A 5 (Sound) × 3 (Block) mixed factorial ANOVA was performed for the response times minus the stimulus durations for the screen choices made during the sound/screen task. The analysis revealed that both main effects were significant (sound, $F(4,160) = 17.75$, $p < .001$, $\eta^2_{\mathrm{p}} = .31$; block, $F(2,320) = 25.17$, $p < .001$, $\eta^2_{\mathrm{p}} = .14$), as was the interaction, $F(8,320) = 2.73$, $p = .006$, $\eta^2_{\mathrm{p}} = .06$). We again report both the main effects and the interaction, but note that the interpretation of these main effects alone may be misleading.

Follow-up analyses for the main effect of block showed that response time was slowest for block 1 ($M = 1142.34$, SD = 368.36) and that there were no differences in response times between block 2 ($M = 1057.39$, SD = 315.66) and block 3 ($M = 1066.91$, SD = 339.22). This follows the same pattern of differences as were seen in the main effect for block on sound/screen response accuracy and for passage comprehension scores.

The main effect for sound condition was examined using Tukey's HSD and speech ($M = 755.75$, SD = 169.48) showed faster response times than all other sound conditions (tones [$M = 1138.49$, SD = 286.30]; sweeps [$M = 1306.93$, SD = 350.77]; instruments [$M = 1131.46$, SD = 276.39]; pitched instruments [$M = 1111.77$, SD = 263.58]). Responses to sweeps were slower than for instruments, and finally, tones, instruments, and pitched-instruments had equivalent response times.

The interaction was examined using the simple effects of sound condition within each block (see Figure 3). This set of analyses revealed that for all three blocks, speech showed faster response times than all other sound conditions. Sweeps showed a slower response time than all other conditions for block 1, but for block 2 and block 3, sweeps were only slower than speech. Instruments, pitched instruments, and tones had equivalent response times across all three blocks.

Follow-up questions for sound conditions

Four one-way between-groups ANOVAs were performed to determine if there were mean differences among the sound conditions for each of the follow-up questions, which used a rating scale of 1, strongly agree to 5, strongly disagree (see Table 2 for means). The results showed no differences among sound conditions for the reported ease of learning which sound corresponded to the screen regions, $F(4,160) = 1.45$, $p = .22$, or for the ease of identifying the correct screen region once the associations were learned, $F(4,160) = 2.09$, $p = .08$. When asked if the sounds were annoying to listen to while listening to the stories, however, participants gave significantly different responses across sound conditions, $F(4,160) = 4.45$, $p = .002$, $\eta^2_{\mathrm{p}} = .10$. Follow-up analyses showed that speech stimuli were

[1] To obtain mean recognition times for the speech stimuli, a pilot study ($n = 9$ additional participants) heard three random presentations of each of the individual speech stimuli and pressed a key when they recognized the number being spoken. Recognition was confirmed after each trial with keypad responses. Mean recognition times for each of the speech stimuli (excluding incorrect recognition trials) were averaged across the three presentations and nine participants for each speech stimulus.

DOI: 10.1002/acp

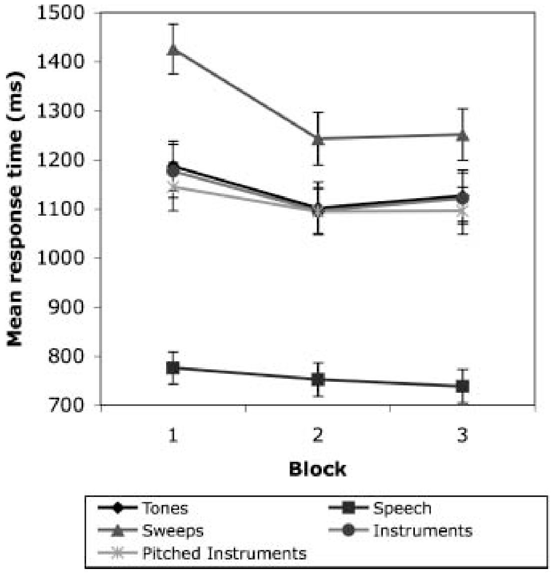

Figure 3. Mean response time minus stimuli durations for screen responses (±SE) for sound conditions by block

Table 2. Mean scores on follow-up questions by sound condition

	Question topic			
	Ease of learning ($n = 33$)	Ease of identification ($n = 33$)	Annoyance ($n = 33$)	Distraction ($n = 33$)
Speech	1.36 (1.00)	1.33 (0.92)	2.73 (1.21)	1.88 (1.14)
Tones	1.67 (1.24)	1.88 (1.14)	2.70 (1.19)	1.88 (0.96)
Sweeps	1.73 (0.88)	1.64 (0.70)	1.85 (1.03)	1.15 (0.36)
Instruments	1.48 (0.91)	1.55 (0.83)	2.06 (1.14)	1.52 (1.00)
Pitched instruments	1.24 (0.75)	1.33 (0.89)	2.00 (1.06)	1.52 (1.18)

Note that responses were made on a scale ranging from 1 (strongly agree) to 5 (strongly disagree), with 3 being neutral. Standard deviations are in parentheses.

perceived as less annoying than sweeps, but that there were no other differences among the conditions. Examination of the means shows that speech was neutral on the scale while the other conditions showed means that tended toward more annoyance. Thus, none of the sounds were perceived as ‘pleasant’ within the context of the dual task. There were also significant differences among the conditions for the question regarding whether hearing the sounds was distracting, $F(4,160) = 3.22$, $p = .014$, $\eta^2_{\mathrm{p}} = .09$, such that participants found sweeps to be more distracting than speech or tones with no other differences among the conditions.

DISCUSSION

The finding that passage comprehension scores for the control group were *not* significantly different from the scores for tones, pitched instruments, instruments or speech clearly

DOI: 10.1002/acp

indicates that the addition of the auditory orienting task did not impose sufficiently high workload to interfere with the concurrent listening task except in the case of the sweeps. Past research has suggested that a policy of ‘pre-emption’ may operate during the concurrent performance of a discrete and a continuous task, whereby the discrete task momentarily interrupts and diverts attentional resources away from the continuous task (Wickens & Liu, 1988). In the current study, any diversion of resources imposed by the discrete sound/screen task did not appreciably affect performance of the continuous, concurrent listening task for the majority of the sound conditions studied here. It is important to keep in mind that the participants did not have to respond simultaneously to the two tasks—responses to the listening task were post hoc; however they did have to maintain the information in memory in order to answer the questions at the end of each block. Multiple resource theory (Wickens, 2002) would predict little interference if concurrent responding taxed different pools of resources (e.g. manual responses and verbal responses draw upon different resources), but future research should examine the extent to which auditory orienting cues can be used during tasks requiring concurrent responses, both manual and verbal, for *both* tasks.

Interestingly, the speech condition for the sound/screen task did not affect performance on the concurrent listening task as compared to the single task baseline, despite the fact that the information for both tasks seemed to rely upon the same underlying verbal processing code. Recent research has suggested that ostensibly verbal processing codes, however, can sometimes evoke underlying spatial representations and interfere with spatial perceptual tasks (Richardson, Spivey, Barsalou, & McRae, 2003). Despite the apparent verbal/phonological nature of the speech numbers, they perhaps may have been quickly translated to their visual spatial display mappings, which would be less prone to interfere with comprehension of the verbal passage. However, it could be that this result was due to the brief duration of the speech stimuli used in the current study (all <650 milliseconds), and this finding may not hold in situations where lengthier concurrent speech would be required. Furthermore, the different voices (male and female) employed for the information streams should have offered a salient cue for perceptual segregation (see Ericson et al., 2003). The relatively comparable accuracy performance for the screen task of the non-speech sound conditions by the third block suggests that non-speech audio may be a viable alternative to verbal auditory cues in scenarios where simultaneous speech presentation is a concern, especially if the user has sufficient practice.

The results for the number of correct responses for the sound/screen task showed a high level of performance (approaching ceiling) for all sound conditions, with speech superior to all conditions except pitched instruments during the first block of trials. By the third block of trials, performance had increased so that only tones and sweeps were still showing fewer correct responses than speech. The initial relative advantage of the speech condition may be due to the known meaning of speech stimuli. An examination of differences in the mean numbers of correct responses in Figure 2 suggests that the redundantly coded (i.e. pitch *and* timbre coded) pitched instruments were easier to use (at least initially) than non-speech audio coded by a single acoustic dimension (i.e. *only* pitch, pitch change, or timbre). It is clear from these data that practice with the instruments led to increases in performance over a very short period of time. This suggests that within contexts where there may be problems using two auditory tasks that both require speech, practice with another type of correctly designed sound may quickly show good performance.

Results for the follow-up questions showed that participants perceived the sound to grid mappings to be easy to learn and easy to identify, regardless of the specific sound stimuli employed. None of the sounds, however, were perceived to be particularly pleasant. When participants were asked about how distracting the sounds were, speech and tones were found to be less distracting than sweeps. In a similar pattern to perceived pleasantness, the overall means suggest that all sounds were fairly distracting, which is not surprising, considering that participants had to respond to a sound cue approximately every 5–6 seconds while listening to the complex content of the passages. We also wish to note that the volume of the sounds should not have been an issue since the participants were allowed to make adjustments during the training period. These results suggest that auditory display designers should continue to examine different types of sounds to determine which are effective for the specific task *and* are pleasant and do not distract from the task.

Taking into consideration all the results in the current study, it is clear that in general the sweeps were the least effective auditory cue, since they took longer to respond to, led to the lowest correct number of screen responses and correct passage scores, and were considered one of the most distracting and annoying stimuli in the set. It seems that the reason for this was that the mappings for the sweeps did not work well for most participants, which is supported anecdotally by open-ended comments where some participants reported that they had a difficult time remembering which tone represented which row in the display. Thus, even though these sounds were produced in accordance with current guidelines for earcons, developers considering manipulating pitch change for auditory information displays should do so cautiously (also see Walker & Ehrenstein, 2000).

The overall results of this study suggest that a system using an LCD display with earcons, would indeed be useful for students with low vision. We have recently piloted such a system with two visually impaired students in upper level psychology courses. Both students were given all five sounds as options for their use. Overall, these students reported that the system worked well and allowed them access to material that they normally would have needed to obtain in some other fashion, such as hard copy given to them prior to class. Neither of the students chose to use the sweeps, which reiterated the results of the current study; however, both students liked having a sound ‘palette’ of options they could try under different classroom conditions.

Although the impetus of the current study was assistive technology in the classroom, the results may be applicable to the design of other interfaces where non-spatialised auditory cues might be used to orient visual attention to a location on a visual display. Focus groups with visually impaired computer users found that this population fears being increasingly shut-out of access to even traditionally low-tech devices and appliances (e.g. stovetops, refrigerators, etc.) as such devices become increasingly menu-driven in their basic operation (Gerber, 2003). Furthermore, as Griffith (1990) argued, such accommodations often stand to benefit all users of an interface and may become even more important with the increasing use of multiple visual displays in computer workstations.

Further research is needed to clarify the extent to which the current findings hold when the complexity or difficulty of the auditory display user’s task increases (see Navon & Gopher, 1979). Given that researchers (e.g. Edworthy, 1998; Kramer, 1994) have long speculated that a major disadvantage of auditory displays may be their interference with other concurrent auditory stimuli, continued empirical investigations of this topic will be imperative for the success of auditory displays. The findings from the current study regarding passage comprehension scores, however, are encouraging and speak to the potential for auditory displays to be deployed in scenarios where concurrent speech must be simultaneously attended to along with the auditory display.

ACKNOWLEDGEMENTS

Part of this research was funded by a National Science Foundation REU Grant (EIA-9911626) and by the Asher Rsearch Fund. We thank Dave Berque, Ben Betz, Seth Kinnett, Nathan Nichols, Adam Peters, and Jing Tun for their assistance with the project and the two anonymous reviewers for their helpful comments.

REFERENCES

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

Berque, D. (2003). Boardtalker: Initial experiences and open problems in prototyping a talking digital whiteboard to assist visually impaired students. *Proceedings of the 2003 International Conference on Auditory Display (ICAD03)* (pp. 296–299), Boston, MA.

Berque, D., Bonebright, T., Kinnett, S., Nichols, N., & Peters, A. (2003). A case study in the design of software that uses auditory cues to help low vision students view notes on a blackboard. *Proceedings of the 9th International Conference on Auditory Display (ICAD03)* (p. 307), Boston, MA.

Betz, B. (2002). Prototyping the development of groupware to help low-vision students view an instructor's blackboard notes. *Proceedings of the World Conference on Educational Multimedia, Hypermedia, and Telecommunications* (pp. 148–153), Chesapeke, VA: AACE.

Blattner, M. M., Sumikawa, D. A., & Greenberg, R. M. (1989). Earcons and icons: Their structure and common design principles. *Human-Computer Interaction*, *4*, 11–44.

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

Brewster, S. A. (1998). Using nonspeech sounds to provide navigation cues. *ACM Transactions on Computer-Human Interaction*, *5*, 224–259.

Brewster, S. A., Raty, V. P., & Kortekangas, A. (1996). Enhancing scanning input with non-speech sounds. *Proceedings of the ACM ASSETS '96*, Vancouver, Canada.

Brewster, S. A., Wright, P. C., & Edwards, A. D. N. (1995). Parallel earcons: Reducing the length of audio messages. *International Journal of Human-Computer Studies*, *43*, 153–175.

Dimigen, G., Roy, A. W. N., Horn, J., & Swan, M. (2001). Integration of visually impaired students into mainstream education: Two case studies. *Journal of Visual Impairment & Blindness*, *95*, 161–164.

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

Edwards, A. D. N. (1989). Soundtrack: An auditory interface for blind users. *Human-Computer Interaction*, *4*, 45–66.

Edworthy, J. (1998). Does sound help us to work better with machines? A commentary on Rautenberg's paper 'About the importance of auditory alarms during the operation of a plant simulator'. *Interacting with Computers*, *10*, 401–409.

Edworthy, J., & Hellier, E. (2006). Complex nonverbal auditory signals and speech warnings. In M. S. Wogalter (Ed.), *Handbook of warnings*. Mahwah, NJ: Lawrence Erlbaum Associates.

Eimer, M. (2001). Crossmodal links in spatial attention between vision, audition, and touch: Evidence from event-related brain potentials. *Neuropsychologia*, *39*, 1292–1303.

Ericson, M. A., Brungart, D. S., & Simpson, B. D. (2003). Factors that influence intelligibility in multitalker speech displays. *International Journal of Aviation Psychology*, *14*, 313–334.

Fang, X., Xu, S., Brzezinski, J., & Chan, S. S. (2006). A study of the feasibility and effectiveness of dual-modal information presentations. *International Journal of Human-Computer Interaction*, *20*, 3–17.

Gerber, E. (2003). The benefits of and barriers to computer use for individuals who are visually impaired. *Journal of Visual Impairment & Blindness*, *97*, 536–550.

GRE: Practicing to take the general test (9th ed.) (1998). Princeton, NJ: Educational Testing Service.

DOI: 10.1002/acp

Griffith, D. (1990). Computer access for persons who are blind or visually impaired: Human factors issues. *Human Factors*, *32*, 467–475.

Hankinson, J., & Edwards, A. (1999). Designing earcons with musical grammars. *ACM SIGCAPH Newsletter*, *65*, 16–20.

Hereford, J., & Winn, W. (1994). Non-speech sound in human-computer interaction: A review and design guidelines. *Journal of Educational Computer Research*, *11*, 211–233.

Ho, C., & Spence, C. (2005). Assessing the effectiveness of various auditory cues in capturing a driver's visual attention. *Journal of Experimental Psychology: Applied*, *11*, 157–174.

Ho, C., & Spence, C. (2006). Verbal interface design: Do verbal directional cues automatically orient visual spatial attention? *Computers in Human Behavior*, *22*, 733–748.

Isaacs, E., Walendowski, A., & Ranganthan, D. (2002). Hubbub: A sound-enhanced mobile instant messenger that supports awareness and opportunistic interactions. *CHI*, *4*, 179–186.

Janata, P., & Childs, E. (2004). Marketbuzz: Sonification of real-time financial data. *Proceedings of the Tenth Meeting of the International Conference on Auditory Display (ICAD04)*, Sydney, Australia.

Kalloniatis, M., & Johnston, A. W. (1994). Visual environmental adaptation problems of partially sighted children. *Journal of Visual Impairment & Blindness*, *88*, 234–243.

Kramer, G. (1994). An introduction to auditory display. In G. Kramer (Ed.), *Auditory display: Sonification, audification, and auditory interfaces* (pp. 1–78). Reading, MA: Addison Wesley.

McDonald, J. J., Teder-Salejarvi, W. A., & Hillyard, S. A. (2000). Involuntary orienting to sound improves visual perception. *Nature*, *407*, 906–908.

Mondor, T. A., & Amirault, K. (1998). Effect of same- and different-modality spatial cues on auditory and visual target identification. *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 745–755.

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

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

Resnikoff, S., Pascolini, D., Etya'ale, D., Kocur, I., Pararajasegaram, R., Pokharel, G. P., et al. (2004). Global data on visual impairment in the year 2002. *Bulletin of the World Health Organization*, *82*, 844–851.

Richardson, D. C., Spivey, M. J., Barsalou, L. W., & McRae, K. (2003). Spatial representations activated during real-time comprehension of verbs. *Cognitive Science*, *27*, 767–780.

Rigas, D., & Alty, J. (2005). The rising pitch metaphor: An empirical study. *International Journal of Human-Computer Studies*, *62*, 1–20.

Schmitt, M., Postma, A., & De Haan, E. (2000). Interactions between exogenous auditory and visual spatial attention. *Quarterly Journal of Experimental Psychology*, *53*, 105–130.

Simpson, B. D., Brungart, D. S., Iyer, N., Gilkey, R. H., & Hamil, J. T. (2006). Detection and localization of speech in the presence of competing speech signals. *Proceedings of the 12th International Conference on Auditory Display (ICAD06)*, London, UK.

Spence, C., McDonald, J. J., & Driver, J. (2004). Exogenous spatial-cuing studies of human cross-modal attention and multisensory integration. In C. Spence, & J. Driver (Eds.), *Cross-modal space and cross-modal attention* (pp. 277–320). Oxford: Oxford University Press.

Stevens, R. D., Edwards, A. D. N., & Harling, P. A. (1997). Access to mathematics for visually disabled students through multimodal interaction. *Human Computer Interaction*, *12*, 47–92.

Tun, P. A., O'Kane, G., & Wingfield, A. (2002). Distraction by competing speech in young and older adult listeners. *Psychology and Aging*, *17*, 453–467.

Walker, B. N., & Ehrenstein, A. (2000). Pitch and pitch change interact in auditory displays. *Journal of Experimental Psychology: Applied*, *6*, 15–30.

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

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

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

DOI: 10.1002/acp