# Spatial audio in small screen device displays

Ashley Walker and Stephen Brewster

Department of Computing Science, University of Glasgow, Scotland G12 8QQ

http://www.dcs.gla.ac.uk/~stephen

email:ashley,stephen@dcs.gla.ac.uk

keywords: interface design, 3D audio, delay affordance, usability testing

Our work addresses problem of (visual) clutter in mobile device interfaces. The solution we propose involves the translation of techniques - from the graphical to the audio domain - for exploiting space in information representation. This article presents an illustrative example in the form of a spatialized *audio* progress bar. In usability tests, participants performed background monitoring tasks significantly more accurately using this spatialized audio (vs. a conventional visual) progress bar. Moreover, their performance in a simultaneously running, visually demanding foreground task was significantly improved in the eyes-free monitoring condition. These results have important implications for the design of multi-tasking interfaces for mobile devices.

# 1 Introduction

Less is more when it comes to mobiles. Devices must keep users in touch without stretching the lining of purse or pocket. In response to this, mobile device size is dropping in inverse proportion to power. Unfortunately, pixels must be shed with pounds, e.g., in 1999, Nokia's best seller (the 3110) featured a 84x64 pixel display (less than *1%* of the space on an average desktop monitor). How will a mobile phone or computer - with a screen small enough to crowd a single graphical button - display the multitude of new applications being developed for them?

To answer this question, it is important to note that much of the increased computing activity aboard mobiles will be associated with a restricted class of activity - namely, background task monitoring.

*"These smart [mobiles] will be just one part of a much bigger wireless business: telemetry. People have used wireless devices for years to monitor expensive pieces of machinery in isolated places. But now that wireless can deliver the Internet, such monitoring is about to get a huge boost. ... Wireless devices can be used to monitor anything from utility meters (thus getting rid of meter readers) to Coke machines (so they can be filled up before they run out). They can be used to signal whether a building is being broken into or pollution has reached a critical level. They can also be used to deliver instructions to other devices: lock yourself out of your car, for example, and you can use your mobile phone to tell the doors to open."* (*The Economist's* October 9, 1999 Telecommunications Survey)

The range background activity - including tasks delegated by the user (e.g., downloads, uploads, synchronizations), as well as activity targeting the user (e.g., email, voice-mail) - has been steadily increasing for years [1]. Examples of this activity include data sharing between mobile and desktop machines, between users in a variety of collaborative work activities (including work-station mediated video and audio connections); autonomous agents performing search tasks for users; distributed computing and borrowed CPU cycles, etc. Technology standards such as *Wireless Application Protocol* and *Bluetooth* are currently minimizing the performance gap between desktop machines and wireless mobiles by evolving methods to deliver wireless data. The latter targets itself at an estimated one billion wireless subscribers (by the year 2004) - a substation proportion of whom will have phones with multimedia capabilities including the ability to retrieve email, and push and pull information from the Internet [2]. How effective these mobile devices will be depends critically on whether users are given effective tools for monitoring and coordinating concurrently running (background) processes and activities. Therefore, the question addressed by this work is: What are the best ways to represent background activity to mobile users without interrupting their foreground task(s)?

We assert that a key strategy for preventing a (visual) information presentation bottleneck aboard mobiles involves the creation of interaction tools which exploit other sensori-motor modalities. In the case of background task monitoring, the auditory system - with its inherent ability to monitor multiple streams of simultaneously presented and dynamically changing data - is a good candidate modality for designing monitoring cues. We naturally use our ears to monitor our environment. Beyond spoken words, ears are tuned to a host of acoustic stimuli that orient us - the reassuring squeak of brakes as they catch; the changing pitch of the wine bottle as it empties; the rhythm, intensity and directionality of approaching traffic. Acoustic guidance is important for interacting with the world of digital information and events as well. Although most interfaces are relatively silent, we use what little sounds computers do make to infer much about their internal state and processes. The sound of a disk-drive spinning is a good example: too much and we are worried about other users over-taxing resources, but if it gets too quiet, the work-context of a networked machine is suspect.

There are several compelling reasons for using audio displays aboard mobile devices, in particular. Of primary importance here is the fact that audio display space is not wed to the disappearing resource of screen space. Moreover, an audio display space is potentially very large, e.g., a *360 degree* sphere surrounding the user's head. This audio bubble provides a natural display for the "personal bubble" - an idea coined by Nokia to describe a user's collection of their very own data that follows them around. A personal audio bubble can be densely packed with information by following the windowing example of partitioning space into many (perceptual) channels. Multiple audio channels, or streams, can be simultaneously listened to via exploitation of cocktail party effect [3]. If this were not enough, digital audio technology for the consumer market (including technology for virtually spatializing sounds) is dramatically increasing in power, while prices are plummeting.

We navigate the largely uncharted audio design space using a GUI design plan for exploiting space: a space-time mapping. Space-time mapping involves displaying temporally extended information along a (more perceptually salient) spatial axis. This paper discusses a prototype interface tool that allows users to monitor a particular type of background activity (i.e., file transfers); however, it is intended as a test bed for exploring general issues associated with background task monitoring. Below we review a selection of literature on the topic of audio task monitoring.

## 1.1 Prior Audio

The current work employs insights from three pockets of literature. Section (**1.1.1**) reviews audio design for background task - looking at examples of systems which provide users with "serendipitous" audio cues for monitoring of collaborative work activities. In Section (**1.1.2**), we look at the use of spatial sound in information displays. Finally, Section (**1.1.3**) compares several existing audio progress bars.

### 1.1.1 Serendipitous audio cues

One of the first examples of serendipitous cuing was used by the *ARKOLA* system. *ARKOLA* employed an "audio ecology" of process sounds by which workers in a simulated bottling plant coordinated activities across their different stations [4]. Cues consisted of alarms and iconic sounds indicating component state and rate. These sounds worked in combination to become patterns. For example, processing machines - which only made sounds when they produced output - helped users to adjust the rate of other machines. Participants explained the utility of this sound via comments such as: "It makes me nervous when the capping machine isn't being rhythmic".

*Ravenscroft Audio Video Environment (RAVE)* equips staff with audio-video channels for establishing teleconferencing connections and monitor activity across the dozens of offices and open plan areas of EuroPARC [5]. *RAVE* is a media space which augments cooperative work in the physical workspace. As a safeguard over user's privacy, *RAVE* uses audio notification to unobtrusively inform a user when the camera in her or his room has become active. For example, a "door opening" sound indicates a user is glancing at another through a camera, "footstep" sounds indicate that a user is monitoring another as part of a camera sweep, etc.

*AudioAura* goes even further to connect a person's activities in the physical world with information culled from the virtual world [6]. As users wearing active badges move around an office equipped with a low-cost network of IR sensors, their movements are combined with information in a database (containing calendar information, email queues, etc.) and sent cues to listeners over wireless headphones. *AudioAura* uses combinations of musical and metaphorical sounds (e.g., sea-scape sounds) whose characteristics (e.g., intensity) are scaled to remotely alert people about the activities of collaborators and collaborative resources as they move into situations in which this information can be employed.

Other examples of serendipitous cuing include *Tangible Bits* - which surrounds people in their office with a wealth of background cues using light, sound and touch [7]. *Nomadic Radio* is a wearable computing platform for monitoring messages via a range of acoustic peripheral awareness cues [8]. Notification is adaptive and context sensitive; with messages being presented as more or less obtrusive depending upon such factors as the importance inferred from context filtering, whether the user is actively engaged in a conversation, and her responses to recent messages.

### 1.1.2 Spatial sound to enhance information search and recall

Several systems use virtual 3D audio to facilitate serendipitous monitoring. Most notably, displays for presenting air or tele-conferencing traffic have exploited an underlying spatial layout to enhance virtual presence. For example, Wenzel and colleagues' cockpit displays represented the bearing of targets as an appropriately spatialized sound so as to facilitate rapid acquisition by other sense modalities (e.g., vision) and inform cognitive tasks such as path (re)planning [9]. In a different application area, AT&T Bell Laboratory's *Virtual Meeting Room* [10] spatialized sampled audio in the form of speech and system sounds (such as keyboard clicks) to provide information about connectivity, presence, focus, and the activity of participants. The *MAW (multi-dimensional audio windows)* teleconferencing tool pushes this virtual presence further [11]. By presenting the user with an exocentric graphical configuration panel in which icons representing speakers can be dragged and rotated such that the spatial configuration of the audio channels reflects the logical (as opposed to purely spatial) organization of the interaction. For example, users can leave pairs of ears scattered about a virtual conference room(s) to facilitate selective listening.

A relatively common use of spatial audio is in displays for sound archive navigation. With spatial audio, the user can browse several clips simultaneously because they are presented from different positions (as in image archive navigation). Spatial audio has also been used to represent information of another sort. For example, *Dynamic Soundscape* - an audio news browsing tool - re-maps temporally extended data (e.g., a news broadcast) onto a spatially extended audio display [12]. Here a virtual news-reader orbits a user's head such that different topics are played at different spatial positions. If a user wants to review any segment of the broadcast, s/he need not rely on temporal recall of its sequence on the audio stream but, rather, simply the position where the topic of interest was heard. *Dynamic Soundscape* was inspired by *AudioStreamer* [13] - another audio browsing tool that exploits spatial memory for navigation through temporally extended data. *AudioStreamer* presents three audio data streams at three fixed locations in a virtual audio space and, like *Dynamic Soundscape*, facilitates selective listening to different streams in response to changes in a user's listening behaviour. Specifically, head-tracking equipment is used to detect when a user leans toward a particular source (in order to hear it better) and responds by amplifying that source relative to other simultaneously playing streams. *Nomadic Radio* uses spatial sound in messaging displays to facilitate differentiation of messages by laying them out along a spatial axis and playing them back ordered by arrival time. It delivers spatial audio to a mobile listener via a light, shoulder mounted speaker array.

### 1.1.3 Audio progress bars

In much audio interfacing work (e.g., *SonicFinder* [14] and the work on earcons of Brewster and colleagues [15,16]), audio enhances visual cues to overcome usability problems with direct manipulation interface components. Exceptions to this include Albers and Bergman's enhancements to the *Mosaic Web Browser's* file transfer facilities [17], Gaver's *copy* command [14] and Crease and Brewster's *audio progress bar* [18]. In the case of the former, users were given iconic audio cues describing file type and size when they moved a mouse over a link. Once a download was initiated, pops and clicks played to indicate data transfer and a breaking glass sound was used to indicate a transfer error. Gaver's *audio copy* took this further by giving users an ability to monitor progress as the change in pitch of a pouring sound (which communicated task progress by the transfer of fluids between two vessels). Crease and Brewster's *audio progress bar* used a non-spatialized earcon to provide a full set of delay affordances for file transfer monitoring. It represented progress via a pair of differentially pitched tones played in rapid succession: one pitch was fixed (to mark the endpoint) and the other varied (with its pitch scaled according to the amount of download remaining). A second motif (with a distinct timbre) was overlayed on the tones to indicate rate as a function of the number of notes played. In experiments where users performed a foreground text transcription task while trying to maximize the amount of sequentially downloaded files, the sonically enhanced progress bar was found to facilitate more rapid response to the end of a download.

## 1.2 Summary

With Crease and Brewster's audio progress bar, questions remain about how much information users extracted from the temporal sequence of sound. It is possible that they were doing the task by simply listening for the completion sound (a rapid succession of tone triplets) rather than monitoring ongoing progress. It is also unclear whether multiple simultaneously playing downloads could be monitored via this earcon design.

These questions motivated the design of the spatialized auditory progress bar described in the next section. This progress bar also derives aspects of its design from the other literature reported here - i.e., serendipitous audio cuing and from the space-time mappings achievable with spatial audio. The strength of this work comes from the integration of these ideas, as well as rigorous testing of them. The majority of psychological research has not been concerned with determining which kind of information is best presented in the different sensory modalities and, consequentially, there are few studies which compare the effectiveness of audio to visual information in facilitating task performance. To remedy this, we present (in Section (**3**)) a usability study comparing our spatialized audio progress bar with a conventional graphical progress bar. The results of that test are given in Section (**4**) and discussed in Section (**5**). Finally, in Section (**6**), we summarize the conclusions of this study.

# 2 Materials: A spatialized audio progress bar

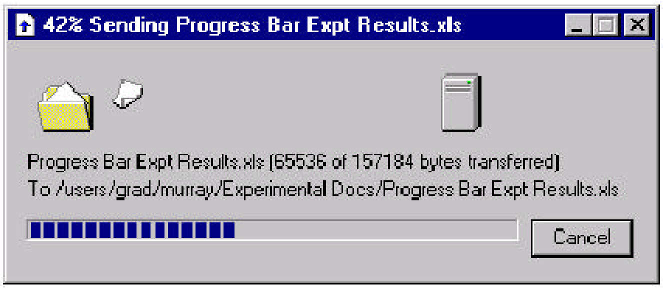
Design of a monitoring cue poses a number of questions. How should users be notified of events? What sensory modalities are available for rendering the different notification modalities. Which modalities are most and least intrusive? Which are most and least informative? In this section, we provide an answer these questions in the form of a design solution.

## 2.1 Motivation

A useful categorization of notification modalities has been suggested by Buxton [19]. This includes (i) alarms and warnings, (ii) encoded messages (communicating quantitative data), and (iii) monitoring and status indicators. These categories are differentiated on the basis of priority, continuity, and frequency. Monitoring/status indicators - the subject of this work - are typically low priority messages characterized by continuously and frequently changing values. It is often the qualitative (rather than the quantitative) character of these messages (e.g., trends) that are of most interest in monitoring tasks.

These aspects of the monitoring and status notification modality constrain cue design and, in turn, the choice of sensory modality. As low priority messages, monitoring and status cues should not force interruption of other (higher priority) tasks. Furthermore, their continuous nature requires that they be communicated by continuous cues (which fade from user's conscious attention). Finally, because they are frequently changing, monitoring and status cues are best represented with a dynamically changing iconic description. In this regard, text (or text-to-speech) messages are typically not a solution as unlikely to keep pace and, consequentially, will lose information or provoke annoyance.

As an example, let's look at how well a conventional visual progress bar adheres to these recommendations. At the heart of most visual progress bars is the 2D iconic bar that fills or moves with task progress. Some visual progress bars, for example, the *Window's NT* progress bar shown in Figure (**1**), back this up with a redundant encoded message containing transfer rates, file size remaining, etc. Although important in some applications, the values of these variables are typically changing too fast to be read without stealing attention from the user's foreground task. Similarly, some progress bars signal an endpoint (stall or successful completion) with a visible (or audible) alarm, which, again, interrupts foreground tasks. In general, however, a more fundamental problem with visual the progress bar is that it is visible and, therefore, steal the eyes and attention away from foreground tasks. Moreover, even the simple iconic element of the visual progress bar is useless when it becomes buried in a cluttered or small screen display.



42% Sending Progress Bar Expt Results.xls

Progress Bar Expt Results.xls (65536 of 157184 bytes transferred)
To /users/grad/murray/Experimental Docs/Progress Bar Expt Results.xls

**Figure 1: Visual progress bar use in Microsoft's *Windows NT* operating system**

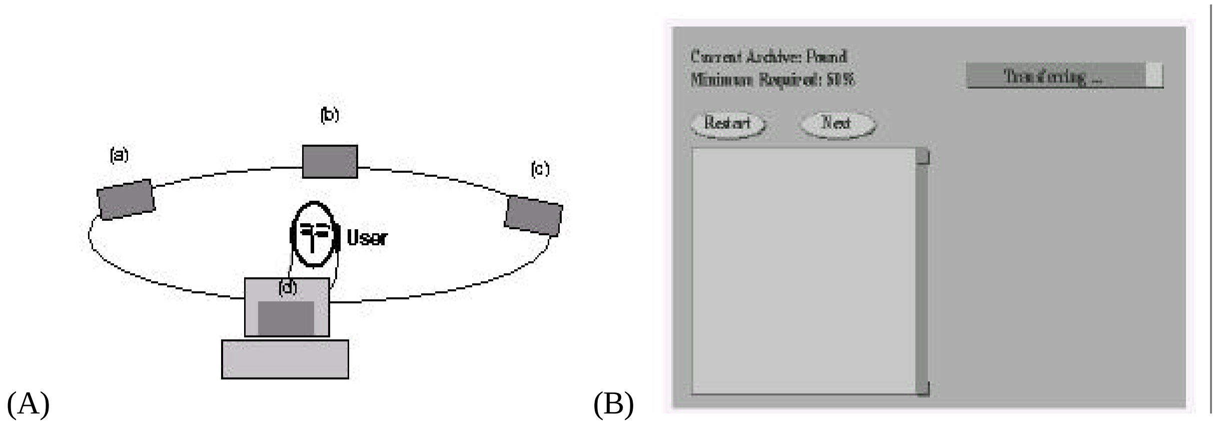
The rest of this section describes a progress bar, i.e., our spatialized audio progress bar, for monitoring background tasks in the (attentional) background. It is deliberately designed as a general tool for monitoring any number of temporally extended background tasks. As such, it does not provide a full exploration of delay affordance cues for the particular task of transferring files [20] and, instead, focuses on communicating progress (percent transferred and rate of progress), as well endpoints (completions and interruptions). It provides these cues unobtrusively - making users aware of these events without interrupting them and forcing attention away from a foreground task.

## 2.2 Design

*An axis is perhaps the first human manifestation; it is the means of every human act. The toddling child moves along an axis, the man striving in the tempest of life traces for himself an axis. The axis is the regulator of architecture.* (Le Corbusier 'Vers une architecture')

The simplicity and clarity of the visual progress bar icon is derived from its use of two basic components: (i) a progress indicator that moves along (ii) a fixed reference axis. Similarly, we built a spatialized audio progress bar using only two spatialized non-speech sounds and allowed the spatial position and presentation rates of these two sounds to communicate the delay affordances of progress, rate and endpoint.

- **Progress.** The first sound (or 'lub' component of the of the progress 'heartbeat') provides the reference. It is played from a fixed target position located in front of the user. Following 'lub' in rapid succession, the 'dub' component is played from a spatial position that communicates task completion by its angular position within a circular orbit centered on the user's head. (See Figure (**2**) (A).)
- **Rate.** The overall rate of progress is perceived as the angular speed of the orbiting sound. The absolute (or size-independent rate) transfer rate is explicitly encoded as the time between the presentation of the two sounds. (This delay is determined via a transfer function running asymptotically to *500*ms with increasing rate and to infinity - i.e., it disappears - in the extreme that the download is stalled).
- **Endpoints.** When a task ends successfully, two identical - and therefore distinct - sounds are played from in front of the user. By contrast to this uniquely symmetrical heartbeat, a stall is heard as a failed heartbeat (i.e., a lone lub sound followed by an infinite inter-component delay).



Current Archive: Pound
Minimum Required: 50%
Transferring ...

**Figure 2: (A) Auditory display space.** File transfers in various states of completion: (a) *25%*, (b) *50%*, (c) *75%*, and (d) *100%*. **(B) Experimental interface.** The interface consisted of the following components: (i) Upper left: a control panel containing information about the current download and buttons for restarting an interrupted download (if the minimum download requirement had not been met at the time of the interrupt) or continuing with the next download. (ii) Lower left: a text entry area into which transcriptions were made. (iii) Upper right: a visual progress bar. In the case of the audio condition (Condition 2), no visual progress bar was presented. Instead, the upper right portion of the interface is blank and the user monitored the download via the spatialized audio progress bar described earlier.

The non-speech sounds used by the audio progress bar described above were spatialized via convolution with empirically-based head-related impulse response (HRIR) models describing the acoustic filtering effects of the pinna and head of participants measured by Brown and colleagues

[21]. To facilitate better externalization of the sound, a simple room model - introducing one delayed and attenuated reverberation - was added to the spatialized sound as a final processing stage. Brown and colleagues report measuring *30 degree* angular resolution, using a head tracker to dynamically re-spatialize sound based on listeners' head motions.

In the experiment reported here, we maximized listeners' static localization discrimination by restricting movement of the sound cue to within the plane containing the user's ears. Within this plane, different motions were tested: movement across frontal sound field in one direction (like the visual progress bar), movement back and forth across the frontal sound field, or an orbital motion. A circular orbit was chosen because it provides a long continuous 2D motion, and the constancy of direction (i.e., a constant clockwise orbit) helps listeners discriminate a sound in the front from one in the back (sounds in the front move right, and *vice versa*). Using the circular orbit, the front-back reversal errors reported during early pilot studies (errors typical of hearing sound spatialized through someone else's ears (i.e., HRIRs) [22]) were virtually eliminated.

In the usability test described in Section (**3**), Java's (*22* kHz) *AudioClip* facilities were used for play-back - leaving *0-11* kHz for the cue. Within this range, short bursts of uniform energy were found to be the most effective non-metaphorical sound with significant energy outside of the speech spectrum. The same *500* ms noise burst was used for the 'lub' (fixed position) and 'dub' (spatialized) component of the progress heartbeat as described above. In an effort to minimize duty cycle without losing acoustic continuity (i.e., streaming) effects, the heartbeat played every *2* seconds. (In future designs the play-back rate could be scaled to the estimated length of the task.)

The next section describes a usability experiment in which the effectiveness of this auditory progress bar was assessed.

# 3 Methods

**Participants.** Sixteen people from University of Glasgow Computing Science Department served as participants. These included *11* men and *5* women, encompassing a range of ages and stages: undergraduates (2), post-graduates (3), research associates (9) and lecturers (2). Participants were not compensated.

**Experimental design.** The experiment was a within-groups design with modality of cue (i.e., audio or visual) as the single independent variable. Each of the sixteen participants performed the task described below using a conventional visual progress bar (Condition 1) and the spatialized audio progress bar (Condition 2). The order of conditions were counter-balanced (with odd number participants performing the visual task first and *vice-versa*). Dependant variables included monitoring performance (accuracy and response time); efficiency in a simultaneously conducted foreground task; and several subjective workload measures.

**Experimental scenario.** Users were set the multi-tasking interfacing scenario of creating a poetry archive on the local file server. This involved transferring files from *20* remote poetry archives, while transcribing contemporary poetry from hard-copies. Users were told to work as fast as possible. To force users to continually monitor the progress of the file transfers (as opposed to responding simply to endpoints), file transfers were periodically interrupted - at which time users were then given the option of restarting the transfer or skipping to the next transfer. Users decided this based on whether the percentage of data transferred before the interruption exceeded the minimum percentage of data required from the present archive. (Workload rating and informal feedback from pilot studies revealed that giving users the ability to stop the downloads themselves foregrounded the download task.)

**Design architecture.** The experiment was run on a *Dell* Pentium II machine (*128* Mb RAM, *400* MHz) with a *21* inch monitor. The interface consumed the whole screen and consisted of the components shown in Figure (**2**) (B).

**Analysis.** As a measure of background performance, response times (i.e., the time it took for a user to click 'restart' or 'next' following a download endpoint) as well as monitoring decisions were collected. The latter were classified as *hits*, *misses*, *false alarms* and *correct rejections*, according to the Table 1. The latter were used to determine the overall percentages of correct responses as well to determine perceptual sensitivity independently of response bias.

| | Decision: ‘Next’ | Decision: ‘Restart’ |
|---|---|---|
| Download interrupted *after* minimum transfer | Hit | Miss |
| Download interrupted *before* minimum transfer | False alarm | Correct rejection |

**Table 1: SDT Definitions**

Foreground task performance was calculated using the total number of words typed and total run-times. Participants were not penalized for spelling or text formatting errors.

Subjective workload assessments - on a modified set of NASA TLX scales [23] - were collected after each condition. The workload ratings included *mental* and *physical demand*, *time pressure*, *effort expended*, *frustration*, *annoyance*, *performance* (i.e., the participant's subjective experience of task proficiency) and *preference* (for either condition).

**Hypotheses.** Using these metrics, we investigated the following hypotheses. *Hypothesis Ia:* A spatialized audio progress bar can be monitored more accurately. *Hypothesis Ib:* A spatialized audio progress bar will yield faster endpoint detection times. *Hypothesis II:* The spatialized audio progress bar facilitates a visually demanding foreground task. *Hypothesis III:* The spatialized audio progress bar requires less workload.

# 4 Results

## 4.1 Hypothesis Ia - Monitoring accuracy

Monitoring accuracy was significantly ($T_{15} = 2.06$, $p = 2.92e\text{-}2$) affected by modality, with users performing better in the audio than the visual condition. The mean percentage of correct decisions (both hits and correct rejections) was *89.77%* and *86.60%* in the audio vs. visual conditions, respectively. Across the *20* trials, this corresponded to an average of *2.03* decision errors in the audio condition and *3.31* in the visual condition. A comparison of errors across participants is shown in Figure (**3**) (A).

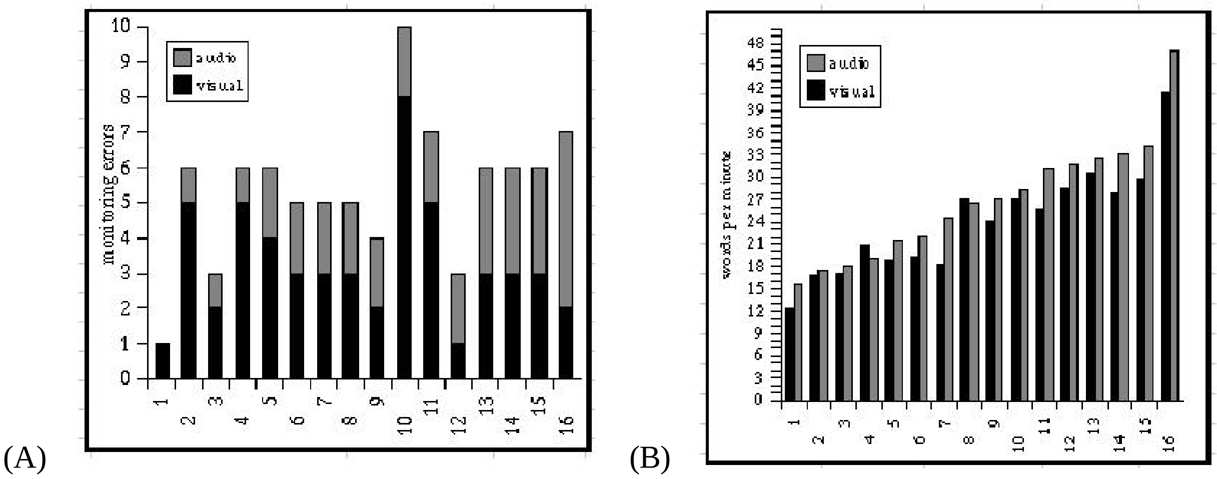


**Figure 3 (A) Performance in background monitoring task.** Accuracy in monitoring task shown as the number of decision errors in the audio vs. visual conditions. Participants are ordered by audio performance along the horizontal axis. **(B) Performance in foreground task.** Typing transcription efficiency measured in words per minute. Participants are ordered by audio performance along the horizontal axis.

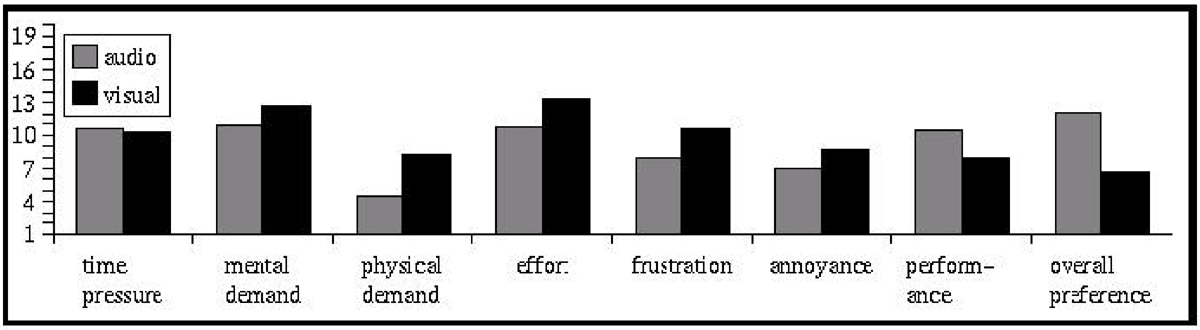


**Figure 4: Performance in foreground task.** Typing transcription efficiency measured in words per minute.

The accuracy results reported above wrap two effects into one - namely, perceptual sensitivity and response bias. We employed Signal Detection Theoretic (SDT) analysis to tease them apart. SDT provides a useful tool with which the experimenter can independently measure perceptual sensitivity (i.e., how well the observer is able to make a correct detection/judgment) from non-sensory effects (i.e., prior learning of a particular probability of a signal and expected gains and losses from decisions). Although SDT was originally used to determine threshold levels in sensory experiments, there are many other areas - such as diagnostics (radiology, weather forecasting, materials testing), pain assessment, recognition memory - to which it has been applied successfully [24]. STD revealed that significant differences in sensitivity, as opposed to response bias, account for the improved monitoring accuracy in the audio condition.

Sensitivity, *A'* **(1)**, is classically measured as the relative frequency of hits to false alarms representing each participants subjective level of stimulation. In this study, it varied significantly with modality (Wilcoxon $T_{11}(nonzero) = 9, p < 0.05$), with participants being more sensitive to audio ($A' = 0.95$) than visual ($A' = 0.93$) cues.

Response bias, *B''*, is affected by non-sensory factors such as habituation and anticipation that predispose a participant to report a particular judgment **(2)**. In this study, response bias, unlike sensitivity, did not vary significantly with modality (Wilcoxon $T_7(nonzero) = 10$ and $p > 0.05$). For the audio and visual conditions, the non-parametric response bias criteria were: $B''= -0.75$ and $B''= -0.65$, respectively. As signal probabilities were equal for first responses, these scores indicate that participants tended to assume that an interrupted transfer had progressed sufficiently and, under the pressure of time, skipped to the next transfer instead of restarting the present one.

### 4.2 Hypothesis Ib - Monitoring response times

Response times were not significantly affected by modality ($T_{15} = 0.50, p_{15} = 3.13e1$). The mean response times were *4.89* seconds (audio) and *5.07* seconds (visual). On the other hand, variance in the response times (*0.83* seconds (audio) and *3.43* seconds (visual)) recorded for the two conditions did vary significantly ($F = 4.12, p = 3.63e\text{-}3$).

### 4.3 Hypothesis II - Foreground task performance

The spatialized audio progress bar facilitated significantly better (i.e., faster) performance in a visually demanding foreground task ($T_{15} = 4.91, p=9.33e\text{-}5$). Users typed more words per minute in the audio condition as compared to the visual condition - with mean speeds of *26.89* vs. *24.09*, respectively. Figure (**3**) (B) displays typing efficiency by condition and participant.

### 4.4 Hypothesis III - Workload

The spatialized audio progress bar required less subjective workload in six of the eight ratings measured. Figure (**4**) displays - for each condition - the mean workload ratings averaged across all participants. Users reported that the audio condition required significantly less *effort* ($T_{15} = 2.73, p = 7.68e\text{-}3$) and *physically demand* ($T_{15} = 3.07, p = 3.85e\text{-}3$), and resulted in significantly lower levels of *frustration* ($T_{15} = 3.39, p=2.03e\text{-}3$) and *annoyance* ($T_{15} = 2.93, p = 5.13e\text{-}3$). Users reported a significantly higher sense of their own *performance* ($T_{15} = 2.83, p = 6.26e\text{-}3$) in the audio condition and a significantly higher overall *preference* ($T_{15} = 5.60, p = 2.51e\text{-}5$) for the audio condition. Users reported similar *mental demand* and *time pressure* in both conditions.

## 5 Discussion

### 5.1 Hypothesis Ia - Monitoring accuracy

As expected, users immersed in a visually demanding foreground task performed a background monitoring task more accurately using audio (vs. visual) cues. On its own, this result has several possible interpretations. Users may have found the audio cue easier to attend to. Alternatively, because participants had prior experience using the visual progress bar in a variety of (generally non-critical) tasks, they may have taken more care with the novel audio progress bar. In order to identify the factors affecting the use of cues in the two modalities, signal detection theory was used to break down the monitoring decision data so as to distinguish between the perceptual sensitivities of each modality and the response biases induced by the use of a particular modality.

The SDT analysis revealed that it was increased sensitivity, as opposed to bias artifacts (e.g., a more serious approach to the auditory progress bar due to familiarity with visual progress bars), which accounts for increased performance in the audio monitoring condition. Moreover, we believe that the benefit of the cue in Condition 2 (audio) is due to it being presented in the alternative audio modality. In support of this, we note that participants ability to resolve progress from both the audio and visual progress bars was more than adequate **(3)**. Moreover, STD analysis on the full data set shows that when adding data from second and subsequent monitoring responses - i.e., data taken from higher workload scenarios - the sensitivity differences increase further. Under these circumstances, visual resources may have been fully committed.

The response bias values measured whether participants performed more liberally or conservatively in either modality. As stated above, we feared that participants might be more cautious in the audio situation because it was new. On the contrary, we found no significant difference between response bias in the audio and visual conditions. Since the probability of a sufficient transfer at interruption was .5 for first interruptions, the conservative (i.e., negative) *B"* values suggest that users came to both the audio and visual conditions with the same expectations - namely, that there was a high probability of successful file transfer (or that, given the time pressure of the experiment, the benefit of skipping to the next download was much higher than the cost of restarting the present one). Again, this is a positive result in that it suggests that users treated the audio progress bar as a real one.

## 5.2 Hypothesis Ib - Monitoring response times

We predicted that users would respond significantly faster to the completion or interruption of the audio (vs. visual) progress bar, if, indeed, the visual foreground task made visual transfer monitoring difficult. This was shown in previous work using a strong endpoint sound and endpoint oriented task [18]. However, in the present study, response times showed no significant difference. The significant differences in the variances on the audio response times - being nearly four times lower than the visual (*0.89* seconds vs. *3.64* seconds) - suggest that users may have responded based on perception in the audio case and a combination of perception and memory in the visual case.

## 5.3 Hypothesis II - Foreground facilitation

Task monitoring with a spatialized audio progress bar significantly facilitates a visually demanding foreground (typing) task. It is unusual to show this sort of facilitation effect. Where usability tests of audio widgets are run, they often simply show the audio cues themselves to be effective or appealing without affecting performance of concurrently running tasks.

## 5.4 Hypothesis III - Workload

The low workload results in the audio condition - combined with the similar response biases in the audio and visual conditions - confirm that participants did not take a different (e.g., more serious) approach to the novel audio condition.

## 5.5 Cue design

Feedback from informal interview and workload ratings indicated that the overall effect of the sound cue used in the audio progress bar was simple and effective. Although participants gave the audio progress bar fairly low annoyance ratings, many of them informally confided that they found the use of noise as a sound cue to be "ugly". (We expected this, as other studies have reported that users found humming and pink noise to be annoying [1].) To its defense, the cue used by the audio progress bar avoids high frequencies, abrupt onsets, in-harmonic timbres, and dramatic changes in frequency and amplitude that contribute to a cue's perceived urgency (which may be a correlate of their propensity to distract or annoy) [25].

The use of noise in this version of the audio progress bar was largely constrained by technical limitations (i.e., the play-back rate of our system and the lack of a head tracker). Having overcome these limitations, we are using feedback from participants to design a new audio progress bar cue which will evoke a more appealing sound-scape and lend itself more easily to use in multiple (simultaneous) task (e.g., downloads) monitoring. In the new design, we minimize demands on perceptual resources by merging multiple download cues into one steam via the creation of an audio cursor which sweeps around the user's head every 2 seconds - playing a "dub" sound from each spatial position representing the state of a download, and a soft, reference ("lub") sound from the others. (This new heart-beat employs everyday sounds - e.g., waves, boys, fog-horns, seagulls - which have been shown to have aesthetic appeal [1,6].) We are currently piloting a study comparing how performance scales with the number of monitoring tasks.

Spatialized sound cues used in the work presented here were delivered over headphones. They might also be delivered via wearable speaker arrays that do not occlude the ears - e.g., like that employed by *Nomadic Radio.* We do not believe; however, the use of headphones to be a hindrance to the acceptance of this technology. Generations of personal stereo users have secured their social acceptance. Moreover, persons most intimately involved in sound intensive computer use - currently game players - chose headphones as a mode of delivery: *"One thing that you have to keep in mind is the presentation method, though. There are a TON of computer systems out there that are being sold 'out of the box' with multi speakers and sub-woofers, but in my experience the majority of people using them actually use headphones."* (personal communication, game industry reviewer, Steve Lieb)

# 6 Conclusions

The mobile computer is becoming the hub of activity taking place both locally and over networks. The increasing rate at which computers access and generate information makes it difficult for users to absorb the flood of data being presented to their small screens. One way to avoid the resulting break-down in efficiency is to distribute the present information among sensory modalities. Although psychological research provides little insight into which types of information are best presented to different sensory modalities, broad characteristics of the different modalities can guide multi-modal interface design and testing. For instance, the auditory system - with its natural ability to attend to multiple streams of continuously changing information - may be employed to relieve the eyes in interfacing tasks involving the monitoring of temporally-extended stream(s) of information.

The aim of the experiment reported here was to determine whether delay affordances which are typically presented visually can be effectively presented in the auditory modality. The present work was motivated by the expectation that a spatialized audio progress bar would afford the same advantages over visual progress bars which have been demonstrated for existing (non-spatialized) audio progress bars [14,17,18]. Moreover, this study showed that the space-time trade-off exploited by a *spatialized* audio progress bar yields additional usability improvements in the form of (i) improved accuracy in a progress monitoring task and (ii) improved efficiency in a simultaneously conducted foreground task.

Spatialization may also prove to be a key ingredient in multi-tasking interface design, as spatialization facilitates segregation of multiple simultaneously playing sound streams [26]. The current article lays the ground work for our future study of the use of multiple audio progress bars.

---

**(1)** Here sensitivity was measured as *A'*, the average of the maximum and minimum probability associated with a particular outcome [27]. We applied this non-parametric measure of sensitivity to a (non-normally distributed) sub-set of the data containing responses to the only first interruption of each file transfer. In transfers proceeding past their first interruption, participants' expectations of further interruptions were artificially elevated due to assumptions about real world causes such as persistent networking problems. For this reason, we eliminated secondary (and later) responses from our analysis. Furthermore, past the first response, signal probabilities vary with participant behaviour, thereby complicating the analysis.;
**(2)** *B''* ranges from -1 (representing a tendency to respond conservatively) to +1 (representing a tendency to report fewer false alarms).;
**(3)** Before the usability test discussed in Section~(**3**), participants were given 5 minutes to familiarize themselves with the audio progress bar's sound cue as it orbited at different rates. Users were asked to identify the quadrant in which an orbiting sound cue stopped. Scores were near perfect for all participants.;

## 7 References

[1] J. Cohen. *Auditory Display: Sonification, audification and auditory interfaces,* chapter Monitoring background activities, pages 499--531. Addison-Wesley, 1994.

[2] Wireless application protocol --- white paper. Wireless Internet Today, October 1999.

[3] B. Arons. A review of the cocktail party effect. *J. Am. Voice I/O Soc.*, 12:35--50, 1992.

[4] W.W. Gaver. Technology affordances. In *Proc. CHI'91*, pages 79--84. ACM Press Addison-Wesley, 1991.

[5] W. Gaver, T. Moran, A MacLean, L Lo:vstrand, P. Dourish, K. Carter, and W. Buxton. Realizing a video environment: Europarc's rave system. In *Proc. CHI'92*, pages 27--35, 1992.

[6] E.D. Mynatt, M. Back, R. Want, M. Baer, and J.B. Ellis. Designing audio aura. In *Proc. CHI'98*, pages 566--573. ACM Press Addison-Wesley, 1998.

[7] H. Ishii and B. Ullmer. Tangible bits: Towards seamless interfaces between people, bits and atoms. In *Proc. CHI'97*. ACM Press Addison-Wesley, 1997.

[8] N. Sawhney and C. Schmandt. Nomadic radio: Scalable and contextual notification for wearable messaging. In *Proc. CHI'99*, pages 96--103. ACM Press, Addison-Wesley, 1999.

[9] E.M. Wenzel. Localization in virtual acoustic displays. *Presence*, 1:80--107, 1992.

[10] D.D. Seligmann, R.T. Mercuri, and J.T. Edmark. Providing assurances in a multimedia interactive environment. In *Proc. CHI'95*. ACM Press Addison-Wesley, 1995.

[11] M. Cohen and F.L. Ludwig. Multidimensional audio window management. *Int. J. Man-Machine Studies*, 34, 1991.

[12] M. Kobayashi and C. Schmandt. Dynamic soundscape: Mapping time to space for audio browsing. In *Proc. CHI'97*, pages 194--201. ACM Press Addison-Wesley, 1997.

[13] C. Schmandt and A. Mullins. AudioStreamer: Exploiting simultaneity for listening. In *Proc. CHI'95*. ACM Press Addison-Wesley, 1995.

[14] W.W. Gaver. The sonic finder: An auditory interface that uses auditory icons. *Human Computer Interaction*, 4:67--94, 1989.

[15] S.A. Brewster. The design of sonically-enhanced widgets. *Interacting with Computers,* 11(2):211--235, 1998.

[16] S.A. Brewster and P.G. Cryer. Maximising screen-space on mobile computing devices. In *Proc. CHI'99*, pages 224--225. ACM Press, Addison-Wesley, 1999.

[17] M.C. Albers and E. Bregman. The audible web: Auditory enhancements for Mosaic. In *CHI Conference Companion*, pages 318--319. ACM Press Addison-Wesley, 1995.

[18] M. Crease and S.A. Brewster. Making progress with sounds --- the design and evaluation of an audio progress bar. In *Proc. ICAD*, 1998.

[19] W. Buxton. Introduction to this special issue on nonspeech audio. *HCI*, 4:1--9, 1989.

[20] A.P. Conn. Time affordances: The time factor in diagnostic usability heuristics. In *Proc. CHI'95*, pages 186--193. ACM Press Addison-Wesley, 1995.

[21] C.P. Brown. Modeling the elevation characteristics of the head-related impulse response. Technical Report 13, San Jose State Univ., 1996.

[22] E.M. Wenzel, M. Arruda, D.J. Kistler, and F.L. Wightman. Localization using nonindividualized head-related transfer functions. *J. Acoust. Soc. Am.*, 94:111--123, 1993.

[23] S. Hart and L. Staveland. Development of nasa-tlx (task load index): Results of empirical and theoretical research. In P. Hancock and N. Meshkati, editors, *Human Mental Workload*, pages 139--183. North Holland B.V., 1988.

[24] Levine and Parkinson. *Experimental Methods in Psychology*. Erlbaum, 1994.

[25] R.D. Paterson, J. Edworthy, M.J. Shailer, M.C. Lower, and P.D. Wheeler. Alarm sounds for medical equipment in intensive care areas and operating theatres. Technical Report AC598, University of Aouthhampton, Auditory Communication and Hearing Unit, 1986.

[26] E.C. Cherry. Some experiments on the recognition of speech. *J. Acoust. Soc. Am.*, 25:975--979, 1953.

[27] J.B. Grier. Nonparametric indexes for sensitivity and bias. *Psychological Bulletin*, pages 339--346, 1971.