Eyes-Free Multitasking: The Effect of Cognitive Load on Mobile Spatial Audio Interfaces

Yolanda Vazquez-Alvarez and Stephen A. Brewster
Glasgow Interactive Systems Group, School of Computing Science
University of Glasgow, Glasgow, G12 8QQ, UK
{yolanda, stephen}@dcs.gla.ac.uk   www.gaime-project.org

ABSTRACT

As mobile devices increase in functionality, users perform more tasks when on the move. Spatial audio interfaces offer a solution for eyes-free interaction. However, such interfaces face a number of challenges when supporting multiple and simultaneous tasks, namely: 1) interference amongst multiple audio streams, and 2) the constraints of cognitive load. We present a comparative study of spatial audio techniques evaluated in a divided- and selective-attention task. A podcast was used for high cognitive load (divided-attention) and classical music for low cognitive load (selective-attention), while interacting with an audio menu. Results showed that spatial audio techniques were preferred when cognitive load was kept low, while a baseline technique using an interruptible single audio stream was significantly less preferred. Conversely, when cognitive load was increased the preferences reversed. Thus, given an appropriate task structure, spatial techniques offer a means of designing effective audio interfaces to support eyes-free mobile multitasking.

Author Keywords

Multitasking, audio interfaces, eyes-free interaction, cognitive load, divided-attention task, selective-attention task.

ACM Classification Keywords

H.5.2. User Interfaces: Evaluation, interaction styles.

General Terms

Design, Experimentation, Human Factors.

INTRODUCTION

How much multitasking are mobile users exposed to in their daily lives? Imagine a typical morning for David as he commutes to work: *David is travelling on the New York City subway system on his way to work and listening to music on his phone, as he always does. While David is in transit, his boss Stephen tries to call him but, as there is no underground cell phone coverage, he leaves a voice mail. As David leaves the station at his usual destination, he lis-*

tens to the voicemail while his favorite music is still playing. He finds out he has to reschedule a meeting with his boss. He then decides to pause his music and starts interacting with his calendar using an audio menu as he keeps on walking. As he starts browsing through his appointments to find a free slot, Stephen calls him back. While taking to Stephen, David browses his calendar and finds a suitable time later on the day for their meeting. The call ends and David continues listening to his music while making his way to the office.

An audio interface like the one illustrated in our example makes eyes-free interactions possible while on the go when visual attention is compromised or the mobile device is out of reach, and enables the user to interact with his mobile devices purely through sound. The extent to which such an audio-only interface is desirable depends on how we deal with cognitive load and multiple streams of information. In our example, we have differing levels of both. From a notification of a voice mail, which might be delivered at the same time David listens to music, to a multitasking extreme of listening to music during a phone call while rearranging a meeting and walking the streets of New York. All of this without getting run over by a yellow cab. How should an audio interface be designed to deal with these competing requirements? In this paper, we consider basic strategies for the presentation of multiple audio streams and examine their usability under varying cognitive load.

Presentation of multiple audio streams

When multiple tasks are supported purely by audio, users must be able to direct their attention selectively to each individual audio stream representing a task. However, when performing multiple tasks at once and conveying them through audio simultaneously, masking of information (i.e. both audio streams are audible and easily confused with each other) occurs as the audio streams overlap. Spatial audio has been used in previous research [1, 2] as a successful technique to segregate multiple audio streams by placing each audio stream at a different location around the user's head, mirroring how humans perceive sounds in real life [3]. What is not yet clear is what spatial audio design might be the most effective for supporting multiple audio streams and how much, if at all, the spatialization of the streams might contribute to an increase or decrease of the user's cognitive load when engaged in a number of simultaneous tasks. Thus, an important issue is how to design audio inter-

faces that will effectively enable users to interact with simultaneous, and often continuous, audio streams whilst at the same time preventing these streams from interfering with each other. This paper presents an evaluation of different spatial and non-spatial audio techniques in such a multitasking scenario.

The importance of cognitive load

Different attention demands impose different amounts of cognitive load on the user. When David is listening to the voicemail left by Stephen, he is focusing attention on the voicemail audio stream while monitoring his music (a selective-attention task). However, when David is talking to Stephen while interacting with his calendar using an audio menu to find a suitable time for their meeting, David is dividing his attention between both audio streams (a divided-attention task). The first task results in less cognitive load, and the second in higher cognitive load. Cognitive load has been described as the amount of mental re-sources needed to perform a given task [4]. As tasks add up, the mental resources needed increase and cognitive load rises. Previous research by Marentakis and Brewster [5] investigating pointing efficiency in deictic spatial audio displays, showed that increased cognitive load resulted in reduced pointing efficiency. Shinn-Cunningham [6] and Best [7] have also investigated how perceived spatial separation of sources and consistency in source locations influences performance on selective- and divided-attention tasks. They found that performance was better when sources were perceived at different locations instead of the same location. However, she adds that "*further experiments are necessary to determine whether spatial attention influences performance differently when competing sources differ from one another in more natural ways*".

When designing audio interfaces it is critical to consider the attention demands expected from the user. This affects the attention required to monitor the information being relayed by the stream and also the attention required to monitor the spatial location of the stream. Spatial audio offers the ability to foreground and background audio streams, for example, moving an audio stream to the side (we will term this *spatial minimization* [8]), while a second stream is played from the front. Spatial minimization could help users alter focus between streams. Thus, when designing mobile audio interfaces to support multitasking, how would different auditory techniques, spatial or non-spatial, perform under different user attention demands? In this paper we quantify the effect of cognitive load on a number of spatial and non-spatial audio techniques during selective- and divided-attention tasks involving user interaction. Our research question was: Are spatial audio techniques able to efficiently support both divided- and selective-attention tasks?

METHODS

Forty-eight participants (26 males, 22 females, aged 18 to 53 years) were recruited. All were native speakers of British English, reported normal hearing and were right-handed. Participants were split equally into two groups: divided- and selective-attention in a between subjects design.

Stimuli

Participants listened to two different streams: one continuous and the other user activated. In the selective-attention group, the continuous stream was a piece of classical music taken from Mozart's Sonata for two pianos K448 in D Major. This specific music piece has been frequently used in spatial-temporal reasoning research. The sonata was divided into different fragments: one for the training session and four others were used in the four different conditions. These fragments were all mono, 16-bit and sampled at 16kHz, and approximately 1.5 minutes long. The participants were told they would have to answer a question on the audio menu tasks to ensure selective attention. In the divided-attention group, the continuous stream was a podcast selected from the BBC Radio 4 programme 'From our own correspondent'. Five different podcasts with a similar journalistic format were chosen. One podcast was used for training the participants and the rest were used in four different test conditions. They were all mono, 16-bit and sampled at 16 kHz, and narrated by a male speaker. In order to ensure divided attention participants were asked to monitor the podcast and told they would have to answer questions on content *as well* as a question on the audio menu tasks. In order to retain coherence, and to allow enough audio material to pose content questions before, during and after the audio menu tasks, the podcasts were longer than the classical musical streams (3 minutes). Our aim was to generate a low cognitive load for the selective-attention group by using classical music, and a high cognitive load in the divided-attention group by using speech.

The user-activated audio stream was a hierarchical audio menu with synthesized audio items. It consisted of a three-item top level: music, appointments and current time. The 'music' item included three items in two different sub-levels: 1) previous track, current track and next track, 2) the song title for each of the items. The 'appointments' item included three items in three different sub-levels: 1) Monday, Tuesday and Wednesday, 2) the times for the appointments, 3) appointment information. The 'current time' item only had one sub-level with time information, (for more details see [8]). All the audio items were mono, 16-bit, sampled at 16 kHz, and were spoken by a female RP voice. All the audio menu items were different for the different conditions. Both the continuous and user-activated streams were normalized to 70% of the audio dynamic range, which equals to a normal conversation typically 60-70dB.

Procedure

Each group was tested in four different conditions: 1) *Baseline*: The continuous stream was paused or *interrupted* while the participant was interacting with the audio menu, 2) *Concurrent*: The continuous stream played while the participant was interacting with the audio menu, 3) *User-activated spatial minimization*: The continuous stream was
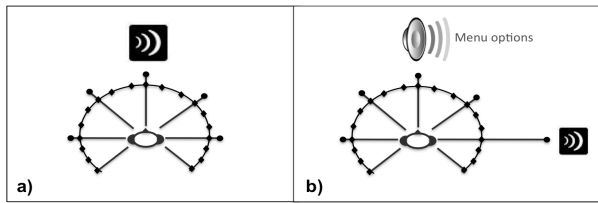
Figure 1. (a) Single continuous stream. Black filled circles show different azimuth locations. (b) User-activated spatial minimization: stream moved from front to right.

located at the origin (0° azimuth) 1m away from the listener in the frontal horizontal plane (see Figure 1), and moved to the right hand-side (90° azimuth) and out to 2m away from the listener when the participant was interacting with the audio menu, and 4) *Fixed spatial minimization*: The continuous stream was fixed to 90° azimuth and 2m away from the listener for the entire duration of this condition.

In all four conditions, participants performed three tasks using a hierarchical audio menu: 1) Finding the next music track title, 2) Checking an appointment for Tuesday, and 3) Finding the current time. The audio menu was always presented at 0° azimuth (in front of the user's nose) and always 1m away in the frontal horizontal plane. All four conditions were tested in the same randomized order in both the divided- and selective-attention groups. Participants were tested in a static lab environment seated on a chair holding the mobile phone in an upright position wearing a pair of headphones. The study was run on a Nokia N95 8GB using the built-in HRTFs and the JAVA JSR-234 Advanced Multimedia Supplements API to position the audio sources. The audio was played over a pair of DT770 PRO – 250 OHM Beyerdynamic headphones.

Participants completed two training sessions before the test conditions. First, a training session was devoted to familiarizing the participants with the audio menu structure in their own time. The second training session used the concurrent condition to familiarize the participants with listening to a continuous audio stream while interacting with the audio menu. For each test condition, participants listened to a continuous audio stream and after approximately 1 minute, the user was prompted with a 25-ms sine wave beep at 1500 Hz to start interacting with the menu and complete the three tasks described previously in any order. To initiate this interaction, the participant pressed the central navigation key on the phone. The arrow keys on the phone were used to browse the menu items. Once the tasks were completed and the audio menu was exited by pressing the central navigation key again, the user continued listening to the continuous audio stream until it was over. A NASA-TLX subjective workload assessment was completed, and, once all four conditions were completed, participants were instructed to rank them in order of preference: '1' being most preferred and '4' the least. The experiment took 30-45 minutes in total. Participants were allowed to rest between conditions.
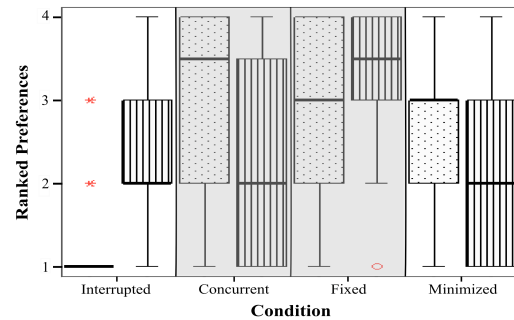

Figure 2. Box plots present ranked preferences per condition and attention group: divided-attention (dotted) and selective-attention (striped). The boxes contain the middle 50% of the data, the horizontal bold black lines show the median and the points outside are suspected outliers. The grey shaded conditions showed no significance.

RESULTS
Ranked Preferences
Figure 2 shows a stacked count for the order of preference for the four conditions compared in the divided- and the selective-attention groups. A non-parametric Kruskal-Wallis for different conditions per attention group showed there was a significant difference in the medians ($\chi2$ (7, N=192)=61.810, p<0.001). Mann-Whitney tests with Bonferroni correction showed a significant difference between the interrupted conditions by group, and also between the user-activated spatial minimization conditions by group. Users' preference for interrupting the continuous stream significantly decreased (two-tailed p<0.0001, total 1st preferences dropped from 20 to 4) and preference for spatially minimizing the continuous stream significantly increased (two-tailed p =0.008, total 1st preferences increased from 2 to 8) when the streamed source was classical music.

Overall Workload
Raw overall workload was calculated from the NASA-TLX questionnaires completed after each condition. A repeated-measures ANOVA with condition type as a within-subjects factor and attention group as a between-subjects factor showed a significant main effect for condition type (F(3,96) =9.786, p<0.001) and attention type (F(1,32)=48.284, p<0.001). There was also an interaction between attention and condition type (F(3,96)=4.34, p < 0.01). As expected, perceived overall workload was significantly higher in the divided-attention group (mean=51.71) than in the selective-attention group (mean=21.33). *Post hoc* Pairwise Comparisons with Bonferroni correction for condition type showed that perceived overall workload during the interrupted condition was significantly lower (p<0.015) than in the rest of conditions for the divided-attention group. No significant differences were found between conditions for the selective-attention group (see Figure 3).

Performance
Total time taken to complete the audio menu tasks was also computed. A repeated-measures ANOVA with condition type as a within-subjects factor and attention group as by-
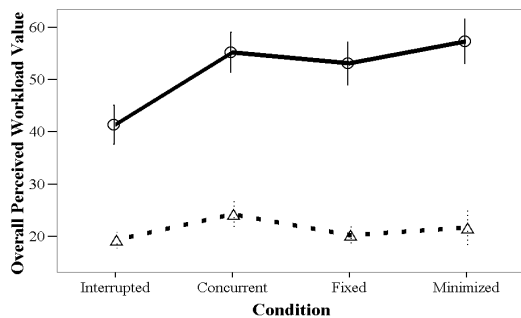
Figure 3. Overall perceived workload per condition and attention group: divided-attention (solid line) and selective-attention (dotted line). Error bars show SD of Mean ± 1.0.
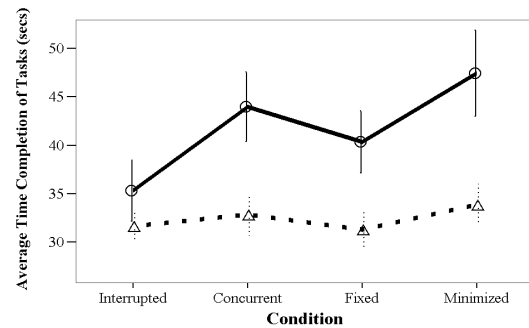

Figure 4. Mean task completion times per condition and attention group: divided-attention (solid line) and selective-attention (dotted line). Error bars show SD of Mean ± 1.0.

subjects factor showed a significant main effect for condition type $(F(3,96)=5.45, p<0.005)$ and attention type $(F(1,32)=7.21, p<0.015)$. There was also an interaction between attention an condition type $(F(3,96)=2.89, p<0.050)$. Task completion times were significantly higher (mean= 41.76 secs) for the divided-attention group than for the selective-attention group (mean=32.03 secs). *Post hoc* Pairwise Comparisons with Bonferroni correction for condition type showed that task completion times for the interrupted condition were significantly lower $(p<0.05)$. Also, *Post hoc* Independent samples T tests across attention type showed that task completion times were significantly higher for the concurrent condition $(t(46)=2.640, p<0.050)$ (mean=43.95 secs) and the minimized condition $(t(46)=2.73, p<0.05)$ (mean=47.42 secs) for the divided-attention group than for the same conditions for the selective-attention group (concurrent: mean=32.31 secs; minimized: mean=33.53 secs). See Figure 4.

DISCUSSION AND CONCLUSIONS

As we might expect, users do not like being put under cognitive load. Participants in this study reported higher workload and took longer to carry out tasks in the divided-attention group. Furthermore, listening to concurrent audio increased the effect of load. The use of spatial techniques had a neglible impact on reducing this effect (although there was a tendency to prefer minimization compared to the other concurrent conditions). In the selective-attention group however, preference results were significantly reversed for interrupted and minimization conditions. This shows that users disliked having the continuous stream interrupted when not under load, and using spatial techniques to separate the continuous stream from the user-activated menu was preferred.

Addressing our research question, we can say that the spatial minimization technique presented in this paper offers an effective means of presenting and interacting with multiple audio streams simultaneously in a selective-attention scenario. However, spatialization techniques were not as effective in the divided-attention task, in which the interaction benefited significantly from the interruption of the continuous stream. In a mobile spatial audio interface design in which interrupting the streams is not an option, it should be

noted that, even in this extreme cognitive load scenario, participants were able to carry out the tasks, although at a reduced ability.

These results suggest a mixed design of audio techniques would be required within our initial usage example. Concurrent streams should be used to play the voicemail notification with spatial audio being used to help separate the information streams. Also, when David decides to access his calendar eyes-free, he could continue listening to the music spatially minimized, as this will not disrupt his task efficiency. It would be nonetheless important to allow him to interrupt his menu interaction at anytime while navigating the streets of New York on foot.

REFERENCES

1. Schmandt, C., and Mullins, A. AudioStreamer: exploiting simultaneity for listening. In Proc. CHI 1995, ACM Press (1995), 218-219.

2. Walker, A. and Brewster, S.A. Spatial audio in small screen device displays. Personal Technologies 4, 2 (2000), 144-154.

3. Bronkhorst, A.W. The cocktail party phenomenon: a review of research on speech intelligibility in multiple-talker conditions. Acustica 86 (2000), 117-128.

4. Draycott, S.G., and Kline, P. Validation of the AGARD STRES Battery of Performance Tests. Human Factors 38, 2 (1996), 347-361.

5. Marentakis, G. and Brewster, S.A. A comparison of feedback cues for enhancing pointing efficiency in interaction with spatial audio displays. In Proc. Mobile-HCI 2005, ACM Press (2005), 55-62.

6. Shinn-Cunningham, B.G. and Ihlefeld, A. Selective and divided attention: extracting information from simultaneous sound sources. In Proc. ICAD 04, (2004).

7. Best, V., Ihlefeld, A. and Shinn-Cunningham, B.G. The effect of auditory spatial layout in a divided attention task. In Proc. ICAD 05, (2005), 17-22.

8. Vazquez-Alvarez, Y. and Brewster, S.A. Designing Spatial Audio Interfaces to Support Multiple Audio Streams. In Proc. MobileHCI 2010, ACM Press (2010), 253-256.