

Effects of Feedback, Mobility and Index of Difficulty on Deictic Spatial Audio Target Acquisition in the Horizontal Plane

Georgios N. Marentakis, Stephen A. Brewster

Glasgow Interactive Systems Group
Department of Computing Science
University of Glasgow, G12 8QQ, UK
{georgios,stephen}@dcs.gla.ac.uk

ABSTRACT

We present the results of an empirical study investigating the effect of feedback, mobility and index of difficulty on a deictic spatial audio target acquisition task in the horizontal plane in front of a user. With audio feedback, spatial audio display elements are found to enable usable deictic interaction that can be described using Fitts law. Feedback does not affect perceived workload or preferred walking speed compared to interaction without feedback. Mobility is found to degrade interaction speed and accuracy by 20%. Participants were able to perform deictic spatial audio target acquisition when mobile while walking at 73% of their preferred walking speed. The proposed feedback design is examined in detail and the effects of variable target widths are quantified. Deictic interaction with a spatial audio display is found to be a feasible solution for future interface designs.

Author Keywords

Gestures, 3D Audio, Mobile interaction, Fitts Law

ACM Classification Keywords

H.5.2 [User Interfaces]. Auditory (non-speech) feedback

INTRODUCTION

Spatial audio displays (displays in which presented audio items are given different spatial locations) enable ‘eyes-free’ interaction, allowing users access to multiple sources of information without needing to look at a screen. Spatial audio rendering can be done either in hardware or software and, unless high display update rates are required, such displays can be rendered on PDAs and wearable computers. In addition, if display control is done by gestures then the display can be highly mobile, since gestures can be performed

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. *CHI 2006*, April 22-28, 2006, Montréal, Québec, Canada. Copyright 2006 ACM 1-59593-178-3/06/0004...\$5.00

based on kinaesthetic feedback. A detailed investigation into selection task design for gesture controlled audio displays is therefore of interest to the HCI community. This is essentially the case since the effect of a number of problems, principally associated with 3D audio fidelity and feedback design on interaction has not been examined.

Spatial audio technology enables people to perceive a sound as emitting from a certain direction in space. One way to accomplish this is by filtering through Head Related Transfer Functions (HRTFs) [21]. HRTFs are measured empirically and capture the properties of the path to the inner ear, including the effects of the outer ear. When applied to a sound HRTFs result in it being perceived as emitting from a given direction in space, outside the head. HRTF filtering can be implemented in real time and thus provides a portable way to reproduce spatial audio.

Gesture recognition can be achieved by means of movement tracking devices. When fitted on parts of the user body, such as the hand or head, they provide movement information that can be subsequently used to recognize gestures. In this study we are interested in deictic interaction, essentially pointing to spatial audio sources.

Pointing has been verified to be an efficient interaction technique and for this reason it forms the basis of direct manipulation UIs. It has been studied extensively within the area of Human Computer Interaction. Fitts’ Law is the most prominent way of characterizing visual pointing actions and the speed/accuracy trade-offs associated with them. In this paper we examine deictic interaction with a spatial audio display in analogy to pointing to a visual target. In particular, the physical act of pointing using a user’s hand to a spatial audio target is examined. We are interested in if and how theories describing visual target acquisition tasks can be applied in pointing to an audio target. Both in real world and in virtual displays, an impression of the direction of an ‘audio target’ can be formed and users can consequently point to the audio target. Due to localization using hearing being less accurate than vision and the fact that the border that separates a target from the background is essentially vague, a major issue that emerges is feedback design. Finally, due to gesture controlled spatial audio displays pro-

viding a truly mobile system design option, we will also examine the effect of mobility on the 3D audio pointing task. Before proceeding with the presentation of an experiment designed to examine the aforementioned issues, we present a review of relevant findings from HCI, psychoacoustics and target acquisition literature.

RELATED WORK

Cohen and Ludwig [5] have discussed how direct manipulation can be transferred to the audio domain. In the resulting concept of ‘Audio Windows’ sounds instead of icons are used to represent spatially arranged display elements and users interact by performing pointing-based interactions either physical or virtual using virtual pointers. Feedback is given using perceptual operators that slightly change sounds as a result of a signal transformation. According to Cohen ‘the idea is to create a just noticeable difference’, an acoustic enhancement that is noticeable but ignorable, unambiguous but unintrusive’ [4]. Such feedback can be provided by filtering, echoes, reverberation or equalization. Cohen’s discussion of audio windows is very interesting since it transfers Shneiderman’s direct manipulation principles into the audio domain. However, a successful direct manipulation type of display is strongly dependent on fast and accurate pointing and therefore the efficiency of each pointing interaction technique has to be evaluated to assess the usability of the task.

In terms of application areas the audio windows concept has been applied in a variety of domains such as for presentation of textual information either from documents, as in Kobayashi and Schmandt [9], or Web content, such as in Goose *et al.* [7]. Gesture controlled menu based interfaces have been proposed by Brewster *et al.* [3] and Savidis *et al.* [19]. A mobile spatial audio display design is Nomadic Radio by Sawhney and Schmandt [20], which provides interaction with a messaging application in a mobile context using speech recognition.

Psychoacoustical aspects & the necessity for feedback

A major problem inherent in spatial audio displays is the limited accuracy of directional hearing which reduces the efficiency of pointing as an interaction technique. Both in the real world and in virtual spatial audio systems, sound localization is not entirely accurate. In our natural environment localization error ranges from $\pm 3.6^\circ$ in the frontal direction, to $\pm 10^\circ$ on the left/right directions and $\pm 5.5^\circ$ to the back of a listener under well controlled conditions and sound sources presented by loudspeakers [2]. In virtual systems localization error (also termed *localization blur*) has been found to be slightly more ranging between 10° and 30° depending on sound direction [21]. Another important problem is that the border that separates target and background is not directly perceivable based on auditory information only. These problems (and others related to spatial audio systems such as front/back and up/down confusions) result in usability problems. Although it may be possible to improve performance by enhancing our knowledge of spa-

tial audio rendering, there is also the possibility of compensating by design. Such a choice is not unnatural and it is justified by the fact that feedback has to be provided anyway to inform the user about the current display state, for example to show whether a certain display element is in focus or has been selected, etc. With appropriate design such feedback can also be used for the additional purpose of assisting users in disambiguating display element position and overcoming speed and accuracy related deficiencies.

The type of feedback to be used in the display has to be such as to compensate for localization error with minimal effect on interaction speed. In [16] a comparison between a number of feedback cues and their combinations for enhancing pointing efficiency in interaction with spatial audio displays was performed and it was found that an external sound as on-target feedback was a fast and efficient way of giving targeting information. In addition, it was found that egocentric presentation is considerably faster but less accurate compared to exocentric (sound position is fixed to the world and updated in real time based on the direction of the user relative to them) presentation. Egocentric presentation refers to the situation where sound position is fixed to the user and remains unchanged no matter what the direction of the user is, whereas in exocentric presentation sound position is fixed to the world and is updated in real time based on the direction of the user relative to the sounds. Based on the results of this study, one can hypothesize that an egocentric display with feedback marked audio elements could provide an efficient design solution. Initial investigations [14,15] prove this approach to be effective in improving selection speed and accuracy. For this reason, we examine this type of feedback design in detail. The acquisition task that we examine is presented in Figure 1. To implement such a feedback design, a certain area of the display has to be assigned to each 3D audio element and the feedback sound should play while the user’s hand is in the target area. With reference to Figure 1, we observe that two parameters are expected to influence performance, distance to the centre of the feedback marked area (A) and width of the area (W). It is worth mentioning that feedback has to be clearly intelligible to avoid problems that can be caused by masking.

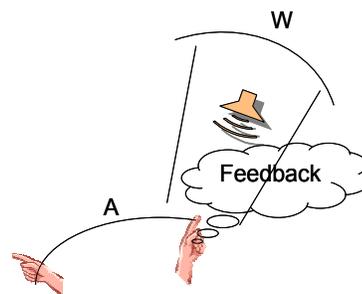


Figure 1. The spatial audio acquisition task under examination.

Feedback should also originate from the target location to assist the user in associating it with the target element.

On-target feedback has been used in visual target acquisition to reduce final positioning time, see Akamatsu *et al.* [1]. Final positioning time is the elapsed time from when a user enters the target area to the moment a selection is made. Akamatsu found on-target feedback resulted in marked differences in final positioning times, with tactile feedback causing a greater reduction than audio or colour indications. Differences were not pronounced but were significant and it can be concluded that feedback can improve final positioning times. In our case however, feedback is given to enhance accuracy and but may reduce speed. The reason for this is misperceptions of target location, localization blur and worse movement support; indeed depending on audio target direction, feedback might not be found exactly where users would expect and centring the arm in the feedback area might not be as efficient as centring the mouse pointer.

To summarize, we consider feedback to be an important issue in compensating for localization errors but anticipate a negative effect on interaction speed that we want to quantify in this study. To design appropriate feedback it is necessary to investigate the effects of target size and distance to target on human performance. We will attempt to perform this in parallel to the theoretical background that already exists for visual pointing tasks. In that way we hope to become able to formulate a methodology that will allow cross-modal comparisons.

Visual target acquisition

A number of models exist for describing visual target acquisition, which include logarithmic, linear and power models. The applicability of each model has been found to depend on the nature of the acquisition task. In spatially constrained tasks, where users are asked to concentrate in finishing the movement inside the target area logarithmic models are deemed to be more appropriate. In temporally constrained tasks where users are required to finish the selection in a certain time, linear models prevail. The difference is mainly due to temporally constrained tasks being performed by a single ballistic movement. In spatially constrained tasks, users perform corrective sub-movements to ensure they are on target. This in effect results in distance to target having a logarithmic (more smooth) effect. Power models emerge in intermediate cases where both parameters are taken into account [17]. In HCI under normal conditions tasks are essentially spatially constrained.

The most prominent logarithmic model for spatially constrained visual target acquisition is Fitts law, according to which time to select a visual target is affected by distance to target and target width in accordance to Equation 1.

$$MT = \alpha + \beta \log_2 \left(\frac{2A}{W} \right) \quad (1)$$

In Equation 1 α and β are constants that can be estimated for example using linear regression, A is the distance to target and W is target width. The log term is called the Index of Difficulty (ID) and its unit is bits, due to the base of the logarithm being 2. The reciprocal of β is the Index of Performance (IP) with units of bits/sec. IP has been associated with the rate of information processing for the movement task under investigation and is therefore treated as measure of the efficiency of different interaction techniques. Nowadays Equation 2 proposed by MacKenzie is used instead of Equation 1, giving a better fit.

$$MT = \alpha + \beta \log_2 (A/W + 1) \quad (2)$$

The equations hold both for linear and for angular movements and A and W can refer to linear or angular distances [8,10]. A number of measures have been developed to characterize pointing efficiency [12]. In this study we will use the measure of *throughput*. Throughput is defined as:

$$Throughput = \frac{\log_2 (A/W_e + 1)}{MT} \quad (3),$$

MT being the movement time. W_e is the effective target width calculated based on the spatial standard deviation of selections. This is essentially different from Equations 1 & 2 where the actual target width is used. Effective target width can be calculated as:

$$W_e = 4.133 \times SD_x \quad (4),$$

SD_x being the standard deviation in the selection coordinates measured along the axis of approach to the target.

Audio target acquisition

Selecting a 3D audio display element based on the direction of the sound event either by a real or virtual pointing gesture is quite similar to homing to a visual target. This is because users are required to perform their movements using directional information. Studies on visual target acquisition can therefore serve as a starting point that can help to identify parameters that affect this type of interaction. However, a different sensory modality is used for event localization and, as described earlier, users have a less precise impression of target location and the border separating target from background. The role of feedback is therefore to denote target size by defining where a target is separated from the background. We decided to base our further analysis on the quantities of distance to target and target width since they have been proven to affect virtually all pointing tasks and serve as a well-founded starting point for such an investigation. We hypothesize that interaction in a spatial audio display is affected by the prominent variables of target width and distance to target in a manner similar to what is stated by Fitts' law.

Experimentation is, however, necessary to verify this hypothesis due to logarithmic models not being appropriate in certain cases such as in temporally constrained visual aimed

movement tasks, steering tasks and tasks where no continuous contact with the target is available. The study of Friedlander *et al.* [6] is a good example. In each trial of their homing to non-visual targets experiment, participants were asked to move in one out of four directions while counting certain steps indicating ring widths in a bull’s-eye menu. Audio and tactile feedback was used to mark ring widths to define target distance in the display. The authors verified that distance to target and target width affected time to select. However, a linear model was found to fit better to movement times. The formula suggested for the approximation of time to target is

$$MT = \alpha + \beta \cdot A/W \quad (5)$$

The difference can be attributed to the lack of continuous contact with the target that might have led to a more steering-like behaviour where participants moved with a constant speed as they were counting rings to reach the target. To summarize, although it is likely that Fitts’ law will apply in our case due to continuous contact to target and the spatially constrained nature of our feedback-enabled task, the relative uncertainty of sound localization may result in a more ballistic-like linear behaviour. We therefore must test for both options.

Mobility

Given the fact that spatial audio displays provide an eyes-free way to interact, they are considered well suited for mobile HCI. Designing for mobile users, however, is considered a challenging task. The two major challenges are the limited display area that is available and the effect of mobility on the control actions of the user. The design approach that we examine in this paper is to use spatial audio for display presentation and deictic gestures for control. Both of these choices help reduce the load on user’s visual attention and are therefore suitable for mobile interaction. Gestures are a very convenient way of communicating in mobile situations. Most people can perform gestures whilst moving. For example, it is very easy to point to something or to raise a hand to greet someone while walking. Empirical evaluation by Pirhonen *et al.* [18] showed gestures to be a more convenient way of interacting with a system compared to common stylus based interaction techniques that require visual attention and inhibit our movement. A very common result in usability studies of systems supporting control based on visual feedback is that users have to interrupt their movement in order to interact with their computers.

Brewster *et al.* [3] found an egocentric spatial audio display with directional nodding gestures for selection could be used successfully while mobile. Users were able to interact with the system whilst walking at an average of 69% of their normal walking speed. Empirical evidence therefore exists that gesture controlled spatial audio displays are usable in mobile contexts. In this paper however, we focus on pointing using the hand. We will try to investigate how walking speed will be affected by target size as well as look

into how mobility will affect selection times and selection accuracy. These results will help in understanding how display design has to be altered to compensate for the effects of mobility.

EXPERIMENT

Our research questions are concerned with the effect of feedback, mobility and Index of Difficulty on deictic spatial egocentric audio target acquisition in the horizontal plane. Our hypotheses are:

- Feedback in the form of audio on-target confirmation will affect accuracy of selections and interaction speed
- Mobility will affect selection time and selection accuracy
- Index of difficulty will affect selection time

An experiment was designed to test the experimental hypotheses. Participants performed the task in Figure 1 according to the design presented in Table 1. Of the independent variables in Table 1, mobility, target width and A/W were tested within subjects, whereas feedback was tested between subjects. Participants were split in two groups, one performing the experiment with on target feedback and the other without. Both groups performed the experimental task standing and walking in a counter-balanced order. When mobile, participants had to walk in figure of eight laps around three traffic cones that had been placed in the lab. The cones were 1.2 meters apart, providing a rather challenging route. This was done to provide a realistic scenario forcing participants to pay attention to their movement.

Walking speed was calculated by dividing the number of laps participants performed with the total time this took. Preferred walking speed in this particular walking route was measured by asking participants to walk five laps without performing the audio selection task. Dependent variables were time to select, selection success ratios, perceived workload (using NASA TLX), percentage preferred walking speed and steps to select the target.

DV	IV	Levels
Time	Feedback	Yes/No
	Target Width	See Table 2
Accuracy	Distance to target	See Table 2
Walking Speed	Mobility	Standing/Mobile
Steps		
Workload		

Table 1. Experimental Design.

Target Width and Distance to Target

To examine the hypotheses on the applicability of Fitts' law on spatial audio target acquisition we decided to test target width and distance to target at the levels presented in Table 2. As can be seen in Table 2, variables A and W were selected so that A/W ratios remained constant for most of the cases. Fitts' law would predict no significant differences in time measurements. One observation is that for the particular task it is hard to obtain high ID values. This is due to the restricted display area in our study (the area in front of the user) and the relatively large target sizes that have to be used in order to provide usable pointing. For this reason, for target widths of 10° and 35°, we decided to test for additional values of ID, namely 3 and 2.48. In the former case, this was done to gain some insight on what is happening at higher IDs and in the latter because using an ID of 2 would result in a distance that would lie to the back of the user whereas our test area was in front. It should be mentioned here that the concept of target width only affects user performance when participants get feedback. For the group where no feedback was given, ID is not expected to affect the results. The relevant variable in this case is distance to target.

Technology

To create a truly mobile test bed we used A small laptop was placed in a rucksack that participants wore on their backs and ran the program used to control the hardware and the stimuli.

W	A	ID	A/W
10°	10°	1	1
	30°	2	3
	70°	3	7
15°	15°	1	1
	27°	1.5	1.83
	45°	2	3
20°	20°	1	1
	36°	1.5	1.83
	60°	2	3
25°	25°	1	1
	46°	1.5	1.83
	75°	2	3
30°	20°	1	1
	55°	1.5	1.83
	90°	2	3
35°	35°	1	1
	64°	1.5	1.83
	87°	1.8	2.48

Table 2. W, A, IDs and A/W Ratios used in the experiment.



Figure 2. A participant performing the experimental task

Each participant was equipped with two MT-9B (www.xsens.com) orientation trackers. One of them was placed in a small belt-mounted case that was placed in the middle of each user's waist. This tracker was used to record user movement and calculate the number of steps taken. The other was held in the right palm of each participant. The difference in orientation readings between the two trackers was used to infer pointing direction of the hand relative to the body. In this way, we had an estimate of pointing direction while participants could freely move in space. HRTF filtering was done on the laptop using the DieselPower Implementation 3D Audio API (www.am3d.com). The API provides an HRTF filtering implementation that uses generic (non-individualized HRTF functions). Localization using non-individualized HRTFs is worse compared to localization using individualized HRTF functions, however the effect in sound localization on the frontal horizontal plane is non-significant. Sennheiser HD 433 headphones were used to present the sounds.

Procedure

Participants were assigned to one of the two groups and were briefed on the experimental task. Participants were instructed that once the experiment started they would experience a sound which would be perceived as playing from a fixed position somewhere between their left and their right and always in front.

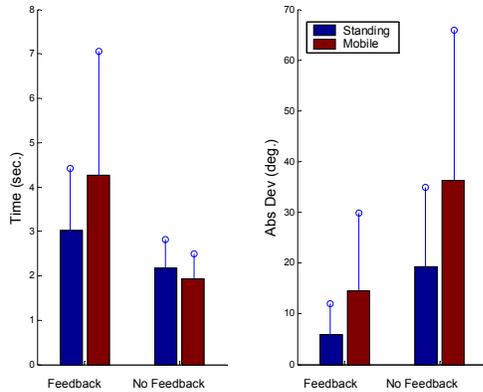


Figure 3. Mean time and accuracy score for standing and mobile participants who did and did not receive feedback

They were told that their task would be to move their hand to the position of the target and perform a downward hand gesture to indicate the sound position. When participants were part of the feedback group they were told to make sure they heard the feedback sound before proceeding with their selection.

After each selection the target sound changed position so that each participant performed 6 selections for each row of Table 2. The next A, W pair was chosen randomly out of the 18 possibilities presented in Table 2. Overall each participant performed 108 selections standing and 108 walking. The target sound was played at the height of the user's nose at a distance of 5 meters. The stimulus was half a second of white noise repeating itself after half a second of silence. Feedback was provided to the appropriate group by means of people cheering when participants were inside the target width that was assigned to each trial. The feedback sound was played from the same direction as the target when participants entered the target area. In total, 24 participants were tested 6 females and 18 males with an age span of 18-42 years (mean 27). All were students from the University of Glasgow and were paid £5 for their participation.

RESULTS

Due to target width having no affect on interaction in the no-feedback case, an overall mobility (2) x feedback (2) analysis of variance was performed first. Mobility was found to have a significant main effect on time to select a target ($F(1,2590)=141.24, p<0.001$). Feedback was also found to have a significant main effect on selection time ($F(1,2590)=1134.909, p<0.001$). Mobility was found to have a significant main effect on the accuracy of selections as measured by the absolute deviation from target, $F(1,2590) = 616.054, p<0.001$. Feedback was found to have a significant main effect on the accuracy of selections, $F(1,2590)=1148.477, p<0.001$. Participants were slower

	T	S
M	$F(1,143)=34.619, p<0.001$	$F(1,11)=122.335, p<0.001$
W	$F(5,715)=38.071, p<0.001$	$F(5,55)=23.861, p<0.001$
A/W	$F(2,286)=11.421, p<0.01$	$F(2,22)=4.690, p<0.02$
MxW	$F(5,715)=17.230, p<0.001$	$F(5,55)=23.861, p<0.01$
MxA/W	$F(2,286)=5.819, p<0.01$	N/S
WxA/W	N/S	N/S

Table 3. ANOVA results for participants that received feedback (T is time and S (%) success ratio).

and less accurate when mobile. They were slower but more accurate when they received feedback. Given the main effects we observed we proceeded with two separate analyses for participants performing the task with and without feedback. Results are illustrated in Figure 3.

Performance with on-target feedback

Time Analysis

Figure 4 presents mean selection times for participants that received feedback as a function of mobility, target width and Index of Difficulty. An overall mobility (2) x width (6) x A/W (3) analysis of variance was performed on the time scores of the participants that received feedback. Results are presented in Table 3. The results confirm a significant main effect of mobility, width and A/W ratio. Interaction between mobility and target width as well as between width and A/W ratio was also significant.

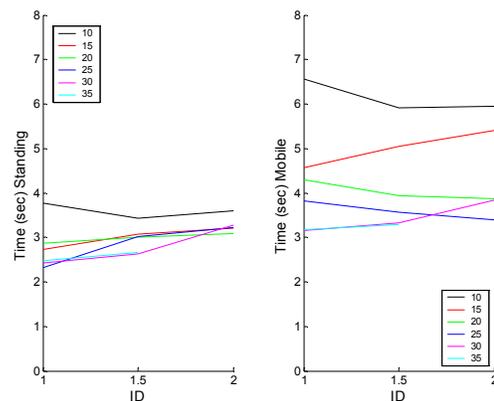


Figure 4. Mean time scores for participants who received on-target feedback as a function of ID and target width.

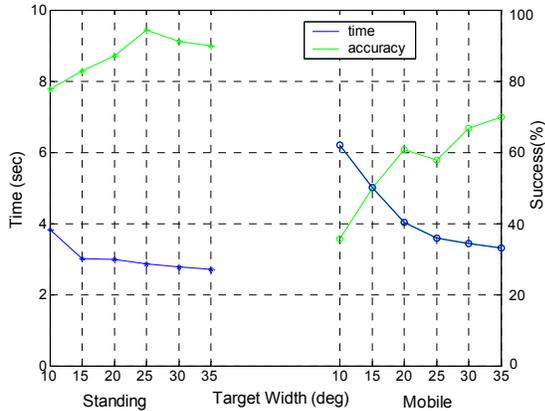


Figure 5. Time & Success scores as a function of target width for standing and mobile users who received on-target feedback averaged over A/W ratios

For standing participants *post-hoc* t-tests using the Bonferroni confidence interval adjustment reveal time scores for target width of 10° to differ significantly from all the rest, with no other differences found. Time scores for all A/W ratios were found to differ significantly for standing participants.

For mobile participants A/W ratios did not have a significant main effect on selection times. Pair-wise comparisons showed selection time for widths of 10°, 15° and 20° to differ significantly from all of the other target widths and between themselves. There was no difference in selection times between target widths of 25° and 30° which, however, differed from all the rest; no difference between 30° and 35° and 25° and 30° target widths were found. Figure 5 shows time scores for all tested target widths averaged over A/W ratios.

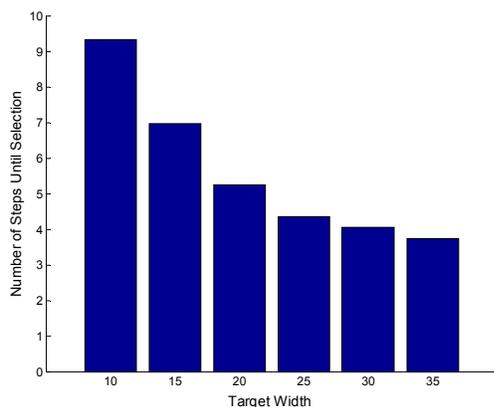


Figure 6. Mean steps until selection as a function of target width.

Success Ratio Analysis

An overall mobility (2) x width (6) x A/W (3) analysis of variance was performed for success scores for participants who received feedback. By success scores, we refer to the percentage of trials participants selected within the feedback marked area. Results are presented in Table 3. The results confirm a significant main effect of mobility, width and A/W ratios. Success ratios increased with target width and decreased when participants when mobile. Interaction between mobility and target width was found to have a significant effect on success scores. *Post hoc* analysis comparing mobile and standing participants showed A/W ratios to have a significant main effect on success ratios for mobile but not for standing users. Figure 5 shows how success ratios varied for the aforementioned cases, averaged over A/W values.

Steps Analysis

A width (6) x A/W (3) analysis of variance was performed on the number of steps taken per selection for (mobile) participants who got feedback. The results showed width to have a significant main effect $F(5,355) = 23.836, p < 0.001$. No effect of A/W ratio was found. Figure 7 presents the results grouped over A/W ratios. Grand mean was 5.6 steps.

Performance without feedback

Without feedback target width was not evident to participants and therefore is not a relevant variable. The factors affecting performance in this case are distance to target and mobility. A within-subjects analysis of variance (mobility (2) x distance (15) (see Table 2) on time scores reveals a significant main effect of both mobility and distance ($F(1,71) = 31.040, p < 0.001$ and $F(17,1207) = 2.497, p < 0.005$). *Post hoc*, pair-wise Bonferroni confidence interval adjusted t-tests showed interaction to be faster for mobile participants compared to standing ones (see Figure 7). Means of the three distance groups that result from averaging time scores every six distances (with respect to Table 2) are presented. A similar analysis on standard deviation from target revealed a significant main effect of mobility with participants being less accurate when mobile, as shown by Bonferroni *post hoc* t-tests. No effect of distance was found in the no feedback case. Figure 8 shows the selection accuracy for each target width when standing and mobile. Participants were significantly less accurate when mobile. It can be observed that very large target widths would be required for effective interaction under these conditions. With respect to steps taken to select overall distance to target produced a significant main effect ($F(17, 1207) = 2.2111, p = 0.003$). On average, participants performed about two (1.74) steps before proceeding to each selection.

Walking Speed Analysis

An analysis was performed to test the effect of feedback on percentage preferred walking speed for mobile participants. No significant difference was found between participants using feedback and no feedback. Both, however, were significantly less than participants' normal walking speed.

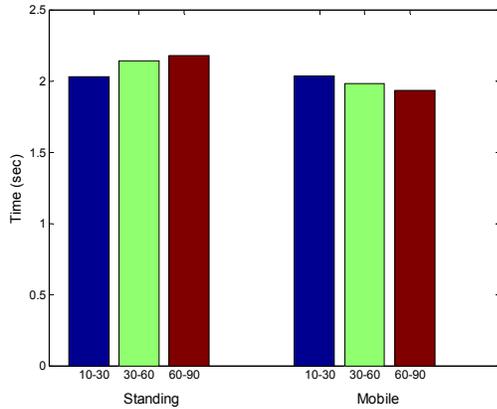


Figure 7. Mean time scores for participants that did not receive on-target feedback as a function of distance to target groups and mobility.

Means were 71% for the no feedback case and 76% for the feedback case, a finding consistent with [3].

Workload Analysis

A mobility (2) x feedback (2) analysis of variance showed mobility to have a significant main effect on overall perceived workload ($F(1,22) = 37.498, p < 0.001$), with mobility increasing perceived workload by about 30%, (Mean Standing Workload = 3.65, Mean Mobile Workload = 4.8). Feedback did not have a significant main effect on overall perceived workload. A more detailed analysis showed mobility to have a significant main effect on mental demand ($F(1,23) = 13.089, p < 0.001$), physical demand ($F(1,23) = 5.741, p < 0.03$), effort expended ($F(1,23) = 12.032, p < 0.005$), performance level achieved ($F(1,23) = 9.304, p < 0.01$) and frustration experienced ($F(1,23) = 13.301, p < 0.001$). Mobility did not affect time pressure and annoyance experienced.

Can Fitts' law be used to describe 3D audio target acquisition?

Is it possible to model spatial audio target acquisition in terms of Fitts' law? This analysis is performed only for participants that received feedback. Linear regression on time measurements was performed for the three models in Equations 1, 2 and 6 for W values in the usable range that is over 20° for standing participants and over 30° for mobile ones. We have chosen these values because no significant difference was observed in time ratings as can be seen in the results section. Regression results are presented in Table 4 and Figure 9. For standing participants, we observe that both logarithmic models correlated significantly, the model of Equation 2 correlating significantly better as a t-test revealed ($t=4.46, p < 0.001$). The linear model, although providing high correlation, did not correlate significantly ($p = 0.058$). Given the significant correlation values we are able to calculate Index of Performance scores for standing participants. This was approximately 1.6 which is about half what

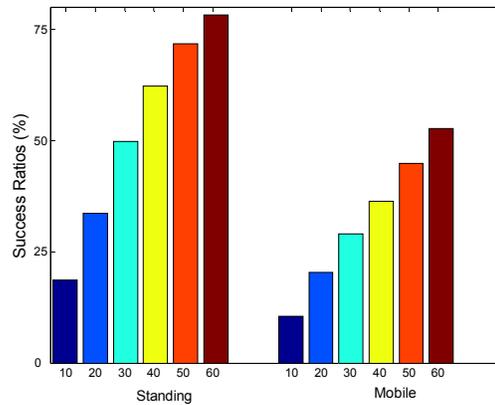


Figure 8. Success ratios as a function of target width for standing and mobile users (no on-target feedback)

has been found in the literature for participants interacting with a visual display using a trackball [13]. For mobile users this calculation was not made since the models did not correlate significantly. Throughput values were also calculated for standing participants giving values of 0.4, 0.53 and 0.63. This is about 3.5 times less than that measured for interaction with virtual pointers in visual displays. Throughput was calculated using A/W values and not the effective target width formulation because effective target widths were unreasonably high and dependent on target size. For example, for target width of 25°, effective target width was 36° and success ratio in the order of 95%, 36° therefore does not provide a better estimate of target width. Throughput increased for increased distance to targets. Indeed, the time measurements reveal that, at the lowest IDs, participants became confused since the feedback area was too close to their initial position. Time scores in throughput calculations include reaction time and selection time (therefore throughput values might be underestimated).

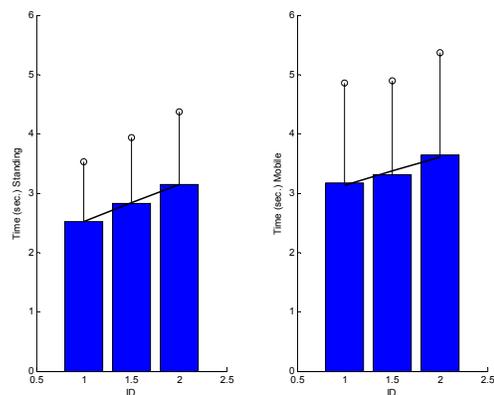


Figure 9. Linear Regression Lines, mean time scores and standard deviations as a function of Index of Difficulty

	R^2_{Lg}	R^2_L	R^2_F	IP
S	0.999*	0.991	0.995*	1.57
M	0.86	0.92	0.82	x

Table 4. Goodness of fit comparisons between the linear and logarithmic models. * denotes that correlation was significant at the $p < 0.05$ level. R^2_{Lg} , R^2_L , R^2_F stand for the R^2 statistic for McKenzie, Linear and Fitts models. IP is the index of performance. S for standing and M for mobile participants.

DISCUSSION

The results of this study verify to a great extent the experimental hypotheses. Mobility was found to cause slower and less accurate interaction. Feedback was found to decrease interaction speed but increase interaction accuracy and interaction was negatively affected by increased Index of Difficulty. The success ratios in the no-feedback condition show that performance in deictic interaction with an ego-centric display without feedback is very poor. It can be observed that such a display can hardly accommodate more than three targets in the area in front of the user. In mobile situations, the maximum number of targets would be two. It has to be said however, that target position was varying randomly in this experiment, in this sense the effects of learning are not taken into account in the results. It might be the case that when interacting with a familiar display one more target might be feasible. Surprisingly, when mobile, participants were faster in their selections when on-target feedback was not given. Participants commented that when mobile they were not able to pay much attention to the target position and they mostly selected on a left right basis. They said that when standing they were able to pay more attention to the target sound and infer more on its position. Consequently, they adopted a more careful strategy for their selections that resulted in increased selection times. We therefore attribute this finding to the negative attitude participants formed towards the system when mobile.

On-target feedback was found to improve performance significantly for standing users and, based on the results of this study, is considered necessary to enable usable gesture interaction with a spatial audio display. Feedback design is, however, critical. For standing users, it was found that time to select is not affected by target width for widths of 15° and more. However, target widths between 20° to 25° were found to be necessary to provide successful selection accuracies of 90% and more. On these grounds, we recommend target widths of 20° or more for standing users. Consequently, such a display could accommodate up to 8 elements in front of a user. On-target feedback was found to be moderately successful in the case of mobile users. It did not affect their walking speed compared to the case of no feedback, nor did it increase the perceived workload. However,

as shown in Figure 5 success rates of more than 75% were not possible to achieve for the target widths that were tested in this study. Although time to select when mobile with a target width of 35° was relatively close to the one observed for standing users, it appears that wider targets than this are necessary to increase success rates. Observing Figure 5 it can be deduced that the performance curve has not saturated and increasing target width further will benefit interaction both from an accuracy and speed point of view. Therefore, mobility was found to influence performance in the feedback case in a negative way. Even with large targets there is a 20% increase in selection times and an approximately 20% decrease in selection accuracy. Increasing target width is a solution as found in this study, however this approach is not optimal because it reduces the number of elements the display can accommodate. Alternative ways to overcome the variability introduced in mobile contexts have to be examined to reduce the negative effect that was found in this study. Appropriate filtering of the movement signals is a promising solution for overcoming mobility problems and provide an experience that will resemble the standing case more closely.

The modalities chosen for display presentation (3D audio) and control (gestures) have been proven in the study to provide a mobile way to interact with the system. Users did not have to stop at all when performing the task, neither with nor without feedback. As can be seen in the number of steps to select the target analysis at appropriate target widths users were able to select a target approximately every three steps, a promising finding given the relatively complicated walking route they had to follow. In addition, the resulting percentage preferred walking speed is close to the preferred one, with a mean of about 73%. Given the random target positions in this experiment, we anticipate even higher figures for users interacting with a familiar system.

According to the results presented, it appears that it is possible to view spatial audio target acquisition in terms of Fitts' law. This is particularly encouraging since it shows that this type of interaction is an efficient one, comparable to interaction with visual displays. This issue is of particular importance since it enables cross-modal comparisons in the context of tasks under examination. In addition, it provides a predictive tool for performance in gesture interaction with spatial audio. Even in the case of mobile participants, we observed high, although not significant, correlations. The complex walking route participants had to follow definitely contributed to this finding. In another context involving a simpler walking route it might prove feasible to calculate the difference in the index of performance between standing and walking conditions. Further design is necessary in order to create pointing tasks that can be described by Fitts law in mobile contexts. In conclusion, this study revealed a number of major factors that affect performance in deictic interaction with a spatial audio display. The selection task that was proposed and examined was found to be usable for standing users and to allow mobile users to interact with a

walking speed close to their normal one, while at the same degrading selection success rates and time at the level of 20%. Given the elementary design that was employed in the experiment the results enable us to be optimistic on the future of gesture interaction with spatial audio in mobile contexts.

CONCLUSIONS

We presented the results of an empirical study that showed mobility, feedback and Index of Difficulty to have a significant effect on spatial audio target acquisition. We found a design using audio feedback marked display elements to result in usable deictic interaction with no effect on walking speed or workload compared to interaction without feedback. We found spatial audio target acquisition to be sufficiently described in terms of Fitts law, when proper target width choices were made. A detailed investigation on the effect of target size on time and accuracy of selections was presented. Participants were able to walk at 73% percent of their normal walking speed, mobility degrading performance by 20%. Deictic interaction with 3D audio displays is shown to be a feasible solution for future human computer interaction designs.

ACKNOWLEDGMENTS

This study was supported by the EPSRC funded Audio-Clouds project (GR/R98105).

REFERENCES

1. Akamatsu, M., MacKenzie, S. I., and Hasbrouc, T., A Comparison of Tactile, Auditory and Visual Feedback in a Pointing Task using a Mouse-Type device. *Ergonomics*, 1995. 38: p. 816-827.
2. Blauert, J., *Spatial Hearing: The psychophysics of human sound localization*. 1999: The MIT Press.
3. Brewster, S., Lumsden, J., Bell, M., Hall, M., and Tasker, S. Multimodal 'Eyes-Free' Interaction Techniques for Wearable Devices. in *ACM CHI*, 2003. Fort Lauderdale, FL: ACM Press, Addison-Wesley. p. 463-480
4. Cohen, M., Throwing, pitching and catching sound: audio windowing models and modes. *Int. J. Man - Machine Studies* (1993), 1993. 39: p. 269 - 304.
5. Cohen, M. and Ludwig, L., Multidimensional Audio Window Management. *International Journal of Man - Machine Studies*, 1991. 34: p. 319-336.
6. Friedlander, N., Schlueter, K., and Mantei, M. Bullseye! When Fitt's Law doesn't fit. in *ACM CHI*, 1998. Los Angeles, CA: ACM Press Addison-Wesley. p. 257-264
7. Goose, S. and Moller, C. A 3D Audio Only Interactive Web Browser: Using Spatialization to Convey Hypermedia Document Structure. in *7th ACM Multimedia*, 1999. Orlando, Florida, United States: ACM Press. p. 363 - 371
8. Jagacinski, R. J. and Monk, D. L., Fitts' law in two dimensions with hand and head movements. *Journal of Motor Behavior*, 1985. 17: p. 77-95.
9. Kobayashi, M. and Schmandt, C. *Dynamic Soundscape: mapping time to space for audio browsing*. in *SIGCHI*, 1997. Atlanta, Georgia, United States: ACM Press. p. 194 - 201
10. Kondraske, G. An angular motion Fitts' Law for human performance modelling and prediction. in *16th Annual Conference on Engineering in Medical and Biology Society*, 1994. p. 307-308
11. MacKenzie, S. and Buxton, W. Extending Fitts' Law to Two-Dimensional Tasks, in *CHI 1992*. Monterey, California, United States. p. 219-226
12. MacKenzie, S. I., Kauppinen, T., and Silfverberg, M., Accuracy Measures for Evaluating Computer Pointing Devices, in *CHI 2001*. Seattle, USA. p. 9-16
13. MacKenzie, S. I., Sellen, A., and Buxton, W. A Comparison of Input Devices in Element Pointing and Dragging Tasks, in *CHI 1991*. New Orleans, Louisiana, United States: ACM Press, p. 161-166.
14. Marentakis, G. and Brewster, S, Effects of reproduction techniques on interaction with a spatial audio display. in *ACM CHI , Late Breaking Results*, 2005. Portland, Oregon. p. 1625-1628
15. Marentakis, G. and Brewster, S, A study on gesture interaction with a 3D Audio Display. in *Mobile HCI*, 2004. Glasgow, UK. p. 55-62
16. Marentakis, G. and Brewster, S, A comparison of feedback cues for enhancing pointing efficiency in gesture interaction with a spatial audio display. in *Mobile HCI*, 2005. Salzburg, Austria: ACM Press. p.180-191
17. Meyer, E., Abrams, A., Kornblum, S., Wright, E., and Smith, J., Optimality in Human Motor Performance: Ideal Control of Rapid Aimed Movements. *Psychological Review*, 1988. 95(3): p. 340-370.
18. Pirhonen, A., Brewster, S., and Holguin, C. Gestural and audio metaphors as a means of control for mobile devices, in *ACM CHI*, 2002. Minneapolis, Minnesota, USA: ACM Press New York, NY, USA. p. 291-298
19. Savidis, A., Stephanidis, C., Korte, A., Rispien, K., and Fellbaum, C, A generic direct-manipulation 3D auditory environment for hierarchical navigation in non-visual interaction, in *ACM ASSETS '96*, 1996. Vancouver, Canada, 1996: ACM Press. p. 117-123
20. Sawhney, N. and Schmandt, C., Nomadic Radio: Speech and Audio Interaction for Contextual Messaging in Nomadic Environments. *ACM Transactions on Computer-Human Interaction*, 2000. 7(3): p. 353-383
21. Wenzel, E., Marianne, A., Kistler, D., and Wightman, L. F., Localization using non-individualized head-related transfer functions, *Journal of the Acoustical Society of America*, 1993. 94(1): p. 111-123