A Comparison of Feedback Cues for Enhancing Pointing Efficiency in Interaction with Spatial Audio Displays

Georgios Marentakis, Stephen A. Brewster
Glasgow Interactive Systems Group, Department of Computing Science
University of Glasgow, Glasgow, G12 8QQ, UK
{georgios,stephen}@dcs.gla.ac.uk www.audioclouds.org

ABSTRACT
An empirical study that compared six different feedback cue types to enhance pointing efficiency in deictic spatial audio displays is presented. Participants were asked to select a sound using a physical pointing gesture, with the help of a loudness cue, a timbre cue and an orientation update cue as well as with combinations of these cues. Display content was varied systematically to investigate the effect of increasing display population. Speed, accuracy and throughput ratings are provided as well as effective target widths that allow for minimal error rates. The results showed direct pointing to be the most efficient interaction technique; however large effective target widths reduce the applicability of this technique. Movement-coupled cues were found to significantly reduce display element size, but resulted in slower interaction and were affected by display content due to the requirement of continuous target attainment. The results show that, with appropriate design, it is possible to overcome interaction uncertainty and provide solutions that are effective in mobile human computer interaction.

Categories and Subject Descriptors
H.5.2 User Interfaces-Auditory *Non-Speech Feedback, Interaction Styles*

General Terms
Performance, Design, Experimentation, Human Factors

Keywords
Spatial Audio Display Design, Gestures, Multimodal Interaction.

1. INTRODUCTION
One of the main advantages of spatial audio displays (displays in which presented audio items are given different spatial locations) is the fact that they enable 'eyes-free' interaction, allowing users access to multiple sources of information without needing to look at a screen. In addition, spatial audio displays are easily portable so overcome some of the problems of limited screen space in many mobile devices. 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 current PDAs and wearable computers. Consequently, we can conclude that as long as such displays can be designed in a way that facilitates interaction they are well suited for mobile situations. However, a number of problems, principally associated with 3D audio fidelity, hinder the widespread use of such systems. A study into ways of improving interaction in such displays is therefore necessary to shed light on design issues and implications of design choices is virtual audio display design.

Mobile HCI'05, September 19–22, 2005, Salzburg, Austria.

Spatial audio technology enables people to perceive a sound as emitting from a certain direction in space by applying certain transformations to the sound signal. One way of accomplishing this is by filtering through Head Related Transfer Functions (HRTFs). 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 signal HRTFs result in the signal being perceived as emitting from a given direction in space. HRTF filtering can be implemented in real time and can thus provide a portable way to produce spatial audio.

Spatial audio displays have been designed for a variety of purposes such as for presentation of textual information either from documents, such as in Kobayashi and Schmandt [10], or Web content, such as in Goose *et al.* [9]. Menu based interfaces have also been proposed as in Brewster *et al.* [4] and in Savidis *et al.* [19]. The former provides the only evaluation of a spatial audio display in a mobile setting and showed that such displays are effective in mobile situations. Another mobile spatial audio display design is Nomadic Radio by Sawhney and Schmandt [20], which provided interaction with a messaging application in a mobile context. Informal, qualitative evaluation justified the proposed design and showed the system to be usable.

Our spatial audio display design is based on the notion of 'audio windows' developed by Cohen and Ludwig [6]. Audio Windows are an application of the direct manipulation design principles, common in graphical user interfaces, to the audio domain. Sounds in the display are used to represent display elements and are given a particular location. Users can interact and control audio display elements by using physical gestures such as pointing, or throwing or using virtual audio pointers controlled by a hardware device. Feedback can be given using perceptual operators that slightly change sounds as a result of a certain 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' [5]. 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 in the audio domain, but the ideas were not evaluated.

A direct manipulation type of display is strongly dependent on fast and accurate pointing, due to the pointing action being a natural and effective way of expressing the location of a target object. Pointing has been studied extensively within the area of Human

Computer Interaction. Fitts' Law is the prominent way of characterizing pointing actions [7, 13] and the speed/accuracy trade-off associated with them. One issue that has been little examined in the literature is the role of the modality used to perform the pointing action and to receive associated feedback from the display. Although localization of objects is done very efficiently using vision, it can also be performed using our sense of hearing, as has been studied in detail by Blauert [3]. Spatial audio user interfaces are based exactly on this ability of our auditory system.

One of the major problems inherent in spatial audio displays is the limited accuracy of directional hearing in virtual audio systems. This limited accuracy reduces the efficiency of pointing as an interaction technique. Both in the real world as in virtual spatial audio systems, sound localization is not entirely accurate. In our natural environment localization error ranges from ±3.6° in the frontal direction, to ±10° on the left/right directions and ±5.5° to the back of a listener under well controlled conditions and sound sources presented by loudspeakers [3]. In virtual systems localization error (also termed *localization blur*) has been found to be in the order of 20° to 30° on average [23]. This particular problem (and others related to spatial audio systems such as front/back and up/down confusions) causes serious problems for the speed and accuracy of interaction. Although it may be possible to improve the performance by enhancing our knowledge of spatial audio rendering, there is also the possibility of compensating by design. One prominent way of doing this is by providing additional feedback in the form of an external sound source whenever the user is on target. As has been found in our previous work [15, 16], this approach is effective and can successfully improve selection speed and accuracy. Such a choice is not unnatural and it is justified by the fact that feedback has to be provided anyway to inform on the current display state, for example to show whether a certain display element is in focus or that it 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.

Another common pointing related issue is final positioning time. Final positioning time is the elapsed time from when a user enters the target area to the moment a selection is made. Even when targeting visual targets, on-target feedback has been found to produce marked differences in final positioning times as found by Akamatsu *et al.* [1]. In the aforementioned study, auditory, tactile, colour and all three combined feedback were compared as a means to indicate that the pointer was over the target. An analysis on final positioning times gave a ranking of tactile, combined, audio, colour and normal. The differences in mean times were not pronounced but were significant and based on this study it can be concluded that feedback can improve final positioning times. Final positioning is also a significant problem in audio displays [11]. In an experimental study Loomis *et al.* asked participants to locate a sound by physically moving to it. Sound position was updated in real time using distance and orientation cues depending on the user's relative position with respect to the target. The authors found that people could quickly get to the target sound source however; there was a significant delay until participants were convinced that they were actually on target.

Feedback can also be provided by adjusting display parameters to give hints on the position of display elements. This can be done using information about user position, obtained through orientation or position tracking equipment. An example of such a technique would be updating the loudness of the display element at which the user is pointing. Given the applicability of such options in display design, it is interesting to evaluate different movement-coupled feedback cues and rate them according to the benefit they bring to interaction.

Another factor that is important from a design point of view is the intelligibility of the audio display elements. When interacting with a spatial audio display a user is faced with a complex audio environment where multiple sounds might coexist (including sounds from the real audio environment surrounding the user). From this point of view, interaction with a spatial audio display is highly associated with divided and selective hearing attention tasks [21]. Divided attention tasks are those in which the user must follow more than one information stream at a time. Selective attention tasks are where attention is focused on only one of multiple information streams. For example, listening simultaneously to two speakers in a teleconference scenario is a divided selection task, since the user is required to understand the 'meaning' conveyed by both of the display elements. On the other hand, the task of selecting a target audio element is a selective attention task, since the user has to focus on the target element with the rest of the display elements acting as distracters. Intelligibility problems in both cases mainly stem from the phenomenon of masking.

Masking is defined as the process or the amount by which the threshold of audibility for one sound is raised by the presence of another sound [17]. When audio display elements are presented simultaneously masking can lead to one or more of them being inaudible if there is significant spectral overlap between them together with marked level differences [17]. In general, masking in spatial audio display is less of a problem due to the fact that both target signals and maskers in a spatial audio display possess spectral and temporal structure that has been proven to assist auditory stream segregation. In addition, in spatial audio displays sounds are presented from different spatial locations. In such cases, the masked threshold is lower compared to when sounds are presented from the same locations. This phenomenon is called *binaural release from masking* and is one of the advantages that spatial audio displays have compared to non-spatial audio displays.

Binaural release from masking has been used to explain the ability of the human auditory system to focus in one of multiple audio streams that are presented simultaneously, known as the 'cocktail party effect' [2]. The individual differences of sound signals between ears have been found to be helpful in reducing the threshold of audibility for sounds presented in the presence of maskers. For all these reasons, performance in selective attention tasks is acceptable in spatial audio displays as long as the levels between display elements do not have big differences [21]. It should be noted that divided attention tasks are more demanding and the benefit from spatial separation is less than in selective attention tasks. As reported in [21], at 0 dB target to masker ratio participants performed at a success rate of 95% in the selective attention task but the success rate in the divided attention task was only slightly more than 70%.

In addition to the aforementioned masking type, also known as energetic masking, there is also the case of informational masking. Informational masking stems from the observation that high levels of masker uncertainty result in higher masked thresholds [12].

Given this observation it is reasonable to assume that consistency in the timbre of the display elements is an aid to a user interacting with a spatial audio display. In addition, spatial separation also improves performance and reduces the effect of informational masking [8]. It has also been found that previous knowledge of the position of the a target is beneficial to intelligibility performance in selective and divided attention tasks so keeping display elements fixed relative to the user may prove beneficial. Therefore, in a familiar spatial audio display the amount of informational masking is expected to be minimal.

Interaction with the display is accomplished in this study using simple physical gestures that are recognized by the system using motion trackers. Physical gestures are a suitable solution for interacting in mobile contexts due to the fact they can be performed without using stylus or similar devices. In an experiment by Pirhonen *et al.* [18], gesture control was found to be superior to stylus based control devices when compared in a mobile setting. In addition, the utilization of data form motion trackers for the recognition of gestures is a promising solution for mobile human computer interaction given the possibility of filtering out the effect of movement, thus providing the user with an experience that is not affected by the variability of movement.

According to the results presented in the review, spatial audio displays seem to favor both types of attention tasks in terms of intelligibility and masking avoidance. However, the effect display 'clutter' has on interaction is not clear as most of the studies focus on intelligibility rather than performance. From this point of view, it is interesting to examine how and whether user performance would be influenced by variable levels of display content (as would happen in any real system).

Two design aspects that justify further investigation have been identified in this literature review. The first is feedback design. Feedback is an important design tool to compensate for the speed and accuracy deficiencies associated with pointing in the audio modality. In particular, we will investigate movement-coupled feedback and compare it with the case of direct pointing since it is expected that such feedback will reduce uncertainty with respect to sound direction. The second is the effect of distracting sounds in the display. Although this effect has been widely investigated from a perceptual point of view, its effects on interaction have not been evaluated in the past. Based on these observations we define the aim of this study as to evaluate a number of prominent feedback cue types and study how interaction assisted by each feedback cue is affected by increasing the number of display elements.

2. FEEDBACK CUES

We are interested in movement-coupled feedback cues. These types of cues continuously update a certain parameter of the audio signal which can then be interpreted by the user to make inferences on his/her pointing direction relative to the direction of the display elements. Consequently, these cues can provide richer information with respect to target direction and thus enhance localization accuracy. Three types of such cues are examined in this study.

A prominent example of a movement-coupled cue is orientation update. This updates the audio scene in real-time based on the orientation of the user relative to the orientation of the display elements. This particular process has also been termed 'active listening'. Such a technique is usually implemented using a head-tracking device that provides the orientation of the user's head and this information is delivered to the spatial audio engine, which updates sound positions in real time. As a consequence, a sound that is defined to be to the left of the user will appear as in front of him when the user's head is facing left.

Orientation update is important from a design point of view. Two displays classes, namely egocentric or exocentric, are defined by whether this feature is active or not. In egocentric displays, orientation update is not used and thus display element positions remain fixed to the user irrespective of orientation. Such a design appears to be a natural choice for applications featuring tasks that are highly repetitive. For example, a mobile menu interface would be best presented by an egocentric display. When mobile, users change orientation often and thus keeping the display elements fixed to them would help them to remember their positions and preserve the pattern of the interaction. In exocentric displays, display element positions are updated in real time relative to the user orientation and appear to be fixed to the world. Exocentric displays are useful in mobile situations in assisting navigation tasks. For example an orientation updated beacon could be used to help a mobile user find his/her way, by orienting to keep the beacon in front. Such an option would not be available by an egocentric display. It is interesting therefore to study performance in using the orientation update cue from a mobile human computer interaction perspective.

From a usability point of view, as found by Wenzel [22, 23], active listening helps to alleviate up/down and front/back confusions in sound direction. However, its effectiveness on localization accuracy has not been validated yet. It is the case however, that such a cue can theoretically improve accuracy if used in the appropriate way. For example, a user interacting in such a display can move until the sound appears to be in front. This action is quite beneficial from an accuracy point of view since sound localization is most accurate for sounds in front of a user. We can expect therefore improvements in selection accuracy when such a technique is used.

Another prominent cue that can be coupled to pointing direction is loudness update. Such a cue can be designed by means of a function that relates the attenuation applied to each display element to the user's distance from the display element. Continuous or discrete attenuation levels can be used, the latter done through mapping of attenuation levels to different ranges of user distance to target. A loudness based cue can guide the user to a target display element location since the loudness of the particular element will increase as the user moves closer to the display element. In addition, such a cue effectively adjusts the target to masker ratios in the display and as such it is helpful in the context of enhancing divided as well as a selective attention. This is also important in mobile settings as it can help overcome problems of masking by display or other real world sounds. At very high attenuation levels the loudness cue is effectively reducing display population, since elements far from the user's pointing direction will be rendered inaudible. This might become problematic since continuous contact with display elements is not preserved.

The last feedback cue of interest is a simple timbre cue. To provide this cue, a different timbre is associated with each display area and only one sound is audible at a time depending on the direction in which the user is pointing. This cue has the advantage

of reducing mental demand since only one display element is audible at a time. However, this cue will also result in loss of continuous contact with all of the display elements; potentially reducing usability.

The rest of this paper presents results of an evaluation of interaction in the presence of the aforementioned three cue types, input being accomplished by means of a physical pointing gesture. We evaluate them in a display with varied display populations to obtain information on how this will affect performance. By the term display populations we refer to the number of display elements that are presented in the display.

3. EVALUATION METHOD

Performance is evaluated by measuring time and accuracy scores in a spatial audio target acquisition task. In addition, we employ two additional standard measures: effective target width and throughput. For a discussion on measures used to evaluate pointing efficiency please refer to [14]. Throughput is defined as:

$$Throughput = \frac{ID_e}{MT} \quad (1),$$

ID being the index of difficulty and MT the movement time. Index of difficulty is defined as:

$$ID_e = \log_2\left(\frac{D}{W_e} + 1\right) \quad (2),$$

W_e is the effective target width calculated based on the standard deviation of measurements and represents the distribution of selection coordinates computed over a sequence of trials. It is calculated as:

$$W_e = 4.133 \times SD_x \quad (3).$$

SD_x is the standard deviation in the selection coordinates measured along the axis of approach to the target. To apply the above formulation in our study we define D to be the angular distance participants had to move to reach the target, measured in degrees. The particular measures have been proposed for evaluating visual target acquisition, however in this paper we attempt to extend their application to spatial audio selection tasks. This is justified because spatial audio is providing a directional cue and therefore the spatial audio target acquisition procedure can be considered similar to visual target acquisition. As far as effective target width is concerned the application is not questionable due to the fact that Equation 3 does not involve any terms that can be thought to relate to modality. With respect to throughput, it might be possible to question the appropriateness of the formulation of Equation 2 as a measure of difficulty of a spatial audio target acquisition task. It is indeed an open question whether the formulation of Equation 2 can be applied to spatial audio selection tasks. However, we proceed with using the formulation and use it uniformly in this study for all feedback cues given, so that it provides useful insight into their effectiveness.

4. EXPERIMENT

We designed an experiment to assess the feedback cues presented in Section 2, as well as combinations of the cues by measuring time and accuracy scores in a physical pointing task in a display with variable distracter populations. The Independent Variables are orientation update (between-subjects), feedback type and distracter population (both within-subjects). Dependent variables are time to complete a trial, angular deviation from target as well as throughput and effective target widths. Participants were split into two groups: one with orientation update enabled in the display and the other without. The feedback type variable was introduced to accommodate the loudness and timbre cue. A void level was used to provide the control condition of direct pointing and to test orientation update alone. The combination of orientation update and feedback type resulted in six different feedback cue combinations. The control condition was provided by the combination of no orientation update and no feedback cue and essentially represents direct pointing. The loudness and the timbre cues were tested with and without the orientation update cue to examine what is the effect of cue combinations. Display population was also tested as a within subjects factor with all participants tested in all available levels of display content. The maximum number of sounds in the display was seven including the target sound and the minimum just one, the target sound. The design of the experiment is presented in detail in Table 1.

The loudness cue was implemented using a continuous bell-shaped attenuation function designed so that when a participant was pointing straight at the location of a display element neighboring elements were played at half their original level. The shape of the function was such that sounds other than the focused and the neighboring ones were played at zero level so that they were inaudible. Attenuation levels were continuous. To implement the timbre cue each display element was assigned an effective area of 20°. When using the timbre cue only the display element in whose effective area the user was pointing was audible. On entering the effective area of a display element, the associated sound was played from the beginning. This type of feedback is similar to the case where a different sound or a variation of an element sound is used to inform the user of a display element being in focus, with the notable difference that no continuous contact exists with the other target elements.

Table 1. Experimental Design.

Orientation Update	*Feedback Type*	*Display Elements*
Yes	None	1,2,3,4,5,6,7
	Loudness	1,2,3,4,5,6,7
	Timbre	1,2,3,4,5,6,7
No	None	1,2,3,4,5,6,7
	Loudness	1,2,3,4,5,6,7
	Timbre	1,2,3,4,5,6,7

4.1 Display Design

Participants were tested in a spatial audio display that is presented in Figure 1. Display content was constrained to a maximum of seven audio elements. Elements were positioned on the arc of a circle of radius of 3m starting from -70° and up to 70° with an inter-element distance of 20° at the level of the user's nose. Interaction was accomplished by means of an Xsens MT-9B orientation tracker (www.xsens.com), which participants held in their hands. Headphones were used to present the sounds.

Figure 1. Experimental Task, Visualization of the hand of a participant alternating between the two targets while tested in the display.

When the display featured orientation update, sound positions were updated automatically based on the direction of the user's hand. The DiselPowerEngine (www.am3d.com) was used for spatial audio rendering of the display elements.

4.2 Experiment Task and Participants

One target sound appeared in the display in each trial and this was a human voice saying 'Woohoo'. To provide a rather uniform sound material, animal sounds were used for the rest of the display elements that were used as distracters. The distracter population was chosen randomly for each particular trial out of a maximum of six sounds. These were the sounds of a kitten meowing, a puppy barking, a horse whining, a cockerel crowing, a cricket chirping and a hen clucking. The sounds were equalized to have roughly the same loudness. Sounds were HRTF-filtered in real time to provide the impression of them emitting from a certain position in space. Sounds were programmed to play as omni-directional sound sources and no directional characteristics were used.

The experimental task was to select the target sound using the feedback cues that were available in each condition. The sounds in the display repeated with a 400msec pause until a selection was made. Onsets and durations were not constrained. To select the target sound participants had to point at its perceived direction with their hand and then rotate the hands slightly downwards to indicate selection. The target sound alternated between the left-most and rightmost display slot in every other trial. This was done in order to minimize searching time and allow for the effects of the distracting sounds and the combinations of the feedback cues to appear. The number of display elements was selected randomly prior to a new trial out of a maximum of seven. Display element positions were filled from the direction of the target according to the number of display elements for each trial. A visualization of the experimental task is provided in Figure 1.

When the display featured the orientation update cue participants were asked to use the updated orientation cues and select a sound when was in front of them. When using the loudness cue participants were asked to select when the sound was at its loudest. When using the timbre cue participants were asked to select when they could hear the target sound. When using combination of cues participants were asked to use both cues on their own initiative. There was a counterbalanced testing order for the within subjects factors. Pointing angle at selection and time to complete trials were recorded throughout the experiment. Sixteen participants were tested. Participants had no previous experience with interacting with an audio display.

5. RESULTS

The results section is organized in two subsections. The first is concerned with the analysis of movement times, the second with the analysis of deviation from target scores and additional observations concerning throughput and effective widths as these were calculated for each feedback cue and their combinations.

Table 2. Between and within subjects effects in time to select measurements.

Independent Variable	*Significance Level*
Orientation Update	F(1,94) = 8.524, p=0.004
Feedback Cue	F(2,188) = 116.541, p<0.001
Number of Display Elements	F(6,564) = 5.731, p<0.001
Update * Feedback Cue	F(2,188) = 74.792, p<0.001
Update * Elements Number	Not Significant
Display Elements * Feedback	Not Significant

5.1 Time Analysis

A (2x3x7) ANOVA with orientation update as a between subjects factor showed a significant main effect of orientation update, display type and number of display elements. The effect of the interaction between the orientation update and the other feedback cues was also significant. The results are presented in Table 2. Mean times as a function of display content for all examined cases can be found in Figure 2.

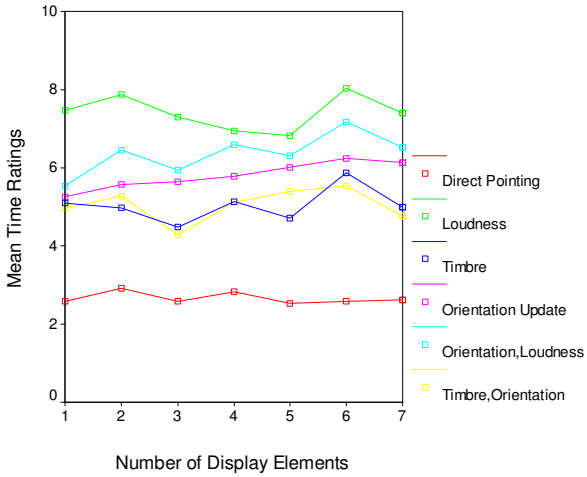

Figure 2. Mean Times to Complete Trials.

Given the main effects observed, *post hoc* t-tests with Bonferroni confidence interval adjustment were performed to check for differences between the different feedback cues. All feedback cues were found to differ significantly with the exception of the two silent interfaces where sounds were presented one at a time. Utilization of the orientation update cue was found to slow interaction. The ordering of feedback cues with respect to speed is therefore: direct pointing, timbre, orientation update, loudness & orientation update and loudness alone (see Figure 2). Interaction with the combination of the orientation and loudness cue resulted in faster interaction compared with the loudness cue alone but slower than

the orientation update cue. It is also interesting to observe that the interface where active listening was enabled was more sensitive to increasing the number of display elements than the interface where orientation update was disabled. The associated curves show a clear increasing trend.

5.2 Accuracy Analysis, Throughput and Effective Target Widths

Accuracy was calculated as the absolute difference between user pointing and target position. A (2x3x7) ANOVA on absolute deviations from target showed a significant main effect of the orientation update cue, display type but not of number of display elements. The interaction between feedback cue and orientation update was significant, as was the interaction of display elements and the three feedback cues. Significance levels can be found in Table 3.

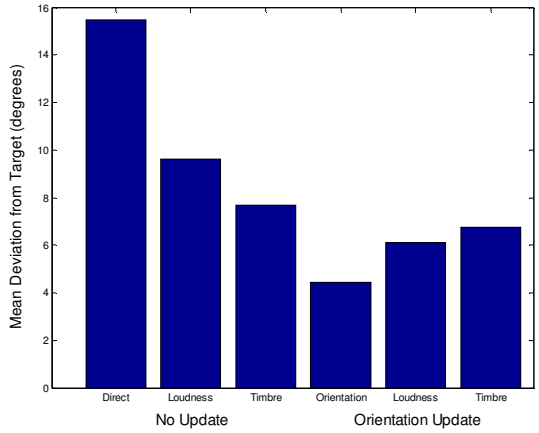

Figure 3. Mean absolute deviation from target means calculated across all display populations.

Given the main effects observed in overall accuracy *post hoc* comparisons were performed for all combinations of feedback cues to order them with respect to accuracy. All feedback cues and combinations were found to differ significantly. Orientation update was found to significantly enhance the accuracy of selections. In general, high standard deviations were observed in user selections. The different feedback cues can be rated with respect to accuracy as: orientation update, loudness and orientation update, timbre, loudness only and finally direct pointing. Mean accuracy ratings for all feedback combinations across all display populations are plotted in Figure 3.To compare the different feedback cues in a more systematic way, we use the measures of *throughput* and *effective target width*. The accuracy ratings in our study exhibited a large amount of between-subject variation. This is due to the influence of throughput and effective width results. To give an indication, the range of throughput and effective width values for the participants of the experiment is provided in Table 4.

The between subject variation can be explained by the skill required by the tasks and the absence of any training. The data presented were measured from participants that had no experience in the sound localization task or in the use of virtual audio feedback cues. It should be noticed that the utilization of such cues is not common in our everyday lives and therefore the relevant skills are not expected to be well developed. Performance will improve through training. The results presented are therefore representative of an untrained population and thus represent a safe approach to design when using the feedback cues under study.

Table 3. Significance levels in accuracy scores.

Independent Variable	*Significance Level*
Orientation Update	$F(1,94) = 854.725$, $p<0.001$
Feedback Cue	$F(2,188) = 43.552$, $p<0.001$
Number of Display Elements	Not Significant
Update * Feedback Cue	$F(2,188) = 36.674$, $p<0.001$
Update * Elements Number	Not Significant
Display Elements * Feedback	$F(12,1128) = 3$, $p<0.001$

Effective target widths for the different feedback types averaged across all subjects and all display populations are presented in Figure 4. The effective target widths when doubled will result in close to perfect selection rates. In terms of target size alone, the different feedback cues are ordered as: orientation update, loudness & orientation, timbre, loudness alone and direct pointing to a static audio target. Throughputs were calculated according to Equation 1, and are presented in Figure 5. We observe that, despite the rather large effective target widths, direct selection proved to be the most efficient when throughput is concerned.

Table 4. Throughput and effective target width variations between participants.

Cue	*Throughput*	*Width*
Direct pointing	0.55-1.33	22 – 54
Loudness	0.29-0.53	21 – 28
Timbre	0.4-0.95	14 – 28
Orientation update	0.51-0.78	8-12
Loudness & Orientation	0.38-0.94	12-19
Timbre & Orientation	0.4-1.18	13-25

The rating of the different feedback cues in terms of throughput is: direct pointing, orientation update, timbre, loudness and orientation, loudness alone. Throughput comparisons are quite useful since they combine accuracy and timing information in one uniform measure. An ANOVA was performed on throughput and target width measures as these were obtained for each participant and for all display populations in the experiment. The result showed that display population was not a significant factor, for this reason throughput and effective width data are presented averaged across all display element populations.

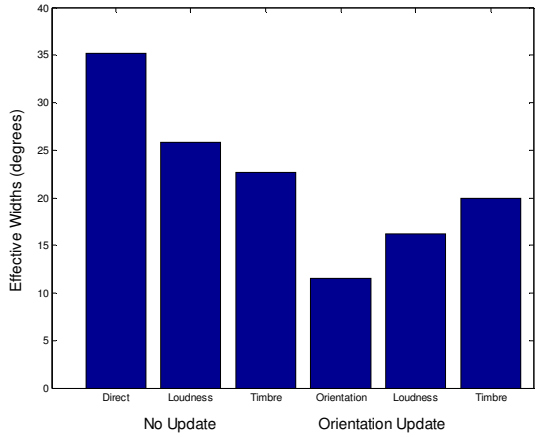

Figure 4. Effective target widths for the different feedback cues calculated across all display populations.

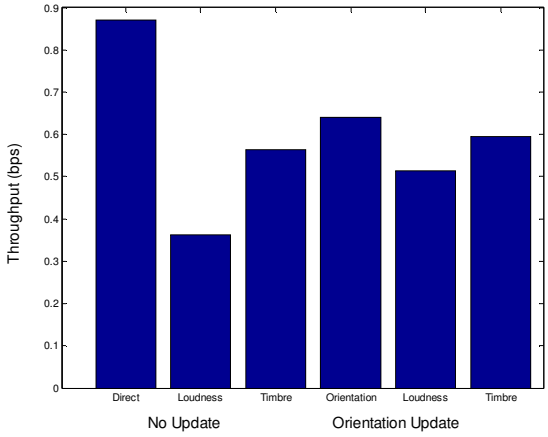

Figure 5. Mean Throughput for the different feedback cues calculated across all display populations.

6. DISCUSSION

One of the major findings of this study is the time/accuracy trade-off that is associated with movement-coupled cues. Although there is an improvement in accuracy, the high timing demands compromise the benefit. This is quite evident in the throughput ratings where the accuracy superiority of the movement-coupled cues was cancelled out by the increased movement times. The reason for the increased time demand is that it requires continuous target attainment due to the fact that each movement users make affects the perceived soundscape and forces them to think about current position with respect to the target. This can result in increased time taken especially when users are close to their final position. The demand for continuous target attainment would be relaxed by providing discrete levels for certain distance to target intervals. Increased final positioning times have been associated with orientation update in other studies, such as [11]. On the other hand, movement-coupled cues were shown to be useful in reducing target size, as is shown by the effective target widths associated with these tasks. The most successful case of the orientation update feedback cue required just one third of the effective width required by direct pointing.

The loudness cue rated reasonably well when used in combination with orientation update, however when used alone was less effective. In this sense, the loudness cue is not very useful in assisting pointing based interactions. However, due to the fact this cue is effectively adjusting focused element to distracter element level ratios, it can be quite useful in assisting selective and divided attention in the display, an option that can be very useful when in mobile settings. The timbre related cue, present in the silent interface, rated quite well in terms of time and accuracy. However, its success was limited due to the lack of continuous contact with the target. This type of cue resulted in a searching action that reduced interaction speed. In terms of throughput, this type of cue was competitive with the orientation update cue.

It is worth considering the results of this experiment with respect to designing a spatial audio pointing task. The results show that the different cues alone are not sufficient due to the time-accuracy trade-offs that were observed.

Solutions to this problem can be sought in cue combinations. As has been found, performance ranges between the limits set by the individual cues when they are combined. For example, combining the orientation update cue (which was more accurate than the loudness cue) with the loudness cue resulted in more accurate performance than the loudness cue alone but was less accurate than the orientation cue alone. A similar trend was observed for the rest of the cue combinations. Combining cues results in a compromise, which can be also used constructively to enhance cues that need to be used but are lacking in a certain interaction aspect. For a task that will be performed repeatedly, combining direct pointing with a timbre cue would result in an interaction that is fast and accurate. The evaluation of such a case can be found in [15], where the above prediction is verified.

The experiment also focused on the effect of distracting sounds on interaction. Interaction using the non-movement coupled cues was not affected by increasing the number of distracting sounds. This is the case in direct pointing and in the timbre cue, where time to select and accuracy of selection was not affected by increasing the number of sounds in the display. For the rest, a rising trend on movement time was observed, however with no significant effect on accuracy. Time to select ranged between 5 sec. for one target and up to 6 sec for 6 or 7 elements in the display in the orientation update case, an increase of 20%. In the loudness case, a similar trend was observed.

The results of this study can be used to design improved spatial audio window applications. They can be useful in predicting performance when using a certain cue in the display and deciding on possible combinations of cues. Depending on the requirements of an application, a designer might use the results to decide on the use of particular cues that can be effective in terms of time to complete, scalable enough and can be performed with low error rates. Due to the mobile aspect of this type of interaction this study can help in the design of usable mobile applications that take advantage of the audio modality and gesture recognition to facilitate interaction and overcome the problems that stem from the variability imposed by movement.

7. CONCLUSIONS

A study comparing feedback cues with the objective of enhancing pointing efficiency in deictic spatial audio displays was presented. Participants were tested in a systematically varied display environment to examine the effect of distracter display elements on interaction. Movement-coupled feedback cues effectively reduced effective target widths, but the efficiency of the cues was found to be compromised due to the reduction in speed caused by the requirement of continuous target attainment these cues impose. Movement-coupled cues were also found to be sensitive to display population, direct pointing cues not being affected significantly. Feedback cue combinations were found to improve the less effective cues but degrade the more effective ones. Lack of continuous contact with the target was found to negatively influence interaction speed. The results reveal that spatial audio display design is challenging, but with appropriate design it is possible to overcome interaction uncertainty and provide solutions that are applicable in mobile human computer interaction.

8. ACKNOWLEDGMENTS

This study was supported by the AudioClouds project (www.audioclouds.org), EPSRC grant number GR/R98105.

9. 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] Arons, B., A Review of the Cocktail Party Effect. Journal of the American Voice I/O Society, 1992. 12: p. 35-50.

[3] Blauert, J., Spatial Hearing: The psychophysics of human sound localization. 1999: The MIT Press.

[4] 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

[5] Cohen , M., Throwing, pitching and catching sound: audio windowing models and modes. Int. J. Man - Machine Studies (1993), 1993. 39: p. 269 - 304.

[6] Cohen , M. and Ludwig , L., Multidimensional Audio Window Management. Internaltional Journal of Man - Machine Studies, 1991. 34: p. 319-336.

[7] Fitts, P. M., The informational capacity of the human motor system in controlling the amplitude of movement. Journal of Experimental Psychology, 1954. 47: p. 381-391.

[8] Freyman, L. R., Balakrishnan, U., and Helfer, S. K., Spatial Release from Informational Masking in Speech Recognition. The Journal of the Acoustical Society of America, 2001. 109(5): p. 2112-2122.

[9] Goose, S. and Moller, C. A 3D Audio Only Interactive Web Browser: Using Spatialization to Convey Hypermedia Document Structure. in 7th ACM international conference on Multimedia, 1999. Orlando, Florida, United States: ACM Press. p. 363 - 371

[10] Kobayashi, M. and Schmandt, C. Dynamic Soundscape: mapping time to space for audio browsing. in SIGCHI conference on Human factors in computing systems, 1997. Atlanta, Georgia, United States: ACM Press. p. 194 - 201

[11] Loomis, J. M., Hebert, C., and Cocinelli, J. G., Active Localization of Virtual Sounds. Journal of the Acoustical Society of America, 1990. 88(4): p. 1757-1764.

[12] Lutfi, A. R., How much masking is informational masking. The Journal of the Acoustical Society of America, 1990. 88(6): p. 2607-2610.

[13] MacKenzie, S. I. and Buxton, W. Extending Fitt's Law to Two-Dimensional Targets. in ACM CHI, 1992. p. 219-226

[14] MacKenzie, S. I., Kauppinen, T., and Silfverberg, M. Accuracy Measure for Evaluating Computer Pointing Devices. in SIGCHI, 2001. Seattle, USA. p. 9-16

[15] Marentakis, G. and Brewster , S. Effects of reproduction techniques on interaction with a spatial audio display. in Vol. II, ACM CHI 2005 , Portland, Oregon. p. 1625-1628.

[16] Marentakis, G. and Brewster , S., A. A study on gesture interaction with a 3D Audio Display. in Mobile HCI, 2004 (Glasgow, UK) Springer LNCS Vol. 3160, p.180-191

[17] Moore , B. C. J., An Introduction to the Psychology of Hearing. 3rd ed. 2001: Academic Press Limited, San Diego, CA. USA.

[18] Pirhonen , A., Brewster , S., and Holguin, C. Gestural and audio metaphors as a means of control for mobile devices. in Conference on Human Factors in Computing Systems., 2002. Minneapolis, Minnesota, USA: ACM Press New York, NY, USA. p. 291-298

[19] Savidis, A., Stephanidis, C., Korte, A., Krispien, 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] Shinn-Cunningham Barbara and Antje, I. Selective and Divided Attention: Extracting Information from Simultaneous Sound Sources. in International Conference on Auditory Display, 2004. Sydney, Australia.

[22] Wenzel M., E. Effect of Increasing System Latency on Localization of Virtual Sounds. in Audio Engineering 16th International Conference on Spatial Sound Reproduction, 1999. Rovaniemi, Finland: New York: Audio Engineering Society. p. 42-50

[23] Wenzel M., E., Marianne, A., Kistler, J. D., and Wightman, L. F., Localization using nonindividualized head-related transfer function. Journal of the Acoustical Society of America, 1993. 94(1): p. 111-123.