A STUDY TOWARD THE DEVELOPMENT OF A SPATIAL, NON-SPEECH AUDITORY INTERFACE FOR TRIGONOMETRIC PROBLEM SOLVING

Flaithrí Neff
Ian Pitt

University College Cork

Cork
Ireland
fn2@cs.ucc.ie

ABSTRACT

There are numerous difficulties for visually disabled students when tackling mathematical problems. This relates more to methods of presentation rather than to any deficiency in the students' abilities. Although presentational advances have been made in some instances, such as for algebraic equations, problems remain when attempting to convey inherently spatial mathematics such as trigonometry or matrices. The linearity of speech and Braille output is not easily mapped to spatial attributes and therefore other methods may prove more useful in this regard.

We suggest the use of non-speech spatial sound to convey an overview of trigonometric shapes. Our aim is to provide a rapid overview without relaying specific information such as angle degrees or side lengths. Later we plan to use speech and virtual navigation to enable the user to extract precise information if required while retaining the ability to revert to an overview at any stage.

Our current concern is how to relay a relatively accurate picture of a trigonometric shape to the blind student using non-speech spatial audio. We therefore examine various non-speech methods of notifying the user to the presence of an angle. We compare various methods for time efficiency and accuracy. We use Microsoft XNA/XACT technology to render the non-speech, spatial sound streams and employ a User Interface Model to consider the psychoacoustic elements involved.

1. INTRODUCTION

For sighted users, an overview of a trigonometric problem often comes in the form of a diagram. This serves to quickly contextualize the problem by displaying both declared and missing information. As a result, goals and appropriate solutions can be identified by the student to solve the problem while retaining the diagram overview for reference. Visually disabled students have neither the facilities for an overview nor a non-linear method of obtaining the precise information required to solve the problem.

Our long-term research goal is to develop a robust, immersive auditory environment with navigation capabilities. We aim to relay an initial overview of inherently spatial mathematics to the user but subsequently allow user-controlled orientation to activate specific speech-based information. The initial step in the implementation of this goal is to determine the most efficient and accurate means of portraying an overview of trigonometric shapes. We examine various methods of relaying non-specific angle information using spatial non-speech sound.

Currently, our sound is output using a 5.1 surround sound system, but we aim to implement our design at a later stage using binaural sound techniques via standard headphones. Future implementations will not only include trigonometry, but also other forms of mathematics that incorporate spatial elements such as matrices.

The main objective in this paper is to examine user response times and accuracy levels relating to different auditory stream designs that represent angles. We hope to isolate the most efficient and accurate design before continuing with further developments of our overall system. We are aware of the considerable cognitive issues involved at this stage of our work and have designed a User Interface Model to help us more fully understand the processes. In this paper, we describe a pilot study in which we compared different designs that implement the rules of our interface model to varying degrees. The aim is both to test our interface model design and also to move forward in our system implementation.

1.1. Review of Math-related Technology for Blind Users

Along with inherently spatial mathematics, linear mathematics can also be problematic for visually disabled students. Speech and Braille tools are often an effective means for solving linear mathematical problems, albeit sometimes slow and burdensome on human memory. LAMBDA [1] is probably one of the most significant attempts to provide blind students with the means to access mathematics via speech and Braille. Consisting of a mathematical editor, it allows a student to progress through a problem using speech and Braille output to relay his/her steps. Also, being MathML compatible, it allows sighted teachers to interpret the LAMBDA code in a more conventional visual manner. However, it is still a linear system, attempting to convey spatial elements in inherently spatial mathematics in a linear fashion. Even in linear mathematics, such as algebra, linear output methods begin to lose their effectiveness when equations become more complex and pose a heavy load on human memory.

The MATHS [2] and MAVIS [3] projects used non-speech sound to relay structural information about an algebraic equation to the user. These methods were often effective at alleviating cognitive overload implicit in delivering algebraic equations using speech, but the non-speech features of the

MATHS and MAVIS projects were sometimes too abstract to provide an accurate interpretation of the structure of a complex equation. Although spatialization was examined in the MATHS project, it was not fully implemented in the final version.

Perhaps the most utilized method in presenting spatial elements in mathematics is via tactile devices. Traditional methods such as German Film and Fuzzy Felt have often formed the basis of digital tactile devices. Projects such as NOMAD [4] demonstrated the need for supportive information (speech and Braille) when presenting abstract shapes on a tactile device. As a result of this, commercial tactile devices such as the IVEO touchpad [5] incorporate speech and do not rely solely on tactile methods. Aural supportive information becomes especially critical with respect to congenitally blind individuals when presenting a three dimensional shape in a two dimensional format.

In contrast, enhanced auditory information can retain its three dimensional qualities using conventional hardware. Also, because visually disabled people regularly interact with speech-based technology and audio games, such constant exposure to computerized audio stimuli may mean that the human auditory system is becoming very accustomed to sonic interpretation in comparison to its tactile counterpart.

2. USER INTERFACE MODEL

Our user interface model [6] [7] outlines the human auditory pathway from the peripheral sensory system to higher cognitive mechanisms. We acknowledge that our understanding of some of the auditory system remains hypothetical and that many of the issues are complex. We are of the opinion, however, that an appropriate model can be founded on contemporary perceptual theory [8] [9] [10] [11] [12] [13] in order to improve auditory interface design.

Our user interface model comprises three primary blocks – a Sensory Filter, a Subtask Attention and Inhibition Manager (SAIM), and a Higher Processing Mechanism.

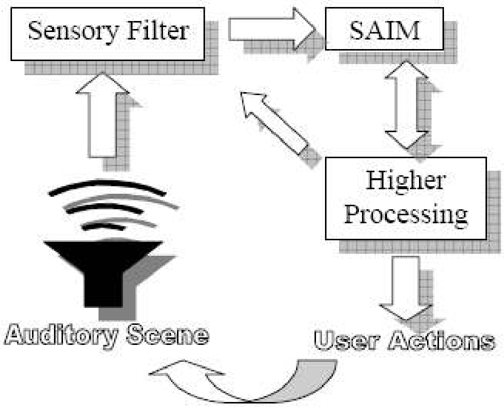

Figure 1. *A summary of the User Interface Model. An auditory scene is presented to the user. The Sensory Filter allows some streams to pass while blocking many others. The SAIM acts as a further filter, determining which stream is allowed into focused and peripheral attention. Depending upon the attention mechanism to which a stream has been allocated, the Higher Processing Mechanism performs processing on each stream – including memory oriented tasks.*

We base the rules of the Auditory Scene in our model on the work of Bregman [9]. A complex auditory scene is segregated into one or more auditory streams, depending on their acoustic makeup. This segregation process is autonomous and performed by the most peripheral mechanisms of the auditory perceptual system.

Depending on the complexity of the auditory scene, the number of streams presented to the auditory system may be vast and therefore some form of filtering is required so as to avoid cognitive overload. The Sensory Filter achieves this by blocking some streams and allowing others to pass. We base the rules of the Sensory Filter on Schema Theory [12], which is highly reliant on user experience. Therefore, the more experience the user has, the more appropriate the schema template will be in relation to the auditory scene at hand. An inexperienced user will present an unsuitable schema template and therefore the Sensory Filter will inaccurately determine which streams pass and which streams don't. A list of many schema templates is stored in memory and controlled by the Higher Processing Mechanism [6] [7]. The schema process depends on experience, building on the schema list dedicated to various scenarios and their variations.

The Subtask Attention and Inhibition Manager acts as another filter and is also controlled by the Higher Processing Mechanism. It is a mechanism for constraining access to vital processing in human memory. Streams that have particular acoustic traits grab attention, even if they are not critical streams. Indeed, this scenario may take focused attention away from the critical stream and therefore the design of sound objects at the initial stages (the auditory scene) needs careful consideration. As outlined in figure 2, only one stream at a time is allocated to Focused Attention while all others are allocated to Peripheral Attention [13]. Focused Attention is tied to the Focal Buffer in the Higher Processing Mechanism, whereas Peripheral Attention is tied to the highly volatile Peripheral Loop. Only the Focal Buffer has access to human memory and therefore any streams in the Peripheral Loop cannot be rehearsed and encoded into memory.

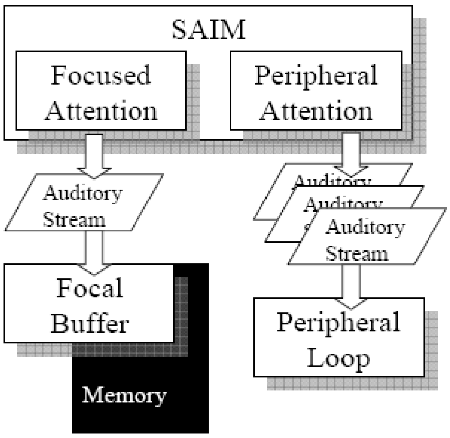

Figure 2. *Streams in Focused Attention go to the Focal Buffer which is linked to Memory and the Rehearsal Process. Stream in the Peripheral Loop are sent to the Peripheral Loop which does not have a link to Memory and therefore cannot be stored in Memory.*

The concept of memory in our model is based on the Changing-State Hypothesis [11]. We integrate this theory into our model by deducing that although streams in the Peripheral Loop are not linked to Memory directly, they may still interfere with the critical stream in Memory by pulling on the top-down rehearsal process.

As is evident from our predictions in our User Interface Model, there are many cognitive issues from peripheral stages of the auditory pathway to central higher processing stages that need to be carefully evaluated when designing an auditory interface for trigonometry.

The entire process depends on the initial design of the auditory entities of our auditory scene (i.e. the sounds we use and map to trigonometric elements in our overview system). If these initial auditory entities do not conform to the systematic arrangement of the auditory perceptual system, then we predict that difficulties will arise at one or more blocks described in our model.

Auditory entities that comply with the arrangement of our auditory system (as outlined in our model) should result in optimal access to cognitive resources. In an idealistic auditory interface design, the most vital information at any given time is relayed by one stream as defined by the user and situation. It should easily pass through the Sensory Filter, be allocated to the most appropriate Schema template, and assigned to Focused Attention. It should subsequently have access to Memory, resulting in its robust storage and access to further Higher Processing. Each sequence of vital information would be streamed in this fashion until a comprehensive interpretation of the problem is assembled. The entire flow therefore hinges on the design of the initial auditory objects in our scene.

3. AUDITORY SCENE DESIGN – PILOT STUDY

In a previous pilot study [14] we examined a number of different designs that conveyed angles in a trigonometric shape. Some of the designs did not comply with our User Interface Model while others complied to varying degrees with our model. Those that did not comply consisted of individual, short sine-tones denoting angles, and we found that the accuracy levels were weakest in these cases. Those that complied more with our model consisted of continuous sine-tones and white noise.

We found that the individual but related sine-tones were not perceived as one stream, but rather three individual competing streams. This broke rules relating to simplifying our auditory scene by promoting fewer streams, and also impacted on rules associated with attention mechanisms. Furthermore, users could not locate the position of sine-tones accurately in 3D space.

Our initial design also mapped pitch to angle size but this feature was not used by subjects. According to our model, such redundant information can cause interference with critical streams, thus undermining our auditory scene design.

With regard to the conditions that complied to varying degrees with our model's predictions, the accuracy levels and localization significantly increased. With continuous traveling sound we were able to enhance the 3D awareness by employing techniques such as Doppler Shift. The traveling sine-tones and traveling white noise did not show significant differences between each other and therefore need further study.

3.1. Pilot Study Introduction`

Following up on the information provided in our previous study, we needed to test various designs incorporating continuous sound consisting of both sine-tones and white noise specifically aimed at angle detection. Therefore, instead of a full triangular shape, we wanted to examine angle detection in isolation. We decided to test groups of traveling sine-tones and white noise that changed direction once (i.e. one angle). We determine that single traveling signal conforms to our model because only one stream is interpreted rather than three in our previous study.

There were a number of reasons for choosing sine-tones and white noise. Using sine-tones in the current pilot study meant that we could relate to our previous pilot study. However, more significantly, we wanted to evaluate performance based on two signal extremes – sine-tone with only the fundamental frequency present and white noise containing a flat, wide frequency spectrum.

Although no vertical localization was required and all trials were based on azimuth placement, we were curious to see if, in this context, the rich white noise signal performed better than the sine-tone for detecting direction of the moving signal. Another aspect that interested us was to determine if subjects reacted better or worse to white noise versus sine-tone when the angle was emphasized using Flanger and Wahwah effects.

		Sine-tone	**White Noise**
Frontal Hemisphere Trials	Without any angle indication	12 angles No effect for angle.	12 angles No effect for angle.
	With angle indication	12 angles. Flanger effect = angle.	12 angles. Flanger effect = angle.
	With emphasized angle indication	12 angles. Wahwah effect = angle.	12 angles. Wahwah effect = angle.
Surround Sound Trials	Without any angle indication	12 angles. No effect for angle.	12 angles. No effect for angle.
	With angle indication	12 angles. Flanger effect = angle.	12 angles. Flanger effect = angle.
	With emphasized angle indication	12 angles. Wahwah effect = angle.	12 angles. Wahwah effect = angle.

Table 1: A summary of the trials conducted. Every sine-tone and white noise condition (absent/present/emphasized textural indication) was tested in frontal hemisphere and surround sound. Twelve varying angles were tested with each condition.

3.2. Pilot Study Overview

Four sighted subjects performed all trials as outlined in Table 1. This was a pilot study with a limited number of subjects intended to inform us on the design of a subsequent, larger-scale study. The subsequent study will comprise more subjects in order to properly evaluate results and obtain statistical significance.

An equal number of sine-tone and white noise trials were conducted. All trials were tested in frontal hemisphere and surround sound scenarios. Each trial comprised twelve different angles presented randomly. User response times and user accuracy were measured on each trial and variation.

User response times were recorded from when the sound began until the sound ended. Angles occurred at 1500ms and those with angle indication had a textural variation starting at 1500ms and ending at 2000ms. Therefore, subjects' time response was compared with this constant (between 1500ms and 2000ms).

Subjects were asked to press a mouse button when they thought an angle was presented (i.e. when the sound changed direction) between sound commencement and sound ending. We used a Logitech G9 laser mouse because of its very fast report rate (1000/second) and the Logitech DIView Application version 4.65.116 [15]. This application records mouse button activity and exports the data to a text file. We examined the text file and noted response times (i.e. mouse button presses).

To assess user accuracy, subjects were required to draw the angle as accurately as possible after only one rendition of each trial. No trial was repeated, and therefore if the subject did not draw the angle, 0% accuracy was recorded. The accuracy levels were evaluated by comparing the exact angle measurement produced by the system with that produced by the subject. This comparison was converted to accuracy percentages in a spreadsheet along with the standard deviation (*σ)* in each case.

The host machine we used was a Dell DIMC521 with a 1.9GHz AMD Athlon Dual Core Processor with 1GB of RAM [16]. To build our virtual environment we used Microsoft XNA Game Studio 2.0 with its associated XACT audio engine and X3DAudio specialization helper library [17]. This ran on Microsoft Windows Vista Home Basic edition. The 5.1 surround sound hardware used was Sigma Tel 9227 audio card [18] and Typhoon speakers [19] (see figure 3).

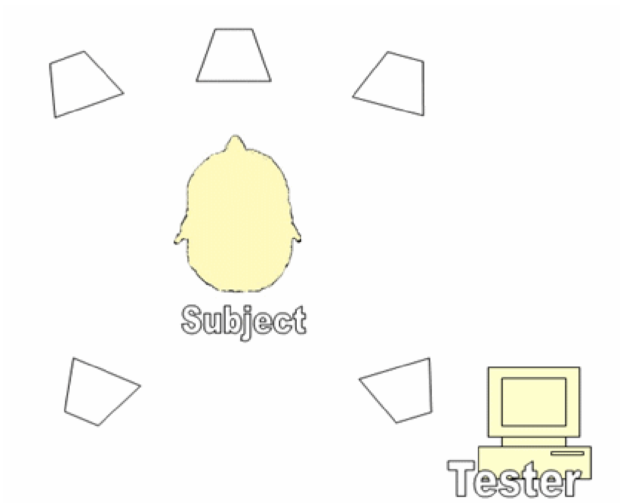

Figure 3. *The subject was positioned in the sweet-spot of a surround sound setup. The sub-woofer was not needed or utilized in this study. 5 satellite speakers were arranged around the subject, each 1 meter from the subject's head. Each speaker was positioned at ear level. The tester sat behind the subject to run the programs, accept data and observe the subject.*

3.3. Response Time - Sine-tone & White Noise - Frontal & Surround – No Angle Indication

Refer to figure 4. Users were required to press the right button on the mouse when they heard a change in direction (i.e. an indication of an angle). Response time was very inaccurate in all cases (sine-tone and white noise, frontal and surround) with most mouse clicks occurring well before the actual angle event. This indicates that the subjects could not clearly define when the continuous sine-tone without textural changes presented an angle. This led to guessing on the part of the subject. Therefore, although compliant with some of the rules of our model, non-textural direction change was not strong enough to pull on focused attention. Although the extra frequencies included in white noise should have helped in terms of localization, it does not seem to be significant enough in this scenario and therefore an extra feature is required to indicate to the user when an angle has been presented.

3.4. Response Time - Sine-tone & White Noise - Frontal & Surround – With Angle Indication

Refer to figure 5. The introduction of a slight textural change in the continuous sine-tone when an angle was presented, significantly improved user response time accuracy in both frontal hemisphere and surround, as shown in figure 5. The textural change was achieved with a slight flanger effect at the point of the direction change (angle). This was also the case for white noise which also included a flanger effect.

3.5. Response Time - Sine-tone & White Noise - Frontal & Surround – With Emphasized Angle Indication

Refer to figure 6. With such a significant difference in accuracy levels in the response times due to the introduction of a textural indication for the angle, we decided to emphasize the angle indication. We altered both the texture and the pitch using a wahwah effect. A further slight improvement was achieved using this technique but needed to be assessed in light of angle accuracy results examined later.

3.6. Accuracy Levels

Subjects were given the task of drawing two lines with direction change (i.e. angle) as presented in all twelve scenarios. Exact measurements were recorded of each angle drawn by each subject and compared with the actual angle presented by the system. Differences in angle degrees were noted between system and subject and converted to accuracy percentages. Figure 7 displays the angle accuracy averages recorded for each scenario.

In terms of accuracy, a sine-tone with strong angle indication using a wahwah effect and presented in the frontal hemisphere fared best with 74% accuracy. However, it was closely followed by a variety of white noise and sine-tone renditions. Some of the more confusing results show that accuracy levels for both white noise and sine-tone without angle indication in surround were quite accurate with 72% and 71% respectively. This result warrants further investigation. It is clear however, that a sine-tone with no angle indication in the frontal hemisphere is very inaccurate, achieving only 29% accuracy.

3.7. Combined Accuracy Levels and Response Times

Having obtained these results we decided to combine accuracy levels with response times. Because accuracy levels produced some inconsistent results, we were curious as to the overall best scenario – highest accuracy levels with best response times. This would mean that certain scenarios that produced high accuracy levels would be discounted because of inaccurate response times. Figure 8 displays the combination of response times with accuracy levels.

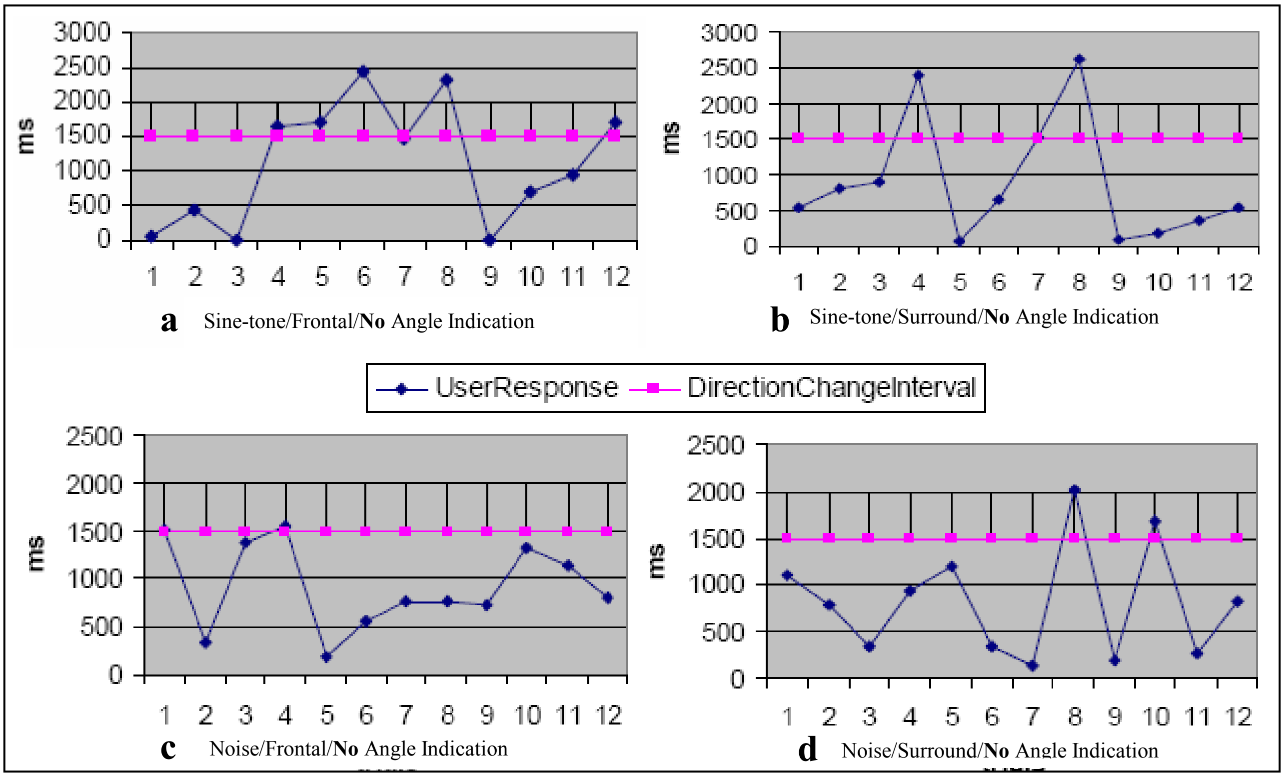

Figure 4. *Accurate response times are between 1500ms and 2000ms. 1500ms is the constant All graphs relate to continuous sound without textural indication to angle occurrence. (a) User response times for continuous sine-tone frontal hemisphere. Most response times were well below 1500ms indicating guessing by the user and/or lack of clarity of the system. σ = 847ms. (b) User response times for continuous sine-tone surround sound. Most response times were again well below 1500ms. σ = 820ms. (c) User response times for continuous white noise frontal hemisphere. Most response times were again well below 1500ms. σ = 435ms. (d) User response times for continuous white noise surround sound. Most response times were again well below 1500ms. σ = 583ms.*

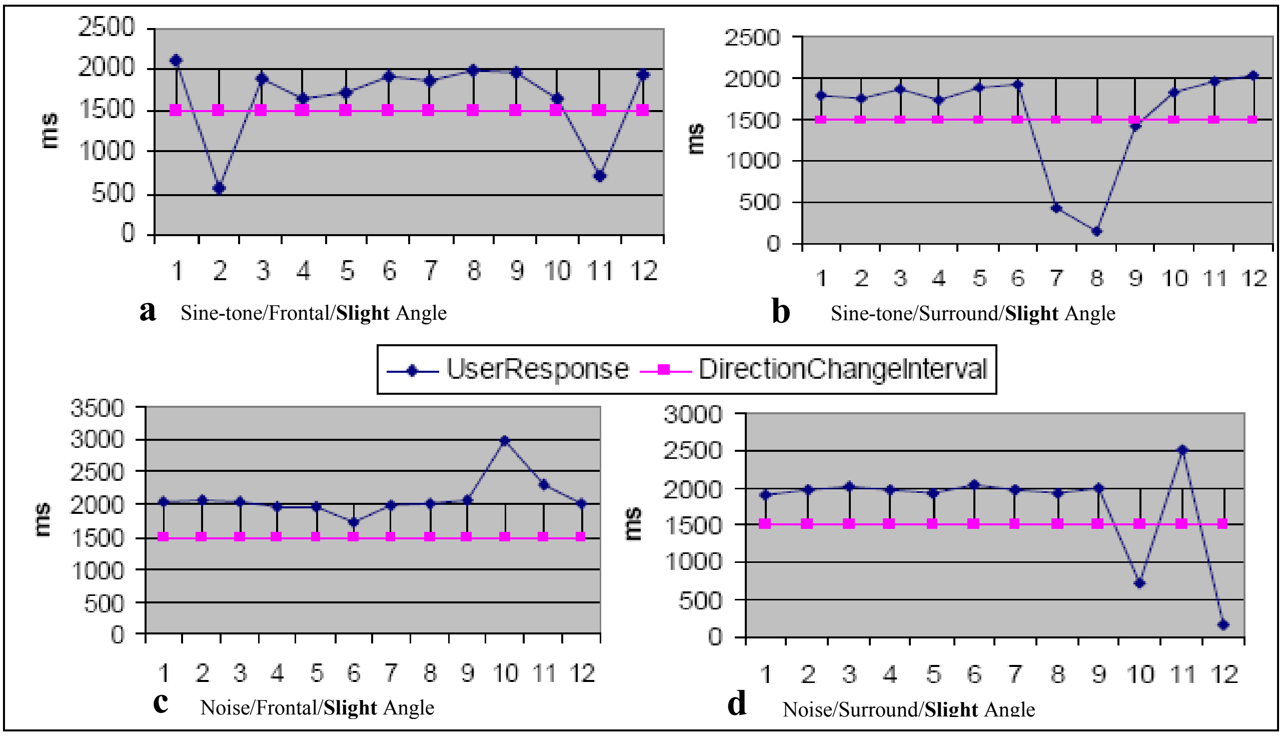

Figure 5. *All graphs relate to continuous sound with textural indication to angle occurrence. User response times are much more accurate than those in figure 4. (a) User response times for continuous sine-tone frontal hemisphere. σ = 479ms. (b) User response times for continuous sine-tone surround sound. σ = 590ms. (c) User response times for continuous white noise frontal hemisphere. σ = 294ms. (d) User response times for continuous white noise surround sound. σ = 621ms.*

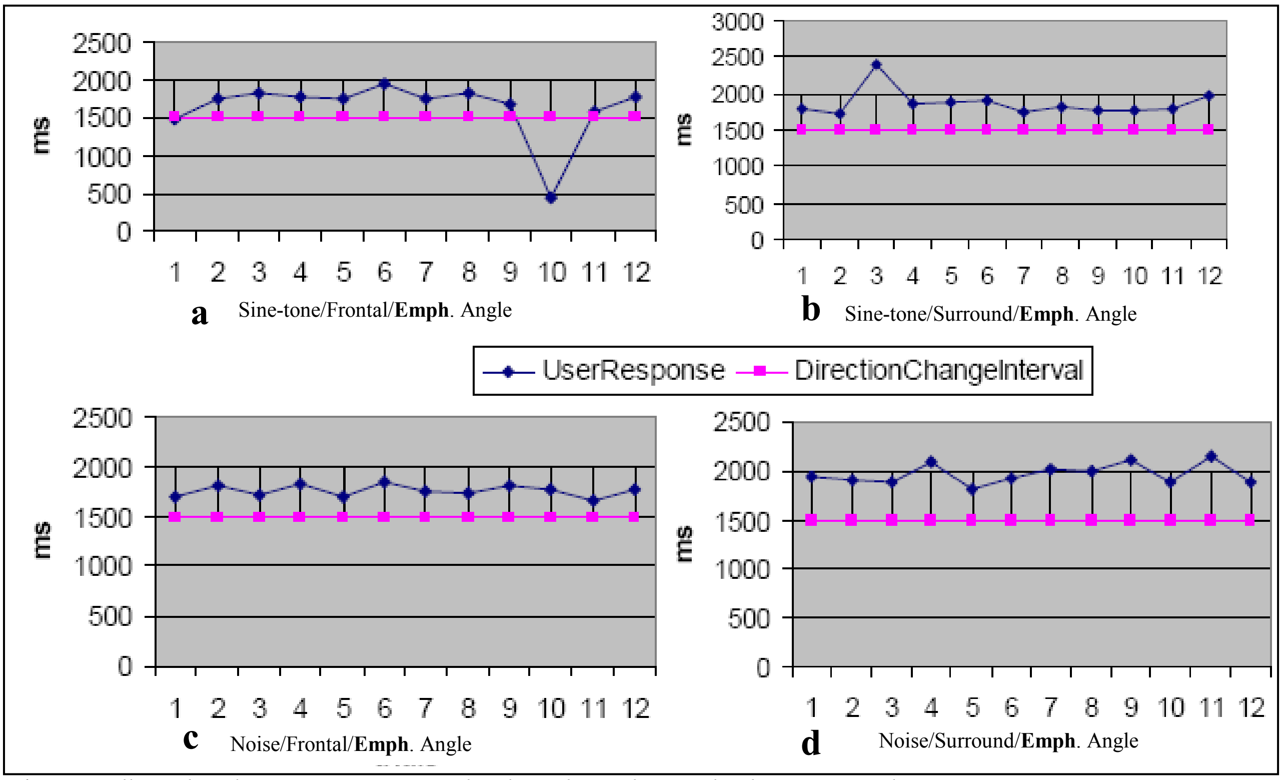

Figure 6. *All graphs relate to continuous sound with emphasized textural indication to angle occurrence. User response times are much more accurate than those in figure 4 but only slightly more accurate than in figure 5. (a) User response times for continuous sine-tone frontal hemisphere. σ = 374ms. (b) User response times for continuous sine-tone surround sound. σ = 170ms. (c) User response times for continuous white noise frontal hemisphere. σ = 56ms. (d) User response times for continuous white noise surround sound. σ = 98ms.*

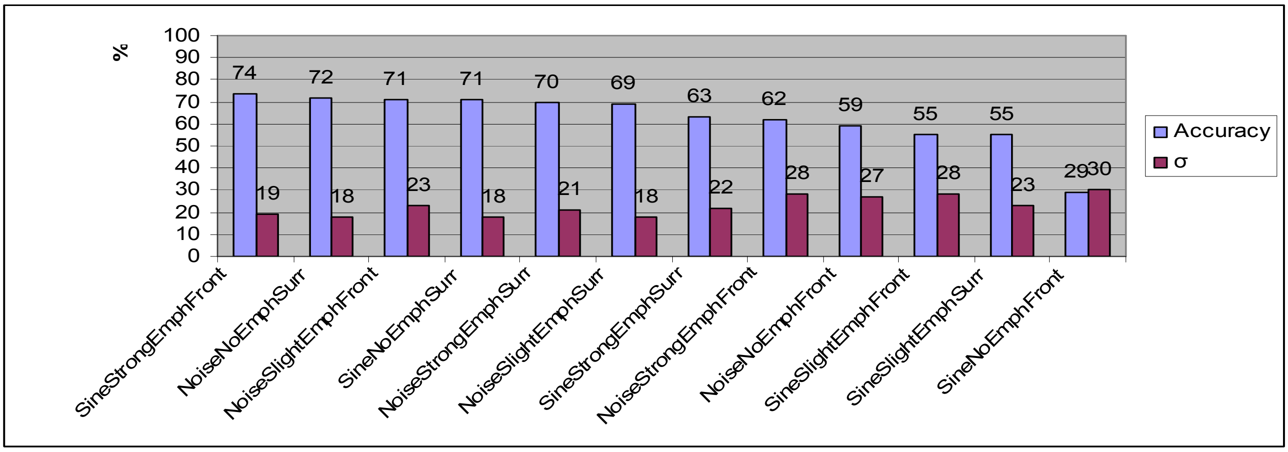

Figure 7. *Accuracy percentages for all scenarios. A sine-tone with strong angle emphasis, frontal hemisphere was the most accurate scenario and with some of the least amount deviance. It is clear that scenarios that were least accurate were also least consistent.*

	Accuracy(%)		Response Time(ms)	
	Average	Deviation	Average	Deviation
Sine StrongEmph Front	74	19	1635.42	374
Noise SlightEmph Front	71	23	2097.17	294
Noise StrongEmph Surround	70	21	1965.67	98
Noise SlightEmph Surround	69	18	1757.58	621
Sine StrongEmph Surround	63	22	1869.75	170
Noise StrongEmph Front	62	28	1757.83	56
Sine SlightEmph Front	55	28	1613.67	479
Sine SlightEmph Surround	55	23	1562.58	590

Table 2. Accuracy and response time results with respective deviations. Some uncorrelated data needing further investigation. However, although no one scenario is obviously the best choice, the top four show the best potential.

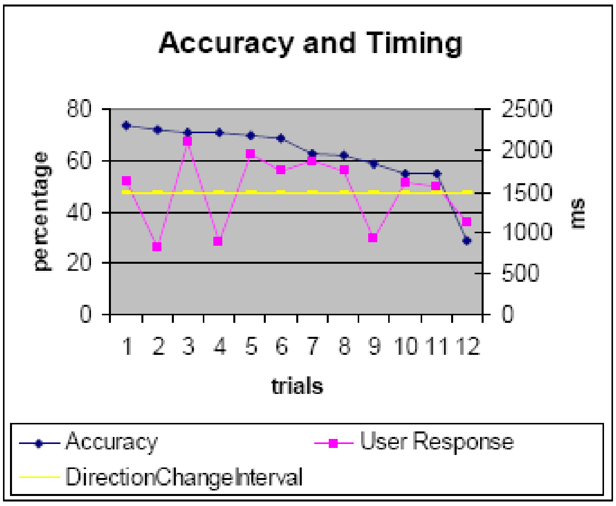

Figure 8. *User response times in ms (right Y-axis) combined with related accuracy percentage levels (left Y-axis). Minimum response time should be 1500ms (constant), therefore, response times below this line indicate errors.*

The results shown in figure 8 therefore discount a number of scenarios because their response times are below the minimum of 1500ms and therefore inconsistent with accuracy levels achieved. The scenarios that are discounted on these grounds are as follows:

- White noise with no angle indication surround.
- Sine-tone with no angle indication surround.
- White noise with no angle indication surround.
- Sine-tone with no angle indication frontal hemisphere.

Therefore, the best scenario in this context with low response time and high accuracy remains a sine-tone with emphasized angle indication in the frontal hemisphere.

3.8. Accuracy Levels, Response Times and Standard Deviance

Discounting the trials that averaged a false time response (i.e. below 1500ms), table 2 depicts scenarios with valid accuracy levels and response times. It also includes the standard deviation of both response times and accuracy levels. Taking the standard deviation of results into account, it is clear that there is no one trial that reveals the best data with respect to accuracy, timing, and the dispersion of accuracy and timing results. For example, the scenario with the highest accuracy levels (refer to table 2) with 74% also has a relatively good accuracy deviation and response time, but has a relatively poor response time deviation result. Some correlated results do appear with regard to accuracy averages and accuracy deviation. As shown in figure 7, scenarios with high accuracy show low deviation, whereas scenarios that show progressively worse accuracy averages display progressively higher accuracy deviation. However, given that there is some imbalance evident when accuracy and response times are combined, we cannot draw a robust conclusion at this stage and a further study is required.

However, as this was a pilot study examining our experimental procedure and eliminating scenarios that displayed the poorest results, it has been successful in allowing us to concentrate our follow-up study on the following scenarios.

- Sine-tone, frontal hemisphere with emphasized angle indication.
- White noise, frontal hemisphere with slight angle indication.
- White noise, surround sound with emphasized angle indication.
- White noise, surround sound with slight angle indication.

Having discounted scenarios that revealed invalid response times, we initially decided that the accuracy cut-off point would be set at 70%. For our purposes, we determine that accuracy is more important than response time and so some trials that had low accuracy but fast response times will also be omitted. However, given that 'NoiseSlightEmphSurround' (table 2) was so close to our initial cut-off point and also had a relatively low accuracy deviation result, we set the cut off point to 69%.

Concentrating on the four scenarios above, we need to examine the reason behind the uncorrelated results between the averages and standard deviations especially concerning response times. Although all trials were ordered randomly, there may be other influences yet to be isolated. These factors will be examined further in a study using a larger set of subjects.

4. CONCLUSIONS

As expected, the scenarios that displayed the best accuracy levels and response times correlated with our model to a greater extent than those that didn't. Trials that had no angle indication in either frontal or surround setup were not accurately interpreted. Trials that did texturally indicate the occurrence of an angle generally allowed for more accurate interpretation of the angle. The acoustic structure of the continuous tone with variation promotes one stream and not several competing streams as was the case in an earlier pilot study that we conducted. The one stream not only simplifies our auditory scene but reduces competition between streams at later stages of our model.

The textural variation in the signal (whether sine-tone or white noise) pulls on focused attention yet retains the perceptual continuation of the signal. This forces the user to note the presence of an angle at that particular point and, because it is part of the same stream, there is no competition between top-down resources.

In terms of Schema formation in our model, the fact that all aurally able individuals use 3D sound localization in our natural environment and visually disabled users interact with sonic devices on a regular basis, appropriate schema templates may naturally exist for our implementation unlike the representation of a 3D object on 2D tactile devices.

The pilot study succeeded in isolating the most promising scenarios for further investigation. It does not explain why there are some uncorrelated results between response time accuracy and deviation as well as some mismatched data between response time averages and accuracy averages. Given the low number of subjects, this may have had an influence in some of the uncorrelated data as individual working styles or work rate (accuracy against response times) may have been accentuated.

Further testing is required to determine some of these unexpected results. We also need to continue trying to improve accuracy results so that a more comprehensive overview of a trigonometric problem is conveyed.

5. FUTURE WORK

Based on current results, we need to re-examine some of the uncorrelated results found between user response times and angle accuracy percentages as well as the fact that the standard deviation also did not fully correlate with the final accuracy and timing results.

We also hope to further simplify the auditory scene by using the subject's head position as a cue. Every triangular shape can be oriented to position one of its sides horizontal to the user (figure 9). Because peripheral hearing is less accurate than frontal hemisphere hearing, we want to test if the user can more accurately determine an angle when it is directly in front rather than to the sides of the user. The subject's head can be used as a constant cue point which means they can quickly build a schema that is accurate in terms of where the angle will change in space, as illustrated in figure 9. If two sides of a triangle can be accurately determined, the third side is merely a case of closing the triangle.

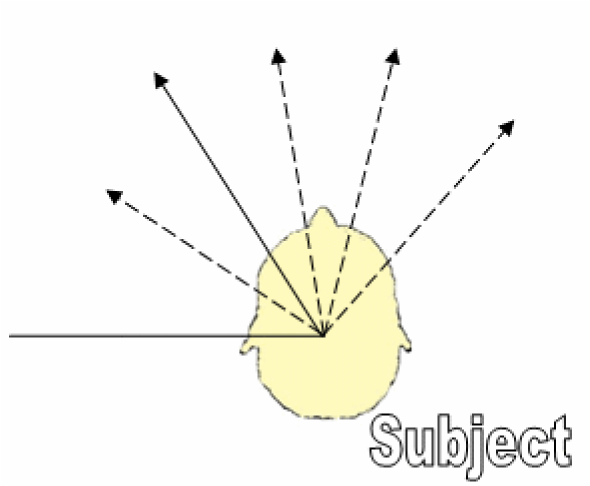

Figure 9. *Using the subject's head position as a cue for angle change. Frontal hemisphere localization is more accurate than peripheral localization.*

Other future implementations will include user activated speech output to gain specific information such as exact angle degrees and side lengths. Therefore, a form of interactive navigation will be investigated using keyboard arrow keys and gamepads. Also, we hope to investigate the potential of simulated human echolocation in relation to navigating the virtual walls and corners of a triangle.

Although surround sound systems for computers are now cheap and easy to setup and use, we hope to finally implement our work using 3D binaural techniques via standard headphones. This is a more mobile setup and less cluttering for everyday use.

6. REFERENCES

[1] A. Edwards, H. McCartney, and F. Fogarolo, "Lambda:: a multimodal approach to making mathematics accessible to blind students," in *Proc. 8th Int. ACM SIGACCESS Conf. on Computers and Accessibility*, Portland, Oregon, USA, 2006, pp. 48-54.

[2] R. Stevens, "Principles for the design of auditory interfaces to present complex information to blind people," DPhil Thesis, University of York, UK, 1996.

[3] A. Karshmer, G. Gupta, S. Geiiger, and C. Weaver, "Reading and writing mathematics: the MAVIS project," in *Proc. Third Int. ACM Conf. on Assistive Technologies*, Marina del Rey, California, USA, 1998, pp. 136-143.

[4] D. Parkes, "Nomad: Enabling access to graphics and text-based information for blind and visually impaired and other disability groups," in *Conf. Proc. World Congress on Technology*, Virginia, USA, 1991.

[5] http://www.viewplus.com/products/touch-audio-learning/IVEO/

[6] F. Neff, A. Kehoe, and I. Pitt, "User Modeling to Support the Development of an Auditory Help System," in *Lecture Notes in Computer Science. Text, Speech and Dialogue*, Springer Berlin/Heidelberg, vol. 4629/2007, pp. 390-397, Aug. 2007.

[7] F. Neff, I. Pitt, and A. Kehoe, "A Consideration of Perceptual Interaction in an Auditory Prolific Mobile Device," in *Proc. Workshop on Spatial Audio for Mobile Devices, Mobile HCI*, Singapore, 2007.

[8] R. Vilimek, and T. Hempel, "Effects of Speech and Non-Speech Sounds on Short-Term Memory and Possible Implications for In-Vehicle Use," in *Int. Conf. on Auditory Display*, Limerick, Ireland, 2005.

[9] A. Bregman, *Auditory Scene Analysis: The Perceptual Organization of Sound*, MIT Press, 2002.

[10] D. Jones and W. Macken, "Irrelevant Tones Produce an Irrelevant Speech Effect: Implications for Phonological Coding in Working Memory," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, Vol. 19, No.2, pp. 369-381, 1993.

[11] D. Jones, C. Madden, and C. Miles, "Privileged access by irrelevant speech to short-term memory: The role of changing state," *Quarterly Journal of Experimental Psychology*, 44A, pp. 645-669, 1992.

[12] A. Baddeley, *Your Memory – A User's Guide*, Prion, London, 1996.

[13] S. Wrigley, and G. Brown, "A model of auditory attention", *Technical Report CS-00-07*, Speech and Hearing Research Group, University of Sheffield, 2000.

[14] F. Neff and I. Pitt, "Using Spatial Non-Speech Sound to Relay Mathematical Problems to Visually Disabled Students," in *International Technology, Education and Development Conference*, Valencia, Spain, March 2008.

[15] http://www.logitech.com/index.cfm/mice_pointers/mice/devices/3053&cl=us,en

[16] http://www.dell.ie

[17] http://creators.xna.com/

[18] http://www.sigmatel.com/

[19] http://www.typhoon.de/en/kat.php?k=13&kidp=103