USE OF GEOSPATIAL DATA SONIFICATION FOR MOBILE AUGMENTED REALITY AUDIO NAVIGATION

Mark Anderson

M.M Music Technology,
New York University,
35 W. 4th St. New York, NY
markmakingmusic@gmail.com

ABSTRACT

This paper presents MeanderMaps: a non-speech Augmented Reality Audio (ARA) application for Apple iPhone that aids in navigation purposes by sonifying geospatial data. Users request directions to a specified location on a Google Map overlay, and MeanderMaps uses spatial auditory cues such as distance and direction to guide him/her to the destination. As the user travels to consecutive waypoints known as *path nodes*, auditory cues indicate whether an incorrect turn has been made or if the user is traveling in the wrong direction. Preliminary findings are reported using qualitative and quantitative methods, evaluating the overall sonification model in addition to individual audio cues that (a) worked, (b) worked somewhat well, and (c) needed to be improved. Future improvements and modifications to MeanderMaps are presented.

1. INTRODUCTION

The use of sonification in mobile ARA applications is becoming increasingly useful for mobile application development. Sonification is defined as the transformation of data relations into perceived relations in an acoustic signal for the purposes of facilitating communication or interpretation. [1] ARA applications rely on sound as the primary mode of interaction, used to present a hybrid overlay of computer generated sound on top of actual photographic or video scenes [2]. Third-party Application Programming Interfaces (APIs) such as Google Maps Coordinate and Google Places provide geospatial data that can be mapped to audio cues to create an interactive navigation system for users. Research into the disciplines of Human-Computer Interaction (HCI) and Sonification has revealed techniques and models embedded in auditory interfaces that are used for navigation purposes. Humans are, after all, poly-sensory beings that use all of their senses to interact and navigate within the physical world. Humans' senses are mutually strengthened by ARA interfaces, particularly those featuring position aware real-time spatial sound, faithfully suggesting direction and metaphorical distance. [2]

Auditory Display is defined as all aspects of HCI systems, including setup, speakers or headphones, modes of interaction within the system, and technical solutions for the gathering, processing, and computing necessary to obtain sound in response to data. Sonification is a core component of this, rendering sound in response to user interactions and data output. In other words, one can use sonification to represent both static and dynamic data as a way to provide useful feedback to the listener. Sonification has been used in many interdisciplinary contexts, ranging from computer science to psychology, from sound design to data mining. [3] Each discipline uses its own methods and models for representing data with interactive sound. The function of each interactive element is typically placed into one of three categories:

1. Alarms, alerts, and warnings
2. Status, process, and monitoring messages
3. Data exploration

The discipline of HCI is concerned with the design, evaluation and implementation of interactive computing systems for human use and with the study of major phenomena surrounding the interaction. It utilizes categories 1 and 2 from the above listing to present auditory feedback based on user events that occur within interactive computer applications. [4] [5] The communication between the user and computer occurs in both directions as illustrated in Figure 1.1.

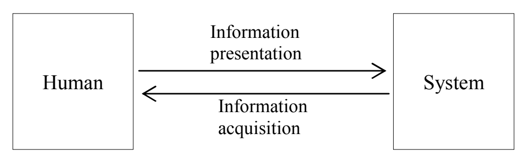

Figure 1.1: Human-Computer Interaction; representation of communication between a user and computer.

The types of auditory feedback presented to the user in HCI are known as *earcons.* Representing visual or event based information using audio messages can be tricky because such earcons must be intuitive enough for the user to understand what information the message represents. Two types of earcons are often used: representational and abstract. Representational earcons are caricatures of naturally occurring sounds such as bumps, scrapes, or files hitting a mailbox. Gaver [6] divided mappings between data and auditory representation into three types: *Symbolic* mappings that rely on social convention for meaning (e.g., applause sound), *Nomic* representations that signify physical components (e.g., the sound of a closing filing cabinet) and *Metaphorical* mappings that represent similarities between data and auditory representation (e.g., falling pitch for a falling object). In contrast, abstract earcons are comprised of motive elements, described as sequences of pitches that create a short, distinctive audio pattern characterized by the simplicity of their rhythm and pitch design. Motives can be comprised of various attributes, such as rhythm and pitch, which are the fixed parameters, and timbre, register, and dynamics, which are variable parameters. Motives used for earcons generally consist of less than four pitches, because otherwise they will take on melodic implications. Gaver found that auditory icons did not necessarily need to be realistic representations of the objects they were portraying but should rather capture the essential features of them. [6]

Both representational and abstract earcons are important for creating a sonification model used for navigation purposes. The goal is to provide as much auditory information as possible so the user's attention is focused on the primary task of moving through the local environment without frequent reference to a graphical user interface (GUI). [7] Non-speech audio was chosen to present a granular feedback mechanism that simultaneously renders distance, orientation, and turning information without the potential distraction of listening for voice commands.

2. METHODS

MeanderMaps was developed using the native iOS framework Cocoa Touch. Several third-party libraries were additionally used to develop the application. Both Google Maps Coordinate and Google Places APIs provided geospatial data via Hypertext Transfer Protocol (HTTP) requests while the Google Maps iOS Software Development Kit (SDK) provided the map overlay and methods for plotting geospatial data such as path nodes and polylines (visual lines connecting path nodes within the map overlay). Geocoding is the process of translating a physical address to geographical coordinates, while reverse geocoding is the process of translating geographical coordinates to a physical address. Both of these processes were needed to translate dropped pin coordinates to physical addresses and from address-searches back to coordinates. The built-in magnetometer sensor provided heading estimation, allowing accurate audio cue placement onto a coordinate position relative to the user location. The OpenAL API was then used to render the audio cues simultaneously to simulate a 3D auditory environment.

The user interface (UI) consists of a map overlay that displays auditory references for each component of the sonification model, in addition to adding a visual component, if needed. The map overlay is centered around the user location, based on his/her heading orientation relative to magnetic North. This is helpful for two reasons; first, the visual rotation of the map overlay gives the user a better idea of heading estimation in relation to the next path node. Furthermore, geospatial data are translated from map coordinates to Cartesian points for OpenAL sound source placement.

The sonification model created for MeanderMaps is described as follows: *direction/orientation* cues provide primary heading orientation to the user. Heading orientation describes the angle of the forward facing direction of the user relative to the next path node. Direction is calculated based on the angle between the user and the location of the nearest path node. The rotation of the map translates the placement of the path node within the map overlay, relative to the user location. As the angle changes between the source and listener, the audio cue is updated in OpenAL to a new position in the auditory space.

The *distance* cue provides an estimation of distance calculated in feet between the user location and the nearest path node. The earcon used is a pulsing sound similar to that of a sonar ping, and its pitch is scaled based on a distance calculation. The values of the pitch are represented as floating point numbers, between the range of 0.0 (normal pitch) and 1.0. The distance between the user location and path node is calculated every 0.15 seconds, to reduce any abrupt shifts in pitch that may occur as the user is navigating.

The *left* and *right* turn cues signal the user to turn in a specified direction. Both contain symbolic and abstract earcon elements, representing the sound of the turn signal heard in vehicles in addition to falling (left turn) or rising (right turn) melodic modes respectively. The complete path between the user location and the final path node may consist of several path nodes (*PN*). The moment in which the path node update cue begins playing is determined when the distance to the next path node is between the range of 100 and 170 *ft*. This range provides an optimal amount of time for the user to react to an upcoming turn, which was tested at a normal walking speed:

$$100.0_{ft} \leq PN_{distance} \leq 170.0_{ft} \quad (1)$$

One of three audio cues is played depending on the upcoming action that is needed from the user: left turn cue (LT_{cue}), right turn cue (RT_{cue}), or path update cue (PU_{cue}). The location of the following path node (PN_2) determines which cue is displayed in the auditory space. The following pseudo-code examples dictate this behavior:

$$\begin{aligned} &if(\angle PN_2 < -35°)\ then\ LT_{cue}\ plays \\ &else\ if(\angle PN_2 > 35°)\ then\ RT_{cue}\ plays \\ &else\ then\ PU_{cue}\ plays \end{aligned} \quad (2)$$

The PU_{cue} is a non-spatial cue because it simply needs to alert the user that he/she should continue on the direct path. Both LT_{cue} and RT_{cue} are spatial cues panned far left and far right respectively, because it is important to make an obvious distinction as to which kind of turn is needed to navigate correctly to PN_2.

Several additional cues are helpful in the MeanderMaps sonification model: the *wrong direction/orientation* cue is an abstract earcon that becomes audible if the user is facing > 90º or < -90º from the next path node. Two audio cues are presented of the same timbre, one octave apart in the direction of exactly 90º or -90º from the path node. As the heading orientation decreases (left orientation) or increases (right orientation), the higher pitched cue decreases in pitch (frequency) and length (time), meant to represent a change in direction. The pitch interval between the two tones adjusts to become a *diminished 5th* interval as the angle reaches a threshold of $-165° < \theta < 165°$. This interval was chosen because of its sensory dissonance based on roughness and level. Experiments contributed by Edworthy et al. suggest that introducing atonality within an interactive auditory environment such as two pitches occurring at the interval of a semitone, triggers a sense of urgency for the listener. [8] As such, the *wrong direction/orientation* cue is a strong indication that the user is facing the wrong direction.

The *path node reached* cue is used to signal the user that a path node has been reached. If the path node does not require a left or right turn, this is simply an indication that the user should continue on the direct path.

The *forward assert cue* provides additional heading information to the user. It is used separately from the *direction/orientation* cue to reassure the user that he/she is traveling in the correct direction. The forward assert cue plays at a 10 second interval if the heading orientation of the user is within a threshold of $-25^\circ < \theta < 25^\circ$ in relation to the upcoming path node.

The following rotation method is used to rotate the map overlay. This method updates once every 0.15 seconds:

```
[_mapView animateToBearing:
_locMan.heading.magneticHeading];
```

Code example 1.1 : Map overlay rotation method towards the user heading.

Each listener (the user) and source (the audio cue) are then translated to points in the Cartesian space by using the following methods:

```
CGPoint newSourcePos =
[_mapView.projection pointForCoordinate:
_firstCoordinate];

CGPoint newListenerPos =
[_mapView.projection pointForCoordinate:
_mapView.myLocation.coordinate];
```

Code example 1.2 : Translation of listener and source coordinates to points in the Cartesian space.

The next example shows the method used to calculate the left and right pan amount for the *direction/orientation* cue based on the angle between the source and the listener:

```
float angle =
[MathHelpers angleBetweenPointSource:
newSourcePos andListener:
newListenerPos];
```

Code example 1.3 : Method call to determine the pan amount for the *direction/orientation* cue given the angle between source and listener.

When the user initially drops a pin on the map, a request is made to the Google Maps Coordinate API. The return object is a JavaScript Object Notation (JSON) object containing directions between the start location and the destination. This JSON object is parsed within the MeanderMaps client code, and polylines are drawn on the map overlay, showing the navigation path between the user and the destination. Current user testing has consisted entirely of walking directions, but there are plans to develop a similar navigation use-case for driving and cycling. Once the polyline is drawn, the path coordinates are converted to Cartesian points, and OpenAL begins rendering the audio. Each audio cue is de-correlated with the rest of the audio cues in the sonification model, meaning that audio cues are not in rhythmic or melodic synchrony with one another. This is important so that audio cues notify the user of an important navigation event.

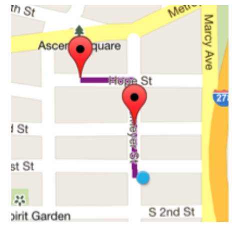

Figure 1.2 : Example of a polyline (purple line) being drawn based on directions acquired from Google Coordinate API.

3. EVALUATION

Qualitative analysis utilized a focus group of six test subjects. After a 10-minute orientation period, each subject was encouraged to use auditory cues (only, if needed) to navigate to a specific location. The following table shows the five destination addresses that comprised the navigation task:

Navigation task order of locations
(start) 35 West 4th Street, New York, NY 10012
(a) 82 West 3rd Street, New York, NY 10012
(b) 149 Bleecker Street, New York, NY 10012
(c) 116 West Houston Street, New York, NY 10012
(d) 3 Washington Square Village, New York, NY 10012
(end) 35 West 4th Street, New York, NY 10012

Figure 1.3 : Navigation task completed by six subjects using MeanderMaps. The starting address was 35 West 4th St., New York, NY.

After the completion of the navigation task, each subject completed a survey evaluating the overall effectiveness of each audio cue, and whether it was possible to complete an unknown navigation task without relying on visual guidance from the device UI. The following table illustrates the results:

Questions	1	2	3	4	5
The sonification model was intuitive for navigational purposes.				16% (1)	83% (5)
The left turn cue properly conveyed its ‘auditory message’.				33% (2)	66% (4)
The right turn cue properly conveyed its ‘auditory message’.				33% (2)	66% (4)
The dissonant tones (wrong direction cue) conveyed that you were traveling in the wrong direction.				33% (2)	66% (4)

The direction cue (forward assert cue) properly conveyed that you were traveling in the correct direction.				66% (4)	33% (2)
The pitch from the distance cue properly indicated my relative distance to the next path node.			49% (3)	33% (2)	16% (1)
The sonification model was disorienting or distracting, requiring the need for visual aid from the device screen.	16% (1)	66% (4)	16% (1)		

Figure 1.4 : Results from Focus Experiment survey. Ratings are given on a scale of 1 – 5, with 1 being the lowest rating (strongly disagree) and 5 being the highest rating (strongly agree). Percentages and numbers of participants with specific ratings are indicated in each rating box.

Each audio cue received positive ratings but the distance cue received the lowest score. The sonification model was only somewhat distracting, and some users required a mild amount of visual aid. The experiment showed encouraging results, suggesting that users were able to navigate to various destinations using the MeanderMaps sonification model.

Quantitative analysis for 26 anonymous test subjects was undertaken using the Flurry Analytics platform during a three-week evaluation period from 03/21/13 to 04/10/13. Flurry Analytics is a mobile analytics tracker, able to collect data points based on user interaction with applications. Test subjects were acquired via two local mailing lists: the NYU Music Technology group and the Augmented Reality New York (ARNY) group. TestFlight, a service that provides application developers a way to distribute applications to beta testers, was used. Once a person expressed interest in testing MeanderMaps, their device identifier was added to the distribution provisioning profile and submitted to TestFlight. The user was then able to download MeanderMaps for evaluation. All interaction between the user and the application was then recorded via Flurry Analytics. Usage data such as *number of active users*, *number of sessions*, *session length*, and *frequency of use* were recorded during the evaluation period. Additionally, three event types recorded the success rate of the user reaching his/her destination:

1. EVENT/DESTINATION_REQUESTED (102 total)
2. EVENT/DESTINATION_REACHED (9 reached)
3. EVENT/PATH_NODE_REACHED (216 reached).

There were 102 total requests, however only 44 of those requests met the minimum criteria that the user arrived at the first path node. These meaningful requests verified the user's intention to navigate to a destination address. Nine of the 44 meaningful requests indicated the user arrived at the final path node, a 20% success rate. The average number of path nodes reached per meaningful request was five, suggesting that the success rate may have been higher than what was recorded in Flurry Analytics. For example, the user may have quit the application before the final path node was recorded.

A supplemental qualitative survey was conducted within the test group in which 12 out of 26 users responded. Survey results showed that 66% of the test population used MeanderMaps more than once. All of the participants were walking to a destination 100% of the time. Additionally, 66% reached a destination all of the time, while 33% reached a destination more than half of the time. This qualitative evaluation shows that users did in fact complete navigation tasks while using MeanderMaps. Seventy-five percent of participants claimed that he/she would use MeanderMaps again in the future. Twenty-five percent indicated MeanderMaps would be a navigation preference in the future, provided that: (a) "It were very reliable, in the sense that you rarely have to look at the screen…I would use it when cycling", (b) "It worked on a 3GS device" and (c), "If I had the option to select different audio tracks because it can get tiring".

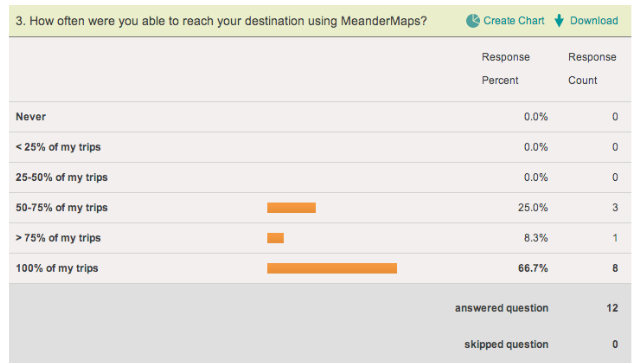

Figure 1.5 : Final survey question #3 asked to 12 of 26 users from the test population: "How often were you able to reach your destination using MeanderMaps"?

4. CONCLUSIONS AND RECOMMENDATIONS

Through testing and iteration, an effective sonification model was chosen for MeanderMaps. Though this model is not perfect, it is a substantial contribution to ARA navigation applications. The research and analysis presented in this paper should provide a general framework that will be helpful for people interested in developing ARA navigation applications. MeanderMaps worked well as a navigation tool, and both qualitative and quantitative results validated that it can be used as an alternative to existing ocular driven navigation tools. Qualitative feedback was beneficial in that it demonstrated the usefulness of MeanderMaps when a user was given a focused navigation task. Quantitative measurements from the larger test population were also important for understanding if novice users could navigate effectively using MeanderMaps.

While it seems that using many different audio cues to sonify geospatial data is helpful, there is a limit to how many of these should be used simultaneously. The initial model proposed in [7] suggests that the most important cues in an auditory navigation system are distance and direction, which is certainly the case. I found that reinforcing these two audio cues with additional cues reduced ambiguity and guesswork for the user. While there is no correct way to sonify geospatial data, the model presented in this paper provides a solid foundation for future sonification models. It is my hope that MeanderMaps will become an integral tool used for everyday navigation purposes.

MeanderMaps continues to grow in scope. A database of audio cues representing a diverse mix of timbre and rhythmic qualities is available for users to customize cues for an individualized navigation experience. Users also have the ability to search for a specific address rather than dropping a pin location on the map overlay. Upcoming features include support for uploading audio cues to the database, helping users personalize the navigation experience even more. Parameters for audio cues will additionally be customizable, such as the distance to the next path node in which left/right turn cues become audible. Further qualitative and quantitative testing will be undergone to improve the sonification model and thus create a better user experience. Lastly, research into the benefits MeanderMaps may have on blind individuals will be explored.

5. REFERENCES

[1] G. Kramer, B. Walker, T. Bonebright, et al. (1999). "The Sonification Report: Status of the Field and Research Agenda. Report prepared for the National Science Foundation by members of the International Community for Auditory Display. Santa Fe, NM. http://www.icad.org/websiteV2.0/References/nsf.html

[2] M. Cohen and J. Villegas, "From whereware to whence- and whitherware: Augmented audio reality for position-aware services," *2011 IEEE International Symposium on VR Innovation*, pp. 273–280, Mar. 2011.

[3] T. Hermann, A. Hunt, and J. G. Neuhoff, *The Sonification Handbook*. COST Office and Logos Verlag Berlin GmbH, 2011.

[4] J. Cohen, "'Kirk Here:' Using Genre Sounds To Monitor Background Activity." *Human Interface Group/ATG, Apple Computer, Inc., Cupertino*, pp. 63–64.

[5] M. Blattner, D. Sumikawa, and R. Greenberg, "Earcons and Icons: Their Structure and Common Design Principles," *Human-Computer Interaction*, vol. 4, no. 1, pp. 11–44, Mar. 1989.

[6] W. Gaver, "Auditory Icons: Using Sound in Computer Interfaces," *Human-Computer Interaction*, vol. 2, no. 2, pp. 167–177, Jun. 1986.

[7] S. Holland and D. Morse, "Audio GPS : spatial audio in a minimal attention interface," pp. 1–6, 2001.

[8] J. Edworthy, S. Loxley and I. Dennis. "Improving Auditory Warning Design: Relationship between Warning Sound Parameters and Perceived Urgency," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 1991, Volume 33, no. 2, pp. 205-231