MOBILE AUDIO DESIGNS MONKEY: AN AUDIO AUGMENTED REALITY DESIGNER'S TOOL

Bruce N. Walker and Kevin Stamper

Sonification Lab,
Department of Psychology,
Georgia Institute of Technology,
654 Cherry Street
Atlanta, Georgia, USA 30332
bruce.walker@psych.gatech.edu, kstamper@cc.gatech.edu

ABSTRACT

Audio Augmented Reality (AR) design is currently a very difficult task. To develop audio for an AR environment a designer must have technical skills which are unrelated to the design process. The designer should be focusing on the creativity, design, and the logic of the AR rather than the details of the audio. To support the design process, an audio AR designers' tool called Mobile Audio Designs (MAD) Monkey was developed. MAD Monkey was developed using the standard User Centered Design process. The stages of the iterative design process are described here, and the features of the resulting system are discussed. Evaluation of the prototype and plans for further development are also enumerated.

1. INTRODUCTION: THE NEED FOR DESIGN TOOLS

Early computer audio consisted of simple beeps and messages were communicated in a Morse code-like fashion. By contrast, today's consumer technology is capable of creating high fidelity computer audio that is presented through surround sound. In response to this, users have responded by creating new ways of using the audio capabilities.[1]

This response has led to the proliferation of audio displays, and the increased use of audio. One particular modern use of sound in computing involves audio in mobile, wearable, and ubiquitous computing, as well as augmented reality applications. Currently, users carry around laptop computers, and are starting to use devices such as the Sidekick, which is among other things, a cell phone, web browser, and email client. Users have also started wearing Bluetooth headsets that provide a constant link to their cellular phones. This provides them constant connectivity, and a speaker with which to display auditory cues. Again, this trend is only going to continue to the point that users will wear full headsets when the infrastructure provides enough data to make such a headset worth wearing.

With computing moving away from the desktop and toward wearable computers, the use of visual interfaces tends to require the mobile user to sacrifice large, central portions their visual field of view, since the display itself occludes those areas. Clearly, this is both distracting and dangerous. In contrast, audio interfaces can be less obtrusive than the visual displays that have been in use. That is, an auditory interface can be engineered so that it is "transparent," maintaining access to the external acoustic environment in ways that a visual display cannot. These auditory display systems can present a significant amount of information from all directions, and because of the fact that humans associate a location with one type of data, can leverage the position of the sound to convey information, and thereby decrease cognitive load. [2]

In the current state of audio Augmented Reality (AR) design, there is a significant problem. In order to develop audio for an AR environment, a designer must have significant programming skills. Most designers do not have such skills, and those that do often find that the programming gets in the way of the audio design process. The design process is then relegated to a secondary position, and suffers because of that. There are some programming interfaces which allow for complex sound design, but these interfaces do not have provide appropriate manipulation and representation of the spatial environment.

In the past, designers of all kinds have realized the need to provide access to several tools through a unified interface. This innovation allows a wider range of designers to use the tools to create products. An example in which several tools were combined into one visual interface is that of software development environments. These tools all used to be separated from one another. The code editor had its interface, which was separate from the compiler's interface, which was separate from the debugger's interface. These tools were sufficient for developing early software, when those programs were small, and the developer was highly skilled in using the interface(s). However, only a select few creative persons also had the technical skill to use the tools. To counter this, software engineers developed the Integrated Development Environment (IDE). This was partly necessary due to increasing program size and complexity, but also to give a wider range of developers access to these tools. The IDE integrates all of the development tools into one larger tool that allows developers to edit and debug code. This speeds and eases the development process because the tools can be aware of each other, which allows each to provide information through the same interface. The designer can focus on the creativity, design, and the logic of the program rather than the details of the code.

2. DESIGN REQUIREMENTS

The design process for an AR is basically: (1) design object (either visual or auditory); (2) place those objects in the environment; and (3) evaluate the aesthetic and functional qualities of the objects as the users interact with them. Designing the object is a very complex task, which deserves its own tool, and a separate analysis. There are many tools which currently provide sophisticated sound design capabilities, so the focus of this software is tasks (2) and (3). Currently, there is one tool that supports the design of an Augmented Reality: Designer's Augmented Reality Toolkit (DART). [3] The DART software focuses on designing for a visual AR display. It supports the three design tasks fairly well for visual objects. For example, the software allows for simple sketches (placeholders) to be placed in the environment, and then later the designer can switch those out for the fully developed object at a later date. DART supports audio as well, but the audio manipulation is cumbersome. The tasks in an audio AR are basically the same as for a visual AR, but there is no need or ability to render a visual interface on the final AR. The visual interface is only necessary and allowable at design time.

In order to design for the state of the art sound system, there are several professional-grade tools in the marketplace. Unfortunately, despite their impressive capabilities, none is really suited for AR and mobile audio design. Digidesign's ProTools is a good example of these tools. It was created as a tool to assist recording engineers mix music albums and soundtracks, but as audio technology got better, ProTools started incorporating surround sound design capabilities. This would seem to be an appropriate tool for developing audio AR experiences. Unfortunately, ProTools (and similar tools) have been designed with the assumption that the listener is stationary or that the listener is not assuming that the sounds are tied to the real physical world. Thus it is clear that there remains a need for an AR design tool that supports audio in an effective way.

Initially, the requirements that arose from a survey of current audio tools were: (1) allow the designer to simultaneously play sets of audio labels; (2) compare those labels through various through visual and auditory displays; (3) identify masked audio or other problems with the environment. After meeting with experts, these changed to: (1) give the designer a sense of the environment for which they are designing; (2) provide an visual representation of the audio; (3) allow the designer to compare sounds based on a region of interest; (4) provide ability to switch audio labels quickly; and (5) allow the designer to browse sounds quickly.

3. MOBILE AUDIO DESIGNS MONKEY

In order to fulfill those requirements, an audio AR designers' tool called Mobile Audio Designs (MAD) Monkey was developed. The development of MAD Monkey was executed using the standard process for User Centered Design. The process consists of a survey of current software, and then a process of iterative design. This consists of a prototype being built, then evaluated by experts, refinement of the prototype and repetition of the process until time or money is depleted. [4]

3.1. Expert Participants

The six experts who participated were chosen due to their experience with audio design or with Augmented Reality environments. These experts were interviewed in order to get their feedback on the prototypes, and to further define the relevant tasks. Expert A was a composer who creates compositions that may be conventionally or unconventionally musical. Expert B was a composer whose primary expertise is in interface design. Expert C was an audio designer who is currently designing AR experiences for mobile devices. Expert D was an AR interface designer who is working with DART. Expert E was a psychologist who works in sonification and psychoacoustical research. Expert F was a composer and professor who works with cutting edge music technology.

In order to determine which features should go in the first round of prototypes, a survey of the current software for audio design was performed. These were analyzed to determine what tasks and interface features would be relevant to audio AR design. Those features were put into the initial paper prototypes, as well as other features that appeared to support the designer's tasks. The prototypes were then evaluated by the experts.

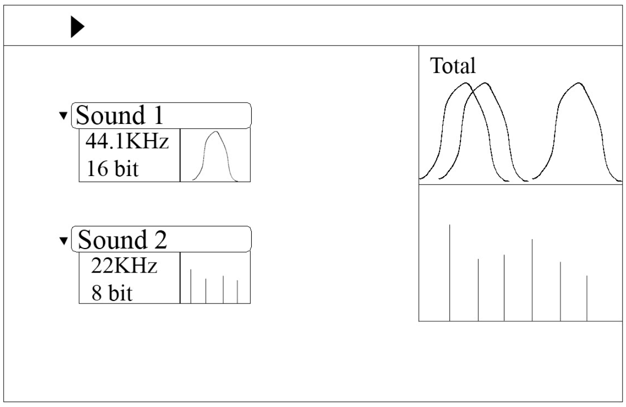

Figure 1. *Initial Design – This design was based on a survey of current audio design interfaces. As a result, it focused exclusively on the audio. There was no representation of physical location.*

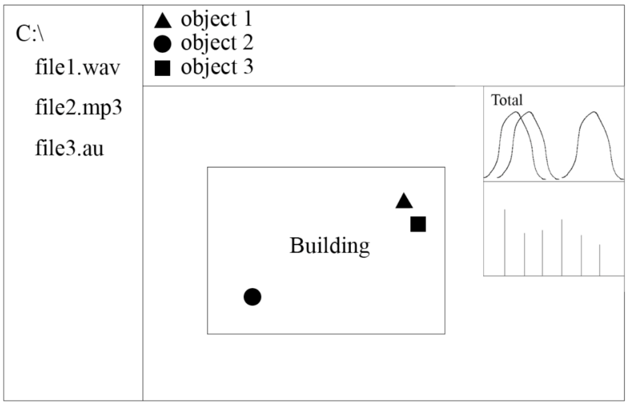

Figure 2. *Second Design – This design incorporated expert feedback, and as a result provided a representation of physical based on the experts' feedback, a second round of evaluation location. It also incorporated access to the designer's sound library, and a more concise representation of the audio in the design.*

3.2. Prototype Iteration

The first design, as depicted in Figure 1, was built with the idea that the system would not be bound to, or leverage the physical environment in which the designer is working. Based on this, the designer had to choose which set of sounds they wanted to hear at any time. This turned out to be an incorrect assumption. Expert A, the composer, provided the perspective that the designer would want to know where the sounds are when he or she is designing them. Expert C, the audio designer, confirmed this, with a discussion of being able to move through the environment as the user would walk. All the experts except for

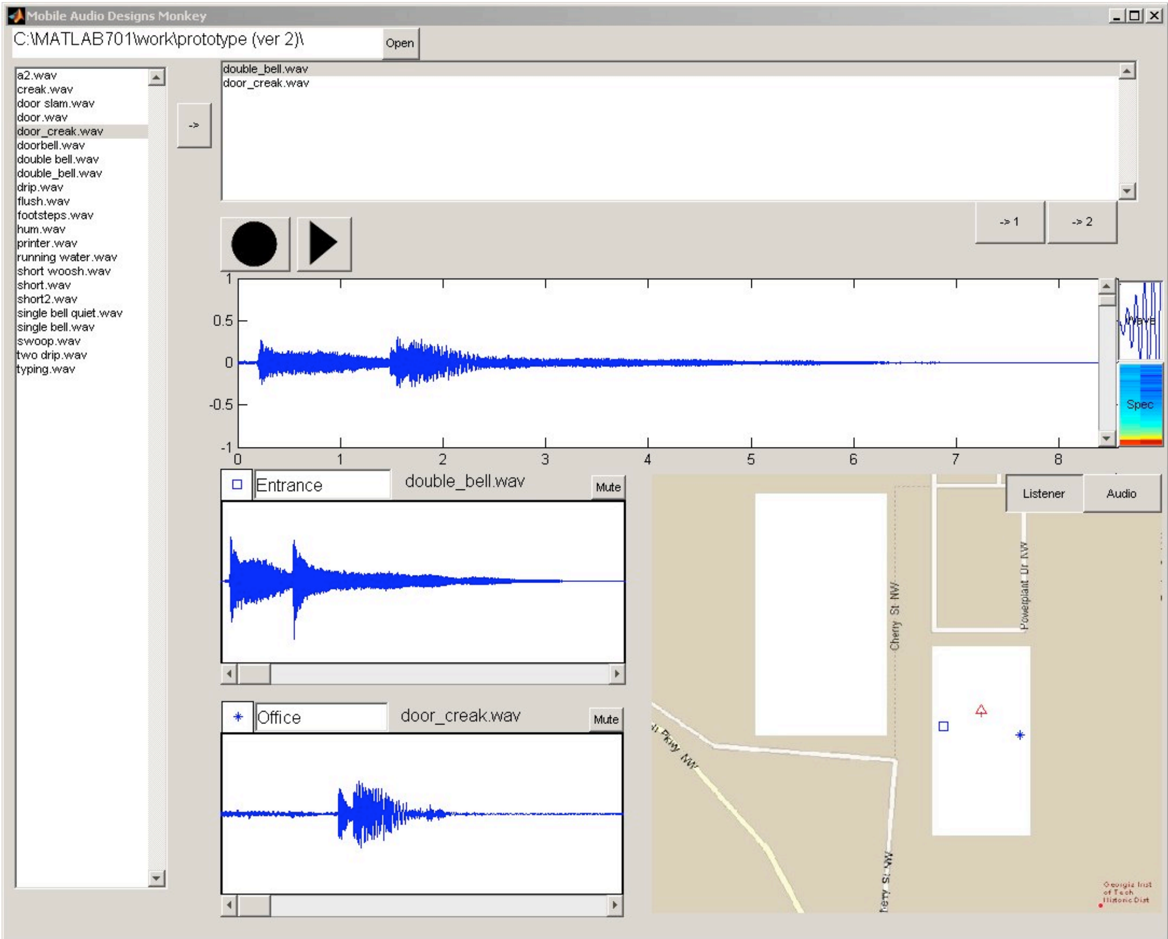

Figure 3. *First Interactive Prototype – This design incorporated a record button for rapid prototyping of audio. It was the first prototype to allow user interaction, and it further refines the visual representation of the audio.*

E, the psychologist, spoke with the expectation of a visual representation of the space. This is not surprising, as expert E mostly spoke of psychoacoustical issues rather than AR tasks.

In order to concisely convey the data about a sound's location, the second iteration (see Figure 2) had a map of the space. Without such a representation, the designer does not know if the sound will be indoors, outdoors, in a noisy lobby, or a quiet hallway. Knowing the location of a sound does not solve all of these problems, as hallways can become noisy depending on the activity in them, but it allows for the designer to know that it will be human noise, and not geese, for example.

Another assumption about the designer's process was that he or she would like a visual representation of the combination of all the sounds that are currently in the designer's focus. This turned out to be correct, though the representations that were chosen by the experts are different from the ones in the initial design. Expert A, the composer, commented that the ratio of the frequencies is more useful to him than the particular distances between the frequencies. Expert E, the psychologist, said that he would like to be able to have a spectrogram as the primary display for each sound to assess issues such as masking.

There was one critical discussion with Expert A, the composer, during which there was discussion about his workflow. This led to the addition of an audio browser and a list of sounds currently in the AR to be added to the prototype.

Knowing that it was necessary to represent the location of the sound caused the second design to change significantly. Providing the designer with the data that is implicitly provided by a map or other representation of the space was very desirable to the experts. Due to the preliminary feedback, the second interface provided an overhead representation of the space. This changed many of the other interface's features. The audio preview function was still available, but was based on regions of interest, rather than arbitrary selection of sounds that the designer chose to hear. The visual representation of the waveform did not change because it was not clear which is the most appropriate. This issue was dealt with by providing multiple representations of the audio. This approach was preferred by the experts, since each method of displaying the sounds supported different experts' needs.

Expert C, the composer, sketched her idea of the layout of the interface, which was very similar to the third prototype (see Figure 3). Expert D, the AR expert, said that users (designers) often need to put a placeholder in the environment instead of the final sound. Based on this feedback, the prototype incorporates a "Record" button. While this may seem unimportant, the ability to easily insert an audio placeholder is a critical feature for designers. Largely, this prototype uses the format of the previous one, but has some minor tweaks. The sound browser is on the left, and the sounds that are currently in focus are at the top of the interface. The interface allows the user to change the visualization of the total audio to view the frequency (spectrogram) or the amplitude (waveform) over time.

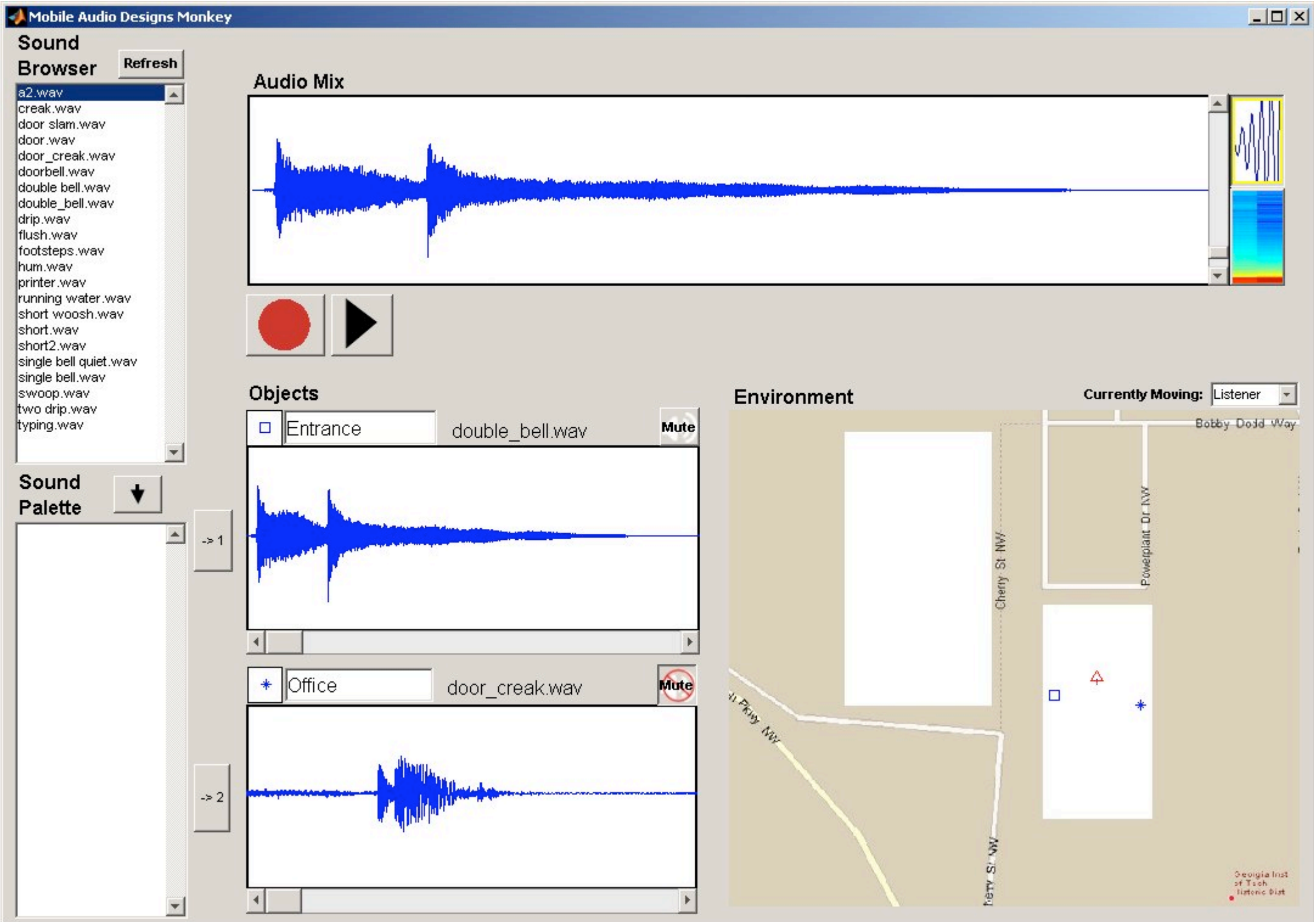

Figure 4. *Final Interactive Prototype – Several interface elements were rearranged in this iteration. Much of the interaction was improved to make the functionality more explicit, as well.*

3.3. MAD Monkey System Features

The system is built in MATLAB, due to its sophisticated audio manipulation capabilities as well as its GUI builder. There is no other system that supports complex transformations of audio while still allowing for a familiar visual interface to be built relatively easily. MATLAB has a substantial number of tools that allow for the display of an audio file. Those that do have significant audio capabilities lack standard GUI development features. The GUI prototyping tools, on the other hand, do not allow for the manipulation and representation of a waveform. MATLAB supports the use of ActiveX controls in its GUI development environment. These controls provide diverse and complex interface functionality, with a minimum of effort.

The system displays the audio as though it were a physical object in the environment. It also allows for the display of the environment, and display of all the audio within a focus region. All of the audio in those regions can be played to give the designer the ability to determine conflicts or dissonance.

This system allows for audio files to be swapped out quickly. It provides access to simple manipulations of the audio, with a focus on those most relevant to augmented reality. Currently, it allows the user to delay the onset of a sound, as that was determined to be the most important manipulation.

3.4. MAD Monkey System Benefits

This system design has many benefits over any existing alternatives. In particular, it (1) gives the designer a sense of the environment for which they are designing; (2) provides a visual representation of the audio; (3) allows the designer to compare sounds based on a region of interest; (4) provides the ability to switch audio labels quickly; and (5) allows designers to browse sounds quickly.

MAD Monkey provides a representation of the environment in which the designer is (virtually) placing the sounds, which allows the designer to leverage any knowledge of the space in the AR design. It provides a visual representation of the audio which will allow the designer to place audio that does not conflict with the audio currently in the system. By incorporating the overhead map of the space, MAD Monkey allows the designer to compare the sounds in their current focus region. The sound cues can be switched very quickly with a simple drag and drop interface. All of the designer's audio files are quickly and easily browsable, with easy access to those sounds for the current AR design.

4. EVALUATION

The process of User Centered Design that has been applied in the development of MAD Monkey largely consists of formative evaluations. Each step in the process incorporates evaluation after the development of each prototype. This reveals problems with (and also showcases features of) the interface and interaction before substantial resources have been used to implement features that may be inappropriate.

A think-aloud protocol was performed on the interactive prototype (Figure 3) using three of the same experts as previous evaluations. These were experts B, D, and E. It may seem that a think-aloud procedure would be inappropriate because this is an evaluation of an audio design interface, but the audio display is intermittent enough to allow the experts to provide feedback. The experts did not comment while they were listening, but provided feedback as soon as the audio stopped.

Expert B felt most comfortable with the interface because she had provided design sketches that were very similar to the prototype. She and expert D suggested that the sound palette have layers, similar to the layers feature in Photoshop®. Layers were not implemented in the MAD Monkey prototypes, but due to those suggestions, the Palette was moved to the left hand side in order to provide a better format for the list and to provide an appropriate place in the interface for future implementation of layers functionality. This also provided better placement for the "->1" and "->2" buttons. Expert D did not understand the functionality of those buttons until he used them. They are now positioned to indicate more clearly the object to which the sound will be sent, when the button is pressed. Expert D also had problems with the lack of visual feedback in the interface. Based on that, many features were reworked to improve the feedback displayed to the user.

Expert E initially had significant problems navigating the interface, because he could not intuitively interpret what many portions of the interface would contain. Labels were added to each portion of the interface in order to indicate its functionality. In previous evaluations, expert A suggested the "Sound Browser" functionality. When expert E used that feature in the think-aloud, he was impressed, and enjoyed the ability to quickly preview sounds.

All of the experts had problems with their interactions with the map. Though one expert managed to find the appropriate use of the "Audio" button in figure 3, none of the experts were satisfied with the interaction. Changing this to a drop-down list combined with the new label communicates whether the user is moving the "Listener", or the "Audio." The experts did not like the "Audio" label, and had several suggestions. Ultimately, the label "Objects" was chosen to replace "Audio." The prototype (Figure 4) that resulted from these evaluations is a prototype from which software engineers could begin building.

5. FUTURE DEVELOPMENT

One of the most important features which will be implemented in the future is a visual representation of a sound's extent. This allows the designer to see which sounds the user will interact with at a given place and time, and focus on those. Other features would include integration with the System for Wearable Audio Navigation and DART, so that the designer can control the 3D engine in real-time. There are several features which help the designer create AR for human perception, such as an automatic removal, or a visual indication of masked frequencies.

There is also a need for tools which generate audio that is specifically designed to be used in an AR environment. Currently, the designer must have significant experience with the audio design in order to design audio which attracts or repels the user. This software would include a library of those sounds, classified into the appropriate categories. In a related feature, instead of presenting a display of the sound's physical properties, the system could present a different representation of the audio. These alternative representations might be: amount of dissonance, emotional content, or other data that is more directly linked to the user's interpretation of the audio, rather than a physical representation.

The system requires the ability to turn on or off the audio associated with complete classes of objects. These may be all the sounds representing drinking fountains, or all the emergency exits, for example. A real-time analysis of the audio that is currently being displayed is also planned for future versions.

6. CONCLUDING REMARKS

Today's computing technology is capable of creating high fidelity audio, which has led to the proliferation of auditory displays, most recently in augmented reality applications. However, only a select few designers have the combination of creativity and technical skills required for effective and compelling design (much less, implementation) of auditory interfaces. The need for an integrated design environment for this kind of interface has become evident. The development of MAD Monkey served first as a look into the requirements of such a software tool, and also yielded a prototype system that could be used and evaluated to further refine the needs of this class of application. An iterative process of User Centered Design revealed the beneficial features, as well as the remaining challenges in creating a useful and usable tool for designing audio augmented reality applications. As auditory displays continue to evolve and proliferate, so too must the design environments. The research and development described in this project should serve as an example of both the methodology and the nature of products that will be effective in this space.

7. WHERE TO GET MAD MONKEY

Source code will be available at
http://sonify.psych.gatech.edu/research/MADMonkey/index.html

8. REFERENCES

[1] C. Roads, *The computer music tutorial*. Cambridge, Mass.: MIT Press, 1995.

[2] E. C. Tolman, "Cognitive Maps in Rats and Men," *Psychological Review*, vol. 55, pp. 198-208, 1948.

[3] B. MacIntyre, M. Gandy, S. Dow, and J. D. Bolter, "DART: A Toolkit for Rapid Design Exploration of Augmented Reality Experiences," presented at User Interface Software and Technology (UIST), Sante Fe, New Mexico, 2004.

[4] A. J. Dix, J. Finlay, G. D. Abowd, and R. Beale, *Human-computer interaction*, 3rd ed. Harlow: Pearson, 2004.