DESIGN OF A GENERATIVE MODEL FOR SOUNDSCAPE CREATION

David Birchfield
Arts, Media and Engineering
Arizona State University
Tempe, AZ 85257
david.birchfield@asu.edu

Nahla Mattar
Arts, Media and Engineering
Arizona State University
Tempe, AZ 85257
nahla.mattar@asu.edu

Hari Sundaram
Arts, Media and Engineering
Arizona State University
Tempe, AZ 85257
hari.sundaram@asu.edu

ABSTRACT

This paper describes the design and preliminary implementation, of a generative model for dynamic, real time soundscape creation. Our model is based on the work of the Acoustic Ecology community and provides a framework for the automated creation of compelling sonic environments that are both real and imagined. We outline extensions to the model that include interaction paradigms, context modeling, sound acquisition, and sound synthesis. Our work is flexible and extensible to a variety of different applications.

1. INTRODUCTION

Beginning in the late 1960's, R. Murray Schafer established a group of researchers at Simon Fraser University to study the presence and effects of sounds in the environment that spawned the World Soundscape Project [10]. This research began as a study documenting the rise of noise pollution and its potentially detrimental effects, but has since grown to encompass the larger field of acoustic design and communication that looks holistically at the impact and structure of sound environments. Barry Truax defines the soundscape as an environment of sound with emphasis on the way it is perceived and understood by individuals or a society. The term may refer to actual environments, abstract constructions, or to artificial environments [15].

Soundscapes and sound collages are capable of creating a strong sense of time, place, and context. This has been documented in analytical work from musicians [14], theorists [2], and anthoropologists [6]. Soundscape collages from the real world have been captured and studied [11], and have been synthesized for artistic purposes [5].

There has been some prior work in the development of computational models for generative soundscapes. In [9] the authors propose an approach to automated soundscape creation in their work on Audio Aura. There they dynamically generate sonic seascapes to convey information in specific applications such as email alerts and ambient displays. However, their model is hindered by a small palette of sound sources and a narrow focus on information display. Cano, *et al* describe a compelling framework for semi-automated soundscape creation [3] that implements effective mechanisms to generate novel sonic environments. However, it is designed to generate reconfigurable, but static soundfiles, while our model facilitates the generation of dynamic, interactive soundscapes that transform in response to meta-level user controls. In this respect our work is more closely aligned with recent developments in gaming and virtual reality. While there has been promising work in this area [12, 16], these models do not specifically address user context adaptation, individual embodiment, or artificial/creative sonic environments that are central to our efforts. In our recent work [1], we have developed generative models for the automated creation of sound environments and music. Although these models have yielded successful musical results, they do not focus on the synthesis of physical/virtual environments.

In the following sections, we present our recent research toward the realization of a generative model for soundscape creation. This model utilizes fundamental concepts drawn from acoustic ecology research, and provides extensions including the integration of user context models. We discuss tools for database interaction, theoretical underpinnings of the model, our preliminary results, and future directions.

2. MODELING USER CONTEXT

A critical aspect of our model is the adaptation of soundscapes to suit the experiences and expectations of the listener. This adaptation requires detailed modeling of the context in which soundscapes are presented. We are currently developing dynamic models of user context that will play a central role in soundscape creation.

Prior work on context includes work in natural language understanding and ubiquitous computing [4]. In our multimodal feedback systems, we require a broader definition of *multimodal context*. The three parallel phases of the context model that depend upon the set of questions are *context acquisition, context representation* and *context evolution*. Research on the architecture of context aware systems has focused on a modular approach to context modeling that attempts to maintain a demarcation in these phases [4]. Based upon the set of questions and their phases, the context can be classified into three types: *Environmental context, Application* *context, and User context*. Our related previous work in this area [13] addresses each of these layers. However, in our current model for soundscape synthesis, we focus on issues pertaining to the user. The user context is related to the meaning and user understanding. It answers the questions such as, *"who", "what are the skills", "what are the interests", "what are the goals", "what is the understanding"* among others. This pertains both to the users who acquire media samples, and to the user who interacts in an application environment. Hence, it is concerned with semantic inter-relationships between concepts and how they evolve in a dynamic situation.

3. DATABASE INTERACTION

Our generative system employs a backend database for storage of annotated soundfiles. We have developed tools and methods to streamline the process of acquisition, uploading, and annotation of sound samples. In this section we describe developed tools for populating the database and mechanisms for search and retrieval.

3.1. Automated Upload Tools

Once sound samples have been recorded and digitized, they must be inserted into the database. To streamline this process, we have developed an application in Java, that recursively searches through a directory tree, locates soundfiles, normalizes the files, extracts basic features, and inserts entries into the database for each file. We use JSyn for sound processing in this application. Presently, we extract only elementary features such as duration, and average amplitude for each sample. In our future work we will incorporate more sophisticated machine listening routines to achiever greater depth during the automated annotation stage.

3.2. User Annotation

In addition to signal level audio features that can be automatically extracted from files, our model requires semantic knowledge about each soundfile. Specifically, each soundfile must be annotated with a reference to its classification (eg. signal, soundmark), location (eg. where was a sound recorded), and keywords that describe its content (eg. bird songs or traffic noise). Such annotations can only be accomplished by human annotators, and we are working to incorporate related work on media annotation to improve the accuracy and efficiency of this annotation process.

In our own annotation system, [13] we address issues relating to both semantics and the end user experience through an incentive-based procedure as follows. Our system uses a combination of low-level as well as WordNet [8] distances to propagate semantics. As the user begins to annotate the media objects, the system creates positive example (or negative example) media sets for the associated WordNet meanings. These are then propagated to the entire database, using low-level features and WordNet distances. The system then determines the media object that is least likely to have been annotated correctly and presents it to the user for relevance feedback and revision.

3.3. Database access

All media objects and annotations are stored on a centrally located server. We employ a MySQL database, and have developed database clients for uploading, annotation, and retrieval in PHP, Java, and Max/MSP. Practical experience has demonstrated that if the sound samples are propagated to a local computer, the database can be searched, and files served, in real time.

4. THEORETICAL FOUNDATIONS

Our model formalizes the basic principles of Acoustic Ecology and offers codified extensions.

4.1. Acoustic Ecology Model

The Acoustic Ecology community has undertaken important and influential work in the empirical study of sound and soundscapes in our communities. Through their work in recording and analyzing real world soundscapes, Schafer and Truax propose a powerful system of semantic sound classification [14, 11].

Four classes of sounds emerge from the work of Shafer: *keynotes, signals, soundmarks, and sound romances*. A *keynote* sound is the tonal center of a soundscape. For example, the *keynote* of an oceanside town is the sound of the beach. The hum of the computer hard drive is the *keynote* for many of our work environments. An infrequent, and sometimes alarming informational sound, is classified as a *signal*. A police siren or email alert are typical examples. A *soundmark* is the acoustic equivalent of a land mark. It sonically distinguishes and identifies a particular location. The carillon of Big Ben or a church's bell are soundmarks. Finally, *sound romances* are those that inspire a feeling of nostalgia or longing in a listener. Often they are anachronistic sounds that have since disappeared. For example, the sound of a wooden wagon wheel on a cobblestone street. As will be described in Section 5, our model explicity utilizes these semantic classifications to construct soundscapes.

4.2. Integration of User Context

As described in Section 2, our model of context includes dynamically updated knowledge about the user that will allow our model to address his/her individual relationship to the sounds present in the generated sonic environments. This contextual knowledge will be used to generate soundscapes that adapt to the particular experiences and interactions of users, and there are two areas where this adaptation can be most effective.

First, for each listener, sound event classifications should evolve over time and are marked by boundaries that are not always distinct. For example, the sound of an ambulance siren is often a clear *signal* sound. However, through repeated exposure to that *signal* – either in the physical world if one lives near a hospital, or through repetition in a synthesized soundscape – the sound of a siren can lose its immediacy and impact as a signal.

Second, the classification of sounds cannot be universal for all users. For example, in Islamic countries the call to prayer sounds five times a day. For native Muslims it would be a *signal* to gather for prayer. For native, but non-Muslims the event will be a *keynote*, and for non-natives the event will be a *soundmark*. As a second example, if the goal of a given soundscape is to convey a sense of nostalgia and familiarity for home, the model might introduce *keynotes* of the ocean, with *soundmarks* such as channel buoy bells. For a listener raised in the plains states of the United States, this ocean metaphor for familiarity would be lost. Rather, keynotes taken from rural environments, with the sound of the wind in trees, and signals such as cow bells would be more effective. User context clearly plays an important role in the perception and understanding of an environment. We are presently working to fully integrate this notion of context into the dynamic model.

4.3. Individual Embodiment in Soundscapes

While the World Soundscape Project provides an excellent basis for the creation of soundscapes that evoke a sense of place, it does not suggest how we might represent people in those places. In recent work, we have examined effective techniques of representation in western common practice music [7], and we borrow concepts from program music and opera to create a sense of embodiment in our soundscapes.

In addition to sampling and uploading sounds from the environment, users can upload and annotate music samples that represent their musical preferences, heritage, and cultural identities. Employing the concept of the *leitmotif* (leading motif), when appropriate, the generative model will represent individual participants in the soundscape, using a selection of their music samples. This representation can be used to convey information about a particular event, to associate a person with a place, or to more broadly introduce the rich communities of the people who populate a location.

4.4. Artificial Sonic Environments

Our model is not intended to simply approximate real world environments, but rather, we seek to move beyond absolute realism, and generate soundscapes that leverage listeners' knowledge of the physical world while engaging their imaginations. Consequently, it is vital that our model be sufficiently flexible to accommodate the creation of wholly artificial environments. For example, in a music composition application, soundscapes can be generated that contain abstract sonic events functioning as *signals* or *keynotes*. Such a soundscape would leverage listeners' familiarity with the functional characteristics and patterns present in real environments, while allowing for creativity and reflection in users. The generation of imagined soundscapes could also be applied to the creation of ambient displays in a virtual or computing environment. This model would allow users to navigate a novel or unfamiliar environment in an intuitive fashion with sonic markers that behave like the physical world.

5. IMPLEMENTATION

Based on the framework detailed thus far, we have implemented a preliminary generative model for soundscape creation in Max/MSP. We have chosen to model three locations from our daily environments that include one quiet indoor office space (*X*), one outdoor urban location (*Y*), and one imagined artificial space (*Z*). The model currently holds approximately 300 soundfiles that we have annotated with a location_id (eg. *X*, *Y*, or *Z*) and class_id (eg. *keynote, soundmark, signal,* or *sound romance).* We have also integrated personal embodiment samples from the music collections of participants in the project.

Five parallel tracks of audio are present in the soundscape: one for each classification proposed by Shafer/Truax, and one for the representation of individuals' sonic presence in the collage. The model contains a table of dynamic probabilities that regulate the density and frequency of sound classes comprising the generated soundscape. For example, *keynotes*, such as the buzz of a flourescent light in an office, serve as the defining sound of a location and are generally present 100% of the time, whereas *signals* are only present 5% of the time by default. However, these probabilities can be readily updated in response to users' behaviors to radically redefine the composition of the soundscape. For example, the current model will shift these probabilities in accordance to the duration that a user has been present in a location. Specifically, when a user first encounters a soundscape location, the probability of the introduction of a *sound romance* is 0% and *soundmark* is 50%. However, to model the notion that a user can grow to know more of the history of a place over time, as a user remains in that location, these ratios will gradually invert. Such dynamism ensures that sonic locations are always evolving and reflect users' behavior.

According to its class activity probability, each track of the model independently queries the sound database to retrieve a soundfile that is associated with the relevant class and current location. Features such as reverberation, an amplitude envelope, and a 5.1 panning envelope are applied to the soundfile to seamlessly weave it into the dynamic texture. In the current model, these attributes are dictated by global parameters, but in future revisions, the model will dynamically assign these mixing attributes to both heighten the sense of physical presence and emphasize particular events.

A dynamic array of probabilities are also employed to define the current location of the soundscape. For example, if a user navigates to the virtual indoor office space, the model will update its location probabilities to *X*=100%, *Y*=0%, *Z*=0%. Importantly, this flexibility ensures that locations can be hybrid spaces. For example, if the user approaches an open window that looks out to the urban environment of space *Y*, the model will shift the location probabilities to *X*=70%, *Y*=30%, and *Z*=0%. A composite soundscape will emerge that includes sounds from both spaces in these proportions. This flexibility allows for dynamic travel between and around sonic locations in the database, and it models the dynamic, overlapping features of our everyday sonic experiences.

6. RESULTS AND FUTURE WORK

We have described relevant theoretical issues critical in the design of a user adaptive model for dynamic soundscape creation, and we have discussed our preliminary implementation of this model.

Our initial evaluation reveals that the model is capable of generating distinctly different sonic environments. Each individual location is dynamic and evolving, with a sonic diversity that approximates the behavior of its real world equivalent. In future work, we hope that expansion of the soundfile database will yield even greater diversity and specificity within a given location. Preliminary tests also demonstrate that the model is capable of generating hybrid soundscapes that accommodate gradual transitions between spaces as a user navigates along a virtual path. Nonetheless, greater emphasis on sound processing and mixing of sounds can help to smooth these transformations and allow for increased exploration at the boundary areas between locations. At present, the model is only user adaptive in that it is shaped by the annotations and biases of those users who have contributed to the soundfile database. This area requires further work to fully integrate our dynamic user context models into the soundscape engine. Finally, we have sought to introduce the notion of personal embodiment in our soundscapes. We have found that our current approach is too coarse to provide satisfactory results. The introduction of music samples in these sound environments tends to overwhelm the soundscape. Further research is required in this area, and we are investigating relevant work in the anthropology of sound to find appropriate solutions.

We are pleased with our preliminary progress in codification of the described acoustic design principles. However, our implementation requires more refinement to negotiate effective layering and organization of sound samples to generate compelling and immersive sound environments. We will also undertake rigorous user studies to fully assess the effectiveness of the model for a broad population of listeners.

7. REFERENCES

[1] Birchfield, D., "Generative Model for the Creation of Musical Emotion, Meaning, and Form", *ACM SIGMM 2003 Workshop on Experiential Telepresence*, Berkeley, CA, 2003.

[2] Bregman, A.,*Auditory Scene Analysis*. MIT Press, 1990.

[3] Cano, P., Fabig, L., Gouyon, F., Koppenberger, M., Loscos, A. and Barbosa, A., "Semi-Automatic Ambiance Generation", *7th International Converence on Digital Audio Effects*, Naples, Italy, 2004.

[4] Dey, A. K. and Abowd, G. D., "Towards a Better Understanding of Context and Context-Awareness", *3rd International Symposium on Wearable Computers*, San Francisco, CA, 1999.

[5] Dunn, D., *Music Language and Environment: Environmental Sound Works 1973-1985*, Innova, 1996.

[6] Feld, S.,*Sound and Sentiment: Birds, Weeping, Poetics, and Sound in Kaluli Expression*. University of Pennsylvania Press, Philadelphia, PA, 1982.

[7] Mattar, N., *Analytical Study to Reffat Granna's Program Music Composition*, Music Department, Helwa University, Cairo, 1998.

[8] Miller, G. A., Beckwith, R. and Fellbaum, C., "Introduction to WordNet: An On-line Lexical Database", *International Journal of Lexicography*, 3 (1993), pp. 235-244.

[9] Mynatt, E., Back, M., Want, R., Baer, M., Ellis, J.B., "Designing Audio Aura", *SIGCHI conference on Human factors in computing systems*, 1998.

[10] Schafer, R. M., ed., *The Music of the Environment Series,* A.R.C. Publications, Vancouver, 1973-78.

[11] Schafer, R. M.,*The Tuning of the World*. Random House, New York, 1977.

[12] Serafin, S. and Serafin, G., "Sound Design to Enhance Presence in Photorealistic Virtual Reality", *International Conference on Auditory Display*, Sydney, Australia, 2004.

[13] Shevade, B. and Sundaram, H., "Vidya: An Experiential Annotation System", *1st ACM Workshop on Experiential Telepresence, in conjunction with ACM Multimedia*, Berkeley, CA, 2003.

[14] Truax, B.,*Acoustic Communication*. Ablex Publishing, 2001.

[15] Truax, B.,*The Handbook for Acoustic Ecology*. A.R.C. Publications, Vancouver, 1978.

[16] Turner, P., McGregor, I., Turner, S. and Carroll, F., "Evaluating Soundscapes as a Means of Creating a Sense of Place", *International Conference on Auditory Display*, Boston, MA, 2003.