AUDIO GRAFFITI: A LOCATION BASED AUDIO-TAGGING AND REMIXING ENVIRONMENT

Zack Settel
University of Montreal
Montréal, Québec, Canada

Mike Wozniewski
Société des arts technologiques
Montréal, Québec, Canada

Nicolas Bouillot, Jeremy R. Cooperstock
Centre for Interdisciplinary Research in Music Media and Technology
McGill University
Montréal, Québec, Canada

ABSTRACT

Technological advances in mobile computing, wireless communications, 3-D audio simulation, global positioning systems (GPS), and distributed data management have opened the door to novel multi-user mobile application spaces. These advances have allowed for a focus on location-based audiovisual content, ranging from navigation displays to distributed gaming, not to mention the serious potential for the arts, including music performance and installations. In our demonstration of *Audio Graffiti*, we explore novel modes of interaction with sound and space. Set in an outdoor "audio augmented" environment, we allow several mobile users to create and explore a gradually evolving wall of audio graffiti. Equipped with headsets and small mobile computers, each participant can 'tag' or 'spray' the wall with their audio, mixing in with pre-existing musical material. Others can walk about, experiencing a mix that changes based on their position and movement, resulting in a collaborative jamming and personal remixing space.

1. INTRODUCTION

The conception and motivation for *Audio Graffiti* comes from the novel modes of interaction with sound and music that are now open to exploration, thanks to key advances in mobile computing hardware, wireless communications, Global Positioning Systems, and network-based content exchange. Mobile computing devices have become a ubiquitous tool, easily found in the pockets of most public users. Artists may now create works that take advantage of these advances and incorporate user context, such as location-awareness (and even orientation-awareness) for richer interaction. A user's spatial context is particularly useful for any application of a spatial nature, particularly those for mixed- and augmented-reality. In such applications, this information may be used to select location-specific audiovisual content for the user, render the material based on the user's 3-D perspective, and allow users to contribute their own content to a particular location. We already see that some of these services are becoming available to the public. For instance, one can access photos and videos that have been 'geotagged' with a particular location using web-based services such as YouTube or Flickr. This content can be presented on the move, with a GPS-enabled device, providing extra information about the environment through which a user travels. In some implementations (e.g. Google Earth with Street View), the visual content can even be rendered for the user's perspective.

Sound and music, however, have yet be deeply integrated in applications employing geotagging or perspective rendering. One contributing factor is the permeating nature of sound, which can be invasive–especially when several sounds are encountered at once, resulting in masking and cacophony. Furthermore, the temporal aspect of sound means that a user can move far away from a sound's location before it has been completely experienced. As an artistic exploration, our demonstration explores novel forms of user interaction for sound and music. In the process, we have seen that concepts for graphical user interfaces (GUIs) reveal new challenges when applied to interfaces for audio content: "how does one zoom-in/out of sounds?"; "what is audio focus?"; "what is sonic foreground vs. background?"; etc. In the development of *Audio Graffiti*, we address these questions, leveraging both our technical experience in spatial audio rendering, and our artistic experience in working with sound in composition and orchestration.

2. THE DEMONSTRATION

As a test bed for interacting with location-based audio, we have developed an installation that focuses on both the tagging of locations with audio and the selective rendering of multiple sounds as a user moves through the piece. The installation's concept and nature of interaction is borrowed from street graffiti, where there is an element of collaborative multi-person authoring that can sometimes produce coherent and artistic organizations of material. *Audio Graffiti* takes place in an "augmented" space, since a virtual layer of audio is superimposed on a real physical wall. Headset-equipped participants may apply live sounds (such as their voice, or instrumental sounds), which gradually erode or fade over time. In addition, the installation's wall is endowed with a set of rhythmically and harmonically related sounds, that provide a base on which users can add their own content. The blend of existing and new sound forms a collaborative, ever-evolving, semi-structured musical texture.

An audio "tag" is created when a stationary user applies sound to the wall. However a moving user can also "spray" sound, so that a temporal stream of audio with a particular physical length, like a piece of audio recording tape, can be captured and fixed to the wall as a "sound contour". The result can then be heard by anyone traveling along the sound's length, effectively scrubbing a playhead along a localized audio buffer. These two methods of sound capture and localization allow for the gradual creation of spatiotemporal music, which can be experienced by a mobile audience using GPS (or other position tracking) and headsets.

3. RELATED WORK

In the most common form, location-based audio is often found as a component of audio tours or sound walks, where users are provided with a device that plays sound files at certain geographical locations. A notable example of this technology is HP Labs' *Mediascape*[1] authoring environment, which allows artists to organize media content on a map and configure hotspots that trigger playback. However, Mediascape does not allow users to record sound and tag it with a location, nor does it support multi-user applications.

Geotagging functionality also exists in many web-based sound archiving projects. For example, ambisonia.com and freesound.org both maintain databases of sound recordings, many of which are tagged with locations. However, these systems use manual methods to specify the location information. We instead focus on applications where individuals are actually present at the location, have the ability to hear audio samples left in their proximity, as well as the ability to contribute their own content.

In terms of presenting audio in situ, projects such as *[murmur]*[2] have allowed users to dial a number to hear a message, while other projects, such as *tejp* [3], have used physical boxes with embedded recording circuitry. Efforts such as the *Hear&There* project [2] have instead used GPS devices to mark the locations of audio recordings, and even provide spatialized rendering of prior recordings during navigation. Likewise, other projects [3] have capitalized on location-based services of 3G cellular networks to provide coarse locations of users' mobile devices.

One of the more similar projects is the Sonic Graffiti project [1], that allowed users to associate audio recordings with graffiti found around the city. They employed an interface in the form of a spray can for composing a tag at a specific location. Unlike their work, we attempt to create one continuous composition environment (in the form of a large graffiti wall), where several tags can be experienced at one location, forming a coherent musical context. Another project, called *Audio Space* [4] allowed users to leave looping audio messages for others to hear. This addition of microphone-based capture is similar to our work, where voice samples are recorded and discovered by moving around the space. However, our proposal offers musical engagement, where even live instrumental sounds can be incorporated. Also, our implementation of "audio contours" provides a spatial interface for playback rather than just the rendering of point sources in one's proximity.

4. APPLICATION AND IMPLEMENTATION

Interaction with the application occurs along a large wall, either indoors or outdoors. Artistic considerations notwithstanding, we chose to constrain the extent of the installation to a single wall primarily to ensure that audio sources are confined to a vertical plane. This simplifies the act of selective listening via navigation, requiring only left/right and closer/further movements with respect to the sources. In turn, this obviates the need for orientation sensing, since we may assume, given the application's "graffiti narrative", that users will generally be facing the wall.

4.1. Audio Interaction Interface

Rather than providing an empty canvas for users to graffiti upon, we seed the wall with several rhythmically and harmonically synchronized sound loops. These sounds are sparse and attenuated; they function as landmarks, offering users background support when temporal (rhythmic) or harmonic interaction is desired. Specifically, these sounds function as a *musical context*, and serve to delineate the wall musically. This context does not fade out over time but remains as a permanent structural feature. We intend to provide several pre-composed contexts that can be activated for different musical sessions.

We propose two modes for applying user-contributed sound to the wall. These offer different methods of interaction to deal with the nature of the content (i.e., tonal, rhythmic, or non-musical), and user movement in space.

Tag Mode: Stationary users are able to record and locate an audio sample or tag, at their current position along

[1]www.mscapers.com
[2]www.murmurtoronto.ca
[3]www.tii.se/reform/projects/pps/tejp

the wall. Once captured, the sample is looped continuously, gradually decaying over time. Since the sound is looped in a potentially rhythmic context, it is essential that the duration of the sample agree with the tempo of the predefined musical context. Specifically, the duration must be set to a whole number ratio of the metronomic period of the material, and all samples must play in rhythmic phase. Thus, looped playback of user recordings can only be triggered at the instant when downbeats occur in the musical context.

Spray Mode: Users in motion can apply sound to a region of the wall, corresponding to their displacements during sound capture. The resulting sound description consists of a sample stream and a physical contour related to the motion, much like sticking a piece of audio tape to the wall . Users moving in front of an existing sound contour automatically cause playback of that sound, since their movement is used to drive (scrub) the read index of the sound contour's sample buffer.[4] The sounds captured in Spray Mode, unlike Tag Mode, are not automatically synchronized to the musical context. Thus, this mode may be more appropriate for material that is independent of rhythm.

In both modes, we provide simple recording controls. In spray mode, audio is applied to the wall while a button is depressed–just like spray paint. However, in tag mode, recording must be synchronized with the phase of all other sounds loops already in the installation. Thus, users press a button to initiate a pickup signal[5] that counts down, beat by beat, to the appropriate moment at which the recording starts, synchronized by a master timing clock. The recording stops automatically after a predefined time period, and looped playback is triggered. Users must be attentive to the particular location on the wall at which they project their audio tags, since proximity to other tags can affect the musical interpretation of the piece, leading to newly perceived syncopations and rhythms.

All user-contributed sounds fade out over time, allowing the audio graffiti to be replaced gradually by new participants. The duration of the fade can be adjusted globally for each session, depending on the number of users and the amount of tagging taking place. This way, the number of simultaneous sounds in the installation is always controlled.

4.2. Listening/Playback

Audio Graffiti is listened to in a location-dependent manner, much the same way as it was recorded. Any audio tags in the vicinity of the user are heard simultaneously. Obviously, this could result in cacophony, which can be averted with proper organization. As described above in Section 4.1, all audio tags are recorded in a manner that ensures they synchronize rhythmically with other tags in their proximity. This allows many samples to be collocated coherently. To further improve a user's ability to differentiate between clustered audio sources, and to discern a potential musical texture, we employ the use of dynamic spatial audio rendering through an "audio zoom" technique. Imagining that each listener receives sound through an audio capture cone, proximity to the wall determines the area of intersection of this cone with the wall, and in turn, the number of audio tags that are heard. The playback intensity is smoothly attenuated toward the edges of the cone to prevent abrupt changes of gain as the cone moves across tags. However, the gain remains constant as the user moves away from the wall. Although this results in a somewhat unnatural spatial rendering, we suggest that this provides a useful mechanism for mixing without affecting sound level.

Audio contours on the other hand, use motion for playback, so we attenuate the signal when the users are stationary and let the sound play only when a minimum threshold speed has been reached. This invites participants to move about the space in order to discover contributed audio, and also provides information about the author's motion. For instance, walking in the opposite direction of the original recording will result in reversed playback.

Together, these features encourage users to engage in "active listening", where particular motions result in a unique mixture of audio. Recordings of these user experiences will no doubt exhibit great variation for the same sonic material. Thus, user motion acts as a sort of remixing tool and different approaches will be seen from each participant. Some will take a more global approach, listening to many sounds from far away, while others will choose specific areas on which to focus.

5. TECHNOLOGICAL CONSIDERATIONS

From our previous experiences with location-based audio [5], we have discovered that GPS accuracy is a significant problem for augmented environments where users move slowly. Consumer-grade devices provide readings with errors in the range of 1-20m. Furthermore, the heading information that is inferred from a user's motion trajectory requires the averaging of several measurements over relatively large distances, and can thus exhibit significant inaccuracies for pedestrian applications. For the purpose of this installation, we offer high-resolution GPS receivers, hoping that this technology will eventually end up in consumer-grade hardware.

To support multiple simultaneous users, we use the Audioscape engine[6] to maintain a 3-D audio model of the in-

[4]The use of an FFT-based phase vocoder allows the pitch of recording to remain constant, independent of playback speed.

[5]Several pickup beats are played through the user's headset, similar to the way that a drummer counts out a tempo to the rest of his band.

[6]www.audioscape.org

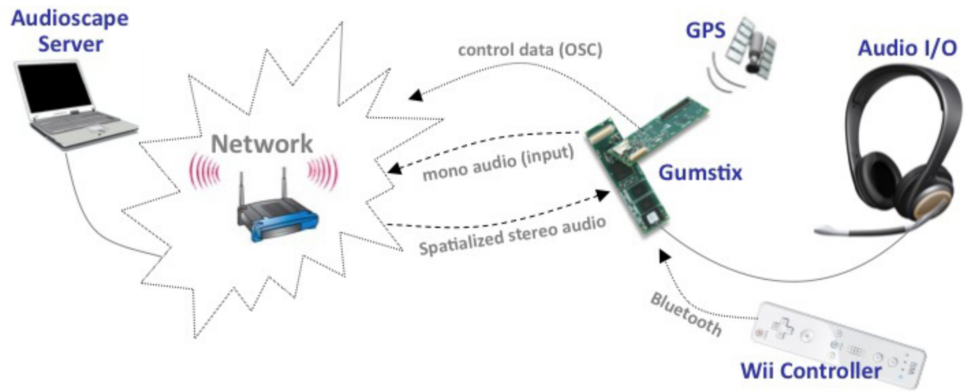

Figure 1. Overview of the Audio Graffiti system

teraction scene. Figure 1 shows how mono channel audio is sent from each client over WiFi, processed by the server, and a binaural stereo rendering is sent back. Control events for audio tagging are provided by a Nintendo Wii controller, combined with GPS data and sent over to the server via OpenSoundControl (OSC). Pure Data[7] is responsible for soundfile playback and providing user-specific phase vocoders.

6. SUMMARY

The Audio Graffiti installation has many artistically and musically interesting features. The installation provides a distributed collaborative jamming environment, and encourages users to explore space in the activity of creating and experiencing music. Participants take on the role of author, contributing to an evolving sonic work, and listener, actively rendering his or her musical perspective on the work via movement, similar to the way one experiences sculpture.

Cacophony is avoided by controlling the duration and start time of user-contributed samples. The fading out over time of user-provided audio allows for an evolving musical piece with enough variety to make a loop-based system enjoyable so long as new users continue populating the wall with their respective audio tags. The addition of sound contours that act as motion-controlled playback buffers adds to the locative aspect, and increases a user's degree of involvement in the listening process.

In the realization of our installation, we have developed sonic interaction techniques that allow musical audio content to be presented in ways that minimize cacophony, provide users with a way to zoom in/out of musical content, and give audio a physically measurable representation. While our demonstration was motivated mainly by artistic exploration, the experience has contributed to further our ideas about "eyes-free" interaction with sound content for browsing, authoring, listening, and navigating. We continue to seek better solutions to dynamic content management for large-scale multi-user audio applications. In the future, we hope to expand this work to larger-scale spaces, adding head orientation sensing in order to direct listening and project audio onto any surface in the surrounding space. With further technological advances, we imagine that location-based audio in augmented spaces can serve as the medium for interaction in vast sonic audioscapes for the public at large.

[7] www.puredata.info

7. ACKNOWLEDGEMENTS

The authors wish to acknowledge the generous support of the NSERC / Canada Council for the Arts: New Media Initiative.

8. REFERENCES

[1] C.-Y. Lee, "Sonic graffiti: Spraying and remixing music on the street," in *International Mobile Music Workshop*, 2007.

[2] J. Rozier, K. Karahalios, and J. Donath, "Hear & There: An augmented reality system of linked audio," in *Intl. Conference on Auditory Display*, 2000.

[3] A. Tanaka and P. Gemeinboeck, "A framework for spatial interaction in locative media," in *Proceedings New Interfaces for Musical Expression (NIME)*. Paris, France: IRCAM, 2006, pp. 26–30.

[4] T. Watson, "Audio space," in *International Conference on Hypermedia and Interactivity in Museums (ICHIM)*, 2005.

[5] M. Wozniewski, N. Bouillot, Z. Settel, and J. R. Cooperstock, "Large-scale mobile audio environments for collaborative musical interaction," in *Intl. Conference on New Interfaces for Musical Expression*, Genova, Italy, 2008.