Eavesdropping: Audience Interaction in Networked Audio Performance

Jack Stockholm
School of Interactive Art and Technology
Faculty of Applied Sciences
Simon Fraser University
Surrey, BC, Canada
jstockho@sfu.ca

Philippe Pasquier
School of Interactive Art and Technology
Faculty of Applied Sciences
Simon Fraser University
Surrey, BC, Canada
pasquier@sfu.ca

ABSTRACT
Eavesdropping is an internet-based, interactive audio system that explores network mediated, musical performance in shared public spaces. The project aims to develop an environment which increases audience interaction and connectedness in a localized, computer-controlled performance. The system is a client-server architecture made of three components: (1) an audio preparation interface, (2) an interactive performance interface, and (3) a machine learning-based conductor. An artificial conductor mixes an acoustic ecology based on mood data entered by participants while learning from their feedback. Technicalities and early evaluation are presented.

Categories and Subject Descriptors
J.5 [**Arts and Humanities**]: *Performing arts.*

General Terms
Algorithms, Performance, Human Factors.

Keywords
Auditory Display, Acoustic Ecology, Artificial Intelligence, Computer Music, Net Art, Reinforcement Learning

1. INTRODUCTION
The pervasive reach of internet access and ubiquitous computing has opened up new means of engagement in electronic performance art environments allowing multiple modes of interaction between audience members and with the art system itself. This paper details Eavesdropping, an internet-based, audio diffusion system designed for public spaces where several computer users are gathered, such as a café. The project is motivated by the intention to produce an engaging web audio art environment that can: raise awareness, increase connectedness and facilitate interaction, between networked participants in the same physical space [1, 2]. This is accomplished through a system which mimics the social acoustic ecology and auditory gesture in public spaces [3, 4] to increase shared experiences by capturing mood data from participants on the network and mapping moods to audio files for playback. The audio files are selected and mixed by an artificial intelligence system which validates the mapping between audio files and moods via interaction with the participants [5]. The resulting audio is played from each participant's private computer, projecting their mood into the public space. This project aims to achieve a balance between audience interaction with the system, personal affinity with the sounds emanating from participants computers and connectedness between the audience members.

MM'08, October 26–31, 2008, Vancouver, British Columbia, Canada.

1.1 Motivations
This work explores several important areas of interactive, networked, performance art. First, it looks at network connected performance in a localized space. It examines the engagement that exist between networked audience members that can see and hear each other in the same physical space and the new articulations with a performance that can be achieved via networked interaction in this space. Second, this project raises questions about the level of interaction required by audience members when co-opting their computers and devices for musical performance. Initial pilots without audience interaction revealed audience frustration at a lack of control over their own machines during performances. As a result, subsequent development has focused on giving participants an interface to affect the outcome of the performance in a meaningful and immediate way without distracting them from the performance. Third, the system explores an audible representation of moods to communicate a participant's state to others in the audience [6]. Participants can inject the performance with their own mood data and can hear the results of other's moods in the room. The server-based meta-creation system shapes the audio to ensure an aesthetically pleasing mix of the audio being performed [7].

The challenge in designing this project is achieving a balance between the various layers of creativity and control. At the system level there are encoding constraints that define how the audio files are tagged, algorithms determining which audio files are chosen to represent participants' moods and to mix with other audio playing in the room, and functions to regulate participant interaction. At the content level, a musician creates a set of audio files with a broad array of characteristics, uploads them in

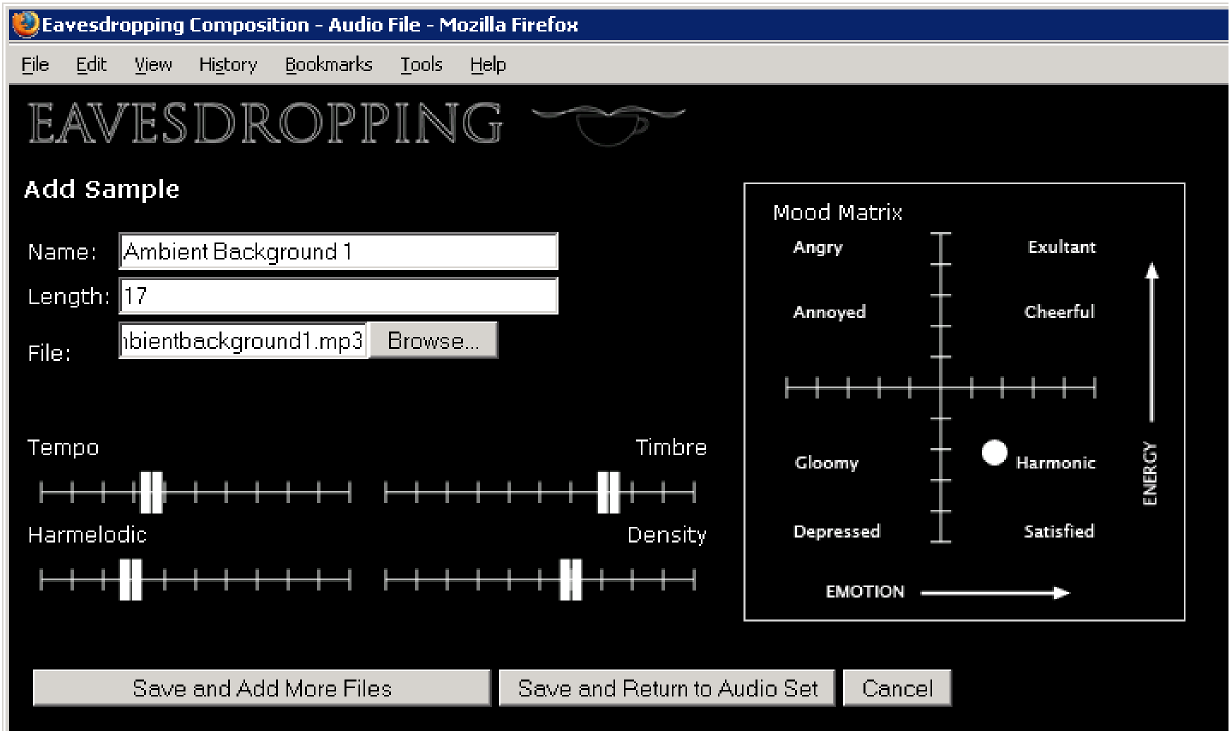

Figure 1. Musician's Interface for Audio File Tagging

advance, and tags them with their respective characteristics. At the audience level, participants require enough interaction with their computer to maintain a sense of control over their machine, while not becoming so focused that they disengage from the audio performance [8-10]. Each layer offers the opportunity for creative input that will affect the performance. The musicians generate the pool of audio samples which will represent participant moods, the participant provides feedback that introduces variety into the performance, and the meta-creation software chooses how to relate the audio to the participant to reinforce their sense of agency as well as to create an aesthetically pleasing performance.

The remainder of this paper examines the theoretical basis for this work as well as detailing the specifics of the design and implementation of the system. Section 2 provides a detailed description of the system design for Eavesdropping, the interface elements critical to the interaction, and the algorithms behind the data input and performance generation. Section 3 briefly discusses the implementation platform, software and some guidelines for creating audio for Eavesdropping Section 4 explores ambient communication in public spaces and representation within networked and data audio arts, and examines an established body of research in addressing increasing participant connectedness and the mappings used in Eavesdropping. Section 5 discusses the development process involved in the creation of the project and initial evaluation of results. Lastly, section 6 includes concluding remarks and proposals for future work.

2. SYSTEM DESCRIPTION

Eavesdropping is an internet-based, client-server architecture made of three components: (1) an audio preparation interface, (2) an interactive performance interface, and (3) a server-based, conductor.

2.1 Audio Preparation Interface

The audio preparation interface (Figure 1) is accessible to musicians via a password protected web page and has two main functions: definition of audio sets, and working with audio files contained within a set. The first function, defining audio sets, allows a musician to create a new audio set or to edit the properties of an existing set. A performance can utilize any one audio set, thereby imparting a specific style or genre as defined by the files the musician has included in the set. The properties available to audio sets are: the name of the set, the creator of the set, and a description identifying the style of the audio files contained in a set. The second function available in this interface, working with the files contained in the set, allows the musician to upload, delete, and tag audio files within a set.

2.1.1 Audio Tagging

A musician with secure access uploads mp3 files to a set via the audio preparation interface. First, the file is given a name for identification by the musician. Second, the length in seconds of the audio file is required. This allows the system to inform a participant's web browser when a file has completed so it can request a new audio file. Finally, the musician tags the file with a variety of representational characteristics which denote formal and abstract properties of the audio. These characteristics were developed to give the system basic information to associate audio files to moods (energy, emotion) as well as to apply specific formal musical properties in layering multiple audio files (tempo, timbre, harmony/melody, and sound density). The system requires each audio file to be defined by all six characteristics.

Four of the characteristics relate to formal musical parameters which allow the system to provide variety in the mix of audio that is being performed. All are abstracted from formal musical definitions and are normalized on a linear, numeric scale from 1 to 100. *Tempo* translates to the speed of the audio represented. *Timbre* lies on a scale from noisy or atonal to smooth or sonorous. *Harmelodic* relates to the clarity of the voicing and ranges from rich harmonies to solo melodies. *Density* identifies whether the file is dense or sparse, indicating whether it would be better played alone or with other samples. *Density* allows the conductor to adapt to the number of participants by selecting samples of minimal *density* when the participants are many or selecting rich, high *density* samples when the participants are few. These characteristics are set by the musician via sliders on a web page as is exhibited on the lower left side of Figure 1.

2.1.2 *Mood Classification*

Mood is represented by a two-axis mood matrix based on Thayer's two-dimensional model of mood [11]. On one dimension, *Emotion* relates to stress state and scales from distressed to carefree. On the other, *Energy* relates to the amount of activity the participant feels within their emotion and scales from inactive to highly active. Therefore audio with a depressed or gloomy mood will have a distressed emotion and low energy, and audio with an angry mood will have a distressed emotion but high energy. Likewise, audio with a contented or satisfied mood exhibits a carefree emotion and a low energy while audio which is ecstatic or exultant exhibits a carefree emotion and high energy.

The audio preparation interface and the interactive performance interface utilize the same grid to identify mood (see the right side of Figure 1 and the left side of Figure 2). This model offers several advantages over common approaches to mood classification in audio using discrete adjective descriptors. First, the model is quite simple to understand for participants as it only applies two scales and therefore requires minimal time investment versus reading through and selecting from a dictionary of mood adjectives. Second, adjective descriptors have been found to have a variety of meanings across a range of participants [12]. The use of a mood map will minimize confusion between terms (though some helper terms have been filled in on the mood matrix to assist in orientation with the map). Lastly, this two-dimensional model offers an evenly-spaced numeric grid which provides the learning system with a simple spatial map to explore audio files with moods that are nearby to the participant's selected mood.

The musician specifies the mood they feel the audio file elicits by moving an indicator dot to the location on the grid which represents that mood.

2.2 Performance Interface

The performance interface is designed to engage the participants by giving them agency over the choice of audio presented during a performance. When a participant first arrives at the Eavesdropping website they first must initiate a new performance or join an existing performance. If the participant opts to initiate a new performance, the system requires selection of an audio set (determining the audio files which will be used for the performance), indication of a duration intended for the performance in minutes, and the entry of some basic information about the location for the performance (venue, city, province, country). This basic information will show up in a list for subsequent participants who wish to join an existing performance and will allow them to select the performance that is happening in their location. The initiator is then shown a simple, one-word password for their specific performance. This allows the initiator to announce the password to the local room to prevent people from other locations joining a performance and skewing the results. Subsequent participants wishing to join a performance merely have to select their location from the list and enter the password that was announced. They will then be taken to the interactive performance interface as is shown in Figure 2 for the duration of the performance. On the left half of Figure 2, a participant indicates her mood on the mood matrix grid, on the right half she indicates whether the audio file sent to her by the system matches her mood.

2.2.1 *Mood Identification*

The performance interface requires each participant to initially set her mood on the two-axis mood matrix by moving a dot to the appropriate location on the grid. Once a participant has set her initial mood, the system will begin sending audio files to her browser based on that mood (the server system is defined in the next section). While each audio file is playing the system will continue to show the mood that was initially set but will prevent the participant from editing their mood. Preventing editing of moods is required for two reasons. First, the intent is that a participant will not become absorbed in the interaction at the expense of attention to the audio performance. Second, aesthetically, a participant will not get a sense of the general state of other participants in the room if everyone is adjusting their mood continuously. For this reason, updates to participant mood are offered only at periodic intervals, currently set to every three minutes. The mood matrix screen shows participants the remaining time until another edit is available so that they do not become frustrated with the inactive interface.

2.2.2 *Mood Reinforcement Question*

While each audio file is played from the participant's browser, she will also be offered a Yes/No mood reinforcement question asking whether the audio that is playing from her computer matches the mood she had indicated (seen on the right side of Figure 2). This interaction is optional. If a participant wishes to merely listen to the performance and observe the room, the performance will continue to play and request new files without any response. If a participant responds by selecting the Yes or No button, the conductor system uses this feedback to reinforce the mood tagging of the audio file.

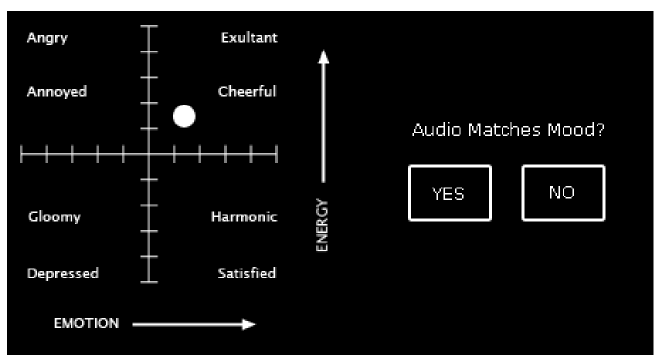

Figure 2. Participant Interaction Interface

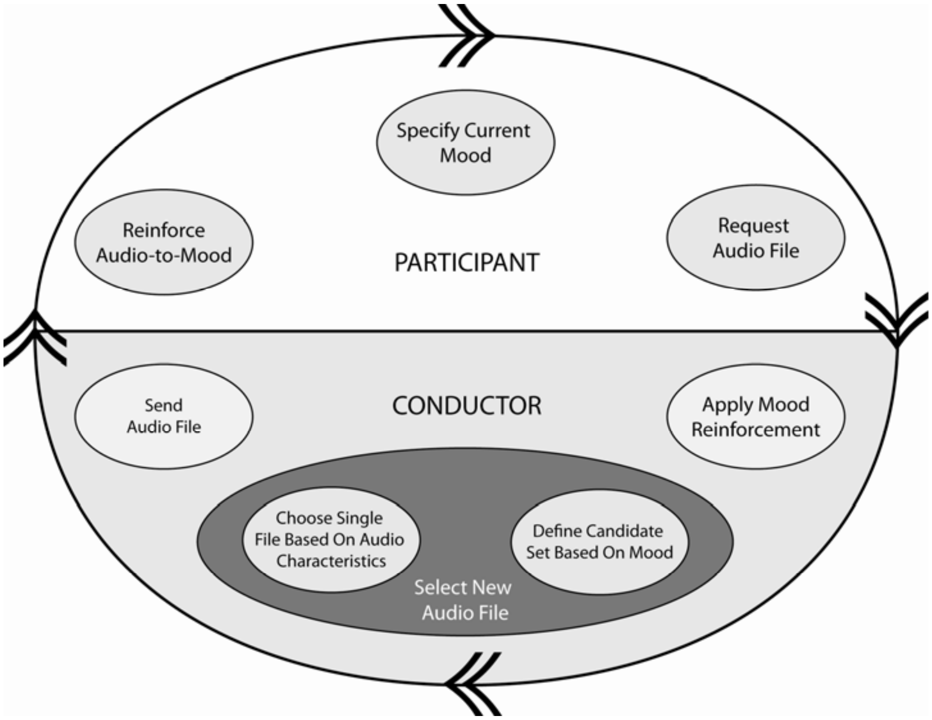

Figure 3. Request-Response Cycle

The initial mood input by participants serves to increase the engagement of the audience by giving them the sense that they are affecting the performance. Reinforcing the mood tagging of audio files further engages participants by offering them the ability to imprint their legacy on future performances. During the performance a participant hears a combination of the audio being played from her own computer and from all other players' computers in the room. Each is aware that the audio being performed relates to the moods others have selected. When interesting audio catches the ear of a participant, they become aware of the person whose computer made the sound. In the relaxed environment of a voluntary art performance this heightened awareness and shared experience is intended to lower barriers to communication and encourage interaction between individuals in the public environment.

2.3 Machine Learning and the Conductor

The centerpiece of the server-based, audio selection engine is a situational-aware Conductor which learns from participant responses and sends an audio file to each participant based on her mood and the mix of the audio sent to other participants. The Conductor first assembles a candidate set of audio files which approximate the mood identified by the participant in the request. The system then chooses one file from that set which will provide an aesthetically pleasing mix of audio characteristics with the other files playing from other participants' computers. Mood reinforcement responses from participants are stored to improve the mapping of audio files to moods and then utilized by the Conductor when selecting the candidate set. A simplified view of the full participant-request and Conductor-response cycle is visible in Figure 3.

2.3.1 Reinforcement Learning

The algorithm managing the mood representation data based on participant responses and to select which files to send to participants to represent their mood is a variation of the popular Q-learning reinforcement learning [13, 14]. In this case, a Q-value table is utilized to record responses as well as for choosing a candidate set of audio files to represent a specific mood.

2.3.2 Mood Storage and Values

Each audio file is associated with a 10*10 data table to store the utility values that have been learned for that file for each possible mood. The two mood characteristics are each normalized on a scale from 1 to 10 offering 10^2 possible moods in the grid. When an audio file is first uploaded into an audio set, this table is pre-seeded with very small random values (meaning that this file is nearly equally appropriate for any mood). The file receives its first reinforcement of a mood from the musician during initial tagging. When the musician selects a mood (for instance, moves the dot to a 2 on the *Emotion* scale and a 3 on the *Energy* scale, seen on the right side of Figure 1), the value in the mood table at that cell (2, 3) is increased by simply adding 0.1 to the value at the mood specified. The higher the Q-value[1] at a specific mood-coordinate for a file, the more likely is that file to get selected to represent that mood. Values in the mood table can range from 0.0 to 1.0. Each time a mood is reinforced for a given file j, a frequency count n_j is also incremented so the system knows how many reinforcements a particular file has received. The total number for reinforcements received for a given audio set N is also stored.

2.3.3 Mood Reinforcement

During the performance a participant's Yes or No response to the question of whether the audio matches their mood determines if a file will be positively or negatively reinforced to represent the mood the participant has indicated. In this case we simply update the Q-value ($Q_{(j)}(x, y)$) for the current mood (x, y) and the current file (j) by adding the value of the learning rate (α) multiplied by the reward value (R) to its existing value.

$$Q_{(j)}(x,y) = Q_{(j)}(x,y) + \alpha R$$

At present the learning rate is set to a constant, 0.1, and the reward value has been set to 1 for positive reinforcements and -1 for

[1] Purists have to keep in mind that these Q-value bear little resemblance to traditional ones as this is a variation on the usual framework. Still, this is the closest we have found in the existing literature.

negative reinforcements. Given that the range of values for each mood falls between 0.0 and 1.0, each file can reach its maximum value with ten positive reinforcements.

2.3.4 Exploitation and Exploration in File Selection

Selection of files to represent a participant's mood utilizes a system that takes into account the fact that in a learning-based model, the best fit for a mood may not be the file that has the highest Q-value for that specific mood. There may be suitable files that have been less reinforced (and thus probably less used) that are worth exploring. In general, reinforcement learning faces the problem of balancing exploitation and exploration. This trade-off, classic in artificial intelligence and machine learning, is about choosing at any given point whether to exploit the file that has the highest degree of confidence to represent a specific mood (in this cases the highest Q-value) or exploring files for which the real Q-value is less known.

A pure exploitation strategy formula would select the file (j) which has the highest Q-value (Q) at the mood location specified (x, y), pondered by its "confidence" (the ratio between the number of reinforcements received for j(n_j) and the total number of feedbacks available for the given audio set so far (N).

$$j \leftarrow \underset{j}{argmax}\ \frac{n_j}{N} Q_j(x,y)$$

In order to achieve some exploration, a set of possible candidates is built containing not only the optimal file (according to the above exploitation function) but also other candidates chosen as follows.

Since we do not want to choose files that are thought by the system to be too far from the user's input, a good compromise implemented in the Conductor is to look at files that are thought to be optimal for moods that are nearby the one requested. For this we select the best files (according to the previous formula for each mood with a Manhattan distance of no more than two from the requested mood. That gives up to 12 (and a minimum of 5) other candidate files to the Conductor to select from.

For reasons that are well beyond the scope of this paper, this constitutes a valid exploration strategy that does not prevent the convergence of the learning system toward the true Q-values (under the assumption that these exist and are static) and thus an optimal mapping.

2.3.5 Audio Characteristic Mixing

Once a candidate set has been identified to match the participant's mood, a single file needs to be chosen based on its ability to mix aesthetically with the other audio files that are playing in the room. This is determined by setting a limit to the combined sum of any individual characteristic of all the audio files playing in the room at the time. Each formal audio characteristic (*Tempo, Timbre, Harmelodic, Density*) is given a constant multiplier (M), determined heuristically through pilot evaluation, to limit how it scales with the number of people (P) involved in a performance (listed in the second column of Figure 5).

If there is only a single participant, the limit value equals the maximum for any characteristic, 100. As new participants join the performance, the limit value is adjusted based on the number of participants and the population multiplier via the following equation:

$$\text{Limit} = 100 * M * P$$

	M	Limit $P = 2$	Limit $P = 3$
Tempo	1	200	300
Timbre	.9	180	270
Harmelodic	.8	160	240
Density	.5	100	150

Figure 5. Population Modifier and Limit Examples

The Conductor sums the values of the each characteristic from all the audio files currently playing and subtracts this amount from the limit values for each characteristic. It then narrows the candidate set provided by the mood selection operation to those files which do not exceed that difference in any characteristic.

As an example, Figure 5 shows the limits for each characteristic for 2 and 3 participants. So if there are two participants playing audio files with densities of 50 and 35, and a third participant joins the performance, the program would narrow the candidate set to only those audio files with densities less than 65.

Once the candidate set has been narrowed to those files which ensure that no limit has been exceeded, the Conductor randomly selects a single file from the remaining set. If no candidates exist that do not exceed the limit in any one characteristic or if the limit value has already been surpassed by the audio files currently playing, the Conductor then looks for the candidates that exceed all three characteristics the least. The selected file is sent to the participant's browser for performance and evaluation.

3. IMPLEMENTATION

3.1 Platform

This project runs on a Windows 2003 Server machine with IIS 6.0 serving the application to the web. The back-end database for the system is MS SQL Server 2000. The server code is written for ASP.NET 1.1 in C# using MS Visual Studio 2005. Client browser code is written in Javascript.

3.2 Software and Browser

The system was designed and tested for Microsoft Internet Explorer and Mozilla Firefox on Windows and Mac operating systems. The intent was to utilize a common framework to function for the widest number of participants requiring the least specialized software. However, in order to consistently deliver audio to all browser configurations on multiple operating systems, the Adobe Flash browser plug-in offered significant advantages over other methods of embedding audio. In initial tests, it was found that most participants from the general population already had the Flash browser plug-in loaded on their systems. The embedded Flash player effectively streams audio from the internet server to the client to minimize delays in playback.

The server tracks individual participants via browser cookies. Each participant gets a unique ID that is transferred with each request. This allows the system to monitor how many people are participating in environments behind proxy servers where several participants may share the same IP address.

All client content is delivered and received via ASP.NET. Manipulation of the client-side, mood-tagging controls (the sliders and the mood matrix) are performed via DHTML tracking of mouse events and manipulation of document elements via style properties.

3.3 Guidelines for Audio Files

Composers uploading files should be aware of some guidelines in designing audio for this system. Files should be mp3s of a standard format that can be played in an Adobe Flash browser plug-in without requiring any special codecs. It should be noted that these will typically be performed from laptop speakers with poor bass response. Volume should be normalized. Due to the variable length of a performance, no file should be longer than approximately two minutes with most files being under one minute in length. Composers will not be able to specifically align the beats of samples. The timing of playback of audio files is determined by when the participants join the performance as well as network latency in the connection. Lastly, the lack of pitch information in the representative data creates an environment where files in various keys could be combined. Composers should therefore design audio that aligns to a complimentary set of keys, or a set of audio that explores a 12-tone range and which appeals to the random, generative possibilities of the system.

It is advised that composers upload a reasonably large number of files (approximately 100 or more) which cover a varied mix of characteristic combinations.[2] This will ensure that the Conductor can select a file that matches the participant moods while providing varied characteristics within the mix of audio in the room.

4. BACKGROUND AND RELATED WORK

In public environments, individuals interact by means of a variety of bodily and auditory cues and gestures. These ambient communication techniques can be directed at specific individuals or may be general expressions of mood meant for anyone who happens to notice. For example, people will shuffle in their seats if they are uncomfortable, they will cough if someone nearby is on a loud cell phone call, they will laugh loudly to attract attention. LaBelle notes that “material presence is determined by the material intervention of social events, physical movements, and the ebb and flow of crowds” [15]. Visitors to public spaces, such as a café, like the audience in a music venue, seek the passive awareness of others to achieve a sense of connectedness born of shared experience [16]. This project highlights the exhibitionism and voyeurism in the public sphere by increasing shared experiences to encourage deeper interaction.

It is now commonplace for public environments to be WiFi internet enabled thus encouraging the use of networked devices such as laptops and PDAs. Simultaneous internet access for several individuals in a public space offers a means of interaction beyond physical proximity; we share the same bandwidth, our data comingles as we engage in private acts on our solitary screens. Data on the network and in databases has become the raw material on the palette of the digital artist. Manovich acknowledges the field of data art as a representational art form in its attempt to portray human beings via their activities on the network [17]. The process of mapping invisible data into phenomena within the realm of human senses then becomes an act of communication and raises questions as to the politics of media representation. What data should be exposed? Which dimensions should be dominant? Does the interface reveal to the participant how their data will be presented? Whitelaw raises the issue that in data mapping, some mediation by the artist between data and its representation is inevitable and as such, data misappropriations, transcodings, and manipulations become a cultural act and statement [18]. This remediation of data becomes more meaningful then when the mappings are motivated and not arbitrary.

In the project, Eavesdropping, the data mapping occurs in the tagging of audio files with characteristics to reflect moods and emotions. Audience participants volunteer their moods and are physically present at the location their moods are projected as audio from their computers. The relationship between music and mood is a well-studied field which has identified two successful roles for music. On one hand music is an effective means to portray emotions that are recognizable by listeners [19], on the other, music has the ability to affect a specific emotional response in listeners [20, 21]. This raises two questions for the audio-to-mood mapping that is applied in the Eavesdropping context. The first asks how we can be certain that our tagging of audio files with moods is consistent within a population of listeners with minimal data misappropriation. The second proposes that as our listeners are immersed in audio, that their mood will likely be affected by the audio and change over time. This then requires that the performance environment can adapt mood mappings to the wider population, and that the interface to capture participant moods offers the flexibility to allow subsequent adjustment of mood.

Initial mappings of moods to audio are applied by a musician uploading audio files into the system, but these moods may be biased toward the musician’s own personal taste or aesthetic. One goal of the project aims to build associations between the participants involved and the audio playing from their system. In order to produce this effect, the system must consistently deliver audio that matches the participant’s defined mood. To increase confidence in the mood data tagged to the audio files and to avoid data misappropriation, the system employs a machine learning agent with a policy to narrow the gap between the mapped characteristics and the actual perceived moods and emotions. This system encourages participants to reinforce the mood mappings of the audio files by responding whether a file represents their mood or not. This has two desirable effects. First, the system will be narrowing in on a set of values which can accurately portray emotions for the widest possible range of participants, and second, the participant will feel engaged in the workings of the system as an agent with control over future outcomes.

[2] The six characteristics define a space resulting in $100^4 * 10^2$ possible combinations for the formal and mood values.

There are many sonic art projects which have used computer networks, identified as *Interconnected Musical Networks* by Weinberg (IMNs), for a variety of different roles [22]. Many early projects focused on multiple musicians in different locations collaboratively performing via network connections, such as the *League of Automatic Music Computers* and *Hub* among many others [23, 24]. In these systems, the goal is to achieve a low-latency means to communicate the actions of the various musicians over the network. In Eavesdropping, the network does not act as a means of communication between performers, but provides audience connectivity to the host server which acts as a conductor for the audio each participant will play. Perfect timing of the audio presented from each participant's computer is not a goal of Eavesdropping. In the social acoustic ecology of physical spaces, people make sounds at irregular moments; their lack of synchronicity often elicits a variety of interesting interpretations. Use of computational and network delay as a performance element can be seen in the work of Chris Shafe and Greg Niemeyer in the sonic installation project *Ping* [25].

A key focus of Eavesdropping is that the networked system is designed for compositions to be performed in a localized environment. Similar projects offer an instrument-based approach to localized network performance by allowing musicians and audience alike to perform together in a collaborative sound space. In Chris Brown's *Talking Drum*, a server-based conductor monitors input from microphones and generates collaborative audio to be performed via four speakers in a localized environment [26]. Barbosa's *Public Sound Object* project provides participants with a visual representation of sound objects on a screen which can be manipulated to affect the pitch, reverberation and amplitude, during their synthesis and playback in a public installation [25].

Eavesdropping brings together pre-recorded audio from musicians to represent moods input by participants and mixed by an artificial conductor. The focus is more on audience interaction than instrument manipulation. Though the interface intents to achieve engagement and personal affinity with audience participants, it should not distract them from the mix of audio in the room and recognition of the moods of other participants. This project differs from similar sound art projects in the careful balance it is attempting to achieve between audience interaction with the system, personal affinity with the sounds emanating from participants computers and connectedness between the audience members.

5. EVALUATION AND DISCUSSION

Eavesdropping has been developed in two stages. In the first implementation the control was weighted significantly more in the hands of the musician than the audience. In that system there was a compositional interface which allowed musicians to create sequences of layered moods that would be assigned to participants and performed at their computers rather than receiving mood input from the participants themselves. In fact, the audience was relegated to the more traditional role of passive recipient of the musical performance. This system worked well as a gallery installation where computers were placed around a room with the audience wandering in to hear the mix of moods in the space. Eavesdropping was presented in this format at an exhibition titled *Other Stories* at Simon Fraser University in the Spring of 2007. This however defied the original intent of the project which was designed to exploit the relationship between a participant and their private computer in a public space.

When this same system was performed in public settings with individuals on laptop computers, the issues with user agency manifest both immediately during the performance as well as in participant frustration expressed during subsequent question and answer sessions. During the performance, participants engaged in all sorts of actions which clearly expressed their intent to be involved in the performance. Some people turned their laptops around to face the majority of other participants. Others raised their laptops above their heads as if to be heard. Some opened multiple browser windows to the system so that their computers were playing multiple sessions. Still others opened music players on their machines and contributed outside sources of audio to the mix.

In question and answer sessions after the performances, many participants asked if there was a future version planned which would allow participant interaction. Despite the fact that an audience is accustomed to passive listening when playing radios or mp3 players or even music from laptops, once the intent is that their laptops are performing for the rest of the room, participants want agency. Likewise, participants have no issues with control when listening to a live performance by others, but when the performance uses their own laptop as the instrument, participants demand control over their machines.

This current version of Eavesdropping attempts to address these issues to bring the system closer to its intent in creating rich shared experiences to improve connectedness between participants. The system offers two roles for participants, one as source material for the performance itself, and two as authority for the improvement of the system. In their first role, the participants inject the performance with something personal, their mood. They know that others in the room have also input their moods and can hear the results of this input. Participants in pilot performances raised the questions a) why would I volunteer my real mood, and b) why would I want to hear others' moods if others are all having a bad day? The first is partially addressed by the design of the system itself. If the participant knows that the system may not be correct in its assignment of moods, and that they have the opportunity to correct the associations, this disarms the association between the participant and the audio. If the sounds are embarrassing, participants can laugh it off as if the system made a bad association.

The second question is one of context. Would this system be utilized as background music or just in the context of a performance? People are willing to tolerate uncomfortable situations as entertainment (scary movies for instance), but not as ambient environment. On the other hand, if some indication of bad moods were present on an ambient level in more public or social situations, this may provide a means for interpersonal relations which could serve to improve the moods of those in need by giving them a voice that is perhaps more accessible in its passivity.

Overall, the engine has performed as expected with full audio delivered to participants in small groups and minimal audio delivered to participants in larger groups. The audio performance is clearly shaped by participant moods, and formal characteristics sounded balanced for both groups of 4-5 and groups of 10-12.

Regarding the aim of improving interpersonal awareness and connectedness, initial performances with Eavesdropping were successful at inspiring conversations to arise between disparate people in the performance environment.

6. CONCLUSION AND FUTURE WORK

This paper describes an internet-based, interactive audio performance system designed to increase connectedness between individuals sharing the same physical space. The system accomplishes this by encouraging participants to share their moods with other participants, represented as audio and played from their laptop computers. Initial performances showed that the system does create the necessary shared experiences to lower barriers to communication amongst strangers in a public setting. The most recent implementation opens up participant interaction to deepen associations between participants and the audio being performed.

Additional studies involving the general public are planned to explore new settings in the reinforcement learning system and to further evaluate participant communication via the new interactive interface [27].

7. ACKNOWLEDGEMENTS

This research is partially funded by a commission from Rhizome.org and the New Museum of Contemporary Art in New York, with additional resources provided by Simon Fraser University. Eavesdropping is also featured in the 404 Festival of Electronic Art in Trieste, Italy in 2008. This system is online at http://www.oddible.com/cafe.

8. REFERENCES

[1] Dourish,P. and Bly,S., Portholes: Supporting awareness in a distributed work group. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Monterey, California, May 3-7, 1992). ACM Press, 1992, 541-547.

[2] Markopoulos,P., Romero,N., Baren,J. v., IJsselsteijn,W., Ruyter,B. d. and Farshchian,B., Keeping in touch with the family: Home and away with the ASTRA awareness system. In *CHI '04 Extended Abstracts on Human Factors in Computing Systems* (Vienna, Austria, April 24-29, 2004). ACM Press, 2004, 1351-1354.

[3] Truax, B. Acoustic communication. Ablex Publishing, London, 2001.

[4] Truax, B. Handbook for acoustic ecology. Cambridge Street Publishing, London, 1999.

[5] Gaver, W. W., What in the world do we hear? an ecological approach to auditory event perception. *Ecological Psychology 5,1 (*1993). 1-29.

[6] Wakkary, R., Hatala, M., Jiang, Y., Droumeva, M., and Hosseini, M. 2008. Making sense of group interaction in an ambient intelligent environment for physical play. In *Proceedings of the 2nd international Conference on Tangible and Embedded interaction* (Bonn, Germany, February 18 - 20, 2008). TEI '08. ACM, New York, NY, 179-186.

[7] Vickers, P. and Hogg, B., "Sonification abstraite/sonification concrete: An 'aesthetic perspective space' for classifying auditory displays in the ars musica domain," presented at International Conference on Auditory Display (ICAD2006), London, UK, 2006.

[8] Schroeder,R. and Axelsson,A., *Avatars at Work and Play: Collaboration and Interaction in Shared Virtual Environments.* Springer, Dordrecht, The Netherlands, 2006.

[9] Dabbish,L. and Kraut,R., Coordinating communication: Awareness displays and interruption. In *CHI '03 Extended Abstracts on Human Factors in Computing Systems* (Ft. Lauderdale, Florida, April 5-10, 2003). ACM Press, 2003, 786-787.

[10] Alty, J. L. Can we use music in computer-human communication? In *Proceedings of the HCI'95 Conference on People and Computers X* (Huddersfield, United Kingdom, August 29-31, 1995). Cambridge University Press, New York, NY, USA, 1995, 409-423.

[11] R. E. Thayer. 1989. The Biopsychology of Mood and Arousal. Oxford University Press.

[12] Liu, D., Lu, L. and Zhang, H. Automatic mood detection from acoustic music data. In Proceedings of the International Symposium on Music Information Retrieval (ISMIR'03) (Baltimore, MD, USA, October 26-30 2003).

[13] Sutton, R. S. 1998. Reinforcement learning: An introduction. Neural Networks, IEEE Transactiions.

[14] Russell, S. and Norvig, P. 1995. Artificial Intelligence: A Modern Approach. New Jersey: Prentice-Hall.

[15] B. LaBelle. 2006. Background Noise: Perspectives on Sound Art. New York and London: Continuum International Publishing Group.

[16] Erickson, T. and Kellogg, W. A. Social translucence: An approach to designing systems that support social processes. ACM Transactions on Computer-Human Interaction 7, 1 (March, 2000), 59-83.

[17] Manovich, L. The anti-sublime in new media. Chair Et Metal / Metal and Flesh (Fall, 2002),

[18] Whitelaw, M. Hearing pure data: Aesthetics and ideals of data-sound, In Unsorted: Thoughts on the Information Arts: An A to Z for Sonic Acts X Altena,A., Ed.

[19] Kivy. P. 1989. Sound Sentiment: An Essay on the Musical Emotions. Philadelphia, PA: Temple University Press.

[20] Berg, J. and Wingstedt, J. Relations between selected musical parameters and expressed emotions: Extending the potential of computer entertainment. In *ACE '05: Proceedings of the 2005 ACM SIGCHI International Conference on Advances in Computer Entertainment Technology* (Valencia, Spain, June 15-17, 2005). ACM, New York, NY, USA, 2005, 164-171.

[21] Scherer,K. R. and Zentner,M. R., Emotional effects of music: Production rules. In *Music and Emotion: Theory and Research.* Juslin,P. N. and Sloboda,J. A., Eds. Oxford University Press, Oxford, England, 2001, 361-392.

[22] Weinberg, G. The aesthetics, history and future challenges of interconnected music networks. In *Proceedings of the*

International Computer Music Association Conference (Goteborg, Sweden, Sept. 2002). 2002, 349-356.

[23] Bischoff, J., Gold, R. and Horton, J. Music for an interactive network of microcomputers. *Computer Music Journal 2, 3* (Dec, 1978), 24-29.

[24] Gresham-Lancaster, S. The aesthetics and history of the hub: The effects of changing technology on network computer music. *Leonardo Music Journal 8 (*1998), 39-44.

[25] Barbosa, A. Displaced soundscapes: A survey of network systems for music and sonic art creation. *Leonardo Music Journal 13 (*2003), 53-59.

[26] Brown, C. Talking drum: A local area network music installation. *Leonardo Music Journal 9 (*1999), 23-28.

[27] van Baren, J., IJsselsteijn, W. A., Romero, N., Markopoulos, P. and de Ruyter, B., Affective benefits in communication: The development and field-testing of a new questionnaire measure. In *PRESENCE 2003* (Aalborg, Denmark, October 2003). 200