Ambient Display using Musical Effects

Luke Barrington[1], Michael J. Lyons[2], Dominique Diegmann[2], Shinji Abe[2]

[1]Electrical & Computer Engineering
University of California, San Diego
9500 Gilman Drive, La Jolla CA 92093
+1-858-822-0077
lbarrington@ucsd.edu

[2]ATR Intelligent Robotics and Communication Labs
2-2-2 Hikaridai, Seika-cho
Soraku-gun, Kyoto 619-0288 Japan
+81-(0)774-95-1433
mlyons@atr.jp

ABSTRACT
The paper presents a novel approach to the peripheral display of information by applying audio effects to an arbitrary selection of music. We examine a specific instance: the communication of information about human affect, and construct a functioning prototype which captures behavioral activity level from the face and maps it to musical effects. Several audio effects are empirically evaluated as to their suitability for ambient display. We report measurements of the ambience, perceived affect, and pleasure of these effects. The findings support the hypothesis that musical effects are a promising method for ambient informational display.

Categories and Subject Descriptors
H.5.2[**User Interfaces**]: Auditory (non-speech) feedback; H.5.2 [**User Interfaces**]: Evaluation/methodology; I.2.10 [**Vision and Scene Understanding**]: Video analysis

General Terms
Design, Experimentation, Human Factors, Theory

Keywords
ambient display, affective computing, musical interface

1. INTRODUCTION
Contemporary life places demands on our ability to fulfill multiple roles in the distinct spheres of our professional, home, and community lives. One may have to juggle childcare, family, professional, and social life while changing demographics involve an increasing fraction of the population in elder care as well. From an information processing viewpoint, this draws on the ability to share one's attentional resources between several quite disjoint responsibilities. The ability to peripherally monitor situations, while accomplishing an unrelated activity could aid greatly in multi-tasking life's various demands. Ambient displays are an active area of research aimed at leveraging one's attentional scope by presenting information in a low-stress fashion in the periphery. Music may provide a natural and flexible medium for ambient display because it is considered a desirable component of many human environments, public and private. Music has low interference with, and may even enhance participation in unrelated activities; most can listen to music while working, relaxing, cleaning the house, chatting, driving, or exercising. Moreover, music is sufficiently complex that it has the potential to provide a display substrate with considerable bandwidth. The relation of music to emotions suggests that it may be well-suited for communicating information about human affect.

This work explores audio effects as a means to construct a music-based ambient display. For concreteness, we have implemented a functioning prototype which represents aspects of behavioral affect in a musical effects code. We have examined a host of audio effects, drawn from several contemporary musical genres, for their use in ambient displays, measuring their potential for ambience, pleasure, and perceived affect.

IUI'06, January 29 – February 1, 2006, Sydney, NSW, Australia.

2. RELATED WORK
While many studies have explored the sonic display of information, three works focusing on ambient aural display are particularly relevant to the current work. The Xerox PARC *Audio Aura* system [1] used pre-composed elements to create soundscapes representing workplace activity levels and specific events. Similarly, the *WISP* system (Weakly Intrusive Ambient Soundscape) [2] used barely audible natural sounds for event notification. More recently, Butz and Jung [3] have proposed a signification display based on related pre-composed musical phrases which combine harmoniously.

These works used pre-selected or composed sonic components. We see this as a potential limitation for circumstances where users may not wish to listen endlessly to the same ambient soundscape. In the approach taken here, we have tried to circumscribe this limitation by using audio effects, applied to the music of the user's choice, to encode the information to be displayed. We thus avoid the need for pre-designed musical phrases. The effects themselves must be chosen ahead of time, however, once chosen these can be applied to arbitrary music, creating an ambient display having great variety.

3. DESIGN
A general schematic of the system we propose is shown in Figure 1. Sensors such as microphones, cameras, or biosensors are used to measure affect cues for an individual or group of people. This scheme could be applied to ambiently communicate affect about a dependent elder, an infant, a pet, a kindergarten classroom, a virtual videogame environment and so on. The prototype we have implemented measures affect from facial activity using a video camera.

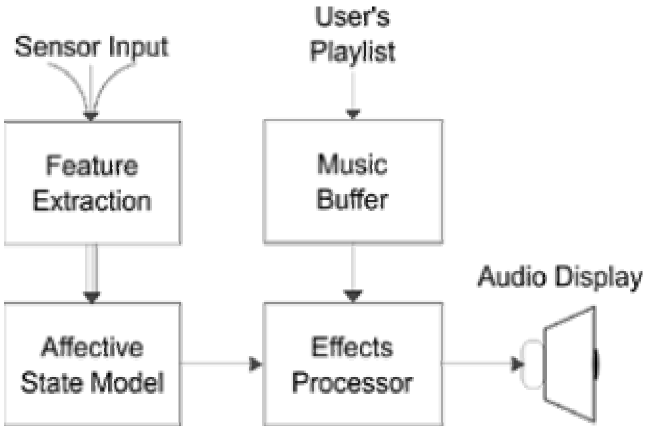

Figure 1. Schematic of the ambient display.

An example scenario this prototype would be suitable for is a dependent elder watching television, while a caregiver or family member carries out daily activities in another location of the home. The ambient affective display is intended to allow the caregiver to maintain continuous awareness of their dependent in a low-stress, pleasant fashion. The system could be generalized to add more cameras as well as other sensors, to cope with a wider range of behavioral contexts.

3.1 Affective Model

Russell [4] proposed a general model of human affect having two primary dimensions: activity level and degree of pleasure. Pleasure is more difficult to reliably measure than activity level with automatic systems. Here we concentrate on assessing and displaying human activity level. Aiming at simple, easily understandable ambient displays [5] we choose a discrete state model of activity having three levels which we label as *relaxed, normal* and *agitated.* A fourth state, corresponding to the situation when the face is not visible in the camera's frame of view is labeled *absent.*

3.2 Sensory Input

While a variety of sensors can be used to capture data related to affective qualities of human behavior, the prototype studied here uses a video camera to capture information from the human head and face, which is known to be an important channel for non-verbal communication and source of affective information.

3.3 Musical Display

Our choice of effects for representing affective states is inspired those used in existing musical genres. For example, reggae and dub music, which are considered to be relaxed, laid back genres, make use of droning basslines and delayed repetition of instrumental or vocal sections. We reasoned that such effects could be used to convey "relaxed" affective states. Conversely, electronica and DJ music often make use of filter-sweeps (where the music is filtered with a gradually varying band-pass filter) to convey a building sense of excitement on the dance floor. DJ music also incorporates turntable effects like skipping or back-spinning the record a few beats. In fact, *turntablism* may be an apt metaphor for our approach to musical display which rather than using specific musical passages to signify information, modifies existing music in a fashion that is both pleasant and informative.

4. IMPLEMENTATION

4.1 Vision System

Previous work from our group [6] described a system for continuous, real-time monitoring of facial movements using combined automatic face detection and optic flow algorithms. This system has been extended to report, at 15 fps, the following three binary signals: presence/absence of the face in the camera's field of view; presence/absence of rigid motion of the head; presence/absence of non-rigid motion of the interior of the face, corresponding to facial expression or speech activity. This information is time-averaged and used to drive the four state behavioral model.

4.2 Activation State Model

The discrete binary signals from the vision system are converted into activation levels, evaluated at a constant interval T (= 0.5s):

$$a(t) = \begin{cases} a(t-T) + n\delta & n > 0 \\ a(t-T) - 2\delta & n = 0 \end{cases}$$

where a is a signal's activation level, n is the number of detections in the interval [t-T, t] and δ (= 0.05) is an accumulator constant.

Three behavior indicators (presence, rigid, non-rigid) determine the affective state. If the presence level decays below a threshold, the absent state is triggered indicating that no face has been detected for a set interval (15s). Given presence, affective state is determined by the rigid and non-rigid motion levels which vary between [0, l]. While these indicators could be used to continuously vary musical effects, for simplicity we choose a discrete model. Threshold levels for each activation value determine the source's state:

$$\begin{array}{ll} a_R(t) + a_{NR}(t) < 0.2 & \Rightarrow relaxed \quad state \\ a_R(t) > 0.8 & \Rightarrow agitated \quad state \\ otherwise, & \quad normal \quad state \end{array}$$

4.3 Music and Effects

The affect model and music effects processing are implemented in MAX/MSP [7]. A music file is loaded from a play-list and, depending on the affective state, is routed through one of the musical effect patches described in Table 1. The effects processed output is sent to the listener's speakers. Effects for representing each affective state are selectable according to user preference. The patches are synchronized to a master position in the music buffer to allow smooth switching between effects which may depend on the current playback position in the song (e.g. skips, rewind).

A beat tracking algorithm [8] was used to estimate the current tempo of the music as well as the phase of individual beats. Synchronizing the onset of effects to beats and setting the effect duration equal to an integer multiple of beat lengths, as a DJ might apply them, enhances the musicality and pleasantness of the effects. Moreover, beat synchronization allows finer control of the sparseness of the display, ensuring that just enough information is transmitted to the user.

Table 1. Musical Effects Used in This Study

Code	Description
N	no effect
LP	low-pass filter
SB	skip backwards by 2 beats
SF	skip forwards by 2 beats
HP	high-pass filter
dub	repeat 8 beat passage delayed 8 beats, with reverb
FS	filter sweep from 100 to 10,000Hz, 12 beat period
RW	play backwards for 1 beat, then play forwards
dub HP	**dub** with low-pass filter
dub LP	**dub** with high-pass filter
T	modulate tempo linearly ±10%, period = 1 beat
FG	add flange effect intermittently on the beat
vol	modulate amplitude linearly ±100%, 12 beat period
FG HP	**FG** with high-pass filter

5. MUSICAL EFFECTS EVALUATION

The prototype system was used to generate samples of music processed with a variety of audio effects, which were used to subjectively evaluate the suitability of the system for communicating affective information in an ambient and pleasurable fashion. Ten subjects were asked to listen to 45 of the musical samples, each lasting 20 seconds. The musical substrate came from three songs of different genres: Bossa Nova, Classical (solo piano), and Electronica. Subjects heard the original clip followed by 14 modified versions in random order, and evaluated these according to three criteria: Activity Level (1=Relaxed, 2=Normal, 3=Agitated); Awareness (1=No effect, 2=Just noticeable, 3=Detectable, 4= Obvious, 5=Dominant); Enjoyment (1=Very pleasant, 2=Pleasant, 3=Neutral, 4=Unpleasant, 5=Very unpleasant). Ratings, averaged across subjects, showed several clear trends which, along with design criteria in [5], can be used to refine the prototype. Averaged ratings for each effect are shown in Figure 2. Significant correlations were found between Activity and Displeasure ($r = 0.89$), Activity and Awareness ($r = 0.96$), and Awareness and Displeasure ($r = 0.76$). Some effects "beat" the overall trends by maintaining desirable levels of pleasure and ambience for relatively agitated states. These results suggest that the following effects were particularly well suited to the ambient display of affect: the low-pass filter (LP), the dub effect plus low pass filter (dub LP), or the filter sweep are suitable for representing a relaxed state, while high-pass filtered dub (dub HP) and rewind (RW) effects well represent an agitated state. In the prototype the absent state was indicated by a gradual fade-out of the volume.

6. CONCLUSIONS

We have described a novel system which employs musical effects to represent and communicate affect ambiently. Empirical study of a variety of musical effects support our hypothesis that musical effects display systems can satisfy many of the important design criteria for ambient displays proposed in [5]. Interestingly, the correlation of effect awareness and displeasure strongly supports the notion that ambience enhances the pleasure of an affective information display.

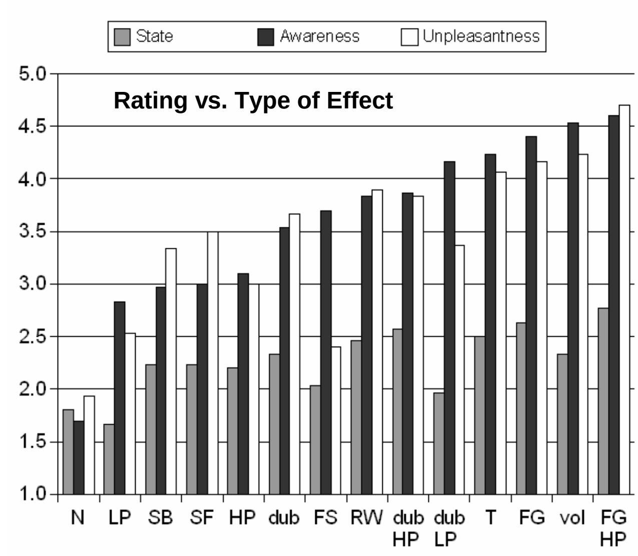

Figure 2. Results of the ambient display evaluation.

This advantage for ambient presentation, as well as the specific measurements for various musical effects may be of general interest to designers of information displays. In future work we plan to continue to test and refine the system as well as explore the display of more complex behavioral models.

7. ACKNOWLEDGMENTS

This work was supported in part by the National Institute of Information and Communications Technology. LB was a recipient of a grant from the NSF/JSPS EAPSI Program.

8. REFERENCES

[1] Mynatt, E., Back, M., Want, R., Baer, M. and Ellis, J. Designing audio aura. *Proceedings, CHI'1998*, 566-573

[2] Kilander, F. and Lonnqvist, P. A whisper in the woods - an ambient soundscape for peripheral awareness of remote processes. *International Conference on Auditory Display.* (2002)

[3] Butz, A. and Jung, R. Seamless user notification in ambient soundscapes *Proceedings, International Conference on Intelligent User Interfaces (*2005) 320-322.

[4] Russell, J.A. Core affect and the psychological construction of emotion. *Psychological Review 110*, 1 (2003), 145-172.

[5] Mankoff, J., Dey, A.K., Hsieh, G., Kientz, J., Lederer, S., Ames, M. Heuristic evaluation of ambient displays. *Proceedings, CHI'2003*, 169-176

[6] Funk, M., Kuwabara, K. and Lyons, M.J. Sonification of Facial Actions for Musical Expression. *International Conference on New Interfaces for Musical Expression* (2005) 127-131.

[7] Max/MSP. Cycling 74. http://www.cycling74.com

[8] Jehan, T. Event-Synchronous Music Analysis/Synthesis *Proceedings, International Conference on Digital Audio Effects* (2004)