Music Monitor: Ambient Musical Data for the Home

Quan T. Tran, Elizabeth D. Mynatt
Everyday Computing Lab, GVU
Georgia Institute of Technology
Atlanta, GA 30332 USA
+1 404 385 1102
{quantt, mynatt}@cc.gatech.edu

ABSTRACT
In this paper, we present a concept prototype, called *Music Monitor*, which illustrates how music can be used to balance attention between two active rooms in the home, with an initial focus between the kitchen and living room. We introduce a simple musical profile to reflect salient information from the monitored activities by identifying instrument voices to connote status information. We also discuss design goals and social implications of this project.

Keywords
Ambient, peripheral, musical data display, smart home.

SCENARIO
Consider occasions when a group of people comes together at a friend's house for a cordial visit or party. Even though the host would like to entertain the guests in the living room or den, he still has to attend periodically to other customary responsibilities of a host such as preparing food and drinks in the kitchen, welcoming guests as they arrive, and so on. With the help of *Music Monitor*, the host would be able to balance his attention better between these parallel activities as he sees appropriate, and would become more engaged in the party.

BACKGROUND
Presented in the background of a multimedia production, musical accompaniment elegantly leverages its inherent dynamics to complement and reinforce the intended message of the video content. Particularly, motion pictures employ distinctive instrumental voices in their musical scores to resonate the developing mood and tone of an unfolding story. Additionally, video games use tempo changes in their music tracks to signify imminent danger as players proceed between different areas. With the same spirit, *Music Monitor* applies these musical techniques toward the home.

DESIGN GOALS
In general, *Music Monitor* enables the homeowner to be aware of two places at one time. The homeowner attends to a task in one room. Meanwhile, he hears music reflecting the activity level in another room. As the user moves between the two rooms, the music switches accordingly to relay the status in the other room. In particular, the dynamic music reflecting the changing activity in the kitchen is continuously being broadcast in the living room and vice versa. It is hoped that the user will attain a balance of attention between the kitchen and the living room, and achieve a heightened awareness about both concurrent activities. In the same spirit, the *ARKola* [1] plant simulation observed that sound seemed to promote a balance of attention between two distinct activities. Namely, the sounds successfully bridged the division of labor and relative inattention to concurrent problems. The study concluded that sounds serve as shared reference points so partners could refer directly to events they couldn't see. The sounds also helped users monitor their own side of the plant and freed them to attend to the other side. Lastly, the ability to hear processes meant that it was possible for participants to travel to the other side of the plant and still continue to monitor their own side. *Music Monitor* intends to extend these benefits of sound found in the *ARKola* simulation to that of ambient music.

In a sense, *Music Monitor* offers an extra set of eyes and ears by keeping itself abreast of interesting events around the house and relaying their status as ambient music in real-time. However, this musical messaging is not meant to replace common household alerts such as fire alarms and doorbells, but rather to complement these by providing gradual, earlier notices of events to alleviate abrupt interruptions so that the user is able to pace his tasks more efficiently. In the same spirit as *Audio Aura* [2], the information provided by *Music Monitor* is designed to be serendipitous. The user appreciates it when he hears it, but he does not rely on it in the same way he relies on traditional house alarms and audio signals built into household appliances. Hence, *Music Monitor* focuses upon events that the homeowner would be interested in knowing the running status of although these events would not be necessarily critical.

The overall objective of *Music Monitor* is to help the homeowner balance his attention between activities in different rooms. Its design strives to present an unobtrusive and appropriate display medium in which to provide sufficient mediation of high-level meaning and serendipity of information. Although the scope of this project encompasses a wide range of computing technology, the pivotal component is the music. The generated music must be informative without losing its inherent acoustic qualities. Thus, the music profile and the corresponding information must be carefully designed.

The Data

In the same spirit of AROMA's [3] abstract representation to yield enough information about the changing monitored events to achieve system awareness but still respect a person's privacy, *Music Monitor* supports general awareness rather than specific activity detection. Thus, four discrete, semantic states (ok, next, warning, alert) are defined to classify the room dynamics depending on their high-level meaning. The *ok* state indicates that all monitored activities are proceeding smoothly and no further attention is needed. The *next* state denotes an interim or pause of activity. A time-dependent variable triggers the *next* state to signal that the current activity had just been completed and the next activity can begin. The *warning* state signals non-critical slips and forewarns imminent *alert* states. The *alert* state signifies that immediate attention is needed. Common activities and events are tracked by default, but custom behaviors could be trained and learned. However, every one of these semantic states requires some type of deductive reasoning to extrapolate the correct meaning from the context clues scattered around the rooms.

In the kitchen, electronic appliances and other applicable objects such as knives could be instrumented with triggers, and kitchen spaces such as the pantry, sink, and countertop can be tracked. From piecing together the relationship of objects to spaces, activities such as cooking, cleaning, and so on can be deduced. Kitchen events triggering the *alert* state consists of accidents requiring immediate action such as burns, fires, spills, injuries, and so on. Non-critical slips such as leaving the refrigerator door open would activate the *warning* state. Egg timers and other clock driven mechanisms including microwaves would prompt the *next* state. The *ok* state consists of the busy-waiting periods of cooking (e.g., allowing water to boil, food to simmer).

In the living room, sensors could be strategically placed in electronic entertainment systems and other household items to surmise the activities in which the guests are participating. Two distinct social structures (one person and more than one person) differentiate the context of possible activities. The *ok* state is in effect whenever the guest engrossed in an activity. The activities in which a single guest is entertained (e.g., browsing literature, listening to music, and playing video or board games) could be extended for a group of participants by raising the activity level threshold (i.e. noise) to accommodate group discussion. The *next* state occurs whenever there is an interim pause of activity. In a group context, noise detection could be used to sense the telltale awkward silence of indecision. For an individual guest, the relative spatial location of tagged physical artifacts with respect to time is used to indicate that the guest is returning an artifact back in storage and is browsing for another item. Noise level and relative spatial location are also used to detect rowdy participants or a restless guest respectively for the *warning* state. The *alert* state signifies when a guest breaks house rules, is bored, or arguing with another guest. In all scenarios given, the system is only concerned with detecting the telltale signs of each activity, so no further details about the participant's involvement are necessary.

The Music

Semantic State	Typical Interesting Events			Musical Instrument
	Kitchen	Living Room		
		1 guest	>1 guest	
OK	Default state	Engrossed in an activity		Woodwind (flute)
Next	Egg timer: next task	Activity finished, interim pause		Strings (violin)
Warning	Non-critical slips, alert threshold	Restless	Rowdy	Brass (trumpet)
Alert	Accident: spill, burn	Bored, broken house rules	Bored, Argu-ments	Drums (wood block)

Table 1: Instrument-State Mapping

The user supplies his favorite music tracks to populate the music library. As each song is played, a simple profile, namely instrument voices and tempo, is dynamically mapped. By default or by user selection, instrument voices are directly associated with the four semantic states defined earlier. Table 1 shows the default instrument-state mapping categories to consist of woodwinds depicting the *ok* state for their calm and airy sound, drums and brass representing *alert* and *warning* states for their loud and abrasive sound, and strings indicating the *next* state because of their distinctive sound. Within each category, the user can list particular instruments as audio cues. Table 1 maps flute as ok, violin as next, trumpet as warning, and wood block as alert. Furthermore, the different instruments within a category (e.g., violin, viola, cello, bass for strings) can be used to identify each event that had triggered that state. This finer granularity of instrument voice mappings would be able to support simultaneous, independent events.

Tempo in the musical piece proportionally corresponds to the activity level of the recent state change in the monitored room. Note that the activity level of a room is essentially directly proportional to the number of simultaneous, independent events within that room. Therefore, changes in the music tempo further support overall awareness of multiple events, and reinforce the potentially subtle variances in particular instrument timbres (e.g. tubas and French horns) with more discernable increases or decreases in the running music tempo.

With a defined musical mapping, the instrument voices are gradually incorporated into the streaming music. An instrument, signifying the most recent state change, is first introduced in the melody line of the streaming music. It then seeps into the harmony and bass lines successively at an increasing rate until the instrument voice encompasses the entire musical piece. This design technique upholds the musical integrity of the audio cues and avoids flashing the state changes as abrupt musical fanfare. The same technique is used to reflect room transitions as the homeowner moves between rooms. The gradual room change keeps in line with the overall instrument composition, so there is no apparent switch in the music band as the host enters or leaves the room. It also keeps in time with the balance of attention given to both rooms since the hostess would need some relay time to transition her mentality and responsibility between the different rooms. Other acoustical techniques are leveraged to transition the music smoothly between the different states. For instance, decrescendo is used to fade out the passing state, and crescendo is used to emphasize the emergence of the new state.

However encoded, the music should still sound to other casual listeners as just music. The musical changes should not be blatant, but rather seem as if the dynamics of the music were originally arranged and intended as such. The musical mapping is simple avoid overloading aural memory. The default grouping of instruments hopes to converge towards an intuitive mapping of musical timbre to the abstract state being represented. We hypothesize that familiarity with the music instruments and music pieces rather than astute musical perception is all that is required to effortlessly understand the encoded music. By using their own music library, users would be able to easily detect the new variances in their favorite songs. To gain familiarity with the instruments, sampling of instrument voices would be helpful to recognize a particular instrument correctly. Long-term use of the system would hone the homeowner's sensitivity to the instruments' timbres.

The User Interface

A control panel provides an interface for three main features: personalization, visual redundancy, and layout manager. Users can personalize the music and data represented by *Music Monitor*. Users supply their music tracks, configure the music profile, and play the resulting music pieces. The control panels can also double as a redundant visual output display of the musical content. The control panel provides a graphical layout of the room so that the user can easily find target sensors to disable or enable. The ideal form factor for this control panel is a painting-size, flat panel wall display. This large display would be appropriate in presenting the visual rendition of the music display as an artistic dynamic painting, and would permit the host to view the visual display from afar. Such a visual display offers a different modality for the user to interpret the encoded information, and provides visual redundancy at times when the musical display is undesirable (e.g. watching television). The large display would also be advantageous in presenting an appropriately scaled layout of the monitored room, and permit the user to interact with the display via direct manipulation (i.e. point and click to select and toggle activation of a sensor).

Providing remote controls also give users quick access to enable or disable the musical mappings onto the ambient music, and to preempt or override misinterpreted musical cues. These controls could easily be embedded as wearable devices woven into the user's apparel such as cuff links or earrings to further integrate the system into everyday context.

INITIAL MUSICAL PROTOTYPE

An initial prototype of the musical mappings and music generation has been implemented. The music pieces were given as MIDI files since this format separates various melody, harmony, and bass lines into independent channels. Microsoft's Music Producer was used to extract the music profile of each MIDI song and to establish a global music style. Then, an interface was prototyped in Visual Basic, leveraging Microsoft's Direct Music [4] to transition instrument voices in real-time. An external MIDI synthesizer was used to render authentic sounding instrument voices.

Preliminary Study

A short laboratory study was conducted to gather cursory feedback and initial opinions about the musical mapping and resulting musical display. Two user groups were identified: incidental listeners (i.e., guests) who were unaware of the musical mapping and informed listeners (i.e. hosts) who were briefed about the musical mapping. Two simple scenarios were used to stage possible home activities: a defined task (i.e., making instant pudding) and a loosely structured activity (e.g., playing card game "Go Fish"). Two slightly contrasting music pieces were selected: a duet with only a melody line and a counter melody line, and an ensemble with several concurrent parts. The hypothesis is that instrument changes in the duet will be more pronounced than those intertwined in the ensemble.

During each scenario, the subjects who played the role of "host" were asked to indicate whenever they perceived the state changes by detecting the change in instrument voices and recognized the current state by interpreting the musical instrument-state mapping. Coarse times were recorded to gauge reaction times for each state change. The analytical results show that the instrument transitions were noted after over half of the music piece had changed. Casual observations revealed that the subjects did notice the instrument transitions as the initial change was introduced into the music, but they would not confirm the transition until the majority of the music was played by the new instrument voice. The exception to this recognition latency was the transition to the wood blocks, a percussion instrument. Its staccato timbre proved to be a sharp contrast to the airy tones of the previously played tuba. On the other hand, the violin to piano transition was very subtle and smooth, especially for the ensemble, because the piano softly crept into the music. The hosts who played cards seemed to detect the transitions in the instrument voices more effortlessly than the hosts who were cooking. They rationalized that they could easily balance their attention since the card game could naturally "play itself out" so that they could pay more attention to the music. One of the cooks was able to perform his cooking very quietly, so he had no difficulty in balancing his attention between listening to the music and cooking. The others had to pause their noisy cooking to confirm that they did indeed hear a state transition, but they commented afterwards that the music does successfully portray gradual transitions rather than abrupt, blatant changes.

The subjects who played the role of "guest" were told that there would be music playing while they played cards and that they would be asked for their opinions about the music afterwards. The guests seemed engrossed in the card game and playful banter of such a "child's card game," but they also were listening to the music and letting the music further contribute to their mood and atmosphere of the activity. In fact, the guests were humming the song's melody during the post-interview. However, they did not notice anything peculiar about the song, and were oblivious to the instrument transitions. They claimed that the music did not interfere or disrupt the game, and that they didn't even pay attention to the music although they could hear it being played.

These preliminary results offer positive support for the essence of the musical mapping. Namely, the instrument transitions in the music were readily discernable and informative to the informed listener, but were not apparent to the incidental listeners.

FUTURE WORK

To realize this concept prototype, further development is needed. First, a complete implementation of the music mapping will further accentuate the initial musical design. More user evaluation in addition to previous psycho-acoustic studies of affective music timbres, instrument voices, tempo, and so on will provide a stronger assessment about the musical mapping. Next, a framework of interesting events needs to be defined with a supporting intelligent agent to piece together these contextual clues and interpret their semantic states correctly. Lastly, a network of sensors will create an active environment in which to actuate the system.

The Smart Home

Music Monitor will be deployed within the context of "smart homes," homes that are equipped with various sensing capabilities to evaluate its social implications.

The social context of the home brings new challenges and perspectives to computing research. First, the traditional goals of productivity and efficiency demanded in the workplace are not coveted in the home. Rather, peace of mind and general awareness would be considered important. Second, the activities within the home differ from traditionally defined productions and tasks in that they have no well-defined beginning or end. Rather, these activities (e.g., chatting with friends, watching television) are unstructured, informal, and continue with the ebb and flow of everyday life. Third, issues of privacy and being monitored are of concern in addition to common courtesy and social propriety. Lastly, although the house presents a relatively constant physical setting, the home occupants perform a myriad of different activities and tasks every day. Thus, the house offers a multifarious and yet stable domain in which to propose innovative and useful interface technology such as the presented *Music Monitor* concept prototype.

REFERENCES

1. Gaver, William W., Smith, Randall B., and Tim O'Shea. *Effective Sounds in Complex Systems: the Arkola Simulation*, in CHI '91.
2. Mynatt, Elizabeth D., Back, Maribeth, Want, Roy, Baer, Michael and Jason B. Ellis. *Designing Audio Aura*, in CHI '98.
3. Pederson, Elin R. and Tomas Sokoler. (1997). *AROMA: abstract representation of presence supporting mutual awareness*, in CHI '97.
4. Microsoft DirectX. http://www.microsoft.com/directx