Acoustics'08 Paris
June 29-July 4, 2008
www.acoustics08-paris.org

Smart sound environments: merging intentional soundscapes, non-speech audio cues and ambient intelligence

Ralf Jung

Universität des Saarlandes, LS Wahlster, FB Informatik, Bldg. E 1 1, Room 1.18, 66123 Saarbrücken, Germany
rjung@cs.uni-sb.de

We introduce an intelligent audio notification system for multi-user environments that provides users with information about events (e.g. important emails) in a more discreet and non-distracting way. The peripheral awareness of individual-related events is done by using non-speech audio cues which can be seamlessly integrated into artificial background soundscapes.
These ambient soundscapes are self-composed with respect to well-known perceptual constraints like auditive Gestalt laws as well as music psychological findings. To follow a hierarchical approach for the notification sounds we use notification instruments, ambient noises and traditional alert signals that are grouped by their level of intrusiveness. Since the notification system also follows a human-centered approach it takes parameters like user preferences, his/her current position in the environment and the type of event into consideration to decide which notification is the appropriate at this time.
In the paper, we will describe the architecture of the personalized ambient audio notification service, compositional constraints as well as some findings of a user study.

1 Introduction

In the vision of ubiquitous computing, the user interfaces are part of our everyday environments. For the acoustic part of the interface this means, that it has to be integrated into or coordinated with existing sounds. Already today, music is used in shopping malls to subtly influence customers, lift their mood, and hence increase consumer spending. The music used in these environments, snidely called "elevator music", is perceived only peripherally. If the background music plays a tune which we particularly like or dislike, it often crosses the border from peripheral to conscious perception and we suddenly become aware of it.
On the one hand we benefit from user adaptive and easy to use systems that provide us with information. But at the same time we are exposed to the danger of a "cognitive overload". The necessary filtering of information should depend on the user´s preferences and result in an appropriate presentation that decreases distractions of the user and other attendees [1]. How users should be provided with information depends on the complexity of data. For elementary data, the use of audio notification often makes more sense than overloading the human visual sense with additional information. Earcons are often used to provide users with information about the system status or ongoing computational processes [2]. They are short non-speech rhythmic sequences of pitch, so called motives, that can be combined to provide more complex information. Above all, notification with audio cues containing information can be done in a more peripheral way that can also be used for monitoring continuous changing data e.g. stock market progression [3].
In our work, we use location awareness to notify users via peripheral audio patterns if a personal event occurs (e.g. receiving an email that matches a user predefined keyword or the notification of a forthcoming appointment). The localization is done by an indoor positioning system that runs on the user´s personal PDA in combination with IR and RFID tag arrays. The notification system automatically starts a preselected ambient soundscape on the audio surround system in the background if the user enters the room with his registered PDA and the service will periodically check whether an event occurs regarding the user. For example, as long as the user stays in the room his email account will be monitored and the subject line of arriving new messages will be compared with optional keywords. If an appropriate email arrives the current position of the user in the room will be figured out and the loudspeaker next to him will play his notification signal that is mixed into the soundscape. Unlike using the PDA loudspeakers for playing the notification signal, the use of the room speakers avoid the source detection of the notification signal. Not only the distraction can be avoided but also the privacy will be increased, because only the target person knows his personal notification instrument that he selected. Since the instrument fits into the composition, other people will perceive the notification as part of the composition and not as a notification cue.

2 Related Work

Zimmermann [4] developed a system for automated music composition in order to assist multi media presentations. In our research we use music composed by humans, but let the system determine the mix of different pre-produced parts.
Mynatt et al. used voice, music, sound effects or a combination of these as notification sounds in their audio aura system [5]. In terms of music they used different short melodies carrying different meanings. They were not integrated into a comprehensive composition playing continuously.
Brewster did several experiments in which he investigated the effectiveness of earcons as navigation cues [6]. Among other things, he found out that earcons can be used to represent hierarchical structures after a short training phase. The results could even be improved by using compound earcons. A detailed introduction to non-speech audio cues for human-computer interfaces and the mentioned earcon experiments can found in Brewster's PhD thesis [7].

3 Musical Constraints

Music can influence the mood of the listeners and/or attract their attention. The efficiency of this approach is well-known from movies in which the mood of the spectators is manipulated by variations and changes of musical elements like rhythm, interval or chord. These elements can be improved and analyzed by their acoustic structure

Category	Parameter	Range	Emotional Impact
Time	Speed	fast - slow	pleasant - calm
	Phrasing	staccato - legato	lively - gently
	Rhythm	firm - smooth	serious - dreamy
	Dynamic	cresc. - decresc.	animated - relax
	Meter	even - odd	dignified - restless
Pitch	Mode	major - minor	bright - plaintive
	Frequency	high - low	exciting - sad
	Melody	ascending - descending	dignified - serene
	Note Range	≥octave - ≤octave	brilliant - mournful
	Harmony	consonant - dissonant	serene - ominous
Texture	Volume	forte - piano	animated - delicate
	Orchestration	instrumentation	majestic - grotesque

Table 1: Musical attributes and their emotional impact.

described by duration, energy, pitch, timbre and harmonic structure [8]. The influence on the affective states, triggered by modifying acoustic structure values, can be manifold and is the goal of functional compositions [9].
In our everyday life, such functional music can also be found in public places like waiting halls, elevators and in shopping malls where ambient music is used to provide a more friendly and comfortable environment for customers. In 1978, Brian Eno was one of the first musicians who designed ambient music (Ambient 1: Music for Airports) which can be heard actively or used for background music depending on the listener who chooses whether he wants to pay attention or not.
This type of music is also known as “Muzak” in which the complexity of the non-speech compositions is reduced to avoid the listener‘s attraction.
The necessary compositional restrictions can be found in the musical components which can be described by their time, pitch and texture. The parameters of these three categories are used to influence how ambience.
Table 1 gives a brief overview of the emotional impact of compositional parametes on the listener’s mood (based on findings described in [10]).

4 Multi-User Notification

We implemented a concrete application for ambient notification. More precisely, we present an event notification service for mobile persons in multi-user environments. To give an example for an event notification, we implemented a service to unobtrusively inform registered persons about incoming emails that passed a keyword matching filter. We will also give more details of the implemented *Indoor Positioning for Location Awareness* system.
Our claim to build a user centered notification service made a variety of demands on the architecture. Figure 1 gives a rough design overview of the four additional elements: *Positioning System*, *email Server*, *Sound Repository* and *SAFIR* (Spatial Audio Framework for Instrumented Rooms) that are used for the Ambient Email Notification service (*AEMN*).

4.1 Architecture

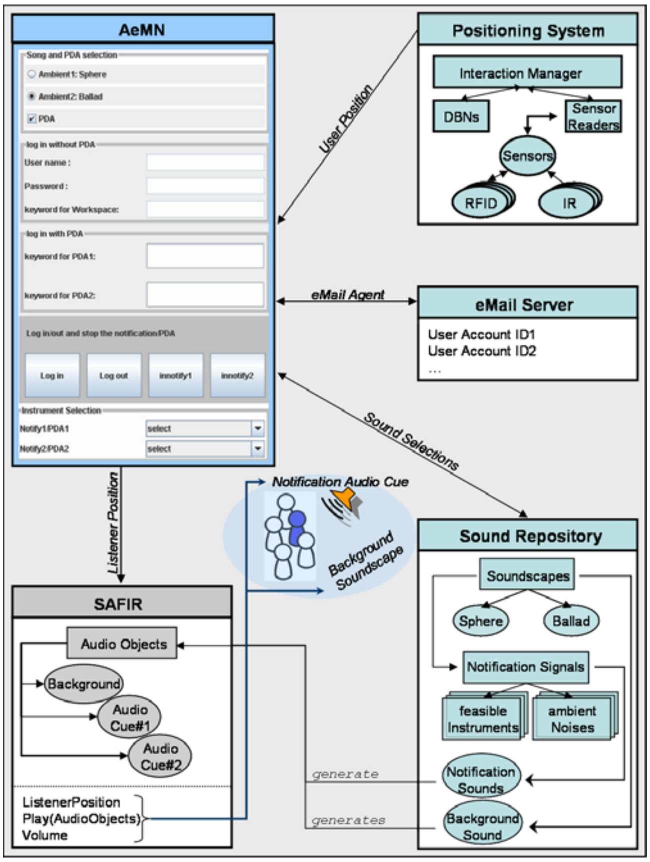

Figure 1: System architecture.

The service has been integrated into our test environment. Among other things, the room is augmented with eight loudspeakers that are mounted at the ceiling and a hidden surround hi-fi system which is connected to a audio server (see Figure 2).
In the configuration phase, the graphical administration interface can be used to choose whether the user wants to use the stationary or the mobile notification mode. The first one makes sense if the user stays at his desk most of the time. He can authenticate himself with his name and his password. The location of his personal desk and email account login information are internally stored in an xml file on the AEMN server. The alternative is the mobile version where the user´s PDA is already registered and used for finding out his current position. We assume that each user has his own PDA, so we don‘t need a manual authentication. In both cases the user has the possibility to enter a personal keyword for filtering incoming messages by their subject line. Incoming messages of registered users are periodically checked by email agents that run on the AEMN server. After authentication, the user can select an ambient soundscape that ought to be played as the background sound. The system checks the sound repository for appropriate notification signals that can be integrated into the background soundscape as notification audio cues.
After selecting his personal notification instrument, e.g.

Figure 2: Instrumented environment.

guitar or drums, the appropriate wav sound file will be retrieved from the sound repository and audio objects will be generated in the spatial audio system SAFIR [11]. AEMN recognizes when one or more registered users enter the room and automatically starts the selected background soundscape and the login process for checking the user accounts on the email server. The selected audio notification cues will be loaded. If a user receives a new email that passed the filter successfully, the coordinates of the current user position that is computed by his PDA will be matched to the spatial audio system coordinates (listener position) and the notification cue will be seamlessly integrated into the soundscape. The loudspeaker that is nearest to the target person plays the notification cue with slightly increased volume to ease the perception. The notification can be stopped by pressing a button on a user interface that is running on the PDA or at the administration interface on the desktop computer.

4.2 User Postion Awareness

There are several indoor-localization systems under research or even commercially available that mainly differ in costs, precision and the used sensors. The users wear senders that are constantly emitting data to these sensors. A centralized server then uses the collected data to calculate the positions of the users. This approach seems to be perfect for location based systems, since the derived positions can easily be distributed to other services.
For the ambient notification system, we use our own egocentric positioning system named LORIOT (Location and ORientation in Indoor and Outdoor environmenTs) [12]. LORIOT is based on infrared beacons and active RFID tags as senders. The corresponding sensors are the built-in infrared port of the user's PDA and an active RFID reader card that is attached to this PDA. Because the system is an egocentric one, the senders (infrared beacons and RFID tags) are installed in the environment and the PDA with the sensors is worn by the user. The active RFID tags have an internal 64 byte memory, of which 56 bytes are freely accessible. We use this memory to store the geo-coordinates of each tag (each tag "knows" its own coordinates). The infrared beacons send out a 16 bit wide identification code. This way all the external data that is needed to calculate the user's position is stored in the environment.
The position calculation is done with the help of a small dynamic Bayesian Network (DBN), which represents the accuracy of the infrared beacons and RFID tags in general. The idea behind this is that detecting an infrared beacon gives a high probability that the user is standing in the vicinity of that particular beacon and thus his coordinates are near to the beacon's coordinates. Receiving an RFID tag gives comparable little evidence that the user is near that tag, due to reflections and damping of the radio signals of the RFID tags. These characteristics are coded into the conditional probability table (CPT) of the DBN. During the runtime of the system, an instance of this DBN is generated for every sender that is received and each instance of the DBN is associated with the geo-coordinates of the corresponded sender. We call these instantiations of the DBN geo-referenced dynamic Bayesian Networks or geoDBNs. Each geoDBN calculates the probability that the user is standing at its coordinates and the actual user position is estimated by calculating the weighted sum of all geoDBNs.
Figure 3 shows the hardware that we use for our location-aware notification service. The audio hardware includes Hi-Fi amplifiers and loudspeakers that are connected to a multi-channel soundcard.

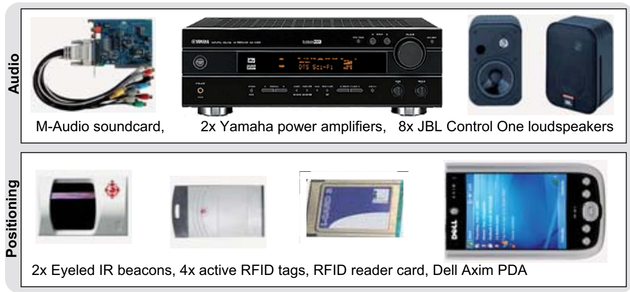

Figure 3: Audio- and positioning hardware.

5 Auditory Cues

The administration of the musical database asks for reasonable categorization parameters for musical soundscapes. We decided to describe each composition with five distinctive attributes which makes it easier for the user and the system to retrieve soundscapes and appropriate notification instruments from our audio database. Here is a short overview of our categorization description called *GEKOS*:

- **G**enre.
  Style of the composition.
  Attributes: pop, jazz, rock, classical etc.

- **E**xpression.
  Predominant impression.
  Attributes: serious, sad, sentimental, serene, happy, exciting or majestic.
- **K**ey.
  Well-defined by circle of fifths.
  Attributes: (Key-)Signature: C, G, D, A, E, H, Fsharp... and (Key-)Gender: major or minor.
- **O**rchestration.
  List of used instruments subdivided into:
  - Wind Instruments: trumpet, horn etc.
  - Percussion Instruments: snare, hi-hat etc.
  - String Instruments: cello, violin, guitar etc.
  - Electronic Instruments: electric bass, sampler etc.
  - Keyboard Instruments: piano, organ etc.
- **S**ignature.
  Metrical structure (beat) of the composition.
  Attributes: even meter (2/2, 2/4, 4/4, 6/4) or odd meter (3/4, 5/4, 6/8, 9/8, 12/8).

In the near future, we will extend our music repository with non-vocal songs from the Real World Computing (RWC) music database which is a copyright-cleared music database. Further information about RWC can be found in: [13].
Since the notification instruments can be seamlessly integrated in the ambient background song, the system works on the edge of human perception. That has the effect that an occurring notification should be perceived after a while. To prevent the effect of ignoring a notification, we also provide a hierarchy of notification signals that are grouped by their "level of intrusiveness".

1. High-Priority. Signals: arousing noises (e.g. beep, siren and bell)
2. Medium-Priority. Signals: ambient noises (e.g. birds, rain, water- and wind noises)
3. Low-Priority. Signals: notification instruments (instruments integrated in background sound)

The notification cues can be mixed into the corresponding soundscape at certain points in time with respect to the composition to avoid disharmony. We are obliged to this restriction to guarantee a fluent integration. Because of the fact that each user can select his personal instrument, other attendees will not be able to associate an instrument to a specific user even if they recognize the new cue. The personal instrument will seamlessly leave the soundscape if the user informs the system by pressing a button on his PDA or desktop GUI that he perceived the notification.

6 Evaluation Study

We recruited 25 participants (five women and twenty men) at ages from 20 to 35 years with mostly academical education.
The study took place in an instrumented room equipped with spatial audio hardware providing output through eight speakers mounted in a circular arrangement under the ceiling (see Fig. 2). In this way, we were able to position the different parts of the soundscape and the specific notification instruments independently of each other.
In addition, we prepared a computer with our test software, which includes a question window, a signal button area and a radio button area for possible answers.
We prepared two recorded and prearranged soundscapes in which the notification instruments (drums, guitars) and a conventional knocking alarm sound, both has been learned by the subject in the introductory phase, appeared randomly.
The task for the subject was to press the corresponding signal button after recognition of a notification signal as soon as possible. To prevent subjects from focusing on the background soundscape and to distract them from the auditive stimulus, they had to answer questions under time pressure. As a result of their limited cognitive resources the subjects perceived the audio signals in a rather peripheral way.
The two soundscapes were played in a row. users were told to push the corresponding signal button as soon as possible when they perceived an audio cue. We recorded the button push actions to get information about the efficiency of audio signals and we measured the elapsed time from the beginning of the notification until the subject press the corresponding signal button (reaction time).

6.1 Study Results

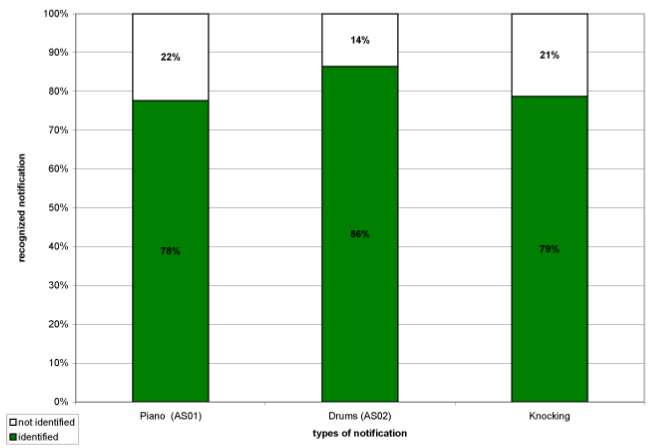

Figure 4: Identified notification cues.

Figure 4 shows recognition rates of each notification signal for all subjects. The standard deviation of the rate for the drum signal with a value of 0,162 is lower than the piano value (0.244) and clearly lower than the knocking deviation (0,286). Over the course of all 25 subjects, there

were 125 piano and 125 drum cues. Of these, 98 piano and 109 drum cues were recognized and identified by the subjects.
Overall, we can say that 79% of all knocking cues, 78% of all piano cues and 86% of all drum cues were correctly identified by the participants.
This suggests that the more rhythmical drums are easier to identify than the a melodical instrument like the piano. Especially the drum notification in AS02 surpassed the knocking sound by seven percent and proved the most efficient of the three notification types.
Our second interest was delay between the notification appearance and the act of pressing the button. Subjects had to perceive the audio signal, identify it and press the corresponding signal button on the screen. We found out that the average reaction time for piano notifications was higher on average (6,59 seconds) than the reaction time for drum (2,1 seconds) and knocking notifications (2,54 seconds).

7 Conclusion

We presented a system for auditory multi-user notification through peripheral audio cues embedded in aesthetic soundscapes. While target subjects can be efficiently notified with our approach, other persons in the same environment will not be distracted because the notification sounds are part of the composition of the background music and will only be recognized as notifications by their target person. Thereby, the type of notification can be selected and associated to events. The unobtrusive notification gives us the chance to follow a low level privacy approach.

References

[1] Brid O'Conaill and David Frohlich, "Timespace in the workplace: dealing with interruptions," in *CHI '95: Conference companion on Human factors in computing systems*, New York, NY, USA, 1995, pp. 262–263, ACM Press.

[2] M. Blattner, D. Sumikawa, and R. Greenberg, "Earcons and icons: Their structure and common design principles," *Human Computer Interaction*, vol. 4, pp. 11–44u, 1989.

[3] B.S. Mauney and B.N. Walker, "Creating functional and livable soundscapes for peripheral monitoring of dynamic data," in *Proceedings of the International Conference on Auditory Display ICAD2004, Sydney*, 2004.

[4] Detlev Zimmermann, *Ein hierarchischer Ansatz zur intentionsbasierten Musikkomposition auf der Basis musikrhetorischer Regel- und Constraintsysteme*, Ph.D. thesis, Universität des Saarlandes, 1998.

[5] Elizabeth D. Mynatt, Maribeth Back, Roy Want, Michael Baer, and Jason B. Ellis, "Designing audio aura," in *CHI '98: Proceedings of the SIGCHI conference on Human factors in computing systems*, New York, NY, USA, 1998, pp. 566–573, ACM Press/Addison-Wesley Publishing Co.

[6] S.A. Brewster, "Using non-speech sounds to provide navigation cues," *ACM Transactions on Computer-Human Interaction*, vol. 5, pp. 224–259, 1998.

[7] S.A. Brewster, *Providing a Structured Method for Integrating Non-Speech Audio into Human-Computer Interfaces*, Ph.D. thesis, University of York, 1994.

[8] K.R. Scherer and M.R. Zentner, *Music and Emotion: Theory and Research*, chapter Emotional Effects of Music: Production Rules, pp. 361–392, New York: Oxford University Press, 2001.

[9] K.R. Scherer, "Psychological model of emotion," in *The Neuropsychology of Emotion*, J. Borod, Ed. 2000, Oxford University Press.

[10] Gordon C. Bruner, "Music, mood, and marketing," *Journal of Marketing*, vol. 54, pp. 94–104, 1990.

[11] M. Schmitz and A. Butz, "Safir: Low-cost spatial audio for instrumented environments," in *Proceedings of the 2nd International Conference on Intelligent Environments, Athens, Greece*, 2006.

[12] T. Schwartz, B. Brandherm, and D. Heckmann, "Calculation of the user-direction in an always best positioned mobile localization system," in *Proceedings of the International Workshop on Artificial Intelligence in Mobile Systems (AIMS)*, Salzburg, Austria, September 2005.

[13] Masataka Goto, Hiroki Hashiguchi, Takuichi Nishimura, and Ryuichi Oka, "Rwc music database: Music genre database and musical instrument sound database," in *Proceedings of the 4th International Conference on Music Information Retrieval (ISMIR 2003)*, 2003, pp. 229–230.