Nomadic Radio: Scaleable and Contextual Notification for Wearable Audio Messaging

Nitin Sawhney and Chris Schmandt
Speech Interface Group, MIT Media Laboratory
20 Ames St., Cambridge, MA 02139
{nitin, geek}@media.mit.edu

ABSTRACT

Mobile workers need seamless access to communication and information services on portable devices. However current solutions overwhelm users with intrusive and ambiguous notifications. In this paper, we describe scaleable auditory techniques and a contextual notification model for providing timely information, while minimizing interruptions. User's actions influence local adaptation in the model. These techniques are demonstrated in *Nomadic Radio*, an audio-only wearable computing platform.

Keywords

Auditory I/O, passive awareness, wearable computing, adaptive interfaces, interruptions, notifications

INTRODUCTION

In today's information-rich environments, people use a number of appliances and portable devices for a variety of tasks in the home, workplace and on the run. Such devices are ubiquitous and each plays a unique functional role in a user's lifestyle. To be effective, these devices need to notify users of changes in their functional state, incoming messages or exceptional conditions. In a typical office environment, the user attends to a plethora of devices with notifications such as calls on telephones, asynchronous messages on pagers, email notification on desktop computers, and reminders on personal organizers or watches. This scenario poses a number of key problems.

Lack of Differentiation in Notification Cues

Every device provides some unique form of notification. In many cases, these are distinct auditory cues. Yet, most cues are generally *binary* in nature, i.e. they convey only the occurrence of a notification and not its urgency or dynamic state. This prevents users from making timely decisions about received messages without having to shift focus of attention (from the primary task) to interact with the device and access the relevant information.

Minimal Awareness of the User and Environment

Such notifications occur without any regard to the user's engagement in her current activity or her focus of attention. This interrupts a conversation or causes an annoying disruption in the user's task and flow of thoughts. To prevent undue embarrassment in social environments, users typically turn off cell-phones and pagers in meetings or lectures. This prevents the user from getting notification of timely messages and frustrates people trying to get in touch with her.

No Learning from Prior Interactions with User

Such systems typically have no mechanism to adapt their behavior based on the positive or negative actions of the user. Pagers continue to buzz and cell-phones do not stop ringing despite the fact that the user may be in a conversation and ignoring the device for some time.

Lack of Coordinated Notifications

All devices compete for a user's undivided attention without any coordination and synchronization of their notifications. If two or more notifications occur within a short time of each other, the user gets confused or frustrated. As people start carrying around many such portable devices, frequent and uncoordinated interruptions inhibit their daily tasks and interactions in social environments.

Given these problems, most devices fail to serve their intended purpose of notification or communication, and thus do not operate in an efficient manner for a majority of their life cycle. New users choose not to adopt such technologies, having observed the obvious problems encountered with their usage. In addition, current users tend to turn off the devices in many situations, inhibiting the optimal operation of such personal devices.

Nature of Interruptions in the Workplace

A recent observational study [4] evaluated the effect of interruptions on the activity of mobile professionals in their workplace. An interruption, defined as an asynchronous and unscheduled interaction, not initiated by the user, results in the recipient discontinuing the current activity. The results revealed several key issues. On average, subjects were interrupted over 4 times per hour, for an average duration slightly over 2 minutes. Hence, nearly 10 minutes per hour

CHI '99 Pittsburgh PA USA
Copyright ACM 1999 0-201-48559-1/99/05...$5.00

was spent on interruptions. Although a majority of the interruptions occurred in a face-to-face setting, 20% were due to telephone calls (no email or pager activity was analyzed in this study). In 64% of the interruptions, the recipient received some benefit from the interaction. This suggests that a blanket approach to prevent interruptions, such as holding all calls at certain times of the day, would prevent beneficial interactions from occurring. However in 41% of the interruptions, the recipients did not resume the work they were doing prior to it. But active use of new communication technologies makes users easily vulnerable to undesirable interruptions.

These interruptions constitute a significant problem for mobile professionals using tools such as pagers, cell-phones and PDAs, by disrupting their time-critical activities. Improved synchronous access using these tools benefits initiators but leaves recipients with little control over the interactions. The study suggests development of improved filtering techniques that are especially light-weight, i.e. don't require more attention from the user and are less disruptive than the interruption itself. By moving interruptions to asynchronous media, messages can be stored for retrieval and delivery at more appropriate times.

NOMADIC RADIO: WEARABLE AUDIO MESSAGING

Personal messaging and communication, demonstrated in ***Nomadic Radio,*** provides a simple and constrained problem domain in which to develop and evaluate a contextual notification model. Messaging requires development of a model that dynamically selects a suitable ***notification strategy*** based on message priority, usage level, and environmental context. Such a system must infer the user's attention by monitoring her current activities such as interactions with the device and conversations in the room. The user's prior responses to notifications must also be taken into consideration to adapt the notifications over time. In this paper, we will consider techniques for ***scaleable auditory presentation*** and an appropriate parameterized approach towards ***contextual notification.***

Several recent projects utilized speech and audio I/O on wearable devices to present information. A prototype augmented audio tour guide [1] played digital audio recordings indexed by the spatial location of visitors in a museum. ***SpeechWear*** [11] enabled users to perform data entry and retrieval using speech recognition and synthesis. ***Audio Aura*** [10] explored the use of background auditory cues to provide serendipitous information coupled with people's physical location in the workplace. In ***Nomadic Radio,*** the user's inferred context rather than actual location is used to decide when and how to deliver scaleable audio notifications. In a recent paper [13], researchers suggest the use of sensors and user modeling to allow wearables to infer when users should be interrupted by incoming messages. They suggest waiting for a break in the conversation to post a message summary on the user's heads-up display. In this paper we describe a primarily non-visual approach to provide timely information to nomadic listeners, based on a variety of contextual cues.

Nomadic Radio is a wearable computing platform that provides a unified audio-only interface to remote services and messages such as email, voice mail, hourly news broadcasts, and personal calendar events. These messages are automatically downloaded to the device throughout the day and users can browse through them using voice commands and tactile input. The system consists of Java-based clients and remote servers (written in C and Perl) that communicate over wireless LAN, and utilize the telephony infrastructure in the Speech Interface group. Simultaneous spatial audio streams are rendered using a HRTF-based Java audio API. Speech I/O is provided via a networked implementation of AT&T ***Watson*** Speech API.

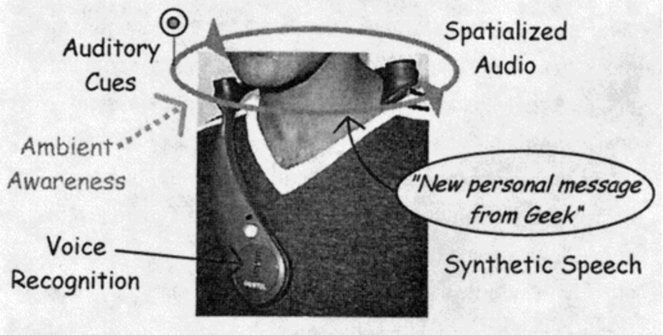


Figure 1: The primary wearable audio device, the *SoundBeam Neckset*. Messages are browsed in a unified manner via speech, auditory cues and spatial audio. The interface is controlled via voice and tactile input.

To provide a hands-free and unobtrusive interface to a nomadic user, the system primarily operates as a wearable audio-only device. The ***SoundBeam Neckset,*** a research prototype patented by Nortel for use in hands-free telephony, was adapted as the primary wearable platform in ***Nomadic Radio.*** It consists of two directional speakers mounted on the user's shoulders, and a directional microphone placed on the chest (see figure 1). Here information and feedback is provided to the user through a combination of auditory cues, spatial audio rendering, and synthetic speech. Integration of a variety of auditory techniques on a wearable device provides hands-free access and navigation as well as lightweight and expressive notification.

An audio-only interface has been incorporated in ***Nomadic Radio,*** and a networked infrastructure for unified messaging has been developed for wearable access [12]. The system currently operates on a Libretto 100 mini-portable PC worn by the user. The key issue addressed in this paper is that of handling interruptions to the listener in a manner that reduces disruption, while providing timely notifications for contextually relevant messages.

USAGE AND NOTIFICATION SCENARIO

The following scenario demonstrates the audio interface and presentation of notifications in *Nomadic Radio* (no voice commands from the user are shown here).

It's 1:15 PM and Jane is wearing *Nomadic Radio*. She has a meeting in a conference room in 15 minutes. The system gives her an early notification via an auditory cue and synthetic speech.

NR: <auditory cue for early event reminder> ***"Jane, you have a scheduled event at 1:30 PM today."*** <pause> ***"Meeting with Motorola sponsors in the conference room for 30 minutes."***

Jane scans her email messages to hear one about the meeting and check who else is coming. A new group message arrives.

NR: <ambient sound speedup and slows down> ***"New group message from Chris Schmandt about lost my glasses?"***

Jane ignores the message and heads over to the conference room. At this point, since Jane has been inactive for some time and the conversation level in the room is higher, the system scales down notifications for all incoming messages. Moments later a timely message arrives (related to an email Jane sent earlier) and the conversation level is lower. The system first plays an auditory cue and gradually speeds up the background sound of water to indicate to Jane that she will hear a summary soon.

NR: <auditory cue for timely message> + <faster ambient sound>

Jane is now engrossed in the meeting so she prevents the system from playing a summary of the message, by pressing a button on *Nomadic Radio* (she does not speak to avoid interrupting the meeting). The sound of water slows down and message playback is aborted. The system recognizes Jane is busy and turns down the notification level of all future messages in the next hour.

Its 1:55 PM and the meeting is nearly over. The system is currently in *sleep* mode. A very important voice message from Jane's daughter arrives. It recognizes the priority of the message and despite the high conversation level and low usage, it plays auditory cues to notify Jane. The ambient sound is speed-up briefly to begin playing a preview of the message in 3.5 seconds.

NR: <audio cue for voice message "telephone ringing" sound> + <VoiceCue of Jane's daughter>

Jane hears her daughter's voice and immediately presses a button to play the message. The system starts playing the full voice message in the foreground (instead of just a preview), two seconds earlier than its computed latency time.

NR: <human voice> ***"Hi mom, its Kathy. Can you pick me up early from school today?"*** <audio cue for end of message>

Jane excuses herself from the meeting and browses her email on *Nomadic Radio* while walking back to get her car keys.

Figure 2: A scenario showing Jane using *Nomadic Radio* to listen to notifications while engaging in other tasks.

SCALEABLE AUDITORY PRESENTATION

A scaleable presentation is necessary for delivering sufficient information while minimizing interruption to the listener. Messages in ***Nomadic Radio are*** scaled dynamically to unfold as seven increasing levels of notification (see figure 3): silence, ambient cues, auditory cues, message summary, preview, full body, and foreground rendering. These are described further below:

Silence for Least Interruption and Conservation

In this mode all auditory cues and speech feedback are turned-off. Messages can be scaled down to silence when the message priority is inferred to be too low for the message to be relevant for playback or awareness to a user, based on her recent usage of the device and the conversation level. This mode also serves to conserve processing, power and memory resources on a portable device or wearable computer.

Ambient Cues for Peripheral Awareness

In ***Nomadic Radio,*** ambient auditory cues are continuously played in the background to provide an awareness of the operational state of the system and ongoing status of messages being downloaded (see figure 4). The sound of flowing water provides an unobtrusive form of ambient awareness that indicates the system is active (silence indicates sleep mode). Such a sound tends to fade into the perceptual background after a short time, so it does not distract the listener. The pitch is increased during file downloads, momentarily foregrounding the ambient sound. A short e-mail message sounds like a splash while a two-minute audio news summary is heard as faster flowing water while being downloaded. This implicitly indicates message size without the need for additional audio cues and prepares the listener to hear (or deactivate) the message before it becomes available. Such ***peripheral awareness*** minimizes cognitive overhead of monitoring incoming messages relative to notifications played as distinct auditory cues, which incur a somewhat higher cost of attention on part of the listener.

Related Work in Auditory Awareness

In ***ARKola*** [5], an audio/visual simulation of a bottling factory, repetitive streams of sounds allowed people to keep track of activity, rate, and functioning of running machines. Without sounds people often overlooked problems; with auditory cues, problems were indicated by the machine's sound ceasing (often ineffective) or via distinct alert sounds. The various auditory ***cues*** (as many as 12 sounds play simultaneously) merged as an auditory texture, allowed people to hear the plant as a complex integrated process. Background sounds were also explored in ***ShareMon*** [3], a prototype application that notified users of file sharing activity. Cohen found that pink noise used to indicate %CPU time was considered "obnoxious", even though users understood the, pitch correlation. However, preliminary reactions to wave sounds were considered positive and even soothing. In ***Audio Aura [IO],*** alarm sounds were eliminated and a number of "harmonically coherent sonic ecologies" were explored, mapping events to auditory, musical or voice-based feedback. Such techniques were used to passively convey the number of email messages received, identity of senders, and abstract representations of group activity.

Auditory Cues for Notification and Identification

In *Nomadic Radio,* auditory cues are a crucial means for conveying awareness, notification and providing necessary assurances in its non-visual interface. Different types of auditory techniques provide distinct feedback, awareness and message information.

Feedback Cues

Several types of audio cues indicate feedback for a number of operational events in *Nomadic Radio:*

1. *Task completion and confirmations* - button pressed, speech understood, connected to servers, finished playing or loaded/deleted messages.
2. *Mode transitions* - switching categories, going to non-speech or ambient mode.
3. *Exceptional conditions* - message not found, lost connection with servers, and errors.

Priority Cues for Notification

In a related project, "email glances" [7] were formulated as a stream of short sounds indicating category, sender and content flags (from keywords in *the* message). In *Nomadic Radio,* message priority inferred from email content filtering provides distinct auditory cues (assigned by the user) for group, personal, timely, and important messages. In addition, auditory cues such as telephone ringing indicate voice mail, whereas an extracted sound of a station identifier indicates a news summary.

VoiceCues for Identification

VoiceCues represent a novel approach for easy identification of the sender of an email, based on a unique auditory signature of the person. *VoiceCues* are created by manually extracting a l-2 second audio sample from the voice messages of callers and associating them with their respective email login. When a new email message arrives, the system queries its database for a related *VoiceCue* for that person before playing it to the user as a notification, along with the priority cues. The authors have found *VoiceCues* to be a remarkably effective method for quickly conveying the sender of the message in a very short duration. This technique reduces the need for synthetic speech feedback, which can often be distracting.

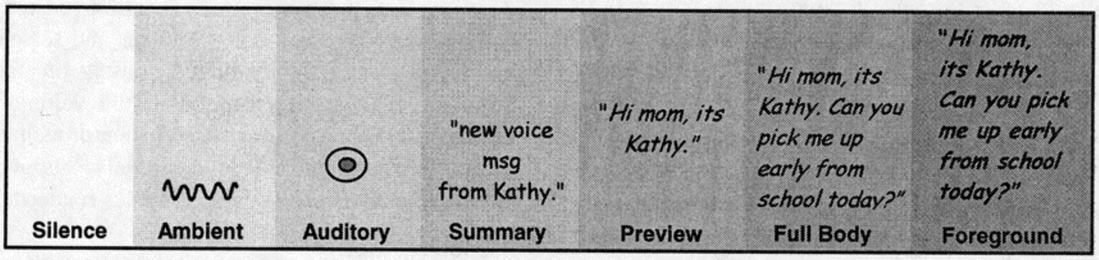

"new voice msg from Kathy."
"Hi mom, its Kathy."
"Hi mom, its Kathy. Can you pick me up early from school today?"
"Hi mom, its Kathy. Can you pick me up early from school today?"
Silence
Ambient
Auditory
Summary
Preview
Full Body
Foreground

Figure 3: Dynamic scaling of an incoming voice message during its life cycle based on the interruptability of the listener. The message is presented at varying levels: from a subtle auditory cue to foreground presentation.

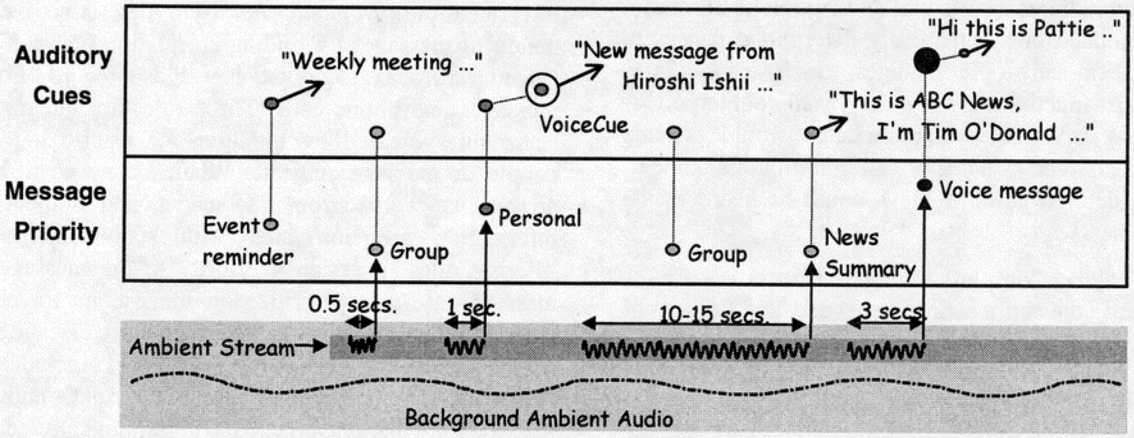


Figure 4: Ambient auditory stream speeded-up while downloading incoming messages. Audio cues indicate priority and *VoiceCues* identify the sender. A few seconds later, the message is foregrounded or spoken as synthetic speech.

Message Summary Generation

A spoken description of an incoming message can present relevant information in a concise manner. Such a description typically utilizes header information in email messages to convey the name of the sender and the subject of the message. In *Nomadic Radio,* message summaries are generated for all messages, including voice-mail, news and calendar events. The summaries are augmented by additional attributes of the message indicating category, order, priority, and duration. For audio sources, like voice messages and news broadcasts, the system plays the first 2.5 seconds of the audio. This identifies the caller and the urgency of the call, inferred from intonation in the caller's voice or provides a station identifier for news summaries.

Message Previews using Content Summarization

Messages are scaled to allow listeners to quickly preview the contents of an email or voice message. In *Nomadic Radio*, a preview for text messages extracts the first 100 characters of the message (a default size that can be user defined). This heuristic generally provides sufficient context for the listener to anticipate the overall message theme and urgency. For email messages, redundant headers and previous replies are eliminated from the preview for effective extraction. Use of text summarization techniques, based on tools such as *ProSum*[1] developed by British Telecom, would allow more flexible means of scaling message content. Natural language parsing techniques used in *ProSum* permit a scaleable summary of an arbitrarily large text document.

A preview for an audio source such as a voice message or news broadcast presents a fifth of the message at a gradually increasing playback rate of up to 1.3 times faster than normal. There are a range of techniques for time-compressing speech without modifying the pitch, however twice the playback rate usually makes the audio incomprehensible. A better representation for content summarization requires a structural description of the audio, based on annotated or automatically determined pauses in speech, speaker and topic changes. Such an *auditory thumbnail* must function similar to its visual counterpart. A preview for a structured voice message would provide pertinent aspects such as name of caller and phone number, whereas a structured news preview would be heard as the hourly headlines.

Full Body: Playing Complete Message Content

This mode plays the entire audio file or reads the full text of the message at the original playback rate. Some parsing of the text is necessary to eliminate redundant header information and format tags. The message is augmented with summary information indicating sender and subject. This message is generally spoken or played in the background of the listener's audio space.

Foreground Rendering via Spatial Proximity

An important message is played in the foreground of the listening space. The audio source of the message is rapidly moved closer to the listener, allowing it to be heard louder, and played there for 4/5th of its duration. The message gradually begins to fade away, moving back to its original position and amplitude for the remaining 1/5th of the duration. The *foregrounding* algorithm ensures that the messages are quickly brought into perceptual focus by pulling them to the listener rapidly. However the messages are pushed back slowly to provide an easy fading effect as the next one is heard. As the message moves its spatial direction is maintained so that the listener can retain a focus on the audio source even if another begins to play.

Hence a range of techniques provide scaleable forms of background awareness, auditory notification, spoken feedback and foreground rendering of incoming messages.

CONTEXTUAL NOTIFICATION

In *Nomadic Radio,* context dynamically scales the notifications for incoming messages. The primary contextual cues used include: *message priority* from email filtering, *usage level* based on time since last user action, and the *likelihood of conversation* estimated from real-time analysis of the auditory scene. In our experience these parameters provide sufficient context to scale notifications, however data from motion or location sensors can also be integrated in such a model. A linear and scaleable auditory notification model is utilized, based on the notion of estimating costs of interruption and the value of information to be delivered to the user. This approach is similar to recent work [6] on using perceptual costs and a focus of attention model for scaleable graphics rendering.

Message Priority

The priority of incoming messages is explicitly determined via content-based email filtering using *CLUES* [9], a filtering and prioritization system. *CLUES* has been integrated into *Nomadic Radio* to determine the timely nature of messages by finding correlation between a user's calendar, rolodex, to-do list, as well as a record of outgoing messages and phone calls. These rules are integrated with static rules created by the user for prioritizing specific people or message subjects. When a new email message arrives, keywords from its sender and subject header information are correlated with static and generated filtering rules to assign a priority to the message. Email messages are also prioritized if the user is traveling and meeting others in the same geographic area (via area codes in the rolodex). The current priorities include: group, personal, very important, most important, and timely. Priorities are parameterized by logarithmically scaling all priorities within a range of 0 to 1. Logarithmic scaling ensures that higher priority messages are weighted higher relative to unimportant or uncategorized messages.

$$\text{Priority}(i) = (\log(i) / \log(\text{Priority Levels}_{\text{Max}}))$$

[1] *http://transend.labs.bt.com/prosum/on_line/*

Usage Level

A user's last interaction with the device determines her usage level. If users are engaged in voice commands to the system or browsing messages on it (or have been in the last few minutes), they are probably more inclined to hear new notifications and speech feedback. Each user action is time-stamped and an active timer compares the time since the last action with default values for state transitions at every clock tick. When a new message arrives, its time of arrival is compared with the $Last\ Action_{Time}$ and scaled based on the $Sleep_{Time}$ (default at 15 minutes). High usage is indicated by values closer to 1 and any message arriving after $Sleep_{Time}$ are assigned a zero usage level. Logarithmic scaling ensures that there is less variance in usage values for recent actions relative to usage levels computed for any duration closer to the $Sleep_{Time}$ (no activity). Hence, the user not responding for 10-60 seconds has less effect on notification than a response delay of over 10-15 minutes.

$$Idle_{Time} = \log\ (\ Current_{Time} - \ Last\ Action_{Time}\ )$$

$$Usage = (\ (\log\ (Sleep_{Time}) - Idle_{Time}\ ) \ /\ \log\ (Sleep_{Time})\ )$$

One problem with using last actions for setting usage levels is that if a user deactivates an annoying message, that action is again time-stamped. Such negative reinforcements continue to increase the usage level and the related notification. Therefore negative actions such as stopping audio playback or deactivating speech are excluded from generating actions for computing the usage.

Likelihood of Conversation

Conversation in the environment can be used to gauge whether the user is in a social context where an interruption is less appropriate. If the system detects the occurrence of more than several speakers over a period of time, that is an indication of a conversational situation.

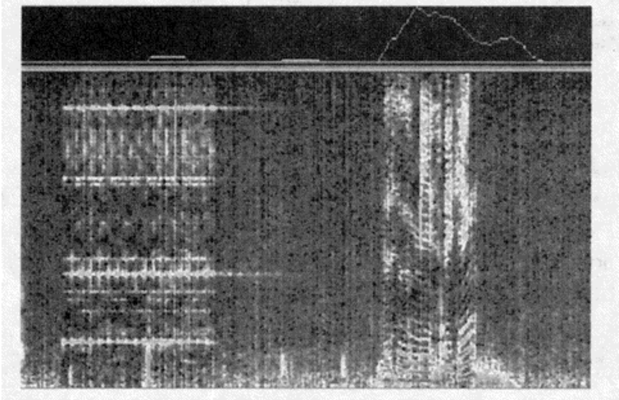


Figure 5: Bottom panel shows a spectrogram (~ 4 secs) with telephone ringing and a speech utterance. The top panel is the output probability (log likelihood) of an HMM trained on speech (which it correctly identified here).

Auditory events are first detected by adaptively thresholding total energy and incorporating constraints on event length and surrounding pauses. The system uses mel-scaled filter-bank coefficients (MFCs) and pitch estimates to discriminate, reasonably well, a variety of speech and non-speech sounds. HMMs (Hidden Markov Models) capture both the temporal characteristics and spectral content of sound events. The techniques for feature extraction and classification of the auditory scene using HMMs are described in a recent workshop paper [2]. The likelihood of speech detected in the environment is computed for each event in a short window of time. In addition, the probabilities are weighted, such that most recent time periods in the window are considered more relevant for computing the overall *Speech Level.* We are evaluating the classifier's effectiveness by training it with a variety of speakers and background sounds.

Notification Level

A weighted average for all three contextual cues provides an overall notification level ($Notify_{Level}$). The conversation level has an inversely proportional relationship with notification i.e. a lower notification must be provided during high conversation.

$$Notify_{Level} = ((Priority \times P_{wt}) + (Usage \times U_{wt}) + \ ((1 - Speech) \times S_{wt}))\ /\ 3$$

Here P_{wt}, U_{wt} and S_{wt} are weights for priority, usage and conversation levels. This notification level must be translated to a discrete scale to play the messages. There are currently 7 notification levels: *foreground, full message*, *preview*, *summary*, *audio cue*, *ambient*, and *silence*. The $Notify_{Level}$ computed must be compared to the thresholds for each of 7 scales to play the message appropriately. The $Notify\ Levels_{Max}$ are scaled by two to produce thresholds with a greater range that accommodates notification levels computed under varying interruption. This provides a reasonable $Notification_{Scale}$ for each message.

$$Threshold_{Level}\,(\ i\ ) = (\ \log\ (\ i\ )\ \ /\ \ \log\ (Notify\ Levels_{Max} \times 2)\ )$$

$$\text{If}\ \ (\ Notify_{Level}\,(\ i\ ) > Threshold_{Level}\,(\ i\ )\ )\ \text{then}$$
$$\text{assign}\ Notification_{Scale} = i,\ \text{where}\ i = \{1\ ..\ Notify\ Levels_{Max} = 7\}$$

Presentation Latency

Latency represents the period of time to wait before playing the message to the listener, after a notification cue is delivered. Latency is computed as a function of the notification level and the maximum window of time *(Latency,&* that a lowest priority message can be delayed for playback. The default maximum latency is set to 20 seconds, but can be modified by the user.

$$Latency\ (\ i\ ) = (\ 1\ -\ Notify_{Level}\ (\ i\ )\ ) \times Latency_{\ Max}$$

A higher $Notify_{Level}$ will cause a shorter latency in message playback and vice versa. An important message will play as a "preview" within 3-4 seconds of arrival, whereas a group message may play as a "summary" after 11-13 seconds of arrival (given high usage and low conversation levels). The use of latency primarily allows a user sufficient time to interrupt and deactivate an undesirable message before it is played.

Dynamic Adaptation of the Notification Model

The user can initially set the weights for the notification model to high, medium, or low (interruption). These weight settings were selected by experimenting with notifications over time using an interactive visualization of message parameters. This allowed us to observe the model, modify weights and infer the effect on notification based on different weighting strategies. Pre-defined weights provide an approximate behavior for the model and help bootstrap the system for novice users. The system also allows the user to dynamically adjust these weights (changing the interruption and notification levels) by their implicit actions while playing or ignoring messages.

The system allows *localized* positive and negative reinforcement of the weights by monitoring the actions of the user during notifications. As a message arrives, the system plays an auditory cue if its computed notification level is above the necessary threshold for auditory cues. It then uses the computed latency interval to wait before playing the appropriate summary or preview of the message. During that time, the user can request the message be played earlier or abort any further notification for the message via speech or button commands. If aborted, all weights are reduced by a fixed percentage (default is 5%), a negative reinforcement. If the user activates the message (positive reinforcement) within 60 seconds after the notification, the playback scale selected by the user is used to increase all weights. If the message is ignored, no change is made to the weights, but the message remains active for 60 seconds during which the user's actions can continue to influence the weights.

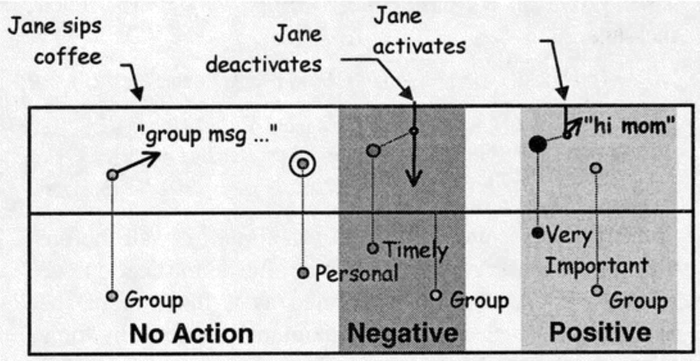
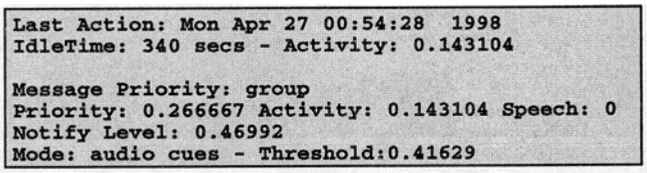


Figure 6: Adaptation of notification weights based on Jane's actions while listening to messages.

Figure 6 shows a zoomed view of the extended scenario introduced earlier, focusing on Jane's actions that reinforce the model. Jane received several messages and ignored most of the group messages and a recent personal message (the weights remain unchanged). While in the meeting, Jane interrupted a timely message to abort its playback. This reduced the weights for future messages, and the ones with low priority (group message) were not notified to Jane. The voice message from Kathy, her daughter, prompted Jane to reinforce the message by playing it. In this case, the weights were increased. Jane was notified of a group message shortly after the voice message, since the system detected higher usage activity. Hence, the system correctly scaled down notifications when Jane did not want to be bothered whereas notifications were scaled up when Jane started to use the system to browse her messages.

EFFECTIVENESS OF THE NOTIFICATION MODEL

The nature of peripheral awareness and unobtrusive notification on a wearable device requires a usage evaluation that must be conducted on an ongoing and long-term basis. However, the predictive effectiveness of the notification model must first be evaluated on a quantitative basis. Hence, all message and notification parameters are captured for such analysis. Lets consider two actual examples of notification computed for email messages with different priorities. Figure 7 shows an auditory cue generated for a group message (low priority).

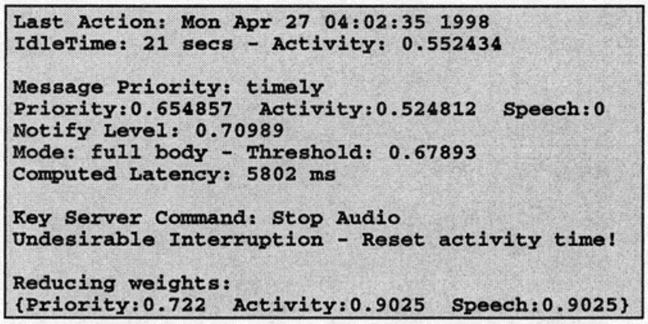


```
Last Action: Mon Apr 27 00:54:28  1998
IdleTime: 340 secs - Activity: 0.143104

Message Priority: group
Priority: 0.266667 Activity: 0.143104 Speech: 0
Notify Level: 0.46992
Mode: audio cues - Threshold:0.41629
```

Figure 7: Notification computed for a group email. The user has been idle; hence it is heard as an auditory cue.

The timely message (in figure 8) received greater priority and consequently a higher notification level for summary playback. A moderate latency time (approx. 6 secs.) was chosen. However when the user interrupted the notification by a button press, the summary playback was aborted. The user's action reduced overall weights by 5%.



```
Last Action: Mon Apr 27 04:02:35 1998
IdleTime: 21 secs - Activity: 0.552434

Message Priority: timely
Priority:0.654857   Activity:0.524812   Speech:0
Notify Level: 0.70989
Mode: full body - Threshold: 0.67893
Computed Latency: 5802 ms

Key Server Command: Stop Audio
Undesirable Interruption - Reset activity time!

Reducing weights:
{Priority:0.722  Activity:0.9025  Speech:0.9025}
```

Figure 8: Notification level and latency computed for a timely email message. The user's action of stopping audio before it plays reduces all the current weights.

Continuous local reinforcement over time should allow the system to reach a state where it is somewhat stable and robust in converging to the user's preferred notification. Currently the user's actions primarily adjust weights for subsequent messages, however effective reinforcement learning requires a model that generalizes a notification policy that maximizes some long-term measure of reinforcement [8]; this will be the focus of our future work.

PRELIMINARY EVALUATION

Although the authors have been using and refining these techniques during system development, a preliminary 2-day evaluation was conducted with a novice user, who had prior experience with mobile phones and 2-way pagers. The user was able to listen to notifications while attending to tasks in parallel such as reading or typing. He managed to have casual discussions with others while hearing notifications; however he preferred turning off all audio during an important meeting with his advisor. People nearby sometimes found the spoken feedback distracting if heard louder, however that also cued them to wait before interrupting the user. The volume on the device was lowered to minimize any disruption to others and maintain the privacy of messages. The user requested an automatic volume gain that adapted to the environmental noise level.

In contrast to speech-only feedback, the user found the unfolding presentation of ambient and auditory cues allowed sufficient time to switch attention to the incoming message. Familiarization with the auditory cues was necessary. He preferred longer and gradual notifications rather than distinct auditory tones. The priority cues were the least useful indicator whereas *VoiceCues* provided obvious benefit. Knowing the actual priority of a message was less important than simply having it presented in the right manner. The user suggested weaving message priority into the ambient audio (as increased pitch). He found the overall auditory scheme somewhat complex, preferring instead a simple notification consisting of ambient awareness, *VoiceCues* and spoken text.

The user stressed that the ambient audio provided the most benefit while requiring least cognitive effort. He wished to hear ambient audio at all times to remain reassured that the system was still operational. An unintended effect discovered was that a "pulsating" audio stream indicated low battery power on the wearable device. A "pause" button was requested, to hold all messages while participating in a conversation, along with subtle but periodic auditory alerts for unread messages waiting in queue. The user felt that *Nomadic Radio* provided appropriate awareness and its expressive qualities justified its use over a pager. A long-term trial with several nomadic users is necessary to further validate these notification techniques.

CONCLUSIONS

We have demonstrated techniques for scaleable auditory presentation and message notification using a variety of contextual cues. The auditory techniques and notification model have been refined based on continuous usage by the authors, however we are currently conducting additional evaluations with several users. Ongoing work explores adaptation of the notification model based on reinforcement from user behavior over time. Our efforts have focused on wearable audio platforms, however these ideas can be readily utilized in consumer devices such as pagers, PDAs and mobile phones to minimize disruptions while providing timely information to users on the move.

ACKNOWLEDGMENTS

Thanks to Brian Clarkson for ongoing work on the audio classifier and Stefan Marti for help with user evaluations. We also thank Lisa Fast and Andre Van Schyndel at Nortel for their support of the project.

REFERENCES

1. Bederson, Benjamin B. Audio Augmented Reality: A Prototype Automated Tour Guide. *Proceedings of CHI '95*, May 1995, pp. 210-211.
2. Clarkson, Brian, Nitin Sawhney and Alex Pentland. Auditory Context Awareness via Wearable Computing, *Workshop on Perceptual User Interfaces*, Nov. 1998.
3. Cohen, J. Monitoring Background Activities. Auditory Display: Sonification, Audification, and Auditory Interfaces. Reading MA: Addison-Wesley, 1994.
4. Conaill, O' Brid and David Frohlich. Timespace in the Workplace: Dealing with Interruptions. *Proceedings of CHI `95*, 1995.
5. Gaver, W.W., R. B. Smith, T. O'Shea. Effective Sounds in Complex Systems: The ARKola Simulation. *Proceedings of CHI '91*, April 28-May 2, 1991.
6. Horvitz, Eric and Jed Lengyel. Perception, Attention, and Resources: A Decision-Theoretic Approach to Graphics Rendering. *Proceedings of Uncertainty in Artificial Intelligence,* Aug. 1-3, 1997, pp. 238-249.
7. Hudson, Scott E. and Ian Smith. Electronic Mail Previews Using Non-Speech Audio. *Proceedings of CHI '96*, April 1996, pp. 237-238.
8. Kaelbling, L.P. and Littman, M.L. Reinforcement Learning: A Survey. *Journal of Artificial Intelligence Research*, vol. 4, 1996, pp. 237-285.
9. Marx, Matthew and Chris Schmandt. CLUES: Dynamic Personalized Message Filtering. *Proceedings of CSCW '96*, pp. 113-121, November 1996.
10. Mynatt, E.D., Back, M., Want, R. Baer, M., and Ellis J.B. Designing Audio Aura. *Proceedings of CHI '98*, April 1998.
11. Rudnicky, Alexander, Reed, S. and Thayer, E. SpeechWear: A mobile speech system. *Proceedings of ICSLP '96*, 1996.
12. Sawhney, Nitin and Chris Schmandt. Speaking and Listening on the Run: Design for Wearable Audio Computing. *Proceedings of the International Symposium on Wearable Computing*, October 1998.
13. Starner, Thad, Mann, S., Rhodes, B., Levine, J., Healey, J., Kirsch, D., Picard, R., and Pentland, A. Augmented Reality through Wearable Computing. Presence, Vol. 6, No. 4, August 1997, pp. 386-398.