Ambient Rhythm – Melodic Sonification of Status Information for IoT-enabled Devices

George Chernyshov
Keio Media Design, Japan
chernyshov@kmd.keio.ac.jp

Jiajun Chen
Keio Media Design, Japan
cyrilchan@kmd.keio.ac.jp

Yenchin Lai
Keio Media Design, Japan
laiyen@kmd.keio.ac.jp

Vontin Noriyasu
Fujitsu Design, Japan
vontin.noriyasu@jp.fujitsu.com

Kai Kunze
Keio Media Design, Japan
kai.kunze@gmail.com

ABSTRACT
In this paper we explore how to embed status information of IoT-enabled devices in the acoustic atmosphere using melodic ambient sounds while limiting obtrusiveness for the user. The user can use arbitrary sound samples to represent the devices he wants to monitor. Our system combines these sound samples into a melodic ambient rhythm that contains information on all the processes or variables that user is monitoring. We focus on continuous rather than binary information (e.g. "monitoring progress status" rather then "new message received"). We evaluate our system in a machine monitoring scenario focusing on 5 distinct machines/processes to monitor with 6 priority levels for each. 9 participants use our system to monitor these processes with an up to 92.44% detection rate, if several levels are combined. Participants had no previous experience with this or similar systems and had only 5-10 minute training session before the tests.

ACM Classification Keywords
H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

Author Keywords
Sonification; Acoustic Atmosphere; Interfaces; Process monitoring; IoT;

INTRODUCTION
With more and more network-enabled computing devices surrounding us, we can easily overcome limitations related to place and interact or monitor distance services. However, the mere amount of services can easily become overwhelming. To deal with this problem is one of the core issues in IoT. The limiting factor is no longer CPU speed or battery power but human attention.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

IoT'16, November 07 - 09, 2016, Stuttgart, Germany

© 2016 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ISBN 978-1-4503-4814-0/16/11…$15.00

DOI: http://dx.doi.org/10.1145/2991561.2991564

How to best deliver information about a multitude of services/devices and how to achieve this in a non-disruptive way depends on the information type to be delivered. For simple information such as individual notifications (e.g. the typical "its time to stand up", "new email" or "download completed") many possibilities exist and have been realized for example in smart watches. A much more difficult question is how to allow the user to `continuously` monitor a process. This amounts to having a progress bar or a couple of discretized steps throughout the entire download process rather then a discrete notification when it has been completed (or failed). There are different reasons why people may opt for continuous monitoring rather then for discrete notifications. First is a fundamental human desire to be able to predict and anticipate rather then wait for things to happen "out of the blue". Thus, a progress bar gives the user a feeling for how much time remains (even if this is not always an accurate predictor). Second, in many situations a temporal pattern in itself may be of interest, not the arrival of a certain discrete event.

In this paper we research how to combine ambient sounds from devices into a melodic audio tune that facilitates monitoring of temporal patterns based on frequency and rhythm changes. This work is motivated by two observations. First, it is well known that people are capable of perceiving subtle, complex information on the audio channel while being mentally and cognitively engaged in complex tasks. [6] Thus, for example, a race car driver can pick up subtle signs of engine trouble in the motor sound while being fully focused on the race. Second, people like to listen to music on their personal devices during many activities: from sports, through work to e.g. reading or learning. Thus, encoding information in melodic rhythm enables users to monitor status updated without "paying attention". In particular, it is much less obtrusive than visually showing an appropriate representation, as the users would need to pay constant attention to the visualization updates (although this alternative may have a good justification in certain situations).

The 3 basic research questions we are tackling in this paper are as follows: 1. can we encode several progress levels of several processes using arbitrary audio samples? 2. how many of these levels and processes are possible to be used using this

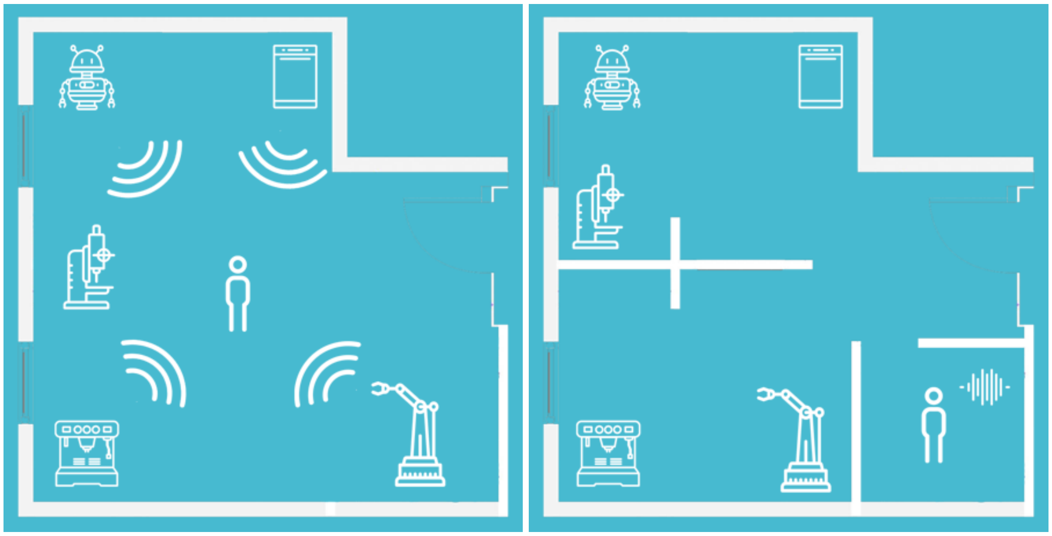


Figure 1. Ambient Rhythm Concept: Melodic Sonification of Status Updates from Several devices to accompany the user.

rhythmic encoding? 3. are people still able to perceive the encoded information?

We follow the basic idea to use arbitrary recorded sound samples for the processes, as the user can pick a sample that best represents the process for him, and then using rhythm changes to represent the progress (or in general temporal change) for the process.

Given this basic idea, we contribute the following in this paper:

1. We evaluate the level of rhythmic changes necessary to recognize process updates.

2. We show to what degree people can perceive temporal patterns through rhythmic changes in the sound sphere.

3. We evaluate to what degree people can recognize the correct process using pre-recorded assigned sound samples in a machine monitoring scenario.

4. We show that users can detect the individual process and its temporal status over 5 processes with 6 levels to up to 92 % in the range of plus minus 2 levels.

RELATED WORK

There's a lot of work in notifications and binary information transfer using audio. Yet so far to our knowledge we are the only researchers applying a rhythm encoding to self-recorded audio samples for continuous information monitoring. In the following, we will look into related work focusing on notifications, sonification and contextual information displays.

Sonic Interface Design and Audio Notifications

Most research in the ubiquitous computing space focuses on conveying notifications to the user [3, 4]. These works mostly focus on discrete events (e.g. a user getting an email, receiving a phone call), they convey status but in a discrete manner. This related work is often not for monitoring a continuous "value" (e.g. the traffic situation on the road, the queue in a super market). Very early work from Gaver et al. introduced sound icons [7] associating meaning to particular actions.

Mynatt et al. used voice, music, sound effects or a combination of these as notification sounds in their audio aura system [11]. In terms of music they used different short melodies carrying different meanings. Sawhney et al. presented Nomadic radio, a wearable computing platform for managing voice and text-based messages using voice and sound [14]. Most of their work focuses on explicit interactions and speech feedback to the user.

Context and Sound

There are many works using sound and audio for localization and communication between devices [12, 1, 15], of which most interesting is research by Madhavapeddy et al. [10]. They also use sound for device to device communication, yet they use a musical composition for transferring data. The users can hear if and potentially what devices communicate with each other. However the focus of the work is still device to device communication.

Sound-based Augmented Reality

There is a lot of work utilizing music and sounds for alternate-reality games and experience sharing [9, 13]. Some of them alter sounds in the environment to make the user appear in a particular place or alter music/sounds based on the user's activities (e.g. changing music speed based on pace [5]). Russel et al. show how to implement a location based augmented reality soundsphere [13]. Their mobile setup using a bone conducting headset is especially interesting and we see this as an attractive extension to the work we present here. For the next setups we will use a similar headset. We see this as complementary to our approach since the sounds and music adoption also convey information, yet the main goals are entertainment.

Information Embedding in Ambient Soundscapes

Butz et al. present work that is close to ours, encoding information in music [2]. However, the music piece in itself is completely generative. Therefore, listening comfort highly depends on the information content. Additionally, Butz et al. just present the concept and do not provide a detailed experimental evaluation. A good overview about information embedding and group awareness is given by Kainulainen et al. [8].

APPROACH

Our main focus is to create an easy to use system that could present the user with more or less continuous information about multiple ongoing processes rather than just discrete notifications. We also want to enable the user to monitor these ongoing processes over a distance (see Figure 1). It is well known that people are capable of perceiving slightest changes in their acoustic atmosphere without diverting too much attention. [6] This skill appeared on the dawn of mankind and helped spotting for example and approaching predator. Good example from modern life would be the ability of a maintenance worker to spot a malfunctioning machine before it breaks down completely. Unfortunately, this skill requires some level of experience with the machine. Otherwise it is impossible to distinguish normal working sounds of the machine from malfunctioning. Other problem is that most of people are not used to this way of monitoring, and constant noise might be too annoying and disruptive for the vast majority of potential users. Also it would provide information only on the machine condition, and not progress or importance of the process. Moreover, some processes that users need to monitor do not produce any noises, e.g. server load, size of a queue in a supermarket, software installation progress, etc.

To solve the issue of silent processes we decided to use the concept of sound icons, by mapping the process to a respective sound sample somehow associated with the process [7]. Since as we mentioned, reading information from the sound of a working machine requires some experience and we wanted to make the system easy to use, we decided to arrange the output in the form of a melodic rhythm. We find it reasonable to assume that only a few to none people had no experience with music, and a vast majority of people can easily spot changes in musical compostions. In order to apply the concept of sound icons to music we decided to arrange the samples representing different machines or processes using rhythm rather than pitch. Even a slight pitch change of a sample may change it to a degree where users can not recognize it or associate it with the process it represents. Considering the above, we used only length of the sample and rhythm to arrange them in a melody. Drum pieces are composed in a very similar fashion, since the drum player has only a limited set of drums and cannot adjust the pitch of each drum dynamically, unlike a guitar player, who can play different pitched notes with different lengths and intervals on each string.

In terms of composing of the rhythm, we set the speed at 120 beats per minute and make 30 loop tracks, 5 tracks representing different emergency levels of bass line and 5 items. In order to contain all the expected cases and prevent notes of different items from overlapping too much, we compose with

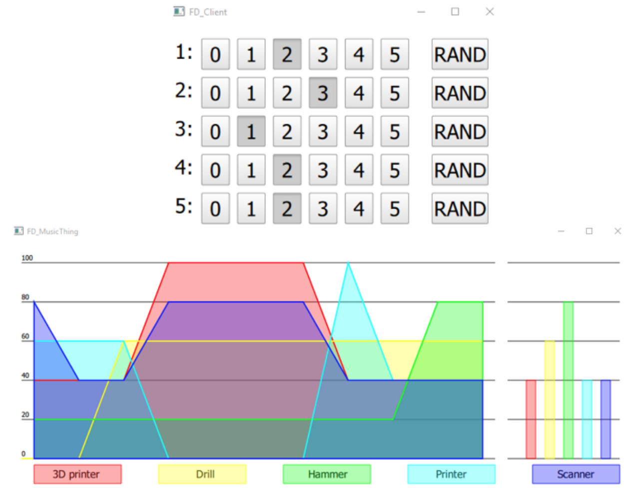


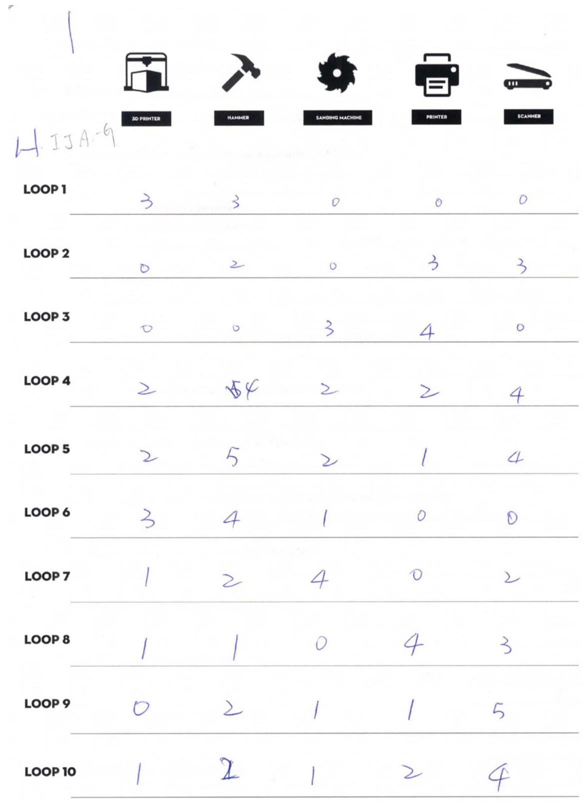
Figure 2. Software used for creating the Ambient Rhythm Sonification. On top there is control panel application with toggles for different levels of process importance. It is used to simulate input from different machines. Below is application that generates the sound and shows graphs representing different processes' importance change over time. The second application receives data regarding the process state or importance over TCP/IP, so it can work with any TCP/IP-enabled IoT device or sensor hub.

only whole notes and sixteenth notes, for that sixteenth notes are short enough to be filled in the loop without overlapping and long enough to be noticed, and whole notes and sixteenth notes of different sound are easy to distinguish because of the notable disparity of length.

Total number of the processes that can be monitored using such approach is highly dependent on samples, and loop composition. The loops we composed for this system are 8 beats long, every beat has 4 notes, which results in up to 32 distinctive notes per loop without overlapping. In case if samples have significantly different pitch, 2 or more samples can start simultaneously and still be easily perceived, which significantly increases the number of possible processes. Changing the tempo would obviously increase or decrease the number of notes in one loop. Thus theoretically it is possible to monitor tens or hundreds of processes. The number of processes highly depends on musical arrangement of the notes, tempo, length of one loop, pitch of samples and user's musical preferences.

The system we developed as the proof of concept can present the user with the information about 5 different processes, with 6 priority levels for each of the processes, from level 0 to 5. Priority levels does not necessarily represent priority of the process, they can represent any time-dependent variable, for example progress of a process, condition of the machine, workload, temperature, queue size, amount of attention required, etc. We also added a bass rhythm to the output that matches the process with the highest level which presumably requires the most attention and is the most important for the user at the given moment. Bass line also has 5 levels to match with the monitored processes. If all the processes are being at level 0, level 1 bass line is played, to indicate that the system is functioning.

Figure 3. Experiments



H. IJA-9

	3D PRINTER	HAMMER	SANDING MACHINE	PRINTER	SCANNER
LOOP 1	3	3	0	0	0
LOOP 2	0	2	0	3	3
LOOP 3	0	0	3	4	0
LOOP 4	2	~~5~~4	2	2	4
LOOP 5	2	5	2	1	4
LOOP 6	3	4	1	0	0
LOOP 7	1	2	4	0	2
LOOP 8	1	1	0	4	3
LOOP 9	0	2	1	1	5
LOOP 10	1	2	1	2	4

Figure 4. Example of a test sheet filled by a participant

For each priority level of each process we played the corresponding sample once or multiple times in a different pattern. In total we used 25 different patterns, five patterns for each of the five processes. For priority level 0, no sample was played at all. Each pattern and bass line is 8 second long and can be played in seamless loops, so the output can be updated every 8 seconds. Each pattern is a MIDI file, so all the samples are easily replaceable. Using MIDI files allows to readjust the system to monitor different processes by simply replacing the sound samples with samples that correspond to a different process.

EXPERIMENTAL EVALUATION

To verify the feasibility of the concept we have conducted the following experiment. Since one of the usage scenarios for the system is remote machinery monitoring, we decided to run the user tests using samples we recorded from various machines we had access to on our campus. Recording actual sounds would simplify mapping the samples to different processes related to those machines. It would also make it much easier for the user to understand which sample represents which machine. For the experiment we used the samples recorded from a 3D printer, sanding machine, laser printer, hammering and a scanner and use our own software to play/adjust the Ambient Rhythm on the fly (see Figure 2).

First participants were introduced to the concept and had a short 5-10 minute training session to get accustomed and comfortable with the setup. During the training session, participants were shown how each level of priority for each of the monitored processes sounds like. During this session only the bass line and samples for only one process were played, all other processes were set to level 0 and did not influence the output. After this step participants were shown how all processes at various levels in the same time sound like. During this part participants were shown what is the priority level for each of the processes.

For the testing part we used 10 predefined situations were all the processes had priority levels varying from 0 to 5. We tried to cover all possible cases, where many processes have high priority at the same time as well as the situations where only a few processes have relatively low priority and the rest are at level 0. Music piece corresponding to each of the situations was played in a loop and participants were asked to write down the priority level for every process. Situations were introduced in a different order for every participant. After finishing the tests participants were asked a set of questions to evaluate their satisfaction and experience with the system, and spot the weak and strong points of the whole concept. The experiments were performed on a MacBook PRO 2014 with Intel HD Audio chip. Participants used the Behringer HPM 1000 headphones.

RESULTS AND DISCUSSION

In total 9 subjects have participated in the experiments. Average age of the participants was 24.44 years, age range was 22-28, median 24. 4 male, 5 female. Every participant was asked to evaluate the priority level of each process out of 5 processes. We presented the subjects with 10 different situations, so every participant had to make 50 evaluations in total, which results in 450 evaluations for all the participants. During the experiment many participants stated that they cannot hear the sample of scanner behind louder samples of other machines, so we have decided to exclude the data regarding the scanner process from the final results. Which lowers the total number of evaluations given by subjects to 400.

Perceived Level

Actual level	0	1	2	3	4	5
0	90.12%	1.23%	2.47%	3.70%	0%	2.47%
1	11.11%	62.50%	18.06%	6.94%	1.39%	0%
2	12.96%	15.74%	58.33%	10.19%	2.78%	0%
3	2.78%	11.11%	30.56%	36.11%	16.67%	2.78%
4	8.33%	0%	2.78%	33.33%	47.22%	8.33%
5	3.70%	0%	7.41%	25.93%	14.81%	48.15%

Figure 5. Confusion matrix for 400 evaluations done during the experiments.

In total 58.22% of evaluations were correct. 81.33% of evaluations fall in ±1 level range, 92.44% are in ±1 level range. It is worth mentioning again that we used 6 levels in total: from 0 to 5. Also users had only 5-10 minutes of training before they started the experimental session.

There is however some indication, that the monitoring is trainable, as we conducted the same test with two additional users who where involved in conducting the experiments and the development of the system and thus had many hours of experience with it. They had close to 100% accurate results. Of course this is not representative.

The survey we conducted after the experiment with every participant has shown that participants are generally satisfied with our system. They noted that using the concept of sound icons the experience is more realistic than with discrete sound notifications. When participants were asked how would they compare the system with traditional visual interface, the dominating opinion was that visual interface is easier to understand, but more distracting then sound. Many subjects noted that during the experiment they were forgetting how different levels of different processes sound like, and some samples were too similar, e.g. sound of printer and 3d printer. Which should have been expected, given that subjects had only a short training session before the experiment. Also participants claimed that it is hard to determine the levels if too many processes are present, which does not correlate with the results of the experiments. Percentage of correct answers for simple situations is nearly the same as the average. Participants asked to make the difference between levels more distinct, and proposed to use tone and volume too. Generally, users were interested with the concept and agreed on its usefulness. They used such words as "interesting" "entertaining" and "funny" to describe the experience. However some subjects defined the sound as not melodic enough or noisy. Which is understandable, since the samples we used were industrial and were recorded from drilling machines, 3D printers, etc.

CONCLUSION AND FUTURE WORK

Overall, we presented a novel way to encode temporal information of several processes and showed in an initial experiment that users can determine process and level from the "Ambient Rhythm". As the sound samples are exchangeable, it is easy to adjust the system to any type of IoT scenario, from monitoring maintenance machines over checking the watering level of plants to getting updates on traffic patterns.

The recognition rates for the whole 6 levels still need to be improved for some participants. Some users found it hard to distinguish especially the high levels. Yet, overall a 3-4 level monitoring leads to respectable results between 80 - 92 %. Also seeing the feedback from our participants and the initial use case, it seems the results can be enormously improved with training.

In the future, we want to look into the training issue that came up during our experiment as well as focusing on making the resulting sound more melodic so it gets closer to music.

This paper presented a simple way to sonify arbitrary devices and processes enableing the user to "take" their sounds with him to monitor their progress. We presented our design rational and lessons learned from the system and we gave first indications on how many processes and levels can be monitored by untrained participants in a maintenance scenario.

REFERENCES

1. Gaetano Borriello, Alan Liu, Tony Offer, Christopher Palistrant, and Richard Sharp. 2005. WALRUS: wireless acoustic location with room-level resolution using ultrasound. In *Proceedings of the 3rd international conference on Mobile systems, applications, and services*. ACM, 191–203.

2. Andreas Butz and Ralf Jung. 2005. Seamless user notification in ambient soundscapes. In *Proceedings of the 10th international conference on Intelligent user interfaces*. ACM, 320–322.

3. Ádám Csapó and György Wersényi. 2013. Overview of auditory representations in human-machine interfaces. *ACM Computing Surveys (CSUR)* 46, 2 (2013), 19.

4. Anind K Dey, Gregory D Abowd, and Daniel Salber. 2001. A conceptual framework and a toolkit for supporting the rapid prototyping of context-aware applications. *Human-computer interaction* 16, 2 (2001), 97–166.

5. Greg T Elliott and Bill Tomlinson. 2006. PersonalSoundtrack: context-aware playlists that adapt to user pace. In *CHI'06 extended abstracts on Human factors in computing systems*. ACM, 736–741.

6. Zwicker E Fastl H. 2006. *Psychoacoustics*. Springer.

7. William W Gaver. 1986. Auditory icons: Using sound in computer interfaces. *Human-computer interaction* 2, 2 (1986), 167–177.

8. Anssi Kainulainen, Markku Turunen, and Jaakko Hakulinen. 2009. Awareness information with speech and sound. In *Awareness Systems*. Springer, 231–256.

9. Kent Lyons, Maribeth Gandy, and Thad Starner. 2000. Guided by voices: An audio augmented reality system. (2000).

10. Anil Madhavapeddy, David Scott, and Richard Sharp. 2003. Context-aware computing with sound. In *UbiComp 2003: Ubiquitous Computing*. Springer, 315–332.
11. Elizabeth D Mynatt, Maribeth Back, Roy Want, Michael Baer, and Jason B Ellis. 1998. Designing audio aura. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM Press/Addison-Wesley Publishing Co., 566–573.
12. Aditya Shekhar Nittala, Xing-Dong Yang, Scott Bateman, Ehud Sharlin, and Saul Greenberg. 2015. PhoneEar: interactions for mobile devices that hear high-frequency sound-encoded data. In *Proceedings of the 7th ACM SIGCHI Symposium on Engineering Interactive Computing Systems*. ACM, 174–179.
13. Spencer Russell, Gershon Dublon, and Joseph A Paradiso. 2016. HearThere: Networked Sensory Prosthetics Through Auditory Augmented Reality. In *Proceedings of the 7th Augmented Human International Conference 2016*. ACM, 20.
14. Nitin Sawhney and Chris Schmandt. 2000. Nomadic radio: speech and audio interaction for contextual messaging in nomadic environments. *ACM transactions on Computer-Human interaction (TOCHI)* 7, 3 (2000), 353–383.
15. Timothy Sohn, William G Griswold, James Scott, Anthony LaMarca, Yatin Chawathe, Ian Smith, and Mike Chen. 2006. Experiences with place lab: an open source toolkit for location-aware computing. In *Proceedings of the 28th international conference on Software engineering*. ACM, 462–471.