

PROCEEDINGS 2008

May 13 – 15, 2008

University of applied arts, Vienna

<http://MMW2008.dieangewandte.at/>

5<sup>th</sup> International Mobile Music Workshop



MobileMusicWorkshop

di:Angewandte  
Universität für angewandte Kunst Wien  
University of Applied Arts Vienna

MobileMusicWorkshop

## INTRO

### 5TH INTERNATIONAL MOBILE MUSIC WORKSHOP 2008 13-15 MAY 2008, VIENNA, AUSTRIA

The Mobile Music Workshop 2008 is the 5th in a series of annual international gatherings that explore the creative, critical and commercial potential of mobile music. They are inspired by the ever-changing social, geographic, ecological, emotional context of using mobile technology for creative ends. We are looking for new ideas and groundbreaking projects on sound in mobile contexts. What new forms of interaction with music and audio lie ahead as locative media, ubiquitous networks, and music access merge into new forms of experiences that shape the everyday? Can they change the way we think about our mobile devices and about walking through the city?

The emerging field of Mobile Music sits at the intersection of ubiquitous computing, portable audio technology and New Interfaces for Musical Expression (NIME). It goes beyond today's personal music players to include creative practices of mobile music making, sharing and mixing. The mobile setting challenges existing notions of interfaces and interaction, stretching music to new creative limits. The workshop has been at the forefront of this innovative area since 2004. Past editions of the event have taken place in Amsterdam, Brighton, Vancouver and Göteborg in collaboration with the Viktoria Institute, STEIM, Waag Society, Futuresonic, NIME and others.

The 2008 edition of the workshop is taking place in Vienna, one of the hotspots in the European for laptop, glitch, and electronic music. Hosted by the University of Applied Arts, it features three evenings of performances and installations, an exhibition in the heart of the city, invited speakers, paper presentations, posters and demo sessions as well as hands-on tutorials.

We invited artists, designers, academic researchers, hackers, industry professionals and practitioners from all areas, including music, technology development, new media, sound-art, music distribution, cultural/media studies and locative media and more to present and discuss projects, prototypes, applications, devices, performances, installations, theoretical and historical considerations.

## PAPERS

The accepted papers present new projects, approaches or reflections exploring the topic of mobile music. They include but are not limited to mobile music systems or enabling technologies, interface design, legal issues, user studies, ethnographic fieldwork, social implications, art pieces and other areas relevant to mobile music.

## POSTERS AND DEMOS

The accepted posters and demos document work-in-progress projects or ideas in similar areas of mobile music technology as the papers.

## ORGANISERS

The 2008 edition is hosted and co-organised by the University of Applied Arts, Vienna, Austria, Nicolaj Kirisits.

The Steering Committee is formed by  
Lalya Gaye (Dånk! Collective and IT-University of Göteborg, Sweden),  
Atau Tanaka (Culture Lab Newcastle, UK),  
Frauke Behrendt (University of Sussex, UK),  
Kristina Andersen (STEIM, The Netherlands).

## CONTACT

info@mobilemusicworkshop.org  
<http://www.mobilemusicworkshop.org>

## TIME SCHEDULE

### DAY 1 – MAY 13

10:00 • Welcome  
10:30 • Keynote  
12:00 – 13:30 • Lunch

#### PAPER1

13:30 • Caressing the skin: Mobile devices and bodily engagement • F. Schröder  
14:00 • MoGMI: Mobile Gesture Music Instrument • A. Dekel / G. Dekel

14:30 – 15:00 • Coffee break

#### PERFORMANCE1

15:00 • Transit • Spat\_Lab  
15:30 • Craving • B. Garnicnig / G. Haider

#### INSTALLATIONS

16:00 – 17:00 • Digital Claiming • Spat\_Lab

### DAY 2 – MAY 14

#### POSTER PRESENTATION

10:00 – 12:00

Mobile Tangible Interfaces as Gestural Instruments • F. Kayali / M. Pichlmair / P. Kotik  
An Augmented Reality Framework for Wireless Mobile Performance • M. Wozniowski / N. Bouillot / J.R. Cooperstock / Z. Settel  
*undersound* and the Above Ground • A. Bassoli / J. Brewer / K. Martin / I. Carreras / D. Tacconi  
Soundfishing • C. Midolo

12:00-13:30 • Lunch

#### PAPER2

13:30 • Some Challenges Related to Music and Movement in Mobile Music Technology • A. R. Jensenius  
14:00 • Real time synaesthetic Sonification of Travelling Landscape • T. Pohle / P. Knees

14:30-15:00 • Coffee break

#### PERFORMANCES2

15:00 • Framework • A. Haberl / K.Filip / A. Faessler / N. Kirisits  
15:30 • Tango Intervention Vienna • L. Robert  
16:00 • IMPROVe - mobile Phone sound improvisation • W. Richard  
16:30 • Collaborative Musical Games with PhonePlay • J. Knowles

19:00 • Community - dinner at Xpedit

### DAY 3 – MAY 15

#### HANDS-ON SESSIONS

10:00 – 12:00

R. Widerberg / Y. Harris / S. Symons

12:00 – 13:30 Lunch

#### PAPER3

13:30 • Developments and Challenges turning Mobile Phones into Generic music Performance Platforms • G. Essl / M. Rohs  
14:00 • A typology for Listening in Place • P. Rebelo / M. Green / F. Hollerweger

14:30-15:00 • Coffee break

#### CLOSING SESSION - PANEL DEBATE

15:00 – 17:00

#### CLOSING PARTY

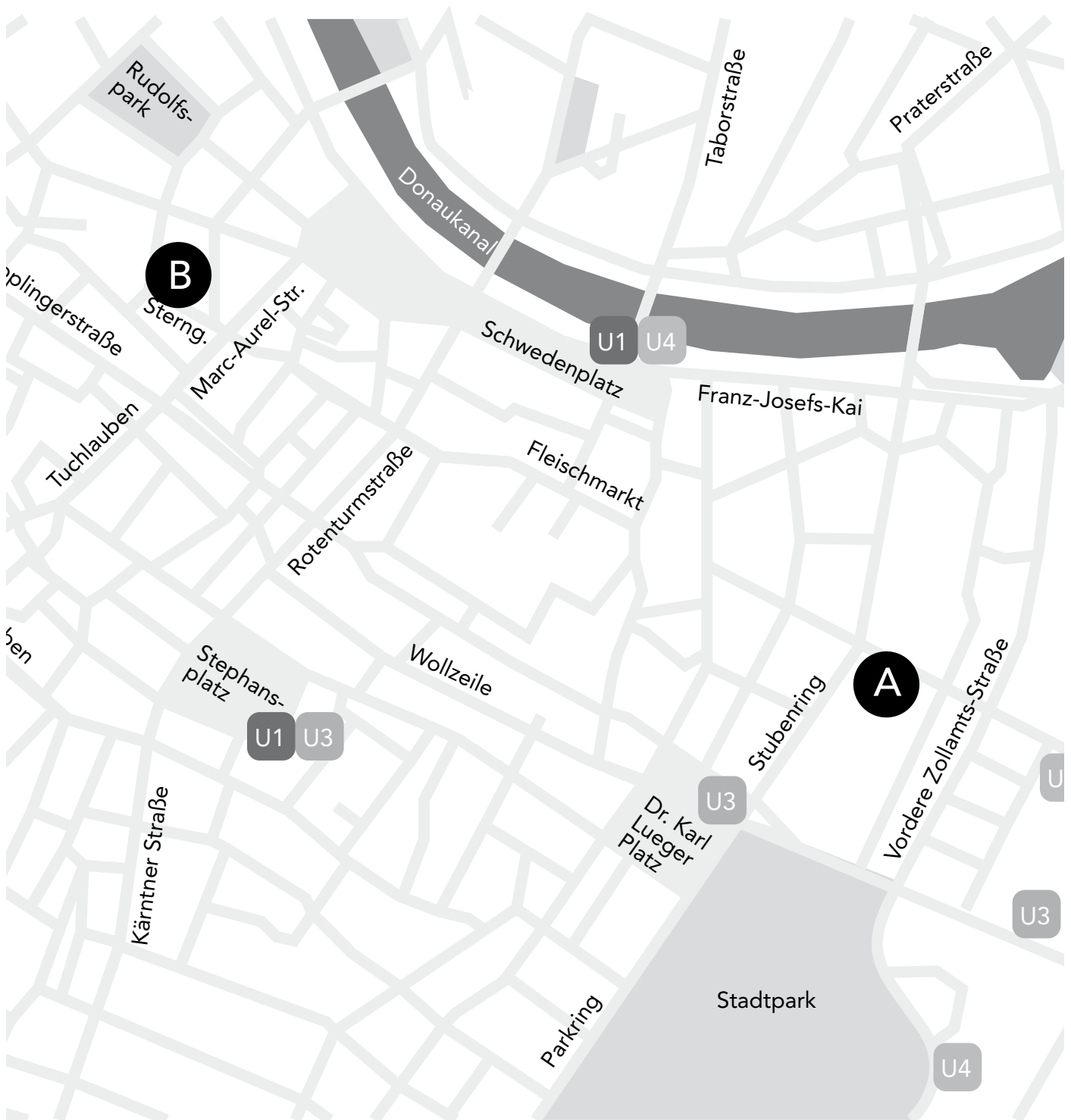
19:00 – ...

#### CONCERTS

20:00 • springfield RVL-003 • J. Perschy / R. Mathy / M. Wyszka  
20:30 • taus • T. Blechman / K. Filip  
21:00 • Institute for transacoustic research • N. Gansterer / M. Meinharter / J. Piringer / E. Reitermayer

## MAP

- A** Main Building University of Applied Arts  
1010 Vienna, Oskar Kokoschka-Platz 2 (Stubenring 3)
- B** Branch Sterngasse, Department for Digital Arts  
1010 Vienna, Sterngasse 13



PAPERS



MobileMusicWorkshop

6 MoGMI: Mobile Gesture Music Instrument • Amnon Dekel / Gilly Dekel

11 Developments and Challenges turning Mobile Phones into Generic music Performance Platforms • G. Essl / G. Wang / M. Rohs

15 A Typology for Listening in Place • P. Rebelo / M. Green / F. Hollerweger

19 Some Challenges Related to Music and Movement in Mobile Music Technology • A. R. Jensenius

23 Real-Time Synaesthetic Sonification of Traveling Landscapes • T. Pohle / P. Knees

26 Caressing the Skin: Mobile devices and bodily engagement • F. Schroeder

# MoGMI: Mobile Gesture Music Instrument

Amnon Dekel  
The Hebrew University Jerusalem  
The Selim and Rachel Benin School of  
Computer Science and Engineering,  
+972548138160  
amnoid@cs.huji.ac.il

Gilly Dekel  
The Hebrew University Jerusalem  
The Selim and Rachel Benin School of  
Computer Science and Engineering,  
+972548138160  
gillyd@huji.ac.il

## ABSTRACT

In this paper we describe the MoGMI project that explores ways of enabling the mobile phone to become a musical instrument for naïve users. Two applications enabled 10 subjects to use physical gesture movements to either record and play back continuous musical pieces using an onboard MIDI player or to play back simple and short digital sound files in real time. A user study explored which accelerometer axis mapping model is preferred by users. Results show that subjects preferred the three axis model in which every axis is mapped to a different dimension of music generation (attack, amplitude, and pitch). This mapping was deemed better by subjects over simpler or more complicated mapping models in three of five dimensions (easier to learn, produces “nicer” music, and in how easy it is to understand the relationship between gestures performed and the music that is subsequently generated).

## Categories and Subject Descriptors

B.4.2 Input/Output Devices, H.5.1 Multimedia Information Systems, H.5.2 User Interfaces H.5.5 Sound and Music Computing

## General Terms

Algorithms, Measurement, Performance, Design, Experimentation, Human Factors

## Keywords

Gesture Based Input and Control, Musical Instruments, Mobile Phone, User Experience, Accelerometer, Python for S60

## 1. INTRODUCTION

Mobile phones have reached the point where they contain powerful media acquisition and processing functions onboard. This enables them for the first time to become more than personal communication and media *consumption* devices and to take a serious part in a person's everyday media *creation*. Apart from being able to shoot good quality photos, phones can now also record high quality sound and even full screen full motion video at TV quality.

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. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

5TH INTERNATIONAL MOBILE MUSIC WORKSHOP 2008  
13-15 May 2008, Vienna, Austria  
Copyright 2008 ACM 1-58113-000-0/00/0004...\$5.00.

## 1.1 Mobile Phone Music Applications:

Mobile phones have been able to play high quality music for a number of years now. Phones can do this so well now that it is safe to say that mobile phones are starting to replace the single use MP3 player. Although this is an interesting area in and of itself, our project is focused on ways to *create* music on phones.

Music making applications on phones have been around for a while. In general they can be categorized in the following way:

### 1.1.1 Phone Keyboard Applications:

These classic applications enable a user to play an instrument by pressing on the number pad which is mapped to specific notes on an instrument (i.e. a piano). This model severely limits the creative possibilities of the player because of the inherent limitation of the keyboard size and the small number of keys.

### 1.1.2 DJ Mixer Applications:

This class of application allows non-musicians to mix prerecorded and user recorded tracks in a form of DJ mixing console. Using a simple interface, users select prerecorded music elements and mix them together for single or multi track playback. The application is simple enough for any phone user to enjoy.

### 1.1.3 Audio Sampling and Synthesis Applications:

This class of application allows non-musicians to generate music by either sampling sound from the world and using it as an instrument or by artificially generating sounds using audio synthesis functions (Elsdon, 2007).

### 1.1.4 Studio Sequencer Applications:

These are complex applications that use the music sequencer model in which a sophisticated multi track interface enables users to record and play back sequences. Additional functions enable mixing, looping and special effects. These applications are too difficult for the casual user. (Elsdon, 2007)

All of the above applications use the standard phone interfaces – the keyboard and on screen buttons only.

## 1.2 Accelerometer Based Applications:

Our project focuses on exploring new ways of interacting with mobile phones and is part of the Mobile Smarts project at the Hebrew University.<sup>1</sup> In the Mobile Gesture Music Instrument (MoGMI) project we are exploring ways of enabling mobile

<sup>1</sup> <http://www.cs.huji.ac.il/~amnoid/mobilesmarts/>

phones with accelerometers<sup>2</sup> to become a new form of music instrument for the casual user.

The family of applications that use on board accelerometers is growing as we speak. In the last few months the development community has been releasing these applications every week. Examples are Rotate-Me (Oueldi, 2007) that automatically rotates the phone graphic user interface to landscape or portrait mode when the phone is rotated, Next Song (Sony-Ericson, 2007) that allows Sony-Ericson users to switch to the next or previous song by twitching their phones, FlipSilent (Tongren, 2007) which allow users to change their phone profile to silent by flipping the phone on its back.

### 1.3 Mobile Gesture Music Instrument:

Our project focuses on enabling the phone to become an easy to use music instrument- that is a device on which a novice user can create music in an intuitive way. By intuitive we mean that the learning process should be very short- no complicated processes and interfaces need to be learned- at most the user needs to make a few choices (i.e. what instrument to play) and then to play them using the physical movement of the phone as the input mechanism. The purpose of the project is to explore if the phone as a physical object can become an enjoyable music creation device. We have opted to use physical gestures as the control mechanism for an on board musical instrument application.

We created two applications that enable slightly different aspects of music making. The first application allows users richer control over the instrument being played as well as its initial amplitude, note pitch and length. Once selected, users move the phone to record movements that are translated into MIDI commands which are later played back using an on-board MIDI player. The second application uses pre-recorded digital sound files which are played back depending on the movement of the phone. Using the second method enables close to real time movement and playback. Note that we strive to enable real time recording and playback of MIDI files also, but technical issues in the implementation of multi-threading in current version of the Python for S60 development environment are delaying this.

### 1.4 Accelerometer Axis Mapping:

An additional dimension which we explored is the mapping of the XYZ axis to different functions. In the first model, dubbed the "3-axis model" the X axis controls instrument attack, the Y axis controls the instrument pitch, and the Z axis it's amplitude. The second model, which we call "all for one", maps all three axis changes to the control of a single instrument pitch, while sound is continuous and amplitude stays the same. The third model (the "multiple model") uses each axis to control a different instrument (similarly to the "all for one" model) which play in parallel. Table 1 shows the different models. Illustration 1 shows the relationship of the three axis motion and the three mapping models.

## 2. OFFLINE MIDI RECORDER/PLAYER

As stated above, current issues with multi-threading in the Python for S60 development environment have forced us to create an offline application which separates input from output. We do not

Table 1. Accelerometer Mapping and Playback Modes

Models	Continuous Playback			Discrete Playback
	3 Axes All for one	All for one All for one	Multiple Different Instruments	Discrete Model
Input / Output Mode	Attack	Continuous Pitch	Instrument 1	Discrete Sound File
	Pitch	Continuous Pitch	Instrument 2	Discrete Sound File
	Amplitude	Continuous Pitch	Instrument 3	Discrete Sound File
Continuous Playback				Discrete Playback

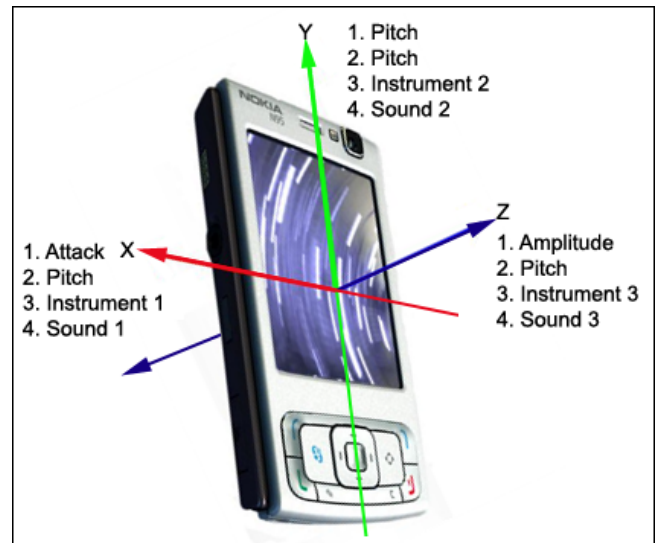


Figure 1: Motion Mapping Models (1-4)

1. Three Axis;
2. All for one;
3. Three Instruments;
4. Discrete Prerecorded Sounds

think that this is a proper method of music generation since the user does not hear what they are playing until after the fact. Once we are able to get a stable multi-threaded version in place, this limitation will be lifted. For the purposes of being able to report our findings in this paper, the offline system has been equipped with a real time graph that gives visual feedback to how the phone is being moved. This makes it easier for users to understand how the system works. Even though users had to compose the music without hearing it, we felt that we could generate information about user's preferences and aesthetic opinions of music generated using the different axis mappings.

### 2.1 Technical Description:

The application was developed on a Nokia N-95 phone with onboard accelerometer. The Symbian OS (v9.2) on this phone allowed us to quickly develop applications using Python for S60 (v1.4.1). We used the aXYZ Accelerometer module for Python written by Cyke64, a Forum Nokia champion. The module gives readings on three axis of movement – x,y,z. For each axis, a threshold of movement was used to recognize 'real' movement as opposed to 'noise'. In each model, movement on each axis was interpreted differently.

**Three Axis Model:** X controls the attack: we first check for a value that is over the X threshold to recognize that the user is

<sup>2</sup> i.e. Nokia N95 & 5500, Sony Ericson W910i, and the list is growing

playing the instrument. If there is no such movement, nothing is played. Y controls the pitch: moving the phone up causes the pitch to rise, and down causes it to go lower. Z controls the volume of the music: Gesturing forwards causes the volume to rise and backwards causes it to weaken.

**All for One Model:** In this model, movement on any axis that crosses the threshold affects the sound in the same way. An increase on any axis causes the pitch to rise, and vice versa. This model uses continuous play, and only the pitch can change.

**Multiple Instrument Model:** In this model, each axis controls a different instrument. The user selects three instruments at the beginning of use, and then 'plays' the entire ensemble by the gesture movement.

In each of the different models, the user is asked to select an instrument from a list. In the third model, each instrument is assigned a different MIDI channel, as there are multiple instruments. At this point, the user can create a piece of music by selecting the composition option. When started, the aXYZ module connects a recorder to the motion sensor. The coordinates are stored to be analyzed later. When the user completes their composing, the sensor is disconnected, and the recording stops. Once the information has been stored, an analyzer goes over the coordinates, and checks for movement – movement here is expressed by a difference in numbers that is larger than the threshold. The MIDI standard allows control of pitch, duration, and volume. The analyzing method goes over the coordinates recorded and writes the equivalent note to the midi file. Once the file is created, the user is free to listen to it, save it, send it to a friend, etc.

### 3. REAL TIME INSTRUMENT

The second application enables real time input and output, but does this by using pre-recorded sound files. This is possible because no intermediate MIDI file needs to be written to and read from at the same time. Although this application allows for close to real time generation of sound relative to the users input gestures, it suffers from having less gestural fluidity and expression. The reason is that this model uses a discreet threshold model in which a specific sound file is played depending on a predetermined threshold being passed in one of the three axis. So while the application feels more natural, it cannot offer continuous and dynamic output changes. Note that Sony-Ericson offers a very similar application on their new 910i phone, and it too suffers from the same limitations.

#### 3.1 Technical Description:

In this variant, we work with a set of prerecorded sound files. To this, we added the interpretation of the gestures, which acted as a trigger for each sound. Each different motion was mapped to one of the sounds and when a threshold was crossed, that sound file would be played.

### 4. USER STUDY

Because of the fact that we had to choose between gestural fluidity and dynamics or real time input and playback, our study has been designed to explore two main issues: first we wanted to learn about how the subjects felt about the relationship between the gestural motions that they moved the phone through and the subsequent music that is generated. Secondly we wanted to learn about how subjects compared the experience of real time

playback (with its more limited gestural language) and offline recording and playback (with its richer gesture language).

#### 4.1 Method

Using a within subjects design, 10 subjects (4 males, 6 females) were placed into one of 4 groups for counterbalancing the effect of the sequence of models to which they were exposed. Subjects were given a short explanation about the applications and how they worked and were allowed to play with them for 2 minutes to get acquainted. The tester then launched each of the applications in sequence (the specific sequence was determined by the group to which they belonged) and handed them to the subjects to use. After each run of the different applications the tester filled in a questionnaire (using a Lickert 5 point scale) in which the subjects were asked their level of agreement with a series of statements (i.e. I felt that the relationship between the movements I created and the music was good). Additionally, in order to gauge how the subjects felt about the results of the different axis mapping models, each subject was asked to rate how much they liked a series of pre-recorded MIDI files, each one generated using one of the different mapping strategies (3-Axis, All for one, Multiple Instruments). We used pre-generated files in this part in order to minimize effects of different gestures creating different musical pieces. This enabled us to have the subjects compare musical compositions which were generated by similar gestures.

#### 4.2 Results

Table 2. Ratings by Model

MODEL		Ease of Learning	Can Recreate Motion	Music Considered Nice	Can Replay Music	Understand how Gesture Affects Music
	3 Axes	4.65	3.3	4.05	2.6	3.45
	All for One	3.55	3.7	2.3	3.1	2.95
	Multiple Instruments	3.65	2.95	2.4	2.1	2.65
	Real Time	3.1	3.15	2.7	2.6	2.95
	AVE	3.7375	3.275	2.8625	2.6	3
	SD	0.653676	0.317543	0.809707	0.408248	0.331662
	P <	* 0.05	0.08	** 0.01	0.09	0.24

\*, \*\*: a significant difference was found between the highest and next to highest value

Table 3. Pre-recorded Music Quality Ratings

MODEL		Quality Preference Rating
	3 Axes	2.9
	All for One	1.9
	Multiple Instruments	2.9
	P <	0.15

Tables 2 and 3 summarize the main results of the study. In Table 2 we see the how subjects rated five dimensions of their experience in using the offline continuous MIDI recording and playback system.



#### 4.2.1 *Ease of Learning:*

Subjects rated how much they thought that the system was easy to learn to use. The single instrument three axis control model was rated as easiest to learn.

#### 4.2.2 *Confidence to be able to recreate gestures:*

Subjects rated how much they felt that they can recreate previous gestures. The All for one model was rated as creating the most confidence to be able to recreate gestures.

#### 4.2.3 *Music Considered "Nice":*

Subjects rated how much they considered the resulting music that was played back as 'nice'. It is obvious that the description 'nice' is very problematic but we chose it as a first and naïve approximation of people's aesthetic preference. The single instrument three axis control model was rated as creating the 'nicest' music.

#### 4.2.4 *Confidence to be able to replay music:*

Similarly to their confidence in being able to recreate specific gestures, here subjects rated how much they could replay a musical piece that they had previously played. The All for One Control model was rated as causing the strongest confidence in being able to replay the music.

#### 4.2.5 *Understand how gestures affect the music:*

Subjects rated how much they understood the relationship between their input gestures and the resulting music. Once again, the single instrument three axis control model was rated as being easiest to understand.

As can be seen, the most preferred model was the single instrument three axis control model (taking top place in three out of five dimensions).

Table 3 summarizes the subjects ratings of the pre-recorded MIDI files. Although the averages point at a preference for the three Axis and the All for One control models, there was no statistically significant difference between any of these. This may be affected by the small sample we used (N=10) as well as the general aesthetic quality of the prerecorded pieces. Note that we did not use the real time system in this case since it created a very different musical expression and quality. In hindsight this might have been a mistake and future work will take this into consideration.

After the testing session a free form discussion took place in which the testers explained the different mapping models and received feedback from the subjects. In all cases subject voiced an interest in continuing to use the application and "to get better at playing music with it". They all voiced their opinion that the application was fun.

### 4.3 Discussion

We start the discussion by reiterating that we think that the interaction model in the Offline MIDI recorder and player is flawed. As stated earlier, this is caused by some technical problems in the way that Python for S60 interacts with the underlying Symbian OS to enable context switching between threads and applications. If the problem had been resolved, we would have used a real time gesture input and music playback interaction model. Because this is not the case, we have opted to move forward with an exploration of people's preference relating to the resulting music as affected by different axis mapping models. We think that these results afford a glimpse into what

type of mapping model will be most preferred by users when the real time system is in fact available.

As seen in the results, the gesture mapping model that was most preferred by subjects was the three axis model. In hindsight this seems logical since this model affords the widest range of gesture dynamics and richness, with all three axis interacting with each other to enable very fine control of pitch, amplitude and attack. Not surprisingly this same model was deemed as more difficult to play (as shown by its lower ratings in subject confidence to be able to recreate gestures or music). It is easy to understand how a more complicated instrument that is richer and affords wider expressive control can be more difficult to play than a simpler instrument that is more limited in its creative range.

This can explain the high marks that the All for one model received. Since in this model all axis are aggregated into affecting only one dimension- musical pitch, it was easier for subjects to feel that they gained a level of control in the short sessions they participated in. This does not mean that the musical results were successful- their opinions hint that the All for one model generates the worst sounding music, precisely because of this unnatural mapping in our opinion.

As for how people felt about the quality of the pre-recorded MIDI sequences- as seen above, we did not find a significant difference, although the data hints that the All for one model creates the worst sounding musical pieces. Once again- we think that this points to people's preferences in hearing pieces that have a gestural logic in them, rather than pieces which use a mathematical algorithm to aggregate all motion into the control of instrument attack.

## 5. FUTURE WORK

As stated above, the major problem with this study is in the inherent limitations of the offline MIDI record and play experience. As these words are typed we are working on trying to find a solution to this problem. We feel confident that a solution will be generated, if by creative programming at our end, if by a third party extension or by the release of a Python for S60 update. Once this is achieved we will be rerunning the study. We also want to improve the Real Time player so that it can work with continuous prerecorded digital sounds and not just discrete short sound bites. Once the above is achieved we plan on wrapping the application in an inviting GUI and publish to the community.

Additional work will revolve around exploring new gesture models. One example is the use of the phone as a conductor's baton that can affect the playback speed as well as volume of the different sections of an orchestra.

## 6. SUMMARY AND CONCLUSIONS

This project explored how the mobile phone can become a standalone musical instrument for naïve non-musician users. We explored different axis mapping models and found that the one considered the richest and best sounding is the three axis model. We think that this finding, when corroborated in a future study that includes an improved user interface and real time playback, can help focus on the place where users will find the best composing experience and musical results for their learning efforts.

## 7. ACKNOWLEDGEMENTS

We would like to thank Paul Wisner and Nokia Research for providing us with a generous equipment grant.

## 8. REFERENCES

- [1] Chang, A., & Ishii, H. (2007). Zstretch: A Stretchy Fabric Music Controller. *NIME07*, (pp. 46-49). June 7-9, New York, NY.
- [2] Elsdon, A. (2007). Mobile Music Creation using PDAs and Smartphones. *4th International Mobile Music Workshop, May 6-8, 2007*, (pp. 59-60). Amsterdam.
- [3] Magnusson, T. H. (2007). THE ACOUSTIC, THE DIGITAL AND THE BODY: A SURVEY ON MUSICAL INSTRUMENTS. *Proceedings of the 2007 Conference on New Interfaces for Musical Expression (NIME07)*, (pp. 94-99). NY, NY.
- [4] Nokia. (2007). *Python for S60*. Retrieved from <http://opensource.nokia.com/projects/pythonfors60/>
- [5] Ouedi, S. (2007). *Rotate Me*. Retrieved 2007, from <http://www.bysamir.fr/rotateme/>
- [6] Rohs, M., Essl, M., & Roth, M. (2006). CaMus: Live Music Performance using Camera Phones and Visual Grid Tracking. *Proceedings of the 2006 International Conference on New Interfaces for Musical Expression (NIME06)*, (pp. 31-36). Paris, France.
- [7] Singh-Benning, M., McGuire, M., & Driessen, P. (2007). Improved Position Tracking of a 3-D Gesture-Based Musical Controller Using a Kalman Filter. *Proceedings of the 2007 Conference on New Interfaces for Musical Expression (NIME07)*, (pp. 334-337). NY, NY.
- [8] Sony-Ericson. (2007). Retrieved from <http://www.sonyericsson.com/cws/products/mobilephones/overview/w910i>
- [9] Tongren. (2007, Dec). Retrieved from <http://www.flipsilent.com/tongren/?q=node/23>

# Developments and Challenges turning Mobile Phones into Generic Music Performance Platforms

Georg Essl  
Deutsche Telekom Laboratories  
TU Berlin  
Ernst-Reuter Platz 7  
10587 Berlin, Germany  
georg.essl@telekom.de

Ge Wang  
Center for Computer Research  
in Music and Acoustics  
Stanford University  
Stanford, CA  
ge@ccrma.stanford.edu

Michael Rohs  
Deutsche Telekom Laboratories  
TU Berlin  
Ernst-Reuter Platz 7  
10587 Berlin, Germany  
michael.rohs@telekom.de

## ABSTRACT

There has been an ongoing effort to turn mobile phones into generic platforms or musical expression. By *generic* we mean useable in a wide range of expressive settings, where the enabling technology has minimal influence on the core artistic expression itself. We describe what has been achieved so far and outline a number of open challenges.

## Keywords

Mobile phone performance, technology, generic musical platform, infrastructure.

## 1. INTRODUCTION

Mobile devices have long been recognized as having potential for musical expression. There has been a rapid development over the last few years and first performances using mobile phones as primary musical instruments have emerged. For example Greg Schiemer's PocketGamelan project has served as the foundation for *Mandala*, a series of mobile-phone based works that have been performed recently [16]. Earlier this year, *MoPhO* – the Mobile Phone Orchestra of CCRMA – was founded [21], performing its first public concert on January 11, 2008 (Figure 1). However, the effort to provide a broad platform for designing and facilitating interactive musical expression on these devices is still very much in its infancy. The ongoing effort described in this paper is part of a larger field of mobile and locative music performance that involved not only mobile phones but also other mobile technologies such as PDAs, GPS and custom made sensing devices [7, 18, 6, 19, 20].

The purpose of this paper is to describe the progress of creating such platforms in the last few years. As so often, the development is mediated by what is technically possible and recent advances in technology of high-end programmable mobile phones have in no small part helped the development of the field. Using the developments so far we want to highlight a number of what we believe to be important open challenges in the field.

### 1.1 What is a generic music platform?

Before starting the discussion it is important to define the goal explicitly: What is “generic”?

*Mobile Music Workshop '08*, May 13–15, 2008, Vienna, Austria.



**Figure 1. The Mobile Phone Orchestra of CCRMA playing the piece *DroneIn/DroneOut* by Ge Wang (2008)**

By *generic* we mean a platform that is not designed with a specific performance in mind (a negative definition) or alternately, a design that is open to flexible, varied use without trying to prefigure artistic intent (a positive definition).

For example a laptop running general-purpose real-time synthesis software is a generic music platform. A laptop running a script written to accommodate a specific piece (e.g. special purpose software to control a motor that moves a speaker), is not generic.

Desktop and laptop computers have a wide range of software available that make them generic music making platforms. A range of sequencing software exist that can control general sound generation engines over MIDI or OSC. In addition to software sequencers and synthesizers, a number of programming languages and environments are available, including Csound, RTCMix, CMusic, CLM, Nyquist, SuperCollider, Max/MSP, Pure Data, and ChucK. While some commercial products may have a musical style in mind (like FruityLoops or Ableton Live) they still are generic within a very broad range and do not intrinsically try to dictate a specific style.

The goal is to have a similar and appropriate level of *genericity* for mobile phones. In other words, a platform should exist that is simultaneously adequately high-level, i.e. abstracting the more

mundane and repetitive development tasks, especially those close to a specific system hardware, and simultaneously universal enough to allow a wide variety of artistic possibilities.

## 2. CURRENT DEVELOPMENTS

Mobile devices come in many different forms and functions: They can be portable games, PDAs, mobile phones, portable media players and so forth. For many of these there have been developments to make them useful for musical performance. The attempt to turn portable gaming platforms into rather generic sounding devices is in fact rather old. Already the original GameBoy inspired a fairly generic music performance platform called nanoloop developed by Wittchow [1].

This example showed already a characteristic of different mobile devices. Often their input is geared to more specific use, like phone dialing on mobile phones, or track selection for digital music players. In the case of gaming platforms, like the Gameboy, joypads and buttons are the primary means of input. It is a regular 16-beat sequencer that can be manipulated on the fly by the game joystick and controller button.

### 2.1 Input Modalities

Here we want to review further examples of technologies that have either been directly proposed for musical use or are related. Again we will attempt at a rough classification of these technologies by type. A detailed review of sensor technologies for mobile music performance can be found in [5].

#### 2.1.1 Hand gesture sensing



**Figure 2. The accelerometer/magnetometer based interaction of ShaMus.**

The hand is a major site of human motor control and most musical instruments rely at least in parts on hand and arm actions. There are a number of technologies that allow sensing of gestures, usually using accelerometers. These platforms include the Mobile Terminal [17], Mesh [11] and XSens [13, see for a review] for iPaks. The Shake [10] which is platform independent and connects via Bluetooth.

The ShaMus project (see Figure 2) incorporates Shake sensors or uses built-in sensors of mobile devices if available (such as the Nokia 5500) [5] to manipulate interactive sound synthesis on the mobile device itself.

An alternative approach to get to hand gestures is the use of the optical system of the camera to track motion. The CaMus system [15] uses both tracking of 2-D markers or optical flow to enable this kind of hand motion sensing.

#### 2.1.2 Gait sensing

Bodily motion has played an important part in some of the mentioned performances. Usually accelerometers are used to sense the gait, from which the pace can then be derived. A possible musical use for gait has been proposed by Elliott in a concept called PersonalSoundtrack [2]. Here the idea is to vary the playback speed of a current sound track to match variation in the pace of a listener. If the pace varies significantly, the system may decide that a different song may be a better match and switch. Gait and pace detection can also be found in commercial products, though usually in the context of sport application such as giving the user feedback on their performance while jogging. Two examples are the Nokia 5500 sport phone ([nds1.nokia.com/phones/files/guides/Nokia\\_5500\\_Sport\\_UG\\_en.pdf](http://nds1.nokia.com/phones/files/guides/Nokia_5500_Sport_UG_en.pdf)) which includes accelerometers for this purpose and the Nike+iPod system embedding a sensor in the running shoe and communicating to the iPod device ([www.apple.com/ipod/nike/](http://www.apple.com/ipod/nike/)).

#### 2.1.3 Touch sensing

Most mobile devices have some number of buttons. These either are part of the standard numeric dialing keypad or are track selection buttons of music players. These can be mapped freely to synthesis algorithms. Some mobile devices, typically PDAs are equipped with a touch-sensitive screen for input. Often these are accompanied by a stylus. Geiger designed a number of interaction patterns on touch screens using a stylus, including 3-string guitar strumming and a 4-pad drum-kit [9]. Recently a commercial product appeared with a similar idea, the software JamSessions by UbiSoft ([www.ubi.com/US/Games/Info.aspx?pid=5560](http://www.ubi.com/US/Games/Info.aspx?pid=5560)) was developed for the Nintendo DS platform allows a single-string strumming interaction with a stylus. The joystick selects from a bank of pre-recorded guitar chords allowing for guitar-chord progressions to be played with a touch-pad strumming gesture.

#### 2.1.4 Using input audio for control

Finally the microphone is an important sensor for mobile devices. It can be used for literal recording as has for example happened in the MoPhive piece by Adnan Marquez-Borbon [21]. It can also be used as a generic sensor [12] where blowing into the microphone is used to excite a wind instrument or police whistle. The great advantage of microphones is its true ubiquity in mobile phones and the good dynamic range and fidelity.

## 2.2 Output Technologies

The main modalities for output on mobile devices are: visual output mainly through a screen, auditory output through speakers, and vibrotactile output via vibrotactile motor display. Often these modalities are used together. A synthesis engine using the speaker output usually also overs visual feedback. Vibrotactile display often relate to visual or auditory cues.

Generic sound synthesis engines that completely run on the mobile itself are only recently emerging. For devices running the ARM port of Linux, a ported version of PD called PDa [8] is available. The open source sound synthesis library STK (Synthesis ToolKit) by Cook and Scavone has been ported to Symbian based mobile devices [4].

An array of sound editing and sequencing programs exist which got recently reviewed by Elsdon [3]. The CaMus system, which uses optical tracing in the plane uses a graphical display in the plane, where sound sources can be placed in a virtual spot and the

distance, height and rotation relative to sound sources allow for interactive manipulation [15]. Visual output also can be an important part of feedback to the performer or the audience during a piece.

There is very limited use of vibrotactile display so far. An interesting recent specific example is Shoogle [22] where vibrotactile combined with auditory display inform the user of the presence of instant messages on a mobile device.

### 3. CHALLENGES

Many questions concerning the generic use of mobile phones as musical instruments remain open. We believe that the most pressing ones are the availability of generic synthesis software, the design of appropriate GUI and editing metaphors for mobile devices, design for the limitation of mobile devices and finally simple yet flexible ad hoc networking.

#### 3.1 More synthesis options

For one we do lack a palette of synthesis and sound rendering architecture. Currently only MobileSTK for Symbian OS [4] and pdA for Linux on mobile devices [8] are available. MobileSTK comes with a basic Symbian-based interface, while pdA retains visual elements from original PD, though these elements are very often only used for event display and not for online authoring.

#### 3.2 Special purpose editing, mapping and manipulation

Generic flexible authoring paradigms are missing. The editing system of CaMus may be the only broadly conceived graphical editing paradigm we have so far, and it is very worthwhile to envision more. CaMus's setup is very camera-centric and hence does not translate easily to gesture-based setups (see Figure 3).

The goal in designing a mobile phone musical instrument involves: 1) Decide which input modalities to use, 2) manipulate them to be good control for synthesis, 3) pick a matching synthesis algorithm. Ideally the composer should have to spend as little energy of any other cursory requirements.

#### 3.3 Limitations set by the nature of the devices

What limits this is specific to mobile devices:

1. Limit and nature of the input capacities – the standard editing interfaces for today's computers are keyboards and mouse/touchpads. Keyboards can transport a lot of textual information quickly and mouses allow to navigate graphical elements. The problem with translating these ideas to mobile devices is that there is no space for a full keyboard and that the screen-space is smaller to warrant other navigation elements than the mouse-bound typical windows-GUI.
2. Limit and nature of visual real estate – the limit of visual real estate is that there is limited space to present information which it may make sense to show graphical patches of Max/MSP or PD as a whole on a standard computer, zooming the total display down to mobile size makes them hard to follow. If one zooms in one loses visual context which leads to excessive scrolling and tedious editing of large structures. A visualization

for mobile devices needs to be much more sensitive to display only what is really crucial and hide what is not important.

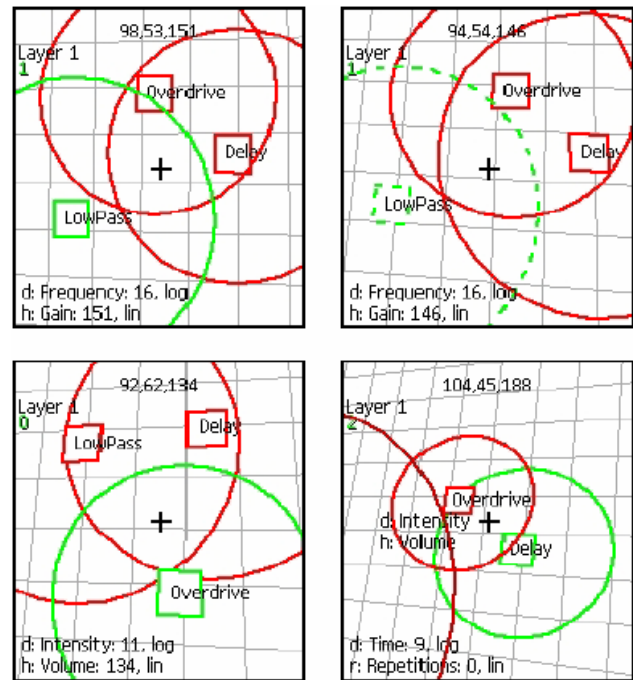


Figure 3. A view of the graphical editing platform of CaMus.

#### 3.4 Flexible ad-hoc networking

This is a very complex topic so we will but mention two basic areas. One is local ad hoc networking for localized performances, others are remote networking for remote performances. Both share that they need to be easy to administer, but certainly there are differences. Local ad hoc networking can hope for sensibly low latencies and may allow non-addressed handshaking, for example handshaking by proximity. Remote networking requires addressed hand-shaking of connections to build.

Ideally we want to be able to exchange broad performance data over these networks, specifically exchange via OSC or other common standards that are performance-centric would be useful.

This area is very much in its infancy. To the best of our knowledge the kinds of networking solutions are so far all special purpose to specific performance and installations. Proposals like the ad hoc networking of the CaMus<sup>2</sup> system is not yet generic either and uses a custom protocol of limited scope [14] (see Figure 4).

### 4. CONCLUSIONS

Mobile phones have reached a point where they have enough interesting sensory capabilities and computational power to serve as generic devices for musical expression. Yet the amount of available software infrastructure is still rather limited. In this paper we discussed a number of early steps in this direction and outlined a few open problems. Mobile phones are very attractive platforms to become generic mobile music instruments.





# A Typology for Listening in Place

Pedro Rebelo  
Sonic Arts Research Centre  
Queen's University Belfast  
Belfast, UK  
p.rebelo@qub.ac.uk

Matt Green  
Sonic Arts Research Centre  
Queen's University Belfast  
Belfast, UK  
mgreen12@qub.ac.uk

Florian Hollerweger  
Sonic Arts Research Centre  
Queen's University Belfast  
Belfast, UK  
fhollerweger01@qub.ac.uk

## ABSTRACT

Sound technologies, particularly mobile and locative media technologies, can provide unique listening experiences within situations that are not themselves exclusive zones for sonic projection, meditation or exploration. This paper seeks to contribute to the understanding of locative sound design by presenting a framework consisting of three spatial archetypes: the Theatre, the Museum and the City. These serve as metaphors through which we can articulate different types of relations between listener, sound and place. The Mobile Music Player has been chosen as an example of a listening condition that both characterises and traverses the Theatre, the Museum and the City listening archetypes.

## Keywords

Locative media, contextual media, listening in place, mobile music player.

## 1. INTRODUCTION

The proliferation of new social conditions in which complex modes of listening are called upon suggests an investigation of the interplay between new technologies, spatial archetypes and the interpretation of sound. We address modes of listening in the context of locative media and suggest three characteristic scenarios that address the relationship between a listener, sound and place. While it is beyond the scope of this paper to pursue a formal typology, we feel it is important to identify conditions of listening in place in order to better understand mobile music design and its experiential implications.

Sound in locative media applications typically depends on the spatial and temporal environment as well as the wider context (environmental and personal condition, social context, etc.) of the listener.<sup>1</sup> This suggests that the exper-

<sup>1</sup>The characteristics of these contexts go beyond the definition of cartesian space, and hence we refer to locative media also as 'contextual media' throughout this paper.

iences which they help to create are deeply personal, but at the same time often have a focus on social interaction. They are networked, mobile and interwoven with everyday routines. The locative media experience is by definition technologically mediated; this is the result of an increasing awareness of the importance of place and people in the development of mobile technologies [12]. Because this mediation can dramatically alter the way in which we engage with content, it is worth investigating how metaphors are developed to address new technologies and how these influence our listening experience.

## 2. LISTENING TO TECHNOLOGY

Beyond mere tool development, technology can serve as a mirror that reflects our understanding of the world, which is evident for example in the work of the Critical Art Ensemble [6]. This process, which usually takes more time than the technological achievement itself, can reveal multi-faceted difficulties in the application of newly developed tools. These can either be due to problems implicit to the technology itself, or pinpoint a lack of understanding on our own behalf, with the line between the two often being hard to draw. Identifying these difficulties and their sources can provide us with a valuable insight into the relation between technology and culture. In this process, metaphors developed through earlier experiences can be useful as a starting point, but at the same time should not prevent us from developing refined and more suitable models for approaching new technologies.

Numerous examples of this practice can be found in music technology: even today, many synthesizers are still structured around the imitation of non-electronic traditional musical instruments. This is perhaps due to the lack of symbolic categories for the rich sonic universe opened up by electronic music. It is hard to talk about things for which we have not yet developed a language. However, if electronic instruments are not being understood as instruments in their own right but as a miraculous and convenient replacement for entire orchestras, they inevitably fall short of our expectations. The early loudspeaker concerts of the 1950s [13] are another example of how an existing context can dominate a technology's pioneering era. Instead of radically questioning our idea of music in the face of the developments of electronic music, the replacement of musicians on stage with loudspeakers maintains recognisable models for music experience. The difficulties which this created for the audience in the reception of this music instigated clichés of electronic music being impersonal and detached from hu-

mans which partly remain until the present day. Eventually, however, they have also led to the development of new listening strategies and art forms (e.g. acousmatic listening, the sound installation, etc.).

Today, locative media and pervasive technologies challenge our concepts of music and its performance on every imaginable level. Music moves out of the safe environment of the concert hall into the open, unpredictable space of the city. Through portable devices, music has long become interwoven with and overlaid on the routines of our everyday lives. By extending these technologies with low-cost sensors, GPS receivers and network capabilities, these devices now become aware of their own environment [1, 2, 9]. The performance of music in the age of contextual media questions the notion of music as object with clearly delineated temporal, spatial and social boundaries. As adaptation to the unpredictability of real-world environments replaces the projection space of a dedicated performance environment, one must re-address the roles of producer and listener. After the gradual individualisation and abstraction of the listening experience, which has been initiated by recording technology and found its peak with the introduction of personal mobile devices [10], locative media offer the chance for a re-integration of the everyday environment into listening.

However, emerging approaches to locative media are often based on metaphors which do not support this re-integration of the everyday; a phenomenon which we will address in more detail in section 4.2. New approaches are required to address the challenges presented by contextual media. While in the first period after the introduction of a new technology, it is inevitable to talk about new means through old language, the development of new suitable metaphors can contribute to an understanding of the complex interface between technology and culture (see Coyne [5] for a discussion on metaphors in technology). We argue that by addressing listening strategies that characterise the relationship between listening and place, we can better comprehend the implications that locative media have for music and sound design.

### 3. LISTENING IN PLACE

Most research that addresses the culture of listening delineates relationships between subject and object. The object (a sound) remains relatively unaffected by the subject (the listener), the interaction between the two being normally described according to intention. The three listening modes proposed by Michael Chion in the context of audio-vision [4] reflect methods for decoding sound which he describes as reduced (Schaeffer [13]), causal and semantic. The tradition of acoustic ecology [14] treats the soundscape as a musical composition in which the listener has an active part, perhaps to the extent that the listener is involved in the composition process. More recently, in ‘Spaces Speak: Are you Listening?’ [3], Blesser and Salter address the issue of sound and space by systematically juxtaposing acoustics, psychoacoustics and musical discourses. Mobility inevitable challenges how we address sound and space. Within this framework, it is worth investigating different types of conditions that identify how one listens in place.

The listening conditions exposed by recent locative and mo-

bile media are arguably more complex, as the listening context shifts from a situation based on intention to one in which the complexity of everyday life permeates the subject-object relationship. As listening becomes increasingly modulated by space, it is necessary to address the role of the listener and associated context. With view to better understand these conditions, we propose a framework which identifies three scenarios with distinct relationships between listener/participant and place. This framework provides not only a method for the analysis of everyday listening situations, but advances a strategy for addressing design issues in the context of sound and locative media. By using the Theatre, the Museum and the City as both archetypal places and metaphors for addressing a social condition, we formalise three distinct types of listening-in-place relationships.

#### 3.1 The Theatre of Listening

The archetypal theatre clearly defines the position of the audience and stage according to the notion of projection. As those on stage embody the role of producers and those in the audience agree to the role of spectators, the listening contract is articulated as one enters the theatre. The threshold is suggested spatially by the doors to the hall and temporally by the curtain call. This mode of listening is characterized by the emphasis on communicating an experience that is notionally equal to all members of the audience and therefore treated as an object that can be projected. This paradigm has been influential in the development of sound projection techniques, instrumental forces (from chamber to orchestral) and architectural acoustics.

#### 3.2 The Museum of Listening

The museum shares with the theatre the clear threshold condition that identifies entrance and engagement. As one enters through its doors, one agrees to inhabit a curator’s world within the safety of the museum walls. The Museum of Listening is distinct from the Theatre because the sense of projection is replaced by the labyrinth of routes that emerge from the overlay of the museum’s own architecture, the exhibition layout and one’s own intentions. Unlike the Theatre, the Museum experience is likely to be fragmented, articulated not by the event, but by the spatial boundaries that differentiate one collection from another. In opposition to the Theatre, the listener in the Museum is mobile. He co-defines the spatial and temporal frame: the Museum is a building/area with clear spatial boundaries and limited opening hours, but the listener decides when to visit and what to explore.

#### 3.3 The City of Listening

The city’s fragmented, dispersed, multiform and migrational characteristics are advanced by de Certeau as an alternative to the readable and planned city [7, p.93]. To listen in the city is to be immersed by all that is not anticipated by the city planner and his ‘visual’ city. The god-like view of the urbanist provides no help in understanding what it is to be in a constant complexity of sound and to be called to articulate a multitude of events. In contrast to the well defined threshold of the theatre and the museum, the city offers no safe boundaries. The listening contract conveniently articulated by doorways and curtain calls is here replaced by a condition of potentially permanent engagement. As in the Museum,



the listener is mobile and free to define her temporal and spatial frame, but there is still an important difference: the boundaries of the city are not clearly defined. The City typically accommodates a variety of simultaneous experiences and temporal conditions.

#### 4. THE MOBILE MUSIC PLAYER

The three scenarios presented above are arguably articulated through a variety of media and situations, often in combination with one another. For the purposes of this discussion we will apply the three archetypes of listening in place to the ubiquitous condition of the Mobile Music Player.

Initially, we shall treat the Mobile Music Player as it is most widely known: the Walkman or iPod. We shall then continue by describing how recent developments in locative and context-aware media can be seen to have incited an evolution of the music player and through doing so have further strengthened the notion that the Theatre and Museum – as well as the City – are ever present and ever referenced metaphors within situated music technologies.

##### 4.1 The Mobile Music Player and the Theatre

The Mobile Music Player (e.g. the iPod) can be seen to make clear reference to the Theatre. The development of this metaphor can easily be tracked: mobile music players have been designed as portable home stereo systems, which in turn follow the idea of the public address system. Public address systems by definition lend themselves to the projection metaphor of the stage in the auditorium [10].

However, if the music player were nothing more than a mere transposition of the Theatre into the city, we would have nothing more than the Ghetto Blaster. The Mobile Music Player demonstrates that consideration has been given to the listener as individual, however superficial this may be. Our desires for mobility, privacy and control within the turbulence of the everyday have been sympathised with. A revision of the theatrical contract has taken place. However, the modification is minimal and the presence of the Theatre is still very much evident.

For some time prior to the distribution of the Mobile Music Player, stereo recordings had removed any great need for the Theatre's projecting distance: A record can retain and recreate the spatial dynamics of a performance space. Hence, the Mobile Music Player was able to create a private listening experience by reducing the required interval to mere millimetres. However, there still remains a separation of stage (the earplug) and audience (the ear). Furthermore, a temporal contract similar to that of the Theatre is still in operation: The music track played by a personal device is of a predetermined length, set by the creator within the studio, it has a definitive start, middle and end. We agree to this contract when we press the play button. However, one may exit this contract at any time: a more immediate and less socially observed stop button has replaced the theatre door.

##### 4.2 The Mobile Music Player and the Museum

Locative media technologies have propagated, and with them a new form of Mobile Music Player has developed. One of the most publicised and well-documented methods for creating such presentations is Hewlett-Packard's 'MediaScape'

software package. An example of its use is 'Riot! 1831' [11]: Within the work, the events of the Bristol riots of 1831 are conveyed through an interactive aural dramatisation. A participant can navigate through the historic scenes by investigation of the actual location of these riots, Queen's square. This is achieved through the use of a GPS positioning unit and a PDA; the sound is projected through headphones.

While environments like the above use the city rather than the white walls of a gallery as a projection surface, we suggest that they still have a great deal in common with the Museum: The spatial boundaries of the experience are still clearly defined by the designer; they choose the dimensions of the work and the location of activity. As one adorns the headphones and enters the mapped space, we agree to inhabit an artificial city, a city of the designer's choosing.

Furthermore, it would not be possible for designers to articulate all that the host city represents. Hence, they must first filter the environment; they must remove the noise. Only the prevalent and steadfast get selected. What remains is a collection of distinct structures adjoined by a channel of sleek pathways: a museum.

##### 4.3 The Mobile Music Player and the City

The theatrical contract, as we have discussed, was adjusted by the designer for use within the urban locale; the Mobile Music Player was thus created. However, it was the consumer who fully appreciated the constitution of the City and fully integrated the music player into a vocabulary of everyday life. Hence, such devices have obtained a greater significance and unique role within urban environments. For instance, beyond its intended function, the Mobile Music Player provides the individual with the means to affect their sense and meaning of an everyday situation. Tia de Nora comments on music used in this manner: "[M]usic is a device or resource to which people turn in order to regulate themselves as aesthetic agents, as feeling, thinking and acting beings in their day-to-day lives." [8, p.45].

The Mobile Music Player is not merely a method for the presentation and contemplation of a musical work, nor is it a mere means to disassociate oneself from the present place in an attempt to escape the tribulations of the city. The musical choice exercised within the everyday, actively chosen to accompany the everyday action, operates as a tool for appropriating our experiences. Through user studies, Williams [16] has identified 11 functions of portable music, most of which probably had not been anticipated by the designers of early mobile music players. They range from aestheticisation (of one's own environment) to boundary demarcation, time management and learning.

In the previous section we introduced a new form of music player that incorporates a sense of location. We suggest that, at present, only the designer has attempted to incorporate this into the City discourse. As we have commented, the success of the iPod/Walkman is as much about how the consumer has positioned the device as the designer. Hence, for location-specific presentations to propagate, they must bear more input from the situated consumer, from the individuals who constitute the audience.

Perhaps an everyday value could arise in a similar manner to that of the musical track. If the consumer becomes the designer of locative media projects or can at least exercise more choice and control, then hopefully a more illustrative reflection upon the residing place can be ascertained. The rigid model of Museum within the city, as described in section 4.2, almost dictates the meaning of locations to the individual; it does not allow the individual to formulate their own understanding of a situation, explore their own sense of place.

## 5. CONCLUSION

We have presented three spatial archetypes through which we address listening in place as a framework for a better understanding of locative media applications. The Theatre, the Museum and the City serve as metaphors through which we can evaluate the relationship of the individual, sound and place. These categories are not to be understood in a dogmatic way; many applications will feel equally comfortable in more than one of these environments, and none of these archetypes should be regarded as superior to any other. However, while mobile media applications are in many ways ‘native’ to the City, they are often addressed through the metaphors of the Museum and the Theatre. We believe that the framework presented in this paper can raise the awareness towards idiosyncrasies of locative media (such as suggested by Tanaka [15]) and therefore make a contribution towards a better understanding of their use. We hope to open a discussion on strategies for musical applications of contextual media, which should ultimately lead to the development of suitable design strategies for specific environments.

In order to address the City and its associated modes of listening, we argue that design strategies for locative media environments need to move beyond cartesian models which support a god-like view of the city [7], but fail to address the on-the-ground complexities that characterise urban environments. To move away from absolute, coordinate-driven, event-based strategies such as those suggested by environments like HP Mediascapes requires a new focus on the creation of conditions rather than events, of behaviours rather than sequences.

When using the city as an environment for listening, the absence of physical boundaries means that context types (e.g. ‘street’, ‘shopping centre’, ‘sports event’) are of greater significance than their particular instances (e.g. ‘Champs-Élysées’, ‘Harrods’, ‘2006 World Cup Final’). With the increasing availability of locative media, it is reasonable to expect further developments in this mode of listening and interaction. With the multiplicity suggested by the City, context awareness rooted in an understanding of aurality can provide listening platforms significantly different from those described in relation to the Theatre and Museum. The absence of boundaries in the City should be understood as an integral part of the design process. By replacing absolute control for multiple conditions that reflect the nature of the City itself, one can begin to address the implications for designing not only the locative and the pervasive, but ultimately the lived.

## 6. ACKNOWLEDGMENTS

We would like to acknowledge the contribution of Hewlett-Packard to this research project, parts of which have also been funded by a SPUR studentship.

## 7. REFERENCES

- [1] F. Axelsson and M. Östergren. Soundpryer: Joint music listening on the road. In *Adjunct Proceedings of the Fourth International Conference on Ubiquitous Computing*, 2002.
- [2] A. Bassoli, C. Cullinan, J. Moore, and S. Agamanolis. TunA: A mobile music experience to foster local interactions. In *Adjunct Proceedings of the Fifth International Conference on Ubiquitous Computing*, Seattle, 2003.
- [3] B. Blesser and L.-R. Salter. *Spaces speak, are you listening?* MIT Press, 2006.
- [4] M. Chion. *Audio-Vision*. Columbia University Press, 1994.
- [5] R. Coyne. *Designing Information Technology in the Postmodern Age*, chapter 7 ‘Metaphors and Machines’. MIT Press, 1995.
- [6] Critical Art Ensemble. *Nomadic Power and Cultural Resistance*. Autonomedia New York, 1994.
- [7] M. de Certeau. *The Practice of Everyday Life*. University of California Press, 1984.
- [8] T. DeNora. Music as a Technology of the Self. *Poetics*, 27(1):31–56, 1999.
- [9] L. Gaye, R. Mazé, and L. E. Holmquist. Sonic city: The urban environment as a musical interface. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, 2003.
- [10] F. Hollerweger. Three strategies for the design of social listening experiences. In *Proceedings of the 2008 Spring Conference of the UK Institute of Acoustics*, pages 609–614.
- [11] J. Reid, R. Hull, K. Cater, and C. Fleuriot. Magic moments in situated mediascapes. *Proceedings of the ACM SIGCHI International Conference on Advances in Computer Entertainment Technology*, pages 290–293, 2005.
- [12] B. Russell. Transcultural media online reader introduction, 2004.
- [13] P. Schaeffer. *Musique Concrète*. Ernst Klett Verlag Stuttgart, 1973. German edition.
- [14] R. M. Schafer. *The Soundscape. Our Sonic Environment and the Tuning of the World*. Destiny Books, 1993.
- [15] A. Tanaka and P. Gemeinboeck. A framework for spatial interaction in locative media. In *Proceedings of the 2006 conference on new interfaces for musical expression*, 2006.
- [16] A. Williams. *Portable Music and its functions*. Peter Lang Publishing, 2006.

# Some Challenges Related to Music and Movement in Mobile Music Technology

Alexander Refsum Jensenius  
Department of Musicology, University of Oslo  
PB 1017 Blindern, 0315 Oslo, Norway  
a.r.jensenius@imv.uio.no

## ABSTRACT

*Mobile music technology* opens many new opportunities in terms of location-aware systems, social interaction etc., but we should not forget that many challenges faced in "immobile" music technology research are also apparent in mobile computing. This paper presents an overview of some challenges related to the design of *action-sound relationships* and *music-movement correspondences*, and suggests how these can be studied and tested in mobile devices.

## Keywords

Music and movement, action-sound

## 1. INTRODUCTION

With the appearance of workshops and conferences, and with the support of an active community, *mobile music technology* has been established as a separate research field during the last decade, located somewhere between ubiquitous computing and new interfaces for musical expression [12]. While this mobility opens up for new and exciting applications, e.g. based on location-aware systems and techno-social interaction, I shall argue that many research questions faced in "immobile" music technologies are equally (or even more) important in mobile applications. This paper outlines some of these challenges, with a focus on the potential conflicts between our music cognition and the new technologies mediating between movement and music.

My point of departure is the idea of an *embodied music cognition* [23], where the body (and its movement) is seen as essential for our experience with, and understanding of, music [7]. Despite the long tradition of neglectance in traditional musicological research, body movement is, by necessity, a very important part of both music performance and perception.<sup>1</sup> Fortunately, the field of music and movement has gained popularity over the last decades. A large

<sup>1</sup>I prefer to use the word *perception* rather than *listening* to account for the multimodal nature of our music cognition.

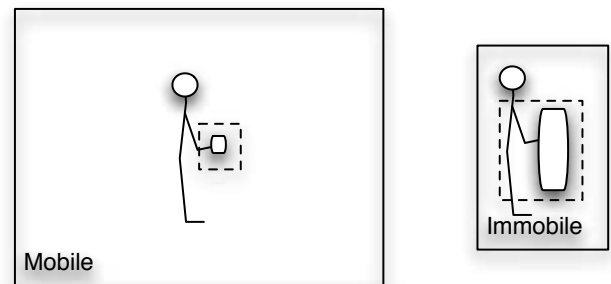
Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

*Mobile Music Workshop* Vienna

Copyright 2008 Copyright remains with the author(s).

extent of this work seem to be focused on the movements of musicians, e.g. the effect of *sound-producing movements* on expressivity and timing [6], the influence of *ancillary movements* on musical phrasing [35], or the importance of *entrainment* in performance [8]. There has also been some research on the movements of perceivers, e.g. [24, 4].

All of the above mentioned studies have been carried out on performers and perceivers that have been confined to a comparably small movement space. This is in line with how music performance and perception have often been seen as "immobile" activities; not in the sense that people do not move, but that the movements are restricted to spaces like a concert stage, or the dance floor of a club. Mobile music technology, on the other hand, is often based on the idea that the person involved is moving around in a comparably large space, e.g. a city. Also, mobile music devices are typically much smaller than immobile devices, so the movements with which the user is interacting with the device are relatively small. Thus, if we use *movement space* to denote the subjective understanding of an area in which it is possible to interact,<sup>2</sup> we may say that mobile devices have a comparably large *external* movement space and comparably small *internal* movement space, as illustrated in Figure 1.



**Figure 1: The differences in external and internal movement spaces in mobile (left: large external, small internal) and immobile (right: small external, large internal) computing.**

This difference in movement space will necessarily influence our understanding of the *action-sound relationships* and *music-movement correspondences* found in a mobile device, but before discussing these concepts further I will have to define some key terms that will be used in the discussion.

<sup>2</sup>See [20] for a longer discussion on various types of *spaces*.

## 2. TERMINOLOGY

I often see that some key terms are used very differently in the literature, so I will start by defining the terms *movement*, *force*, *action* and *gesture*.<sup>3</sup> First, we may start with two physical concepts:

- *Movement/motion*: displacement of an object in space, e.g. moving an arm in the air.
- *Force*: push or pull of an object, e.g. pushing the button on a device.

From mechanics we know that movement and force are related, e.g. a push may result in movement, but not necessarily. I will argue that movement and force are objective entities, that can be measured and quantified with for example accelerometers and force sensing resistors.

While movement and force refer to mechanics, I will argue that the terms *action* and *gesture* refer to cognition:

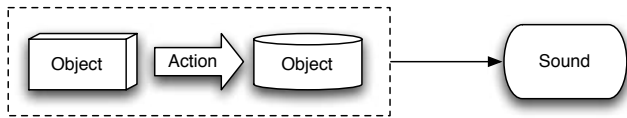
- *Action*: a *chunk* of several related movements and forces, e.g. opening a door or playing a chord on a piano. They make up a coherent unit, and also seem to be a basic building block in our cognitive system [28], and falling within the idea of a *present now* [32].
- *Gesture*: the meaning (semantics) of an action, e.g. saying goodbye when waving the hand in the air.

In the following I will mainly use the word *action*, since I will be focusing on chunks of movement and force, and their relationships to sound.

## 3. ACTION-SOUND

Recent studies suggest that our experience with *action-sound couplings*, based on relationships between actions and objects and the resultant sounds, guide the way we think about both actions and sounds [15, 20]. This is based on the motor theory of perception [25], which suggests that we mentally simulate how a sound was produced while listening. Such close connections have been tested in a series of psychological studies of sound-source perception [14] and psychomechanics [27], and have to a large extent been neurophysiologically explained by the findings of *mirror neurons* in the ventral premotor cortex of the brain [9, 22, 29].

Borrowing a term from the ecological psychology of Gibson [13], we may say that the objects and actions involved in an interaction *afford* specific sounds based on their mechanical and acoustical properties (Figure 2).



**Figure 2: The mechanical and acoustical properties of objects and actions guide our experience of the appearing sound.**

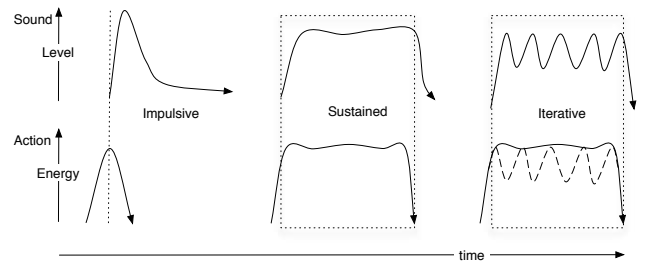
This means that we through our life-long experience with objects and actions have built up an understanding of how

<sup>3</sup>Please refer to [20] for a literature review and detailed discussion of these terms.

various types of materials and shapes sound like when excited in various ways. From this follows our ability to "see" the action of a sound we only hear, and "hear" the sound of an action we only see. Combining terminology from Schaeffer [31], and Cadoz [2], we may talk about three different *action-sound types* [16]:

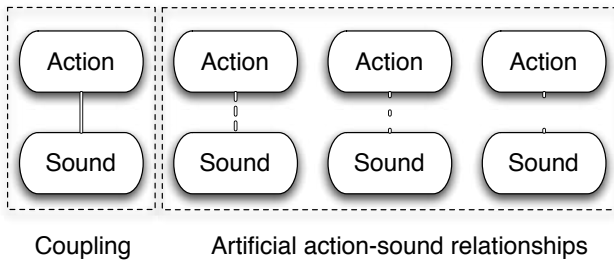
- *Impulsive*: the excitation is based on a *discontinuous* energy transfer, resulting in a rapid sonic attack with a decaying resonance. This is typical of percussion, keyboard and plucked instruments.
- *Sustained*: the excitation is based on a *continuous* energy transfer, resulting in a continuously changing sound. This is typical of wind and bowed string instruments.
- *Iterative*: the excitation is based on a series of rapid and discontinuous energy transfers, resulting in sounds with a series of successive attacks that are so rapid that they tend to fuse, i.e. are not perceived individually. This is typical of some percussion instruments, such as guiro and cabasa, but may also be produced by a series of rapid attacks on other instruments, for example quick finger movements on a guitar.

Each of these categories can be identified with a specific action and sound profile, as illustrated in Figure 3.



**Figure 3: Sketch of action energy and sound levels for the different types of sound-producing actions (the dotted lines/boxes suggest the duration of contact during excitation).**

Mapping action to sound has emerged as one of the most important research topics in the development of digital musical instruments [1, 19], and in general human-computer interaction [10, 11, 30]. However, to be able to create better artificial *relationships* between action and sound, I believe it is important to understand more about natural *action-sound couplings*. I prefer to differentiate between *couplings* found in nature and the *relationships* created artificially in technological devices. This is because I believe that a relationship can never be as solid (cognitively) as a coupling (Figure 4). Take the simple example of an electronic doorbell. Even though the action-sound relationship in the doorbell has been working the exact same way for 20 years, and you have never experienced it failing, you can never be absolutely certain that it will always work. If the power is out, there will be no sound. This type of uncertainty will never occur with a coupling. If you are dropping a glass in the floor, you know that there will be sound. The sound may be different than what you expected, but there will certainly be sound.



**Figure 4:** Artificial action-sound *relationships* (right) can never be as solid as a natural action-sound coupling (left).

## 4. GOING MOBILE

In the Musical Gestures project,<sup>4</sup> we studied how the abovementioned theories on action-sound couplings and relationships influence our cognition of music. This included a series of observation studies of how people move to music, e.g. free dance to music [5], air instrument performance [18], and sound tracing [17]. Much work still remains before the relationships between music and movement are better understood, but we did find that there were large consistencies in how people, ranging from musical novices to experts, moved to music. We also found that properties of the action-sound couplings (or relationships) seen/heard/imagined in the music were important for the motoric response to the music.

How music changes movement, and how movement can be used to change music in mobile devices, is the topic of our new 3 year research project called *Sensing Music-related Actions*. After working on music and movement in a primarily immobile setting for several years, we are now excited to start exploring the topic in a mobile setting. One reason for this is the promising results in an observation study of how people walk to music [33]. This study revealed that people walk faster on music than on metronome stimuli, and that walking on music can be modeled as a resonance phenomenon such as suggested by van Noorden and Moelants [34]. Such a resonance phenomenon has also been seen in 10 hour recordings of people’s movement patterns, which showed movement peaks with a periodicity of around 2 Hz [26].

Our approach will be based on the ideas of an embodied music cognition, where the limitations and possibilities of our cognitive system is used as the point of departure for understanding action-sound couplings and creating action-sound relationships. We will set up observation studies where people’s movement patterns will be measured and compared to the musical sound, and investigate how it is possible to create *active music devices* based on the actions of the user.

It will be particularly interesting to explore the differences in movement spaces as mentioned in the beginning of the paper. Since the internal and external movement spaces in mobile music technology differs so much from that of immobile devices, we will probably have to rethink how we capture and process movement and force data. Today’s mobile technology seem too much focused on duplicating the functionality of immobile technologies, where the focus is on capturing movement and force data from a device. In a mo-

bile setting, however, we will probably have to focus more on the actions of the user rather than the device. This calls for developing new and better sensor technologies that better capture complex body movement, and accompanying segmentation methods that can be used to find the associated actions and gestures.

Finally, we believe it is important to develop solutions for measuring and understanding everything from low- to high-level features. There has been an increasing interest in finding relationships between low-level and high-level features, i.e. going directly from motion capture data to expressive features [3] or emotional response [21]. We believe it is also important to understand more about mid-level features, i.e. *action units*. This also requires a greater conceptual understanding of relationships between continuous body movement and the semantics, i.e. the *gesture*, of the movement.

Answering such questions will, hopefully, provide further knowledge about how we can develop better mobile (and immobile) music technologies.

## 5. ACKNOWLEDGMENTS

The Musical Gestures Group for collaboration over the years, the Norwegian Research Council for financial support.

## 6. REFERENCES

- [1] D. Arfib, J.-M. Couturier, L. Kessours, and V. Verfaillie. Strategies of mapping between gesture data and synthesis model parameters using perceptual spaces. *Organised Sound*, 7(2):135–152, 2002.
- [2] C. Cadoz. Instrumental gesture and musical composition. In *Proceedings of the 1998 International Computer Music Conference*, pages 60–73, Den Haag, Netherlands, 1988.
- [3] A. Camurri, G. De Poli, A. Friberg, M. Leman, and G. Volpe. The MEGA project: Analysis and synthesis of multisensory expressive gesture in performing art applications. *Journal of New Music Research*, 34(1):5–21, 2005.
- [4] A. Camurri, B. Mazzarino, and G. Volpe. Analysis of expressive gestures in human movement: the eyesweb expressive gesture processing library. In *Proceedings of the XIV Colloquium on Musical Informatics (XIV CIM 2003)*, Firenze, Italy, May 8-9-10, 2003, 2003.
- [5] C. Casciato, A. R. Jensenius, and M. M. Wanderley. Studying free dance movement to music. In *Proceedings of ESCOM 2005 Performance Matters! Conference*, Porto, Portugal, 2005.
- [6] E. Clarke. The perception of expressive timing in music. *Psychological Research*, 51(1):2–9, 1989.
- [7] E. F. Clarke. *Ways of Listening: An Ecological Approach to the Perception of Musical Meaning*. Oxford: Oxford University Press, 2005.
- [8] M. Clayton, R. Sager, and U. Will. In time with the music: the concept of entrainment and its significance for ethnomusicology. In *European Meetings in Ethnomusicology (ESEM Counterpoint 1)*, pages 3–75, 2005.
- [9] V. Gallese, L. Fadiga, L. Fogassi, and G. Rizzolatti. Action recognition in the premotor cortex. *Brain*, 119(2):593–609, 1996.

<sup>4</sup><http://musicalgestures.uio.no>

- [10] W. W. Gaver. Auditory icons: Using sound in computer interfaces. *Human-Computer Interaction*, 2:167–177, 1986.
- [11] W. W. Gaver. The SonicFinder: An interface that uses auditory icons. *Human-Computer Interaction*, 4(1):67–94, 1989.
- [12] L. Gaye, L. Holmquist, F. Behrendt, and A. Tanaka. Mobile music technology: Report on an emerging field. In *NIME '06: Proceedings of the 2006 International Conference on New Interfaces for Musical Expression, June 4–8, Paris, France, 2006*. Paris: IRCAM – Centre Pompidou.
- [13] J. J. Gibson. *The Ecological Approach to Visual Perception*. Houghton-Mifflin, New York, 1979.
- [14] B. L. Giordano. *Sound source perception in impact sounds*. PhD thesis, University of Padova, Padova, Italy, 2005.
- [15] R. I. Godøy. Gestural imagery in the service of musical imagery. In A. Camurri and G. Volpe, editors, *Gesture-Based Communication in Human-Computer Interaction: 5th In-ternational Gesture Workshop, GW 2003, Genova, Italy, April 15-17, 2003, Selected Revised Papers*, volume LNAI 2915, pages 55–62. Berlin Heidelberg: Springer-Verlag, 2004.
- [16] R. I. Godøy. Gestural-sonorous objects: embodied extensions of schaeffer’s conceptual apparatus. *Organised Sound*, 11(2):149–157, 2006.
- [17] R. I. Godøy, E. Haga, and A. R. Jensenius. Exploring music-related gestures by sound-tracing. - a preliminary study. In *2nd ConGAS International Symposium on Gesture Interfaces for Multimedia Systems, May 9-10 2006, Leeds, UK, 2006*.
- [18] R. I. Godøy, E. Haga, and A. R. Jensenius. Playing “air instruments”: Mimicry of sound-producing gestures by novices and experts. In S. Gibet, N. Courty, and J.-F. Kamp, editors, *Gesture in Human-Computer Interaction and Simulation: 6th International Gesture Workshop, GW 2005, Berder Island, France, May 18-20, 2005, Revised Selected Papers*, volume 3881/2006, pages 256–267. Berlin Heidelberg: Springer-Verlag, 2006.
- [19] A. Hunt, M. M. Wanderley, and M. Paradis. The importance of parameter mapping in electronic instrument design. In *NIME '02: Proceedings of the 2002 International Conference on New Interfaces for Musical Expression*, Dublin, Ireland, 2002. Dublin: Media Lab Europe.
- [20] A. R. Jensenius. *Action-Sound : Developing Methods and Tools to Study Music-Related Bodily Movement*. PhD thesis, University of Oslo, 2007.
- [21] P. N. Juslin. Five facets of musical expression: A psychologist’s perspective on music performance. *Psychology of Music*, 31(3):273–302, 2003.
- [22] C. Keysers, E. Kohler, M. A. Umiltá, L. Nanetti, L. Fogassi, and V. Gallese. Audiovisual mirror neurons and action recognition. *Experimental Brain Research*, 153(4):628–636, 2003.
- [23] M. Leman. *Embodied Music Cognition and Mediation Technology*. The MIT Press, Cambridge, MA, 2007.
- [24] M. Leman, V. Vermeulen, L. D. Voogdt, A. Camurri, B. Mazzarino, and G. Volpe. Relationships between musical audio, perceived qualities and motoric responses - a pilot study. In R. Bresin, editor, *Proceedings of the Stockholm Music Acoustics Conference, August 6-9, 2003 (SMAC 03), Stockholm, Sweden, Stockholm, Sweden, 2003*.
- [25] A. M. Liberman and I. G. Mattingly. The motor theory of speech perception revised. *Cognition*, 21:1–36, 1985.
- [26] H. MacDougall and S. Moore. Marching to the beat of the same drummer: the spontaneous tempo of human locomotion. *Journal of Applied Physiology*, 99(3):1164–1173, 2005.
- [27] S. McAdams. The psychomechanics of real and simulated sound sources. *The Journal of the Acoustical Society of America*, 107:2792, 2000.
- [28] G. A. Miller. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63:81–97, 1956.
- [29] I. Molnar-Szakacs and K. Overy. Music and mirror neurons: from motion to ‘e’motion. *Social Cognitive and Affective Neuroscience*, 1(3):235–241, 2006.
- [30] D. Rocchesso and F. Fontana. *The Sounding Object*. Florence: Edizioni di Mondo Estremo, 2003.
- [31] P. Schaeffer. *Traité des objets musicaux*. Paris: Editions du Seuil, 1966.
- [32] D. N. Stern. *The Present Moment in Psychotherapy and Everyday Life*. W. W. Norton & Company, January 2004.
- [33] F. Styns, L. van Noorden, D. Moelants, and M. Leman. Walking on music. *Human Movement Science*, 26(5):769–785, 2007.
- [34] L. van Noorden and D. Moelants. Resonance in the Perception of Musical Pulse. *Journal of New Music Research*, 28(1):43–66, 1999.
- [35] M. M. Wanderley. *Performer-Instrument Interaction: Applications to Gestural Control of Sound Synthesis*. Phd-thesis, Université Pierre et Marie Curie, Paris VI, Paris, France, 2001.

# Real-Time Synaesthetic Sonification of Traveling Landscapes

Tim Pohle

Dept. of Computational Perception  
Johannes Kepler University  
Linz, Austria  
tim.pohle@jku.at

Peter Knees

Dept. of Computational Perception  
Johannes Kepler University  
Linz, Austria  
peter.knees@jku.at

## ABSTRACT

When traveling on a train, many people enjoy looking out of the window at the landscape passing by. We present an application that translates the perceived movement of the landscape and other occurring events such as passing trains into music. The continuously changing view outside the window is captured with a camera and translated into midi events that are replayed instantaneously. This allows for a reflection of the visual impression, adding a sound dimension to the visual experience and deepening the state of contemplation. The application can both be run on mobile phones (with built-in camera) and on laptops (with a connected Web-cam). We present and discuss different approaches to translate the video signal into midi events.

## 1. INTRODUCTION AND CONTEXT

Looking out of the window and watching the landscape passing by is a common thing to do on train journeys. Another popular activity is listening to music on a mobile player. However, the impressions from these two tasks -- although they are often performed simultaneously -- are not corresponding with each other, i.e. there is no synaesthetic experience. In this paper, we present an application that aims to create these synaesthetic experiences by capturing images of the outside and translating them to sounds that correspond to the visual impressions.

A major inspiration for this work was the music video for the track "Star Guitar" by "The Chemical Brothers" directed by Michel Gondry [1]. The video gives the impression of a continuous shot filmed from a passenger's perspective on a speeding train. The train passes through visually rich towns, industrial areas, and countryside. The crux of the video is that all buildings and objects passing by appear exactly in sync with the various beats and musical elements of the track. While in this video the visual elements were composed based on the musical structure, in this work, we try to compose music in real-time based on the visual structure of the passing real-world landscape. For the resulting compositions, the elements surrounding the tracks can be considered the score which is going to be interpreted based on outside conditions such as weather and lighting, the speed of the train, and the quality of the camera. Thus, every journey will yield a unique composition.

In the past, several other approaches that aim at automatically composing music based on visual content have been presented. Most of them directly map the two dimensions of images onto two acoustic dimensions, i.e. the position of pixels on the y-axis is

often interpreted as the pitch of the corresponding sound, while the x-axis is interpreted as time domain. A more sophisticated approach is presented in the work of Lauri Gröhn [2]. Based on a cell-automaton like concept, images are filtered by removing pixels in an iterative process. Different tracks for the compositions can be obtained by partitioning the image and different movements by applying slightly different graphical filters. A large number of impressive examples is made available on the Web site and the high number of on-line visits suggest that also that a wide audience is considering the results to exhibit some sort of synaesthetic correspondence. With our approach presented in this paper, we try to go even one step further by not only sonifying static images but real-time video instantaneously captured with a camera.

## 2. GENERAL IDEA

The general idea of our application is to capture the passing landscape with a camera and transform the visual data to sound, probably even to music. The landscape can either be recorded with the built-in camera of a mobile phone or with a Web-cam connected to a laptop. The data captured by the camera is given as a series of images (frames). Each frame is an array of pixels. From each captured frame, we take the middle column of the pixels and use this data to create and modify an audio stream.

The user interface of the application is divided into two main areas (cf. Figure 2). In the right half of the screen, the current frame is shown as delivered by the video camera. The left half contains a kind of history of the middle column of the picture. This history is updated at a constant rate. Every time a new frame is processed, the data contained in the left part is copied one column to the left. The (now empty) rightmost column closest to the red line is assigned the values from the last frame's middle column.

## 3. TRANSFORMATION APPROACHES

For the suggested application, we tried out several approaches to use the video data for sound creation. These are presented in this section after briefly discussing the used color space models.

In this work, two color representations are used. First, in the  $(r,g,b)$  representation the red, green and blue components of each pixel are measured independently, and represented each as a value in the range  $[0...1]$ . The values measured by the camera are given in this representation. For a perceptually more meaningful representation, the  $(r,g,b)$  representation can be transformed into the  $(h,s,v)$  representation, where *hue* (i.e., color), *saturation* (i.e.,

color intensity, ranging from 0 which is white / gray / black to 1 which is “screaming” color) and *value* (related to perceived brightness, e.g. sun / shadow) are given independently.

### 3.1 Filter Approach

In the first approach to transform video data into sound, the  $(r,g,b)$ -values in the middle column of the current video frame are transformed to grayscale by taking their mean, so that each pixel has only one value associated instead of three. These values then are interpreted as the characteristic of an audio filter. The band associated with the bottom pixel (index  $i=0$ ) has a center frequency  $f_0$ , and the bands associated with the other pixels have center frequencies of integer multiples of  $f_0$  (i.e., band  $i$  has center frequency  $i \cdot f_0$ ).

Such a filter can be implemented by applying an *inverse FFT* (iFFT, inverse Fast Fourier Transformation) on the pixel values. The output values of the iFFT are then used as taps in a FIR filter (*Finite Impulse Response*). We use this filter to impose the desired spectrum on pink noise.

### 3.2 Piano Roll Approach

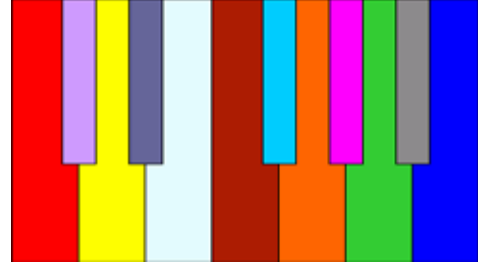
The second approach we tried is quite common. It is based on interpreting the middle column of the current video frame as a short fragment of a piano roll [5]. The piano roll was invented at the *fin de siècle*. It allows for operating player pianos without a pianist being present.

In our straightforward approach, the top pixels of the video frame column are associated with high pitches, and bottom pixels are associated with the lowest notes. To generate music, the interpreted data is sent to a MIDI instrument. The brightness ( $v$ -value) of a pixel is interpreted as the volume, while its color ( $h$ -value) is mapped to the available midi instruments (MIDI *program* number). To come closer to human perception, connected regions of similar color (or alternatively, similar brightness) are treated as one entity. To this end, we applied edge detection and region merging algorithms. If sound and pitch do not change in consecutive frames, no new MIDI event is generated.

### 3.3 Color-based Approach

Some synaesthetes have color associations when listening to sounds or tones. The Russian composer and pianist Alexander Nikolajewitsch Skrjabin was a synaesthete who created a mapping between piano tones and colors (cf. Figure 1). His composition technique sometimes is referred to as a precursor of the twelve tone technique [4].

We adopt this color mapping in the following way: As the range of the given mapping is only one octave, the middle pixel column of the current video frame is divided into  $n=4$  parts of equal height. Each of these parts then is used to generate tones played in a different octave. The pixels of a part are transformed to a pitch by calculating the cosine distance of each contained pixel's  $(h,s,v)$ -value to all of the twelve colors of the color piano. These values then are subsumed into a twelve-binned histogram over all pixels. The fullest bin is the pitch that is played in this octave. Additionally, if the second fullest bin is nearly as full as the



fullest **Figure 1: Tone-to-color mapping of Skrjabin's Clavier à lumières (cf. [4,6])**

(cutoff value 0.75) then this tone is also played. The velocity is taken from the maximum  $v$ -value of all pixels. The corresponding notes are played on a midi sound generator with a piano sound.

In many cases, colors do not change significantly between consecutive frames. To avoid repetitively playing such notes at every frame, these notes are held if the change is below a certain threshold. However, the piano has a sound that decays and vanishes after some time. Thus, this could result in a situation where all sound is gone, for example when the train stops at a station, or when the passing landscape does only change slightly. Therefore, if a note is constantly held for more than  $m=7$  frames, it is repeated. In some cases, this results in repetitive patterns that are perceived as musical themes. To avoid dominance of such patterns over the resulting overall sound, notes repeated this way are played with less and less velocity, until a minimum velocity value is reached, which is used for all consecutive repetitions.

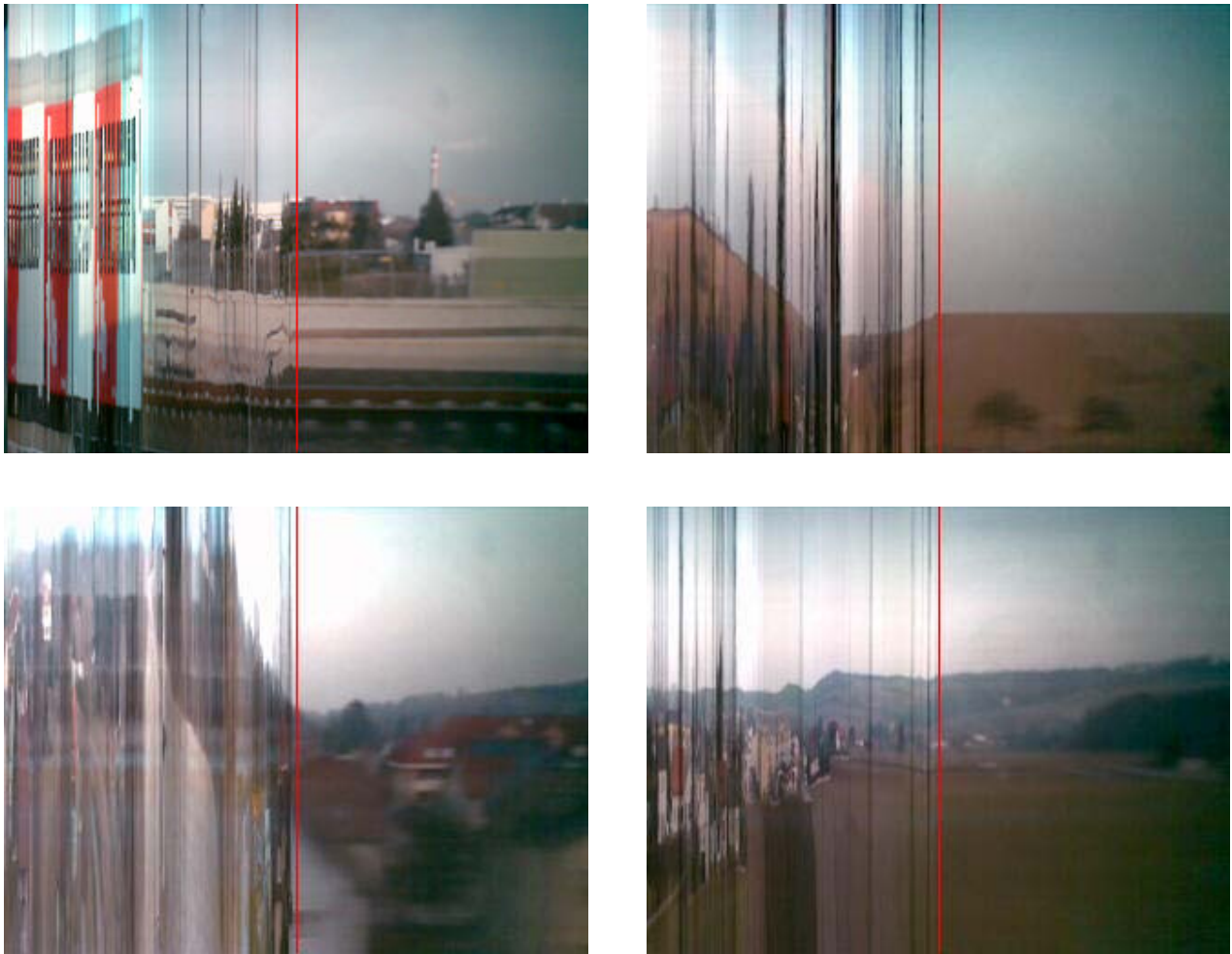
## 4. DISCUSSION

In our implementation, the filter approach did not produce convincing results. Probably the most important thing is that the resulting sounds are noisy and whistling. Such sounds are associated with trains anyway, so producing them by technical means does not add so much to the experience already available without any equipment. Also, the implementation seemed not to be sufficiently fast for a real time usage on mobile devices since calculation of an iFFT for each frame together with the transitions between the frames turned out to be too computationally expensive.

The Piano Roll Approach is more promising. However, creating algorithms that reasonably map the regions perceived in the landscape by humans to sounds (both in the  $x$ - and  $y$ -Dimension) turned out to be a task beyond the scope of this work. Although we tried to reduce the amount of notes and note onsets by region finding algorithms and by holding non-changing notes, the resulting sounds are very complex even for landscapes with a very simplistic appearance.

The Color-based Approach yields in our opinion by far the best results. Due to a steady rate of seven frames per second, there is a clearly noticeable basic rhythm pattern in the music, which the listener may associate with the steady progression of the train. Depending on the landscape, notes in some bands are played in fast repetition or movements, while in other bands they sound only sporadic. The resulting harmonies are quite pleasurable,





**Figure 2: Four example screenshots taken from the mobile version of the software running on a Nokia 6120. The right half of the screen displays the current image taken from the camera. The left half consists of the sequence of recently sonified pixel columns. The left part also exhibits some interesting effects caused by the movement of the train. Since frame rate and position of the camera are both static, proximity of objects and slope and velocity of the train result in characteristic visual effects. For example, objects that “move” at high speeds are displayed very narrow, whereas objects filmed at low speeds appear stretched. Note that similar effects can also be observed using the tx-transform technique by Martin Reinhart and Virgil Widrich (cf. [3]).**

which might be a result of the color distribution in the mapping from colors to pitches. Also, a changing landscape is reflected in the resulting music, while the overall feeling remains the same. An example video of a sonified train journey sequence can be found at <http://www.cp.jku.at/people/pohle/trainpiano.wmv>.

## 5. ACKNOWLEDGMENTS

This research is supported by the Austrian Fonds zur Förderung der Wissenschaftlichen Forschung (FWF) under project number L112-N04.

## 6. REFERENCES

- [1] Michel Gondry. 2002. Music video for “The Chemical Brothers – Star Guitar”.
- [2] Lauri Gröhn. Sound of Paintings. URL: <http://www.synesthesia.fi/> (last access: 08-Feb-2008).
- [3] Martin Reinhart and Virgil Widrich. tx-transform. URL: <http://www.tx-transform.com/> (last access: 08-Feb-2008).
- [4] Wikipedia (English). Alexander Scriabin. URL: [http://en.wikipedia.org/wiki/Alexander\\_Scriabin](http://en.wikipedia.org/wiki/Alexander_Scriabin) (last access: 08-Feb-2008).
- [5] Wikipedia (English). Piano roll. URL: [http://en.wikipedia.org/wiki/Piano\\_roll](http://en.wikipedia.org/wiki/Piano_roll) (last access: 08-Feb-2008).
- [6] Wikipedia (German). Alexander Nikolajewitsch Skrjabin. URL: [http://de.wikipedia.org/wiki/Alexander\\_Skrjabin](http://de.wikipedia.org/wiki/Alexander_Skrjabin) (last access: 08-Feb-2008).

# Caressing the Skin: Mobile devices and bodily engagement

Franziska Schroeder  
 AHRC Research Fellow  
 Sonic Arts Research Centre  
 Queen's University Belfast  
 BT7 1NN  
 ++44 90971024  
 f.schroeder@qub.ac.uk

## ABSTRACT

This text examines mobile devices by looking at the tactile interaction of the human body with the technological device. I show that the body is rendered performative by engaging with a device and I draw on a performer's interaction with a musical instrument to support this argument. This tactile interaction also exposes the tension between the bodily intimate, as experienced through the skin, and the more distant, as represented by the technological device. I argue that recent design aesthetics are driven by the urge to bring the device closer to the bodily intimate, closer to the skin.

I show that the complexities of the human touch as fronted by the skin that allow the human body to navigate the world in intricate ways become central to these design aesthetics. For this argument I examine touch by looking closely at the skin and at the ways that the skin has been understood over several centuries. The skin will be examined with view to its essential position to the perception of self, aided by the psycho-analytical interpretation developed by Daniel Anzieu [1]. It will become clear that, historically, the skin was mainly seen as a covering that kept the body together. It was then exposed in the Medieval period as an organ of interchange, more akin to a permeable membrane. With the release of the taboo of cutting the skin in the European Renaissance the perception of the skin alters immensely, and the skin is finally exposed as an entire environment, as a meeting place for the other senses [8].

In this paper I highlight that the multi-touch interfaces of recent mobile devices allow for the multiplying functions of the skin to come into being by engaging the body in various gestural moves, by providing conditions for participation, rather than by simply presenting functions of control that are still highly characteristic of many design aesthetics.

## Keywords

The Skin, Touch, Mobile Devices, Bodily Engagement

## INTRODUCTION

We are 'performing animals'<sup>1</sup> and are arguably becoming more so through engaging with recent mobile technologies. A musical instrument, a mobile technology in its own right, is testimony to such performative engagement with a device. These days we also perform the world by ways of our technological devices. Indeed, our mobile devices themselves are becoming more performative in nature, and increasingly more demanding of our attention and constant interaction. The mobile phone has ceased to be the simple 'transmitter of sounds from far' [the French term telephone consists of *télé* = "far" and *phone* = "sound". It was coined c.1830 (6)]. Instead our mobile devices have changed that they now also want us to play, work, organise, phone, schedule, plan, listen, and watch, urging us to be constantly alert, to be continually available, always connected, and also to be constantly on the move. In short, we find ourselves more and more in a constant condition of performativity, in which we are lulled into a state of needing to be fiddling, feeling, and indeed *wanting* to touch. The new mobile device has not only redefined the tactile intimacy with which we engage with the device, but also, the mobile device is redefining our notion of 'home' and place. Hence it is not surprising that in the process of being able to leave our home more easily and thus being drawn into the public eye away from a state of being intimately with oneself, mobile design aesthetics are continually looking to seduce us into a tactile engagement with the device. The devices are positioned in a way as to pull the human body towards itself, towards the intimate

<sup>1</sup> Lewis Mumford argued that we are not only *homo sapiens* or *homo ludens*, but *homo faber*, man the maker, that is maker of tools as well as maker of his own self [7, p.9]. In 1987 Victor Turner, in his seminal work *The Anthropology of Performance*, goes even further when he posits man as *homo performans*, as a performing animal. He argues that "in performing, [man] reveals himself to himself" [16, p.81].

self-touch. The recent Apple iPhone [2] represents not only a good example of an attempt to satisfy the need to touch towards which the body is drawn, but also, it creates a constantly-lurking performance potential to which the body is exposed to.

We perform the world by sight and sound, and it is worth noting that only more recently has the sense of touch become part of the discussion of how we sense the world through technology. As sight and sound become complemented by the sense of touch, we are more and more able to address complex multimodal interactions. It is thus of particular interest that recent developments in mobile technology design highlight the potential for a type of interaction that is centered in the complexity of touch rather than in the ability to push buttons. The EU project “Tangible Acoustic Interfaces for Computer Human Interaction” [14] challenged human-computer interaction by creating tangible interfaces that would allow for communication via augmented physical surfaces for example. These researchers investigated tactile impacts, such as finger tapping, nail clicking and knocking. These types of tactile two-way interactions are only rudimentarily echoed in devices such as in the Apple iPhone that is enriched with metaphors of tapping, dragging, flicking and pinching. Most commonly, mobile devices do not allow for haptic feedback as all feedback is either aural and, or visual.

## 1. Touching the Music

I propose that a mobile device should be culturally and technologically specific in the same way that a musical instrument is. I see musical instruments as an ideal research framework for investigating notions of resistance and threshold conditions. Although is beyond the scope of this paper to go deeply into investigations of the types of engagements that a performer can have with a musical instrument, I want to propose that a mobile device should suggest its own types of interaction and engagement, and like a musical instrument, must invite and ask for bodily participation. I have argued elsewhere [12] that a performer of a musical instrument does not adapt to her instrument, but that she is involved in a process of constant negotiation of the specifics as offered by the instrument, and that through this engagement she intuitively develops a sense of the instrument’s malleability, resistance and fragility.

This tactile involvement of the human body presents an important move away from the notion of the performer-instrument interaction that has often been misunderstood as one in which the instrument forms an extension to the performer’s body, in which the performing body has been misread as one that is extended by certain technologies (a musical instrument for example). The idea of extension, in which the instrument is seen as a technological extension to the body, as something that “extends” (from the Latin word

“tendere” = “to stretch” and “ex” = “out”) from the body erroneously brings to the fore the notion of transferal of body onto the instrument. The engagement with an instrument is still often seen as a transfer of information from one’s body to the instrument, as a transfer from the body to the world. The formula “from-to” mistakenly becomes of importance and I have questioned this in the past [11].

In contrast to the notion of extension we may be better positioned to see the engagement of instrument or device and body as one in which the human body is free to explore the resistances that are offered by the device. This is a type of participation, in which the instrument or the device itself suggest specific ideas of their textures and in which the human body becomes acquainted with the “thing” at hand by being able to test boundaries, negotiate subtleties and uncover certain threshold conditions. This is a type of engagement in which the device is negotiated by drawing on multiple functions, as found in the sense of touch.

We know that touch is all-pervasive in our interaction with the world [12] and it is also not surprising that the need to be touched is an innate property of humans and animals alike. It is therefore highly interesting to see that with the new ‘Apple iPod touch’, based on the same multi-touch interface as Apple’s iPhone, music is not solely something to be “called” for, something to recall with the push of a button, but the new device lets one tap, drag, flick, glide and pinch, and ultimately ‘touch’ the music [3]. Already Pierre Schaeffer, one of the pioneers of electronic music and the man often quoted as the inventor of *musique concrète* said, “we listen to music with our hands” [10]. And indeed it is in touch, by means of the differences in the skins’ textures, by means of the skin itself that we can keep in touch with ourselves [5]. Steven Connor in his elaborate writings on the skin [5] highlights the importance of the skin, and I thus do not merely want rush past it carelessly. If it is the skin that allows one to keep in touch with oneself it surely deserves some respectful touch.



**Figure 1:** the Apple iPod touch. [www.apple.com/ipodtouch](http://www.apple.com/ipodtouch)

## 2. Keeping in Touch

The reason for being able to keep in touch with oneself is that one has developed a sense of self in the first place. Connor, taking his cue from the psychoanalytical interpretation developed by Daniel Anzieu, reminds us of the importance of this development of the self through skin contact in early childhood. The first skin contact of the infant with the carer, with the one who strokes, cuddles and feeds the newborn, is integral for the development of the self in later life. Hence, there is often talk of the “peau-moi” or the skin-self; the skin as a border between self and not-self. While at the beginning of the infant’s development there is a symbiotic relationship with the carer, in which the skins of both infant and carer merge together, the infant soon distances itself from others, and the process of individuation commences. This takes place during what the French psychoanalyst Jacques Lacan titles the “mirror stage”<sup>2</sup>. Thus, the skin takes on “a function of individuation for the self, which transmits the feeling to the self that it is a single being” [1].

The skin is essential to the perception of self, and it is worth pointing out that for centuries the skin has been subject to tender fascination and caressing obsession. For a long time, however, there existed an inattention to the skin, as Connor argues. The skin was mainly seen as a kind of covering that kept the body together; a covering that maintained the integrity of the body and therewith also brought forth and emphasized the notion of inside and outside.

The Greek physician and anatomist Galen (c. 129–c. 200), although emphasizing the importance of the human anatomy, instructed his students in his work “De anatomiciis administrationibus” on dissecting a body and, as Connor convincingly argues, he did so by paying no attention to the skin [5, p.13]. When the taboo of cutting the skin in the European Renaissance was released it was primarily to gain access to the interior of the body. The skin was something one needed to get past; the focus was towards the incision and retraction of the skin in order to get inside the body. Jonathan Sawday refers to this moment of early-modern Europe as the “culture of dissection” [9, p.3]<sup>3</sup>.

<sup>2</sup> Lacan had proposed that the human infant goes through such a mirror stage, the primordial experience of identification, in which the infant identifies with an external image of the body (as seen in a mirror or as represented by his mother). This image, as it gives rise to the mental representation of an “I”, an ideal image of him- or herself, is only an *imago*, an image, which is external to the infant. It gives rise to the infant’s perception of “self” while at the same time, establishes the ego as fundamentally dependent upon external objects. The foundation for all subsequent identifications is laid in this mirror stage: the baby looks at something external and starts perceiving itself as a separate being. The “I” comes into being as the result of an encounter with an “other”.

<sup>3</sup> This was particularly triggered by the Belgian physician Andreas Vesalius (1514 -1564). Vesalius’ seminal publication of 1543, entitled

Throughout the Medieval period the skin became to be seen as an organ of interchange, a permeable membrane, not something simply to cut into in order to get past, to the inside. The skin was understood to be “traversable in two directions” [5, p.21] and the skin’s function (sweating for example) was taken into consideration as crucial for maintaining the body’s well-being.

If at first the skin was seen as a covering, it was consequently thought of more as an organ in itself, with its own structure and functions<sup>4</sup>.

The writings of Michel Serres [8] finally exposed the multiplying functions of the skin, where the skin was seen as the most various of organs. Rather than understood as surface, membrane or interface, the skin was thought of as an entire environment. The skin, according to Serres, is a meeting place for the other senses; it is what he calls a *milieu* of the other senses, the “milieu of the milieux” [8, p.97]. The skin is integral to the sense of touch and according to Serres, “in the skin, through the skin, the world and body touch, defining their common border. Contingency means mutual touching: world and body meet and caress in the skin” [8, p.97]. Thus, the skin has the inherent power to facilitate physical mingling and to delineate bodily borders, and it is the skin that is so closely linked to the notion of the intimate. The gap between what is close to the body - the skin - and the more distant - the device - is an interesting area for reconsidering design aesthetics.

## 3. Interfaces That Do Not Touch

Without overstating too much it can be said that recent mobile technology design approaches are moving into a direction where the highly complex interactions of our bodies with the world are being considered as essential to the design aesthetics. In the same way that the skin has ceased to be a pure covering of the body solely maintaining its’ integrity, and instead has become an entire environment and a meeting place for the other senses, mobile devices are ceasing to be apparatuses on which to simply press buttons. Design concepts are steering away from metaphors of extension and are abstaining from simply providing

“De humanis corporis fabrica Libri septem” - “On the fabric of the human body in seven book” (for an online English translation, see: <http://vesalius.northwestern.edu>) was one of the first textbooks of human anatomy to describe in great detail the organs and structure of the human body and included incredibly detailed illustrations of human dissections. The book was revolutionary in that prior to this moment, the medical profession had relied on inferences from animal dissections since dissections of the human body had not been possible [9].

<sup>4</sup> The Viennese military surgeon Joseph Plenck in his work of 1776, entitled “Doctrina de Morbis Cutaneis qua hi morbid in suas classes, genera & species rediguntur”, paid great attention to the skin, in particular to the diseases of the skin. Plenck’s work is often quoted as having provided the foundations for the knowledge of modern dermatology.

functions of control by which the body becomes reduced to the crudest of tactile interaction – the pressing of buttons.

In this vein, a recent design approach that attempted to go beyond the crude interaction of pressing buttons brought to the fore the Hug shirt [15]. This shirt is made to sense the strength of touch, the skins' temperature as well as the heartbeat rate and a sender's hug length. It thus aims to recreate the sensation of touch, warmth of a hug in order to send it to a distant being which must also be wearing a Hug shirt. The shirt delivers data from the inbuilt sensors' Bluetooth and actuators to one's mobile phone and sends this data to another user's phone and shirt.

The Hug shirt builds on previous interfaces as employed by Stahl Stenslie already in 1993. Stenslie's "CyberSM" performance project, a kind of cybersex performance, was a performance environment in which tactile stimuli were transmitted in real time in the world of cyberspace via a sensor suit. Stenslie's cybersex, real-time, multi-sensory, stimuli and tactile feedback communication system for two performers engaged performers in a quasi-fetishistic relation with their sensor-suits. In the work "Inter\_Skin" [13] for example, two participants wore suits made of rubber and latex that had various stimulators and effectors, including electrical stimulators and heat pads, mounted in and on it. The suits, connected over international telephone lines, were placed on the bodies of the performers, covering in particular erogenous zones like breast and genitals. After building their own 3D virtual identity from a body data bank, the performers were able to send remote tactile messages to each other via their suits.



**Figure 2:** Stahl Stenslie, »CyberSM«, 1993 - 1994  
<http://www.medienkunstnetz.de/werke/cybersm> [April 2008]  
 Fotografie | © Stahl Stenslie

Above examples are testimony to the fact that some design aesthetics offer us conditions for engagement and participation in addressing complexities of touch and the body's demand for the negotiation of subtleties. Indeed, these inherent ambiguities of the human body are poignantly echoed in Apple's ambiguous promise: "Now there's even more to touch" [4].

More interestingly though, because mobile technologies allow us to withdraw from the type of touch that is dependent on the actual contact with another person - we can be *in touch* with another without having *to touch* - the new mobile device needs to compensate by satisfying humans' urge for the skin's caress. There literally *needs* to be 'more to touch'. We want to touch, we need to touch and we want to be touched. To a certain extent, the multi-touch interfaces of recent mobile devices tenderly make room for the multiplying functions of the skin to come into being by engaging the body in taps, pinches, drags, flicks and glides. In this type of one-way tactile interaction the body is being drawn into a constantly-lurking performance potential. It is drawn out of its home, where the bodily intimate is being urged towards the distant device. It is thus not surprising that in this process of the body's alienation from itself, mobile devices are being designed in ways that, like humans, constantly position themselves to be touched, because, unlike humans, they ultimately cannot touch themselves.

#### 4. ACKNOWLEDGMENTS

The author is supported by the Arts and Humanities Research Council UK (AHRC), on a three-year Fellowship in the Creative and Performing Arts.

## 5. REFERENCES

1. Anzieu, Daniel, (1991). *Das Haut-Ich*. Frankfurt am Main, Suhrkamp Verlag.
2. Apple iphone: Available: [www.apple.com/iphone](http://www.apple.com/iphone) [April 2008].
3. Apple ipod: [www.apple.com/ipodtouch/features.html#multitouch](http://www.apple.com/ipodtouch/features.html#multitouch) [January 2008].
4. Apple ipod touch: Available [www.apple.com/ipodtouch](http://www.apple.com/ipodtouch) [April 2008].
5. Connor, Steven (2004). *The Book of Skin*. Cornell University Press, New York
6. Harper, D., 2001-last update, online etymology dictionary. Available: [www.etymonline.com](http://www.etymonline.com) [April 2008].
7. Mumford, L., (1967). *The Myth of the Machine*. London: M. Secker & Warburg Limited.
8. Serres, Michel, (1998). *Les Cinq Sens*. Paris: Hachette.
9. Sawday, Jonathan, (1995). *The Body Emblazoned: Dissection and the Human Body in Renaissance Culture*. London/New York: Routledge.
10. Schaeffer, P., 1971. *A propos des ordinateurs*. La Revue Musicale, trans. Peter Nelson, pp. 214-215. Paris.
11. Schroeder, F. (2005). *The Touching of the Touch – performance as itching and scratching a quasi-incestuous object*. Extensions: The Online Journal for Embodied Technology, Vol.2 (2005): Mediated Bodies: Locating Corporeality in a Pixilated World. Available: [www.wac.ucla.edu/extensionsjournal](http://www.wac.ucla.edu/extensionsjournal) [April 2008].
12. Schroeder, F. and Rebelo, P. (2006). *Wearable Music in Engaging Technologies*. Springer international journal AI & Society: The Journal of Human-Centred Systems, Journal no. 146. ISSN: 0951-5666 (print version) and ISSN: 1435-5655 (electronic version). Springer London. Available: <http://www.springerlink.com/content/x361120325nr2876/?p=130005a4330d438c8aa2e72115f2e5af&pi=8> [February 2008]. Previously published in the 2005 Proceedings of the Wearable Futures Conference, University of Wales, Newport/UK (DVD version of 2007).
13. Stahl Stensile (1996). *Wiring the flesh: towards the limits and possibilities of the virtual body*. Available: [www.stenslie.net](http://www.stenslie.net) [April, 2008].
14. Tai-Chi: *Tangible Acoustic Interfaces for Computer Human Interaction*. Available: [www.taichi.cf.ac.uk](http://www.taichi.cf.ac.uk) [April 2008].
15. The Hug Shirt (2006). Available: [www.cutecircuit.com](http://www.cutecircuit.com) [April 2008].
16. Turner, V., 1987. *The Anthropology of Performance*. AJ Publishing Company.

POSTERS



MobileMusicWorkshop

32 *undersound* and the Above Ground • A. Bassoli / J. Brewer / K. Martin / I. Carreras / D. Tacconi

36 An Augmented Reality Framework for Wireless Mobile Performance • M. Wozniowski / N. Bouillot / Z. Settel / J.R. Cooperstock

38 Mobile Tangible Interfaces as Gestural Instruments • F. Kayali / M. Pichlmair / P. Kotik

41 *soundFishing* • C. Midolo



# *undersound* and the Above Ground

Arianna Bassoli  
ISIG, London School  
of Economics  
London, UK WC2 2AE

a.bassoli@lse.ac.uk

Johanna Brewer  
Donald Bren School of  
Information and Computer  
Sciences, UC Irvine  
92697-3440

johannab@ics.uci.edu

Karen Martin  
The Bartlett School of  
Graduate Studies  
London, UK WC1E 6BT

karen.martin@ucl.ac.uk

Iacopo Carreras & David Tacconi  
CREATE-NET  
Via alla Cascata 56/C  
Trento, Italy

iacopo.carreras@create-net.org  
david.tacconi@create-net.org

## ABSTRACT

This paper presents the design of a mobile music sharing application, *undersound*, targeted towards the London Underground, explores the design and implementation challenges, and suggests a small-scale experiment within a bar environment to test some of the technical and interactional aspects of the application.

## Categories and Subject Descriptors

H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

## Keywords

Mobile music, peer-to-peer, music sharing, situated design

## 1. INTRODUCTION

As the field of mobile music grows, the practice of sharing music from one portable device to the next is becoming more of an everyday reality (e.g. [7]). In turn, we will describe work we have been doing with a peer-to-peer music sharing system that runs on mobile phones. This work focuses on incorporating current technologies that have a wide penetration to try and reach a large audience. Further, we are attempting to allow the use of these technologies to remain firmly grounded in real world practices. Thus, our work is less experimental and more experiential. In this paper we will discuss one long-term design project that falls into this category, and propose a very new one which it has recently inspired.

## 2. UNDERSOUND

This section will describe our living experiences with designing and developing a mobile music sharing application geared toward a very high number of users in a widely distributed setting. This section intends to report both on the design itself, and on the practicalities faced when trying to create a highly-mobile and technically challenging system that does not require users to purchase new technology.

### 2.1 Design of *undersound*

With *undersound* we have sought to create a very situated design. We decided to focus on a culturally specific experience, the one of riding the London Underground, and design a system that could be integrated with such an experience. We approached the Underground both from a personal and experiential perspective and through an ethnographic study that included in-depth interviews and photographic documentation (see [2], [4]). The rationale behind this was to understand different aspects of

traveling by ‘Tube’, and the relationships of individuals to one other, with technology and with the surrounding space.

*undersound* is a mobile music application that allows local unsigned musicians to upload Creative Commons copyrighted song at the Underground stations, and people to download and share the songs while traveling. Songs can be downloaded on mobile phones from the platforms and shared with other *undersound* users using Bluetooth (see Figure 1). Finally, passengers can see the results of the music itself ‘traveling’ through the Underground on public displays located within the stations where this information is displayed in a symbolic way. As the application is part of a EU-funded project called BIONETS, its implementation is currently being done within the scope of that project, which, as a whole, seeks to foster the development of a biologically-inspired wireless network.

We will now explain more about the design itself in connection to the predominant themes it is addressing: music-on-the-go, mnemonic places and social travelers.



Figure 1: Peer-to-peer music sharing in *undersound*

### 2.2 Music on-the-go

There are two distinct manners in which music is accessed and consumed within the Underground. On one hand, many people enter the space of the Tube with their portable music players and use them throughout their whole journeys. On the other hand, buskers play legally under the tunnels that connect different Underground lines, often entertaining, sometimes interacting with, and occasionally harassing, commuters while they walk from one side to another of the Underground network. *undersound* mirrors



both experiences in that it allows people to listen to their own music in private through their personal players, and allows musicians to distribute their music in specific parts of the Underground, like the buskers. In addition to this, *undersound* allows a bottom-up approach to music consumption, where songs are rated depending on how many times they are downloaded, shared and listened to. The application offers another way to distribute and consume music, and at the same time promotes Creative Commons practices, encouraging a share and share-alike approach. Finally, it offers to musicians a much wider diffusion of music than buskers can currently obtain, as listening is no longer confined to a singular temporal event occurring at a particular location. But, like the distinctive locations the buskers frequent, the fact that songs can be uploaded only once in the network allows for the creation of a strong link between music and place. Musicians are in fact only allowed to upload their songs once and they have therefore to choose a location for their music to be stored. The intention is that musicians would choose not only popular stations to upload their songs but also choose locations that are symbolic and meaningful to them, over time, allowing musical characters to emerge for the various stations.

### 2.3 Mnemonic Places

The upload limit is therefore meant to foster the creation of a stronger link between music and location. In his account of the commuting experience in the Paris Metro, Augé recalls how many of his personal memories are connected to specific stops of the Metro [1]; although not directly exposed to the overground, Underground stations are far from being quite similar locations spread around – or better under – the city. They are gateways to specific locations where people have lived experiences that often are recalled by the simple act of going through that station. The same way Bull noted how people using portable music players often create mnemonic narratives that combine music and location [5], we wanted to create a strong personal connection between the music that commuters download and listen to while traveling, and the locations they traverse while using the system.

In addition to this, *undersound* has been designed to increase people's awareness of their surroundings, especially in an environment where people are often passing by quickly to arrive as soon as possible to their destination. Because of the one-to-one relationship between songs and locations people can only obtain the songs either by finding them through others or going personally to the location. This can provide a motivation for people to explore not only new stations but also new locations within the city.

### 2.4 Social Travelers

Apart from fostering people's awareness of location, with *undersound* we also wanted to provide a tool that makes people interact, even subtly, with other co-located travelers. Through our ethnographic study we noted how people often listen to music or read newspapers and books to avoid social contact and to create a bubble around them, to cocoon [6] – to form a perceived private space within an overly-crowded public domain (see Figure 2). However, commuters are also often curious about each other and even express signs of socialization when they, for instance, leave their newspapers on the seat for other people to read after they have left the Tube. As newspapers in the Underground have become sort of 'common goods' and a means through which passengers acknowledge and subtly interact with one another,

*undersound* treats music as well as both a channel for interactions and as a common good that can be shared between commuters. Accordingly, *undersound* users can see on their mobile phones who else is in range and browse through their *undersound* music, with the opportunity to download songs or also to send messages to each other. We previously conducted research on mobile music sharing through a project called tunA which presented similar features to the peer-to-peer side of *undersound*, with the difference that in tunA users could only stream, in a synchronized way, music but not download the songs [3]. The Creative Commons license under which songs are registered within *undersound* allows music to not only be browsed but also freely shared.



Figure 2: Alone-together on the Underground

This interactional aspect of *undersound* does not force travelers to communicate with each other, but rather provides the opportunity to do so, or merely to discover new music through one another while traveling by Underground. Through the design we attempted, then, to strike a balance between allowing for moments of personal isolation but still accounting for, rather than imposing, times of socialization.

### 2.5 Implementing *undersound*

With an understanding of the design of *undersound* from the previous section, we will now turn to a discussion of the experience of implementing the project in this section. Because *undersound* is part of the larger BIONETS project, within the project itself, it is intended to act as a showcase for lower-level technical aspects. This means that it must run on a specific networking architecture and be developed as a modular service that can evolve with time. Because of these challenges, so far only a laptop version of the *undersound* demo has been implemented as a proof of concept for BIONETS. However, we are currently working on a version of the prototype that can run on mobile phones, which has been the original implementation plan for the design. In terms of evaluation, because *undersound* is an example of a situated design it would make sense to evaluate the application within the context it has been designed for, the London Underground.

It is, however, very challenging to gain permission to install an application such as *undersound*, as a research project, within the Tube. We have been in discussions with various members of

Transport for London about this opportunity, and have decided that once the full prototype is ready we will submit it as a proposal for the Platform for Art, a program that has shown increasing support for the presence of art installations within the Underground.

However, we have also been made aware of the difficulties that permanent installations in the Underground can pose – first, because of accessing things like power, but perhaps more importantly, because of security concerns. During the process of learning more about the real potential for such an installation, we realized that biggest hurdle would be the installation of permanent servers. Along with member of the BIONETS team, we then began to make steps towards a completely mobile version of the system, which relied on giving Nokia N800s to station workers who could instead of acting as permanent station servers, at least be a more high-powered node. However, because *undersound* is only a small part of a much larger EU project, this initiative is not yet underway.

We have been eager, however, to test any part of our design in a real-world deployment. While PC-based tests work well for the exploring the technical challenges of the network, it is still difficult to gauge the interactional component of the design. In the next section we will report on the steps we have been taking to develop a similar project which takes its inspiration from *undersound*, but will be deployed in a much smaller-scale setting, and has as its aim to build and deploy a system we could test in the near future. Hopefully, this will allow us to prepare more effectively for the challenges of mounting a field trial in the Underground itself.

### 3. A BAR EXPERIMENT

Given the complexities of developing a new interface within the scope of a much broader project, we decided it might be advantageous to start by creating a much smaller-scale, yet still demo-ready, interface. This effort is still in its early stages. We have begun to design an interface that stems from some of the same themes that *undersound* presents, while attempting to work on a much smaller scale. In this section, then, we would like to outline the design as it is taking shape, to continue a discussion on where interfaces like *undersound* might inspire future work. While we hope to have a demo of this ready in the next few months, here we will talk briefly about the design concept.

With this project we have decided to greatly narrow the scope to a single venue, and are focused on a more contained user experience. Because of the much smaller-scale, we wanted to more directly support immediate social interactions, more similarly to projects like Jukola [8] or MobiLenin [9], rather than ones drawn out over long time periods between a massive group of people. In order to do so, we decided to design the interface for a medium-sized bar which features a DJ. Further, rather than focusing on the broad exchange of music, we focus the interaction around playing a game which involves listening to—and interacting with the digital objects that represent—music.

What follows is a typical description of one such game-night at the bar. The resident DJ has already chosen his playlist for the night. In this list he has chosen several Creative Commons licensed tracks which have come from local musicians. Before the game-night, artists were able to upload their music in a contest to be featured in the event.

These tracks from local musicians are then distributed on a series of mobile phones. These phones are given to the bar staff, effectively rendering them mobile repositories for the music. Patrons of the bar can opt to play the game by signing up and taking another mobile phone from us. Then, they are grouped into teams. The DJ begins his set, and after a few tracks announces that he will need the game players to find the next track in his playlist. A notification will go up on a publicly situated display about which track needs to be found. Team members will then have to make their way through the bar to find the person who has the song on their phone and negotiate with them to be allowed to download it. While they will be able to search over Bluetooth for the track, they will still need to, possibly, employ a bit of social negotiation to convince the member of the bar staff to authorize the transfer.

Once a team manages to acquire the song they must upload it to the DJ so that he can play it. The DJ will announce that in order for the music to continue, a team must be successful. Hopefully, this will encourage other bar patrons to become involved in order to keep the music playing. After one success, the DJ will then play another track or two, before announcing that the next Creative Commons licensed track must be recovered.

The game will continue in rounds until all the songs have been found. In order to further integrate the other non-playing bar-goers, when a song is recovered and uploaded by one of the teams, it will also become publicly available for download within the bar. That way, anyone who enjoys the track can take it home with them. Over the course of the night, a tally will be kept of how many people downloaded each track, and at the end of the night this chart will be displayed on the public display. Further, these rankings will be incorporated into the final score for the teams. Thus, winning the game depends not only on recovering the most tracks, but recovering the most *popular* tracks. At first, it might be difficult to know which tracks will be popular, but over time as the chart is kept, regular patrons will begin to learn what the crowd favors, and the crowd too will be able to see the trends in their taste.

Though this system is on a much smaller-scale than *undersound* we attempted to embody in it a similar set of principles, allowing us to test both the technology and a bit of the user-experience at the same time. In a similar way local unsigned musicians are encouraged to promote their music using our system and therefore adopting an alternative way to distribute their work. Such music is being rated according to public appreciation in the same way it is in *undersound*, through embodied user interactions like downloading. Further, the ideology behind Creative Commons licensing is supported in this project as well. In addition to this, the relationship between music and location is highlighted, although not in terms of a diverse set of locales, but rather as a representation of what is happening in a particular place over time.

Here, however, the focus is more strongly on the communal and (semi)public aspect of accessing and consuming music, at least within the context of the game-night itself. Though with *undersound* the personal experience of music is also strongly highlighted, we believe this project is a good opportunity to explore the other, more public, aspects of music consumption that *undersound* addresses. Indeed, in our studies, we found that both these aspects of music listening are complex, intertwined, and need to be explored in their own right. Though this game more

strongly encourages social interaction, this was done to reflect the more intimate nature of the venue we are situating our system in.

#### 4. CONCLUSION

This paper has described the design of *undersound* and the ways in which it has addressed certain aspects of a specific experience, traveling by the London Underground. Reflecting on the real-world constraints of such a design for the scope of a large-scale research project, we have described how the arc of our development process has led to the creation of a small-scale experiment in a bar environment. While the *undersound* prototype is being developed, the bar experiment is being planned for the following months and will help to shape the scope of future steps towards the implementation and interaction design of *undersound* itself.

#### 5. ACKNOWLEDGMENTS

We would like to thank all the BIONETS partners, especially Sun/TechIdeas for their hard work on the *undersound* prototype. This work was supported in part by the National Science Foundation under awards 0133749, 0205724, 0326105, 0527729, and 0524033, by a grant from Intel Corporation, by BT, and by BIONETS.

#### 6. REFERENCES

- [1] Augé, M. 2002. *In the Metro*. University of Minnesota Press.
- [2] Bassoli, A., Brewer, J., Martin, K., Dourish, P., and Mainwaring, S. 2007. *Underground Aesthetics: Rethinking Urban Computing*. Special issue (Urban Computing) of *IEEE Pervasive Computing*, 6(3), 39-45.
- [3] Bassoli, A., Moore, J., and Agamanolis, S. 2006. *tunA: Socialising Music Sharing on the Move*. In O'Hara and Brown (eds), *Consuming Music Together: Social and Collaborative Aspects of Music Consumption Technologies*. Springer.
- [4] Brewer, J., Mainwaring, S., and Dourish, P. 2008. *Aesthetic Journeys*" In *Proc. of DIS 2008*, (Cape Town, South Africa).
- [5] Bull, M. 2000. *Sounding Out the City: Personal Stereos and the Management of Everyday Life*. Oxford, UK: Berg.
- [6] Mainwaring, S., Anderson, K. and Chang, M. 2005. *Living for the Global City: Mobile Kits, Urban Interfaces, and Ubicomp*. In *Proc. of UbiComp 2005* (Tokyo, Japan).
- [7] [www.mishare.com](http://www.mishare.com)
- [8] O'Hara, K., Lipson, M., Jansen, M., Unger, A., Jeffries, H. and Macer, P. 2004. *Jukola: democratic music choice in a public space*. In *Proc. of DIS 2004* (Cambridge, MA), 145-154.
- [9] Scheible, J. and Ojala, T. 2005. *MobiLenin – Combining A Multi-Track Music Video, Personal Mobile Phones and A Public Display into Multi-User Interactive Entertainment*. ACM Multimedia 2005 conference, Interactive Art Program, Singapore.

# An Augmented Reality Framework for Wireless Mobile Performance

Mike Wozniowski &  
Nicolas Bouillot  
Centre for Intelligent Machines  
McGill University  
Montréal, Québec, Canada  
{mikewoz,nicolas}@cim.mcgill.ca

Zack Settel  
Université de Montréal  
Montréal, Québec, Canada  
zs@sympatico.ca

Jeremy R. Cooperstock  
Centre for Intelligent Machines  
McGill University  
Montréal, Québec, Canada  
jer@cim.mcgill.ca

## ABSTRACT

We demonstrate that musical performance can take place in a large-scale augmented reality setting. With the use of mobile computers equipped with GPS receivers, we allow a performer to navigate through an outdoor space while interacting with an overlaid virtual audio environment. The scene is segregated into zones, with attractive forces that keep the virtual representation of the performer locked in place, thus overcoming the inaccuracies of GPS technology. Each zone is designed with particular musical potential, provided by a spatial arrangement of interactive audio elements that surround the user in that location. A subjective 3-D audio rendering is provided via headphones, and users are able to input audio at their locations, steering their sound towards sound effects of interest. An objective 3-D rendering of the entire scene can be provided to an audience in a concert hall or gallery space nearby.

## 1. INTRODUCTION

Large-scale outdoor spaces offer an interesting interaction space for musical performance, where all participants involved are free to explore sound in a random and nonlinear fashion. Whereas most traditional music is composed and arranged in time, mobile musical applications need to consider the spatial aspect as well. Rather than just focusing on *when* sonic events occur, the composer must also consider *where* they should be located in space. This is a difficult task, since sounds must evolve over time to be heard, while users, who trigger such events with their motions, may do so in a nondeterministic fashion. Composers must therefore arrange their sonic scores in a coherent spatial fashion. There need to be boundaries between sounds that do not mix well, and transitions between adjacent sounds that work in both directions. In a sense, the composer must lay out a mix in a topographical fashion, becoming a sort of sonic cartographer who maps out the score in both space and time.

If we then consider the ability to add live audio input into the scene, a rich venue for live performance emerges, where dynamic sound sources are driven by mobile performers. However, a significant challenge arises, since users will need to transmit audio wirelessly while maintaining synchronization with everything in the scene. There are few tools available to artists for accomplishing all of these tasks. As a result, we have expanded the Audioscape framework<sup>1</sup> to support this kind of mobile interaction. In related work, we have designed an adaptive audio streaming protocol that can transmit sound between multiple individuals on an ad-hoc network, with very low latency [9]. Furthermore, we have explored the use of Global Positioning Systems (GPS) to track multiple users in an outdoor environment and immerse them in an overlaid virtual audio scene. The initial prototype that we developed, seen in Figure 1, allowed two performers to navigate about a physical space and encounter various audio elements such as sound loops and virtual acoustic enclosures. Each participant had a subjectively rendered audio display that allowed for a unique experience as they travelled through the shared virtual scene.



Figure 1: Mobile performers

## 2. RELATED WORK

To our knowledge, the interaction we aim to achieve has not been supported by any single system, though researchers have explored various subtasks related to this challenge. The *Hear&There* project [4] allowed users to record audio at a given GPS coordinate, while providing a spatial rendering of

<sup>1</sup>[www.audioscape.org](http://www.audioscape.org)

other recordings as they navigated. Unfortunately, this was limited to a single-person experience, where the state of the augmented reality scene was only maintained on one computer. Tanaka proposed a peer-to-peer wireless networking strategy to allow multiple musicians to simultaneously share sound using hand held computers [6]. The system did not incorporate position awareness, but other work by the author [7] capitalized on location-based services of 3G cellular networks to provide coarse locations of users' mobile devices.

Position is, however, not the only type of data that has been explored in mobile applications. Projects including *GpsTunes* [5] and *Melodious Walkabout* [3] have used heading information to provide audio cues that guide individuals in specific directions. In fact, spatial audio and simulated acoustics can provide users with a wealth of information about the superimposed virtual audio scene. However, very few projects have used orientation information to *steer* audio propagation through virtual space. In our work (see [www.audioscape.org](http://www.audioscape.org) for an overview), we have provided users with the ability to precisely control the direction in which they may emit or capture sound. Thus, by walking and turning, performers can steer their instrumental sound (e.g., harmonica) towards specific virtual effects units for processing. Virtual space thus becomes the medium for musical interaction, and the organization of musical pieces becomes spatial in nature.

### 3. APPROACH

From our experiences with the initial prototype we created, we discovered that GPS accuracy is a significant problem for augmented audio scenes where users move slowly. Consumer-grade devices provide readings with errors of about 5m in the best case [8] and 100m [1] in the worst case. Furthermore, the heading information that is inferred from a user's trajectory of motion requires the averaging of several measurements over relatively large distances, and can thus exhibit large latency and inaccuracies in pedestrian applications. For spatial audio applications, these delays and errors can deteriorate the quality of the experience. In particular, head tracker latency is most noticeable in augmented reality applications, since a listener can compare virtual sounds to reference sounds in the real environment. In these cases, latencies as low as 25ms can be detected, and can then begin to impair performance in localization tasks [2].

As a result, we propose a new strategy for the organization of virtual sound elements in an augmented reality scene where users move slowly. Our approach divides physical space into *interaction zones*, where musical material is clustered around a single location. When GPS readings indicate that a user has entered within a threshold distance of a such a location, their virtual position in the scene will be 'pulled' to that location over a short period of time, where they remain fixed until they enter another zone. In a sense, we create a discrete number of attractor locations that help to minimize the effects of GPS errors, and thus lock users into positions where interesting sonic material is present.

### 4. DEMONSTRATION

Our demonstration allows a user to explore a sonically augmented space, such as a small park or playing field. While roaming in that space, the user can enter different kinds of

virtual zones, each one with a specific potential for sonic interaction – like a musical instrument. For instance, when inside an echo zone, sounds emitted by the user (or a previous user) will be heard echoing for several minutes, and musical layers can be built up. In a remixing zone, the user is surrounded by a number of synchronized playback voices which can be selectively listened to using the orientation of the listener's head.

The user is outfitted with a headset (plus mounted orientation sensor) and GPS receiver, all connected to a tiny wearable computer. An WiFi connection is established with a local (laptop) server, that maintains bidirectional audio streams and receives control signals. The audio scene is updated based on a user's current location, and spatial audio is rendered specifically for that user. For onlookers, or people waiting their turn, an additional audiovisual representation of the current scene can be displayed.

### 5. ACKNOWLEDGEMENTS

The initial prototype was produced in co-production with The Banff New Media Institute in Canada. The authors also wish to acknowledge the generous support of the NSERC / Canada Council for the Arts: New Media Initiative.

### 6. REFERENCES

- [1] R. Bajaj, S. L. Ranaweera, and D. P. Agrawal. GPS: Location-tracking technology. *Computer*, 35(4):92–94, 2002.
- [2] D. S. Brungart and A. J. Kordik. The detectability of headtracker latency in virtual audio displays. In *Proceedings of International conference on Auditory Display (ICAD)*, pages 37–42, 2005.
- [3] R. Etter. Implicit navigation with contextualized personal audio contents. In *Adjunct Proceedings of the Third International Conference on Pervasive Computing*, pages 43–49, 2005.
- [4] J. Rozier, K. Karahalios, and J. Donath. Hear & There: An augmented reality system of linked audio. In *Proceedings of International Conference on Auditory Display (ICAD)*, 2000.
- [5] S. Strachan, P. Eslambolchilar, R. Murray-Smith, S. Hughes, and S. O'Modhrain. GpsTunes: Controlling navigation via audio feedback. In *International Conference on Human Computer Interaction with Mobile devices & services (MobileHCI)*, pages 275–278, New York, 2005. ACM.
- [6] A. Tanaka. Mobile music making. In *Proceedings of New Interfaces for Musical Interaction (NIME)*, 2004.
- [7] A. Tanaka and P. Gemeinboeck. A framework for spatial interaction in locative media. In *Proceedings New Interfaces for Musical Expression (NIME)*, pages 26–30, Paris, France, 2006. IRCAM.
- [8] M. Wing, A. Eklund, and L. Kellogg. Consumer-grade global positioning system (GPS) accuracy and reliability. *Journal of Forestry*, 103(4):169–173, 2005.
- [9] M. Wozniowski, N. Bouillot, Z. Settel, and J. R. Cooperstock. Large-scale mobile audio environments for collaborative musical interaction. In *International Conference on New Interfaces for Musical Expression*, Genova, Italy, 2008.

# Mobile Tangible Interfaces as Gestural Instruments

**Fares Kayali**

Institute of Design &  
Assessment of Technology  
Vienna University of  
Technology  
Favoritenstraße 9-11, 1040  
Vienna  
fares@igw.tuwien.ac.at

**Martin Pichlmair**

Institute of Design &  
Assessment of Technology  
Vienna University of  
Technology  
Favoritenstraße 9-11, 1040  
Vienna  
pi@igw.tuwien.ac.at

**Petr Kotik**

Institute of Design &  
Assessment of Technology  
Vienna University of  
Technology  
Favoritenstraße 9-11, 1040  
Vienna  
petr.j.kotik@gmail.com

## ABSTRACT

In this paper we describe gestures for the interaction with tangible mobile interfaces. These were derived from three prototype instruments we developed over the last year. They were implemented for the Nintendo DS platform and offer different approaches to gestural interaction with music. Our research resulted in a number of suitable gestures for musical expression with mobile tangible interfaces.

### Keywords

Gestures, Tangible Interfaces, Music, Music-based Games, Prototypes, Experimental Design.

## 1. INTRODUCTION

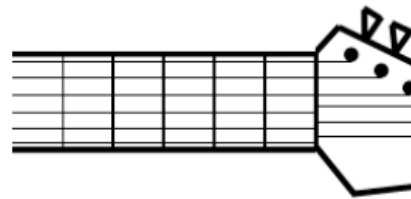
From the strumming of a guitar's strings to the beating of a drum's decks, traditional musical instruments are played by performing gestures shaped by the physical representation of the instrument. Since the musical output of digital instruments is not defined by their physical appearance, their interface can be structured more freely. Tangible interfaces put this kind of flexibility into practice. Popular examples are the Jazzmutant ›Lemur‹ [6], the ›ReacTable‹ [7,8], and acoustic environments like Masaki Fujihata's ›A small Fish‹ [3]. These instruments use a screen to offer a sound environment. Players can use physical objects to affect the music ›ReacTable‹ and ›A small Fish‹ produce. The ›Lemur‹ offers a multi-touch screen operated with the fingers. The haptic quality of the interaction significantly shapes the musical expression. With the introduction of the Nintendo DS portable console, tangible interfaces successfully penetrated mainstream culture. Playing music is an activity common to a number of Nintendo DS games - Toshio Iwai's Electroplankton [5] is maybe the most well-engineered example.

## 2. THE PROTOTYPES

Gestures are the result of a cognitive process that combines a sequence of actions to a single mental unit [2]. The kind of action

involved is determined by the interface. In the case of touch-screen interaction the interface usually displays information on the audio environment. The player interacts with widgets - active components of the user interface. The actions a player can invoke are: tapping, dragging, and releasing a widget. Since tangible interfaces feature no cursor, hovering over a widget usually is not supported (unless capacitive touch screens are used). In order to explore gestures for musical interaction we proceeded exploratively. We implemented a number of prototypical mini-games featuring distinct musical environments. Mini-games offer small and restricted spaces for the exploration of interactivity. Chaim Gingold [4] notes that they specifically allow to expose the essence of the interaction. We chose the Nintendo DS platform for its tangible user interface. The implementations are realised using the homebrew development libraries ›NDSLlib‹ [11] and ›PALib‹ [12].

The first prototype is a very simplified guitar (Fig. 1). Strumming and grabbing chords are abstracted to a single gesture. The player strums the individual frets of the guitar with the DS' stylus, triggering pre-recorded chords. The principle of a guitar on the Nintendo DS has later been commercially released as the full fledged guitar simulator ›Jam Sessions‹ [14].



**Fig.1: The guitar prototype**

The second prototype is a synthesiser instrument that is almost solely played with the stylus. The touchscreen is used as playing field. The notes of a pentatonic scale are mapped on the screen according to an invisible grid. The pitch increases from left to right, while the vertical location determines the duty cycle. A noise generator can be triggered by holding a modifier button. The player plays the instrument by either tapping the screen for individual tones or by sweeping across it to produce continuous sounds.

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. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

*Mobile Music Workshop*, May 13-15, 2008, Vienna, Austria.  
Copyright 2008



The third prototype, Thumbtack (Fig. 2), has less instrumental character than the previous two. The player acts in a playful musical environment. Four moving widgets (that look like thumbtacks) can be played with using the stylus to hold, drag and throw them around. The widgets obey simple physical rules. Each of them has a unique sonic characteristic. Every collision among the widgets or with the border of the playing field triggers a distinct sound. The player is thereby enticed into playfully creating lasting rhythmical patterns.



**Fig.2: Thumbtack prototype**

### 3. GESTURES

After building the above prototypes we examined them regarding gestures of interaction. The gestures extracted from the Thumbtack prototype all reference widgets that act as mediation between the player input and the musical result. These widgets are also described by Golan Levin [10]: »A contemporary design pattern for screen-based computer music is built on the metaphor of a group of virtual objects (or “widgets”) which can be manipulated, stretched, collided, etc. by a performer in order to shape or compose music.« This concept is used to describe several of the gestures identified in the following list:

- Holding: Touch and hold a widget. Holding it fixates the current position of the widget.
- Throwing: Touch the widget, drag it in a direction and release it. The widget moves according to the direction and velocity of the stylus' movement.
- Pushing: Touch and drag widget A against widget B. Widget B is pushed away according to the laws of reflection.
- Reflecting: Touch and drag widget A to a specific location. Hold it, while widget B approaches and collides. This way, widget B gets reflected while widget A stays in place.
- Dragging: Touch a widget and drag it to a different place. Release it after holding it still for a short time. If the widget is released while moving, it gets thrown.
- Sticking: Drag a widget over the edge of the screen. The border of the screen can be used to stick a widget under it. This way, a widget can be quickly moved out of the way.
- Confine: Using the above mentioned reflection, widget A can be used to capture and confine widget B. Widget B constantly bounces between widget A and the screen border.

- Strumming: The guitar frets can be strummed by making a sweeping movement.
- Tapping: The synthesiser can be played rhythmically by tapping on different locations of the touch-screen.
- Sweeping: Moving the stylus over the touch-screen results in a continuously changing flow of sound.

In Thumbtack the actions of the widgets form the musical output following Adriano Abbado's assumption that »a sound can be abstracted as an aural object« [1]. We already described the concept of »sound agents« in video games in [13] as one of the interactive qualities of music-based games. Sound agents are aural objects that act as interactive gameplay elements. Many of the described gestures (e.g. throwing or pushing) depend on the use of an object (the sound agent) to be understandable and reproducible. The sound agents are an abstract musical interface that is accessed with the above described gestures.

### 4. CONCLUSION

All three prototypes we built are very limited simulations of existing systems. The implemented guitar is an abstraction of a real guitar. The synthesiser is a simplified version of the Korg Kaoss Pad [9]. Thumbtack takes elements from billiard and adds an enriched audio layer. Similarly, the gestures involved in playing the first two prototypes are derived from the gestures used to play their physical counterparts. In the case of Thumbtack, the situation is slightly more complicated; some physical rules of the original game are replicated, but the interaction is vastly different. These circumstances lead to completely new conceptions about how to interact with the instrument. The player interacts indirectly via sound agents. The actual sound emerges according to rules laid out by the simulation, the game mechanics, and how the game is played. This way, a great deal of the control over the sound output remains in the hands of the designer. This article gives insights on a specific case where gestures are used to interact with widgets in the context of music-based videogames. Further research on understanding how to use gestures in conjunction with sound agents will facilitate the design of mobile instruments.

### 5. REFERENCES

- [1] Abbado, A. 1988. Perceptual Correspondences of Abstract Animation and Synthetic Sound. M.S. Thesis, MIT Media Laboratory.
- [2] Buxton, William A.S. 1986. Chunking and Phrasing and the Design of Human-Computer Dialogs. In: Information Processing '86: Proceedings of the IFIP 10th World Computer Congress, North Holland Publishers, Amsterdam.
- [3] Fujihata, M. 2001. Understanding the World (Interactive Art Category Jury Statement). In: Leopoldseider, H. & Schöpf, C. eds. (2001): Cyberarts 2001 - International Compendium of the Prix Ars Electronica. Springer-Verlag, Wien & New York, pp. 80-85.
- [4] Gingold, Chaim 2005. What WarioWare can teach us about Game Design, in Game Studies – the international journal of computer games research, volume 5, issue 1. DOI=<http://www.gamestudies.org/0501/gingold/>
- [5] Iwai, T. 2005. Electrophlankton, Nintendo. (Nintendo DS) DOI=<http://www.electrophlankton.com>
- [6] Jazzmutant. Lemur. DOI=[http://www.jazzmutant.com/lemur\\_overview.php](http://www.jazzmutant.com/lemur_overview.php)

- [7] Jordà, S., Kaltenbrunner, M., Geiger G. & Bencina R. 2005. The ReacTable\*, In Proceedings of the International Computer Music Conference (ICMC) 2005, Barcelona.
- [8] Kaltenbrunner, M., Geiger, G. & Jordà, S. 2004. Dynamic Patches for Live Musical Performance, In Proceedings of the 2004 Conference on New Interfaces for Musical Expression (NIME04), Hamamatsu, Japan, 06/03-06/05.
- [9] Korg. Kaoss Pad. DOI=[http://www.korg.com/gear/info.asp?a\\_prod\\_no=KP2](http://www.korg.com/gear/info.asp?a_prod_no=KP2)
- [10] Levin, G. 2000. Painterly Interfaces for Audiovisual Performance, Master Thesis, MIT. DOI=<http://acg.media.mit.edu/people/golan/thesis/thesis300.pdf>
- [11] NDSlib. DOI=<http://sourceforge.net/projects/ndslib/>
- [12] PALib. DOI=<http://palib.info>
- [13] Pichlmair, M. & Kayali, F. 2007. Playing Music: On the Principles of Interactivity in Music Video Games. Tokyo, Japan : Situated Play, Proceedings of DiGRA 2007 Conference, 2007.
- [14] Plato. 2007. Jam Sessions, Ubisoft. (Nintendo DS) DOI=<http://jamsessionsgame.uk.ubi.com/>



# soundFishing

Claudio Lucio Midolo

1064 Myrtle avenue, New York, NY  
1 718 809 71 46

claudio.midolo@gmail.com

## ABSTRACT

The aim of this paper is to explain the design investigation behind the creation of the soundFishing interface: a portable, semi-autonomous, digital tool that is able to analyze the sonic environment around us and extract some particular sounds out of it. The main purpose of this project is to draw attention to those every-day life sonic perceptions that we usually don't pay too much attention to, and to enable the individual to rediscover both the power and the value that they carry. The development of this concept and the research project is here outlined.

## 1. INTRODUCTION

The initial inspiration for the soundFishing project can be tracked back to a subway routine trip from Brooklyn to Manhattan that took place in early October 2007. Usually the train is very crowded, but that morning I found myself alone and, since I had nothing else to do but wait for the train to stop at my destination, I started listening to those environmental sounds I usually don't pay attention to. At the beginning they seemed to be just random audio events caused by the motion of the train, but the more I was paying attention to them, the subtler the pattern became and, in the end, they really merged into a strange, yet fascinating musical piece made of rhythmical accelerations, repetitions and vibrations.

## 2. MOTIVATION

That experience had a double impact: first it revealed the real value carried by those ephemeral sonic perceptions, then it also showed that all that "sound matter" is often silenced: it usually goes wasted because we are constantly surrounded by it, leaving us unable to really understand its value. The project tries to find a solution to this issue saving these sound fragments from oblivion, rescuing them from the world's indifference and letting them tell a different story about it: a tale about the places we live in, the people we meet and the experiences we are going through every day. They speak to us about something that maybe we didn't know and never noticed before.

## 3. CONCEPTS

The following concepts represent the core of the soundFishing project and its development:

### Sounds as an intimate diary

The audio captured from the environment will build up a sonic diary of the events that take place during the user's everyday life.

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. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

The basic difference between a traditional textual diary and this diary of sounds is that the former is created consciously by the user who has the power to personally intervene on it, deciding what and when to write something, while the diary produced by the soundFishing interface is composed "unconsciously" by the user who has just the power to set the basic logical rules that will control the capturing of the sound events: the user can not decide explicitly what and when to record. The control loss embedded in the tool's functionality can result into a surprise effect and induce curiosity towards an otherwise rather obvious final output.

### Generative sampling - Automation

The nature of the interface will be intimately procedural and algorithmic as the user will define a set of rules that will regulate the recording process. The interface works on its own without any direct control: it operates as an autonomous audio filtering agent continuously browsing the environment for events to happen. Once found the sonic event in compliance with the user's instruction, the device starts to capture the sounds. This process can be linked to Manovich's concept of Automation: "The numerical coding of media (principle 1) and the modular structure of a media object (principle 2) allow for the automation of many operations involved in media creation, manipulation and access. Thus human intentionality can, in part, be removed from the creative process.", "The Internet, which can be thought of as one huge distributed media database, also crystallized the basic condition of the new information society: overabundance of information of all kinds. One response was an idea of software 'agents' designed to automate searching for relevant information. Some agents act as filters that deliver small amounts of information..." The concept of the agent Manovich refers to in his book "The Language of New Media" is very similar to that behind the soundFishing interface: they are both media softwares that analyze and filter a particular environment. In Manovich the filter is applied to a virtual environment, such as the Internet, whereas the soundFishing interface acts on a sonic layer, in order to extract some valuable sound fragments out of it.

### Multiplicity

Generative processes lead to multiplicity. In order to capture the essence of this concept, I consider relevant the differences between the following two statements: "I want to record the sound of the police car siren that is now patrolling the street." and "I want to record all and only the loud sounds that I'll come across in my daily routine today." A huge difference lies between these two statements: the first sentence leads to a simple, but rather obvious, result. On the contrary, the second statement opens up to many possible results, giving the user a glimpse of the almost infinitely wide spectrum of possibilities that we come across in our daily experiences and depicting only a tiny portion of the space of potential. In a comment posted to the teemingvoid Blog on October 29, 2007, Mitchell Whitelaw noted: "Multiplicity here is a way to get a perceptual grasp on something quite abstract -

that space of possibility. We get a visual ‘feel’ for that space, but also a sense of its vastness, a sense of what lies beyond the visualization”, “... Multiplicity refers to the specific space of potential in any single system, by actualizing a subset of points within it.”

### Expanded cinema

"The computer liberates man from specialization and amplifies intelligence." Youngblood compares computer processing to human neural processing, where logic and intelligence are the brain's software. According to him computer software will become more important than hardware and that in the future super-computers will design ever more advanced computers. His vision of the future is represented by the Aesthetic Machine: "Aesthetic application of technology is the only means of achieving new consciousness to match our environment." It is also stated that according to Youngblood creativity will be shared between man and machine. This idea can be supported by the 1010ap-fm01 case, as it is explained on the homonymous website: "fm transposes non-metaphoric systems and grammar theory (of computer languages, abstraction and data containers) to the realm of expanded cinema. The base proposal concerns the development of a scripting language, data structures, and suitable file system for the automated production and grammatical expression of endless cinema.". According to this specific point of view, the soundFishing interface can become an extension of the human ear and memory, allowing a more powerful perception of the sonic environment and a more effective memorization of sounds in the form of digital samples. These samples can then feed another generative system which assembles them algorithmically to produce further sonic experiences.

## 4. PRECEDENTS

*The Dictaphone:* A sound recording device most commonly used to record speech for later playback or to be typed into print.

*Microsoft SenseCam:* A wearable digital device equipped with an array of sensors able to feel the surrounding environment and autonomously take photographs of the user's life.

*Remembrance Agent:* An autonomous software agent which enhance user's memory analyzing the current context suggesting in real time a list of documents relevant to the current task.

*Forget-Me-Not:* A device who is constantly carried by the user, which captures important data in its context in order to build a database of memories organized similarly to the way the human mind naturally structures human episodic memory.

*Sonic City:* A portable device designed to analyze the environment which surrounds the user, extract meaningful data from it in order to compose electronic music in real time, the urban environment becomes an interface for music creation.

These project are considered both technical and conceptual precedents as soundFishing shares with them many key ideas such as the real time analysis of the context which surrounds the user, the filtering of the incoming data in order to extract valuable information, the enhancement of user's memory and creative possibilities.

## 5. METHODOLOGY

A three stages prototyping process has been followed, according to a hierarchy based on building ease, portability and power. These values have been chosen in order to successfully build and test the interface during a one month time span. The work carried out during the first stage of the project is based on Processing, a

very powerful and versatile programming environment based on Java. Processing, among many other functions, is ideal as a rapid prototyping tool. This environment allowed me to build a working software prototype which embodies the main features of the soundFishing tool through a very short period of time. During the second stage of this project an hardware circuit will be built using a microcontroller and an audio recorder chip, achieving good portability and power. Finally, over the third stage, I planned to take the prototype to the maximum portability and power hacking a classic iPod Mp3 player - an already existing audio device which in theory could have given me the power to store a huge amount of sound data into a compact, comfortable and common object.

## 6. IMPLEMENTATION

The setting for the first prototype was a laptop running a Java applet built in Processing, an external microphone attached to it and a bag to carry them around; the logical rule implemented at this stage was telling the interface to capture all the “loud” sounds relatively to the default volume that characterized the environment. To summarize the process, the user sets a rule – in this case based on sound volume - editing a configuration file which enables him to choose both the duration of the total recordings, the duration of the volume buffer and, more importantly, the interface tolerance in relation to the volume: in order for the sounds to be captured, the lower this final parameter is, the louder the sounds have to be in relation to the default volume that characterizes the environment. While the software is running, it listens to the sound input coming from the microphone and continuously calculates the default environmental volume in order to define and adapt the threshold in relation to which a sound is considered a loud event and hence is recorded.

## 7. EVALUATION

Due to time constrain only the results of the first prototype stage are available in form of digital audio files, these are some sounds “fished” in Brooklyn during the first prototyping stage:

<http://de.posi.to/podcast/ps2p.rss>

Although the recording quality is not excellent and the form factor of the device is cumbersome, the basic rule system worked very well, recording just the sound events that matched the rule set by the user. In order to transform this project into a working tool ready to be distributed to the public, many efforts must be put in shrinking the device to make it wearable, so that the user perceives what he is carrying around not as something detached and cumbersome, but as something intimate and easy to wear. According to this, mobile phones can be considered an interesting platform to work on as they already embody the technical and computational features needed to let the soundFishing interface run as a software application, possibly embedded in their hardwares. The rule system has to be refined so that many different audio parameters can control the recording process, not just the amplitude but also the frequencies, thus the final output can carry a wider variety spectrum. Finally a system to access, manage and arrange the audio fragments is desirable, so that the user can create new audio experiences from the samples captured from his or hers life experiences. The sonic snapshots can be valuable also to other people as creative assets. Musicians and audio producers are always looking for interesting sounds and the output of the soundFishing interface can be appreciated also by these professionals.

## 8. CONCLUSION

I believe that this device is much more than an automatic sound recorder: it could, at a first glance, look similar to the old

Dictaphone, but only if its technology is considered inattentively. The motivations that have led me to build the soundFishing interface and the context I imagine it can be used for as an asset can explain its features and its nature, bearing in mind the evolution that the digital devices are nowadays experiencing. This process is transforming them into objects as intimate as our personal diaries or as personal as our favorite garments. I started this project with some clear ideas on my mind and a problem to solve: we are surrounded by sounds, sometimes they are awful and annoying, often they are sublime and inspiring, but in both cases we are losing them, not just because they are volatile in their nature, but because we usually take them for granted. We consider sound a common and unremarkable matter, therefore we don't travel around waiting to record a sound that could interest or move us. The problem lies precisely here: maybe that "common" sound can be valuable to us or to another person, it can make us laugh or remind us of an important experience or tell us something more about our life. So why not try to save them from oblivion? The key to really grab the essence of this project is held by the concept of curiosity, a virtue that can turn something usual and useless into something unique and meaningful, a powerful entity that can open the door of knowledge to all of us.

## 9. REFERENCES

- [1] Manovich, L. 2002. *The Language of New Media* (Cambridge: The MIT Press, 2002).
- [2] Youngblood, G. 1971. *Expanded Cinema* (New York: EP Dutton, 1971), 180-2.
- [3] Rhodes, B. J. 1996. Remembrance Agent: A continuously running automated information retrieval system. *Proceedings of The First International Conference on The Practical Application Of Intelligent Agents and Multi Agent Technology (PAAM '96)*, London, The Practical Application Company Ltd, pp. 487-95.
- [4] Lamming, M., Flynn M., 1994. Forget-me-not: Intimate Computing in Support of Human Memory. *Proceedings of FRIEND21, '94 International Symposium on Next Generation Human Interface*, Meguro Gajoen, Japan.
- [5] Whitelaw, M. 2007. More is More: Multiplicity and Generative Art. <http://teemingvoid.blogspot.com/2007/10/more-is-more-multiplicity-and.html>
- [6] Gaye L., Holmquist L. E., Mazé R. 2003. *Sonic City: The Urban Environment as a Musical Interface*. NIME 2003, Montreal, Canada.

## CREDITS

### ORGANISATION

Nicolaj Kirisits (University of Applied Arts, Vienna, Austria),  
Lalya Gaye (Dänk! Collective and IT-University of Göteborg, Sweden),  
Atau Tanaka (Culture Lab Newcastle, UK),  
Frauke Behrendt (University of Sussex, UK),  
Kristina Andersen (STEIM, The Netherlands)

### THANKS

to rector Dr. Gerald Bast and the University of applied arts Vienna

### DESIGN

Bernhard Faiss (isebuki, Austria)

Unless stated otherwise all rights remain with the authors.



Workshop

May 13 – 15, 2008

University of applied arts, Vienna

<http://MMW2008.dieangewandte.at/>

# 5<sup>th</sup> International Mobile Music Workshop

[www.mobilemusicworkshop.org](http://www.mobilemusicworkshop.org)

Keynotes • Hands-on Sessions • Performances • Papers • Posters • Demos

