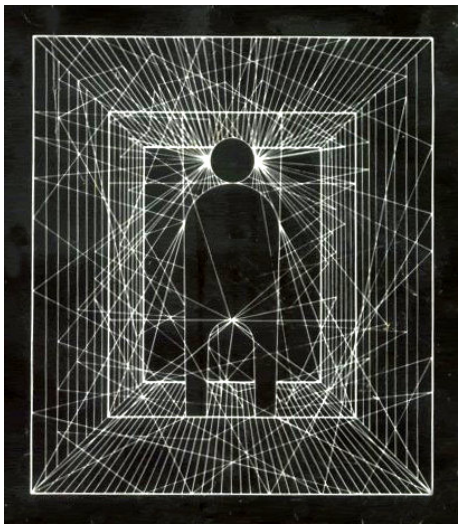
TELEFUNKER - CREATURES OF TIME AND PLACE

Sound localisation and spatialisation in a sonic environment



Project report

Oliver ’Olsen’ Wolf
Media Arts and Technology
Queen Mary University London

Placement project at IRCAM
Paris – August 2014

Dedicated to our position in the universe.

Titlepageimage: Cover from *Effects of noise on man* Kryter (1970)

Abstract

About Telefunker

This work is focusing on the usage of sound localisation and spatialisation techniques in a multi-user system employing mobile devices.

As part of the work a set-up was designed to generate playful performative interactions based on sound. The system analyses the topology of sounds appearing in the environment and controls the topology of synthesised sounds in response. The design and development aims for an easy to set-up collaborative environment based on web technology.

In a scenario two or more participants make and react to generated sounds in respect to timing and position in the set-up. The set-up is formed by the participants by placing their devices at a fixed location in a room. Once set up, participants generate a series of percussive sounds, like clapping, chirruping or snipping with the fingers in the space. These sounds are re-synthesised spatially on the devices.
Through the arrangement the participants can interact with the topological reproduction and transformation of their actions. Because of the spatialisation and the emerging rhythms, the participants are situated in a network of further interactions evolving over time.

The work entails a sample accurate audio clock synchronisation on independent clients in a web based environment to analyse the events and a multi channel sound synthesis system based on the Web Audio API.

Keywords: Sound perception, auditive localisation and spatialisation, active and passive audition, location based systems.

The Turing machine and the von Neumann computer (Rechner) were conceived as reproductions of the human calculator, not of the thinker.
—Oswald Wiener

Acknowledgements

This work was carried out during a five month intern-ship in the ISMM group at IRCAM in Paris.

I would like to thank the ISMM group for having me and in particular Gérard Assayag from IRCAM and the Media Arts and Technology program at Queen Mary University London for making this intern-ship possible.

Furthermore I would like to thank Geraint Wiggins for his supervision, Norbert Schnell for his creative contributions and enduring support, Victor Saiz for guiding me through the 36 chambers of JavaScript and Diemo Schwarz for his support and the inspiring music.

This work was funded by the Engineering and Physical Sciences Research Council (EPSRC) as part of the Centre for Doctoral Training in Media and Arts Technology at Queen Mary University of London.

Contents

List of Figures

1 Introduction

1.1 MOTIVATION

The work described in this report was carried out during a 5 month intern-ship within the *{Sound Music Movement} Interaction – ISMM* Team at the Institute de Recherche et Coordination Acoustique/Musique (IRCAM) in Paris France. The ISMM Team conducts research and development on interactive music systems, gesture and sound modeling, interactive music synthesis, gesture capture systems and motion interfaces[1].

The work itself was part of ISMMs *Collaborative Situated Media – CoSiMa* project[2]. The CoSiMa project aims at developing a platform for collaborative and collective interaction based on recent mobile and web technologies. Amongst others applications include: collective audiovisual performances, collaborative games, and interactive fictions in the framework of artistic projects.

My intention was to explore the phenomena of phonotactics and self-assembly in a compositional framework. On the one hand phonotaxis, found in natures crickets (Müller and Robert, 2001) as well as in robotics (Reeve and Webb, 2003), describes the localization and movement of an agent towards a sound source. On the other hand the process of self-assembly, exploring the effects of an agents morphology to compose a structure (Miyashita et al., 2009).
The initial idea was to combine agents acoustic localization capabil-

1 http://ismm.ircam.fr/
2 http://cosima.ircam.fr/

ities (phonotactic behaviour) with the capacity of generating sounds in relation to spatial position to other agents. This should result in a constant transformation of their (sound) morphology.
One objective was to exploit the complexity of the agents behaviour in a world or environment that is not simply available but produced by their activity. I consider this as simple cognitive architecture raising questions of artificial curiosity and its dynamics in a computer based musical context.

1.2 AIMS AND RESEARCH QUESTION

The aim of my placement within the ISMM team as part of the CoSiMa project, was to explore sound localisation and spatialisation techniques using mobile devices and to design a framework to generate performative interactions based on sound localisation.
The work carried out during the period of the project was split into an audio-analysis part on the one hand, with the objective of learning about sound localisation and in what way it can be applied to mobile devices and on the other how the devices can be used for multichannel audio-synthesis. Thereto developing scenarios generating a sound based performative interaction resulting from the topology of localized sounds by the audio-analysis.

My research question was how localisation and spatialisation of sound can be accomplished in a set-up using multiple mobile devices, in particular considering the inherent constrains of those devices: how localisation solely based on acoustic information captured by multiple devices with a single 'ear'/built-in microphone can be achieved (Audio-Analysis). And how these devices could respond collectively to this input in a meaningful way (Audio-Synthesis).

1.3 STRUCTURE OF THE REPORT

The report comprises *Related Works* to contextualise the work in the fields of psychoacoustics of humans and animals, models of spatial perception abstracted from nature found in robotics and mobile systems as well as enactive approaches to sound perception like phonotaxis and active audition.
The *Development* provides an overview of the framework and a detailed description of the set-up. Furthermore it addresses the various challenges that arose during the project and how those affected the development of the final work.
The *Conclusion* discusses the outcome in relation to the context established in the *Related Works* section. What are the limitations and what are the contributions of the work to the field of sound localisation in a multi-client set-up based on mobile devices.

2

Related Work

SOUND LOCALISATION

The aim of my literature review is to set up a framework for sound localisation. What are the defining parameters of sound perception in humans (Blauert, 1997; Moore, 2012) and animals (Klump, 1995). Furthermore how the models and frameworks evolved from the research of spatial acoustic orientation are taken into consideration and how those findings are implemented and explored in biomimetic and robotic applications as well as in mobile computing environments.
Additionally the attention is drawn onto enactive approaches of sound perception like phonotaxis and active audition.

2.1 SPATIAL ACOUSTIC ORIENTATION

"*The interactions between humans (and animals) and the world of sound is called psychoacoustics. It encompasses all studies of the perception of sound, as well as the production of speech.*"
(McGraw-Hill, 2005)

This section establishes the principles of spatial hearing defining parameters for sound localisation. The focus is the scientific study of the sound perception of humans and animals (i.e. psychoacoustics) and in particular the perception of space through sound.

2.1.1 *Principles of Spatial Sound Perception*

Localisation and Lateralisation

The position of the head, pinna and cochlea are the main tools to localize and lateralize sound. In his well cited book about the psychology of hearing, Moore (2012) determines *localisation* as the reference to judgement of direction and distance of a sound source. *Lateralization* is the apparent location of the sound source within the head.

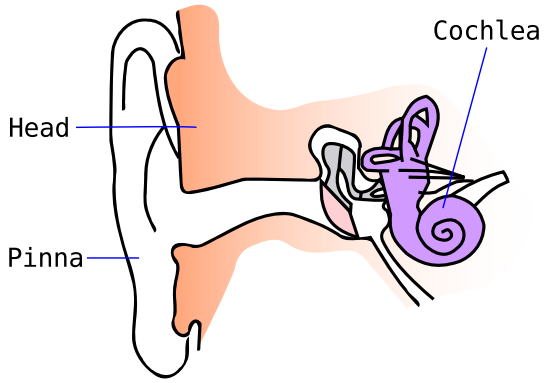


Figure 1: Anatomy of the ear (Fig. reproduced from [1])

The head and pinna together form a complex direction-dependent filter. The filter action is estimated by measuring the spectrum of the sound sources and the spectrum of the sound reaching the eardrum. The relation of those two gives what is called the head-related transfer function (HRTF).
Head movements play an important role to determine the direction of sound. Movements of the head produce changes in the spectral pattern for each ear. These changes provide cues for the extent and direction of the movement. The spectral changes can be used to judge the location of a sound source and are important for the judgement of location in the vertical direction and for front-back discrimination (See Figure 2). The cochlea decomposes the signal into frequency bands (See also video illustration[2]). And the brain uses time and intensity of the signal to determine the localisation of sound-sources (see next Section).
Situations where sound reaches both ears are referred to as Binaural but stimuli can also be observed mono-aurally (stimulus to one ear only) and it can be either the same for both ears, diotic – or different for each ear, dichotic.

As Moore (2012) concludes, our acuity in locating sounds is greatest in the horizontal plane (azimuth α), fairly good in the vertical plane (elevation δ) and poorest for distance.

1 Link to Wikimedia source

2 Video illustrating auditory transduction

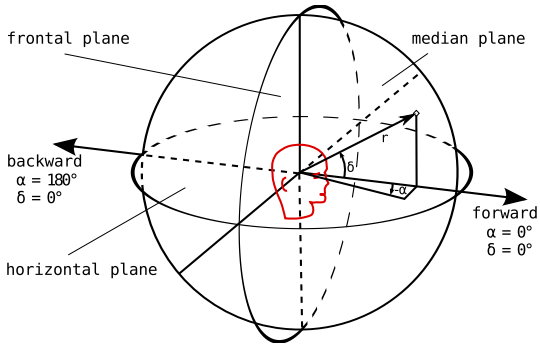


Figure 2: Coordinate system to define sound positions relative to the head (Fig. reproduced after (Moore, 2012))

Time and Intensity

The main mechanism applied by the hearing system to locate sound are time or Interaural Time Difference (ITD) and intensity or Interaural Intensity Difference (IID) of the signal: The performance and resolution of how well a subject can detect changes in localisation depends on the changes in the IID and/or the ITD, that occur when the angle of a source is changed, resulting in the Localisation Blur (LB), "*the amount of displacement of the position of the sound source that is recognised by 50 percent of experimental subjects as a change in the position of the auditory event*" (Blauert, 1997) as the establishing factor of spatial hearing.

Limits of Human Spatial Hearing

As shown by Blauert (1997) resolution for the binaural system for sinusoids on the horizontal plane is best for sounds that come from directly ahead (0° azimuth) and then the LB increases with the displacement of the sound source from the forward direction. Around 0° azimuth it varies around 1°-4° and it is becoming worst between 90°-270° with a 9°-10° deviation but improves at 180° to 5.5°.

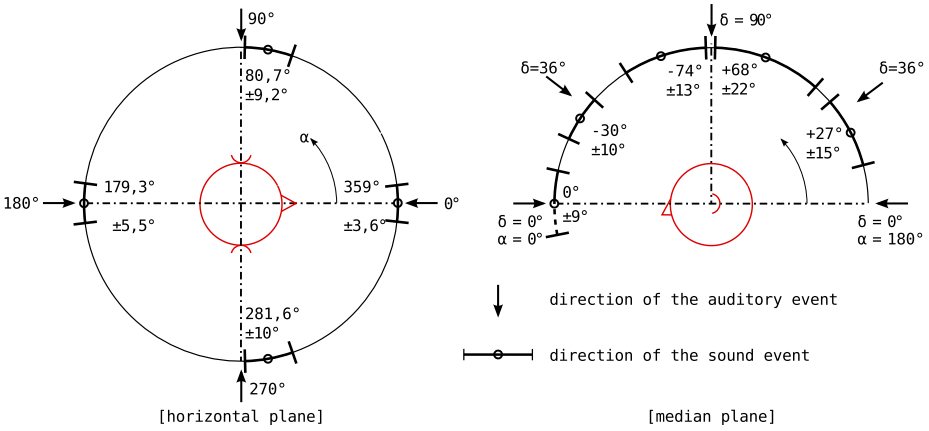


Figure 3: Spacial blurs around the human head (Fig. reproduced after (Blauert, 1997))

On the median plane the interaural phase difference information is absent, making it hard for subjects to locate sounds i.e. to differentiate if a sound is coming from ahead or behind. In the absence of IID cues,

the frequency content becomes important for the localisation process. In particular the frequency response of the outer ear, the Pinna.
The localisation blur on the median plan for speech has been established by Blauert as 17° for speech of a person unfamiliar to the listener, 9° if the person is familiar and 4° for white noise.

2.1.2 *Active Audition and Phonotaxis*

In his studies on *Spatial Hearing* Blauert (1997) already emphasized the role of the head movements to improve the ability to localize sound. This activity as a sensory-motory loop (Noë, 2004) between movement and sound was moreover studied in the enactive approaches of *Active Audition* and *Phonotaxis* that emphasize the role of the subjects motion in the perception of the environment.

Positioning information based on acoustics is successfully used in nature by animals like owls, bats, fishes, frogs, toads, crickets and flies (Gerhardt, 1995). The latter species Ormia Ochracea is due to its accuracy even considered as a world champion[3] in localising sound sources.
Insights into biological localisation capabilities are based on the studies of animals' ability to localize and move towards sound – *phonotaxis*. Phonotactic behaviour found in nature can be surprisingly accurate and persistent even without sight, as shown in experiments by Müller and Robert (2001), Mason et al. (2001) or Schöneich and Berthold (2010). The fly Ormia Ochracea is often taken as a model organism. The acoustic properties of its unusual ears allow for extraordinary direction sensitivity (Kaiser, 2003): A mechanical coupling of the two eardrums (Tympana) and a time- and population-encoding within the organ of hearing (Bulba acousticae). The time-encoding takes place through the sound intensity dependent latency-time of the neurons. The population encoding happens through the amount of enervated neurons and also dependent on the sound intensity.
Even though their organs for hearing are very close to each other (0.5mm) they can localize the host, using its 5Khz chirping as orientation, with an 2° accuracy.

2.2 EXISTING APPLICATIONS

Biological evolution e.g. vocal communication in animals, humans language and its connection to music or the effects of social organisations have not only shaped humankind's musical behaviour as depicted by Wallin et al. (2001). Inspired by the human and animal sound localisation capabilities various technologies are developed to provide directional sensing.
This section focuses on biomimetic and robotic applications, the usage of location awareness in mobile computing devices and sound-localisation as a sensory-motor activity, as found in the embodied approaches of *phonotaxis* and *active audition*.

3 'Lauschangriff: Fliegen haben die besseren Ohren' Spiegel magazine 5.4.2001

Figure 4: SAM – Sound Activated Mobile Edward Ihnatowicz (1968)

2.2.1 *Biomimetic and Robotic Applications*

How natural processes could be effective to the development of devices is described by Cariani (1993). Refering to Gordon Pasks series of electrical devices based on *organic analogues*, Cariani presents how they emerge sensory capabilities similar to natural processes and evolve sensitivity to sound or magnetic fields.
By *Building ears for robots*, Huang et al. (1997) applied and explored the perceptual sound component grouping method for sound separation to distinguish the echo free onsets, essential for human sound localisation.
The paper *A Biomimetic Apparatus for Sound-source Localization* by Handzel et al. (2003) reports an successful implementation of acoustic sensing based on a principle observed in nature. In their study they accomplish directional sensing with a biomimetic approach imitating human interaural spatial perception. This is achieved by using only two ears situated on a sphere and an algorithm modelling the head as a sphere. They show how this method outperforms a commonly used audition technique with multiple microphones placed in an array. They conclude that the biomimetic localisation around a sphere based on phase and level difference (IPD-ILD) outperforms a standard cross-correlation calculation between the microphones placed in an array. Although, they admit that – at the cost of using longer data samples – a cross-correlation based technique can perform nearly as well.

Active Audition

Active audition takes advantage of the movement capabilities of the listening platform – in comparison, passive audition is based on analysing wave phenomena of sound but also radio, radar and sonar (with static sensors).
Similar to biological systems, active audition can be almost as accurate as in humans (Berglund et al., 2008), by adaptively pinpointing the sweet spot towards a sound source in order to achieve optimal listening capabilities.
As Berglund et al. report, the term appears to originate with Reid et al. 1999, describing a system using two omnidirectional microphones on a pan and tilt-able microphone assembly. Berglund furthermore notes that active audition in synergy of visual and auditory cues has been research by the SIG group since 2000. The group aims to create a robotic receptionist capable of interacting with several humans at once in a noise office environment. This is carried out by using FFT to extract interaural phase and intensity difference from a stereo signal.
An early artistic application is the interactive sculpture SAM (1968) by the Polish Artist Edward Ihnatowicz (1968) shown in Figure 4.

Phonotaxis

Barbara Webb is exploring the "*the astonishing variety of sensory-motory tasks that species face*". She pursued various studies on phonotactic behaviour (Webb, 1995; Reeve and Webb, 2003) and how it is effected by the morphology of the auditory apparatus (Lund et al., 1997).
Building a sound-seeking robotic cricket her aim was to understand a biological sensory-motor system by examining an artificial system in a real world environment that uses sensing to control actions (phonotaxis). Exploring the navigation of the physical model under a variety of conditions, Webb illustrated the adaptivity, for example the robustness towards perturbation like noise or distortion of information.
In the experiment a subsumption architecture (Braitenberg, 1987) (see also (Brooks, 1990)) was adopted, resulting in a distributed structure without centralized control, featuring sensory transduction, neural processing and motor control. For Webb this simulates the sensory-motor conditions and adaptive behaviours found in nature. Furthermore it emphasizes the importance of physical embodiment as a solution and constraint on behaviour, here the successful autonomous exploration of an auditory scene.

2.2.2 *Location Awareness and Mobile Devices*

Similar to animals, most scenarios using sound localisation in a multi-client set-up based on mobile devices use audible chirp signals (Girod et al., 2006), Beeps (Lopes et al., 2006; Peng et al., 2007; Qiu et al., 2011) or Ping signals (Kim et al., 2014) to localise their 'mates' and use multi-modal sensing as a combination of internal or external sensors in addition to acoustic signals.

Why Sound – The Advantage of Sonic Positioning

Aside the commonly known GPS, other sensory components in mobile phones can be used to determine position. Schlienger (2012) studies on the suitability of positioning systems and Filonenko et al. (2010) work provide an overview of multi-modal sensing methods for *Indoor and Outdoor Location Based Systems*. The performance of various positioning methods are thoroughly evaluated and compared in Figure 5.

Principle	System	Accuracy	Area	Cost to User	Availability	Ubi	Costs
RF	Satellite navigation	low	global	low	market	yes	NA
	Pseudolites	medium	local	low	planned	no	high
	Ultra wide band	high	indoors	high	market	no	medium
	WLAN	low	local	very low	market	yes	low
	Wireless sensor net	medium-low	scaleable	low	market/DIY	no	low
	Bluetooth	low	20m	very low	DIY	yes	low
Inertial	Gyro/Accelerometer	0.5-20%*	1-100m	low	market/DIY	yes	low
Optical	Infrared, wii	medium	scaleable	low	market/DIY	yes	low
	MoCap, hi-end	very high	scaleable	high	market	no	high
	MoCap, low-end	medium	scaleable	medium	market	no	low
Magnetic	Magnetic field	high	1-20m	medium	market	no	medium
	Induction	NA	NA	NA	no	no	NA
Sonic	Ultrasonic	high	scaleable	medium	market	no	medium
	Acoustic tracking	high	scaleable	very low	DIY	yes	low

*No absolute measure, DIY:Do It Yourself, WLAN:Wireless Local Area Network, NA:Not Available, Ubi.:Ubiquitousness

Figure 5: Overview of positioning systems (Figure reproduced after (Schlienger and Tervo, 2014))

Comparatively, accuracy of sonic systems is among the best of the listed possibilities. Moreover a loudspeaker and microphone is part of off-the-shelf mobile phones which makes them very suitable for positioning. In comparison to sound, as indicated by Filonenko, to measure the time-of-arrival of RF signals, specialised equipment is required as RF-signals travel significantly faster.
Measuring the time-of-arrival of sound is possible with conventional mobile hardware as Borriello et al. (2005) showed already in 2005 by emitting 21Khz from a mobile phone speaker and receiving it with a conventional microphone. Peng et al. (2007) presented the possibility of utilizing sound in order to measure the distance between two mobile phones using time-of-arrival. In his *Investigating ultrasonic positioning on mobile phones* Filonenko et al. (2010) combines these two principles for trilateration of an inaudible ultrasound signal using a static microphone array.
As he deduces, the problem with ultrasound or sound in general is that it has a distance based attenuation curve, and high frequencies can be easily blocked by furniture. The sound frequencies presents a choice between efficiency and usability, anything above 8Khz attenuates too quickly on the other hand sounds above 20Khz are not audible.

As Filonenko et al. conclude, sound positioning can potentially offer sub-meter accuracy. Especially using ultrasound, as the signal needs to be as distinct as possible in order to cover long distances. Additionally it needs to be sharp and loud to resist reverberation and clearly identify time-of-arrival to be used for trilateration. Although reverberation noise and multipath can be an issue.

A better robustness to multipath conditions and reverb noises particularly found in indoor environments is achieved through auditory maps (Martinson and Schultz, 2009) or fingerprints (Pourhomayoun and Fowler, 2012). By collecting information gathered by movement and processing it through a spatial likelihood filter to compare the cross-correlation values with corresponding positional values on the pre-built (through different data collection strategies) map.

Qiu et al. (2011) use a combination of sensory data and sound. They define alignment regions to either use information from accelerometers and digital compass when the acoustic signal is distorted/out of range or acoustic information from two microphones and one speaker per phone for accuracy. This results in a localisation resolution of 13.9cm error for 90% and 4.9cm for 50% of the estimations.

Almost all approaches either use further infrastructure like external microphones and multi-modal sensing, or rely on data collection.
In Janson et al. (2010) approach, the localisation is based on ambient signals without the need of external infrastructure. In the specified application scenario they localize arbitrary devices using only the random environmental noise peaks and ellipsoid Time-Difference-of-Arrival (TDoA) method. Thus they determine the relative distances in a triangle of devices. Additionally they synchronise their clocks via network to exchange time marks. With this set-up a positioning precision on the order of 10cm is reported.
Similarly Hoflinger et al. (2012) present an *Acoustic Self-calibrating System for Indoor Smartphone Tracking* solely based on sound. Generating an acoustic chirp signal received by self-made sound receivers connected to a WiFi network. Through correlation and TDoA calculation an error margin of less than 30cm is reported.

Kim et al. (2014) focus on sonic measurement between mobile devices to discover the geometrical location of the devices in physical space. In comparison to others their 'Ping-Pong' system is carrying out the measurements in a musical way and was designed to be a musical piece in itself.
With their Ping-Pong method they report sub-metric accuracy using pairwise distance estimation through the round-trip of an audio signal. A two note sequence is sent for reconnaissance and confirmation to the receiver to estimate distance based on travel time. The pairwise distance is then converted into relative position to get the geometry of the smart-phones collocated in a shared physical space (Herrera and Kim, 2014).

Their method uses an onset and a pitch detector running in parallel and a comb-filter tuned to the specific frequency emitted by the counterpart. The distance measurement is musically the inter-onset interval between the notes on the score distributed amongst the devices. Problems addressed are the control of the intensity parameter on different devices. Louder sounds mean better signal-to-noise ratio and below some threshold, the measurement is impossible. Other problems arise from the interaction of sound and space like reverberations.

2.3 SUMMARY

In this chapter, an introduction to psychoacoustic as the interactions between humans (and animals) and the world of sound was presented. Models of spatial perception abstracted from nature with a focus on the basic parameters of perceptual models in humans and animals and their restraints where covered. Furthermore applications of sound localisation and spatialisation emerging from those biological and cognitive forces were introduced as connections between biological solutions and conventional engineering approaches.

Considering a variety of applied methods (in robotics and mobile computing environments) for a mechanical and algorithmic reproduction of location awareness, sound was presented as the preferred positioning method in comparison to other sensor modalities. Furthermore the concepts of active audition and phonotaxis were presented as enactive approaches to the design of adaptive interactive systems that rely on the benefits of sensory-motor action-perception cycles for dynamic source localisation.

3

Development

Based on the concepts presented in Chapter 2 and the aim of creating a sound-scape as a collaborative environment – using sound localisation and spatialisation techniques on mobile devices – an set-up was developed, whose design is explained in this chapter.

Before the actual set-up is illustrated in the two stages of *Audio-Analysis* and *Audio-Synthesis*, the *Context and Scenario* section describes *'Collective Sound Checks'* by the CoSiMa group as the context of this project. The *Overall Scenario* covers the set-up of the environment as well as the involved technologies. The *Implementation* unfolds to what extent the constraints of the hardware and the envisioned set-up of the system presented a challenge for the development and how those affected particular *Design Choices* of our set-up. The *Audio-Analysis* section exposes different practices of obtaining and analysing positioning information within the set-up, the *Audio-Synthesis* describes a variety of aesthetically appealing scenarios resynthesising the analysis with a focus on spatial transformation of the interaction between the devices and participants in the environment.

3.1 CONTEXT AND SCENARIO

3.1.1 *Application context - Collective Sound Checks*

The envisaged context for the application was given by *Collective Sound Checks*, a series of events that took place during the placement project at the Centre Pompidou in the framework of the CoSiMa project.
However, the idea of *Collective Sound Checks* can be put in an everyday situation as it features a casual coming together of participants to form a collaborative environment for playing and performing spontaneously with sound, using attendees smart-phones.
To participate in a *Collective Sound Check* participants do not have to install any application on their devices. Using recent browser technology with the Webaudio API provides an easy to use set-up without further technological requirements of the prospective performers.
To start performing the participants visit a web site running on a server in the same network and depending on the acoustic properties of the space a small loudspeaker is hooked up to each device.

Figure 6: *Collective Sound Check* event at Centre Pompidou[1]

3.1.2 *Overall Application Scenario - Telefunker*

Telefunker was developed in the above mentioned context as a particular scenario to explore sound localisation and spatialisation techniques on mobile devices by generating performative interactions based on sound localisation.
In the Telefunker scenario two or more participants generate and react to timing and position of sound events in a collaborative environment. To form the environment the participants visit a website instructing them to place their mobile devices at fixed locations in the room (See Figure 7). Once the environment is settled, participants generate a series of percussive sounds within this space, such as clapping, chirruping or snipping with the fingers. The percussive sounds

1 CoSiMa Collective Sound Check 'Matrix' formation at Centre Pompidou July 2014

are then analysed and synthesized – analysis to gather position information of events occurring in the space, followed by the synthesis, replaying a transformation of the topology of events based on the position information gathered through the analysis.

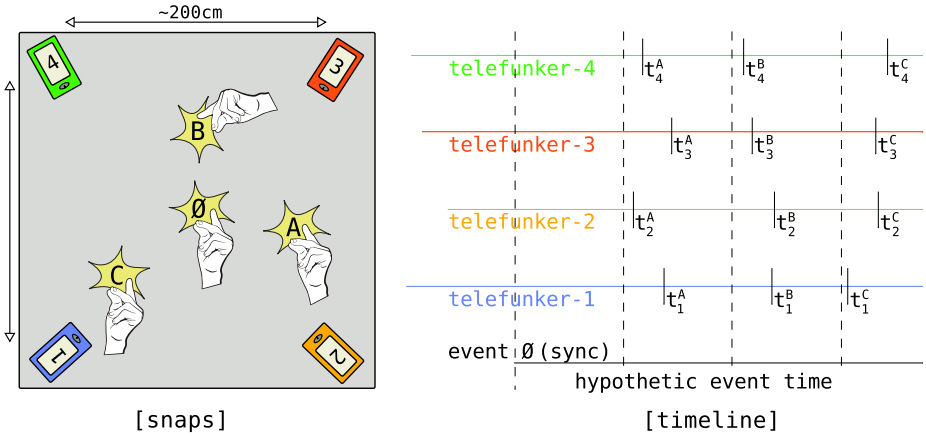


Figure 7: Set-up of the environment and events over time

3.2 IMPLEMENTATION

To fit the requirements of the CoSiMa project of providing an 'easy setup' without any further application to be installed on the participants' mobile devices and a maximum compatibility on different operating systems, the application was implemented in Javascript in a web based client/server set-up as illustrated in Figure 8.

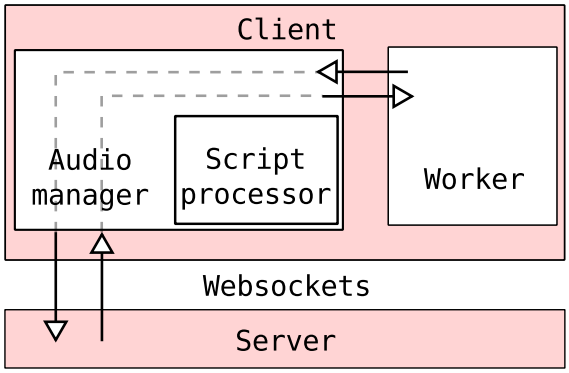


Figure 8: Client-Server work-flow

Two or more mobile web-clients communicate and synchronize via a web server[2]. For analysis, processing and synthesis of the audio signals on the clients the Web Audio API[3] is used. Webaudio API is implemented in most recent browsers and is designed to be used in conjunction with other APIs and elements on the web.
The getUserMedia() method, part of Webaudio API, enables access to media devices, in our case the microphone, through the browser. In

2 NPM command-line http server

3 W3C-Web Audio API specifications

current off-the-shelf mobile phones this feature is only enabled in Android devices.
WebAudio APIs ScriptProcessorNode interface allows the generation, processing, or analysis of the signal coming from the microphone or audio in general. WebRTC[4] provides Real-Time Communication capabilities in the browser, essential for our audio analysis.
For data exchange and to add low latency bidirectional client-server communication the independent TCP-based protocol of WebSockets is used[5].

3.2.1 *Principal Challenges*

Part of the research and development in the CoSiMa project is based on the usage of the Webaudio API. Recent technology like the Webaudio API is subject to constant changes and improvements. The usage of this upcoming web standard brought up some principle challenges in conjunction with the deployed technology and the set-up itself. These challenges are addressed here.

No Common Temporal Reference

Each mobile phone visiting the web page of the project creates an instance of a web-client starting an individual clock. As a consequence the clients do not have a common temporal reference. Hence, the difference between the measured times of an event detected between two clients can be decomposed into three components:

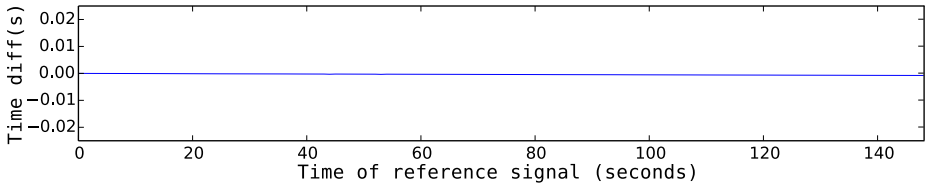
- The actual difference in travel time of the event's sound towards the two microphones.
- The overall latency (hardware & software) between the hypothetical arrival time at the microphone and the corresponding samples occurring in the ScriptProcessor input buffer of each web-client.
- The difference between the individual clocks used for measuring the event on each instance of the web-client.

Latency
The latency time is the overall system latency between the audio input and the sample occurring in the script-processor node, where the event times are estimated based on a sample accurate clock.
The overall latency can be considered identical on identical hard- and software systems and are eliminated in the calculation of time differences across different clients as well as across events on the same client (see Section 3.3 for more details).
We have estimated the overall input/output latency of our low cost mobile phone (Startaddict III) running Chrome (36.0.0) at 1.1 seconds. Assuming symmetrical processing between input and output, the input latency can be considered as half of this amount.

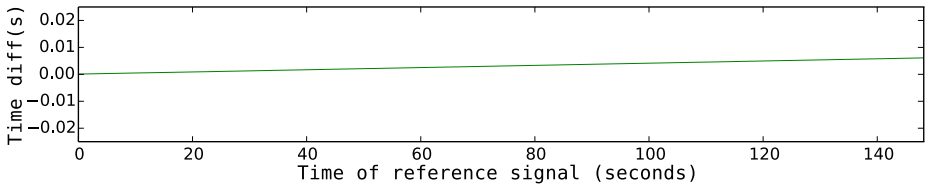
4 WebRTC website

5 W3C-Web Sockets API specifications

Synchronisation
The absence of a common clock synchronisation for the clients was an anticipated challenge. To calculate the position of a sound based on Time-of-Arrival (ToA), we not only need two ears or microphones to perceive the differences between the ToA (see Chapter 2), we also need a common clock to calculate the discrepancy between two incoming signals.
As current off-the-shelf mobile devices hold only one microphone accessible by the web-browser, at least two devices are necessary to perform sound localisation. However, each devices clock starts running individually as soon as the audio-context is established. Synchronisation of the clocks over the network is not an option due to the imprecision of wireless networks. In WLAN networks a precision of 0.1 ms is reported by Hoflinger et al. (2012) which results in an localisation error of 3.4cm.
Hence we decided to use the audio signal itself as a sync signal. With the first event, an audio stimulus like a snap or clap performed at equal distance between the devices, is used to sync the individual clocks running on each web-client. The ToA of this initial sync event is used for calibration. The difference in ToA of future events is calculated relatively to the ToA of the initial sync.
As an advantage of this method we obtained a synchronised clock on each device. However, over a longer period, individual drifts of the internal clock on each device was encountered as shown in Figures 9, 10.



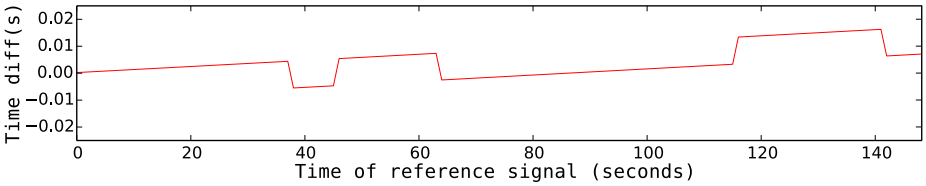
(a) Clock drift on Linux Computer



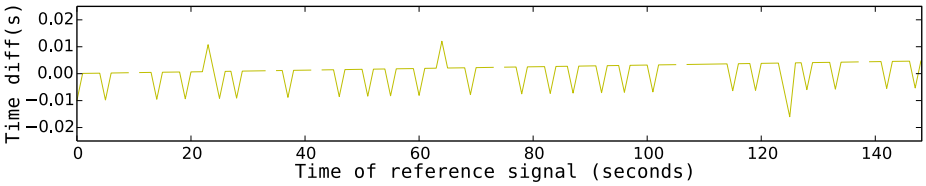
(b) Clock drift on OSX Computer

Figure 9: Measurements with unit sample reference signal on computers

The measured clock drifts shown in the Figures 9, 10 result from the following experiment: Over a time period of 150 seconds an incoming signal was measured on various devices. The signal was a unit sample played back on Audacity in one second intervals. The time of the incoming signal on the devices was measured using the maximum detection algorithm. The times of the maxima where stored locally in an array on each device. After 150 maxima the local array was reported to the server via the websocket.
With this method measurements could be done in parallel. The web-



(a) Clock drift and drop-outs on Nexus 3 mobile device



(b) Clock drift and drop-outs on Staraddict III mobile device

Figure 10: Measurements with unit sample reference signal on mobile phones

client was running simultaneously on two computers, Linux and OSX and two smart-phones, a Nexus 3 and Staraddict III. No other applications where running on the cell phones to maximize the computation power.
The outcome of those measurements are plotted in the Figures 9, 10. The y-axis is showing the differences between the measured maxima times by subtracting the reference time (0-150sec). The drifts and drop-outs (see description below in 3.3.1) we observed, gave us the evidence that it is impossible to maintain an accurate calculation of position over a longer period of time throughout different devices.

Based on this fact we discarded the idea of an absolute time and used each stimulus/event as a synchronization to the next, ending up with only having relative times, the Time-of-Arrival differences between two events. Although as a possibility to deal with this drift we took into consideration a resynchronisation triggered by the server to keep the drift negligible.

Limited Computation Power

Another challenge was the limited computation power encountered on the mobile devices available during the project. We mainly had access to three different models running an Android OS with a recent version of Google Chrome (Version 36.0). Our main test devices were several Android Staraddict III, one Samsung Google Nexus 3 and one Samsung LG Google Nexus 4. Significant performance lacks where encountered with the first two models resulting in unreliable timing of the audio input streams and drop-outs (See Fig. 10).
We have developed various strategies to cope with these limit an minimize the computation on the clients, that are described below.

3.2.2 *Design Choices*

To mitigate the challenges given in the previous section we made some careful design decisions. The two main modification are described here.
One of them was to concentrate on the interaction through percussive events. Thus the analysis of the localisation could be carried out by onset or peak detection in the signal and based on these, a time difference calculation between the clients. This event detection based on onsets or maxima furthermore entails filter properties featuring resistance to background noise.
Another design principle we agreed on was immobile devices. The set-up instructions provide a fixed position for the mobile device to be taken before starting the actual performance (Illustrated in Figure 7). Although a constant movement of the devices would have added another interesting layer in our aim of a performative interaction between the devices and users, we leave this part for future developments.

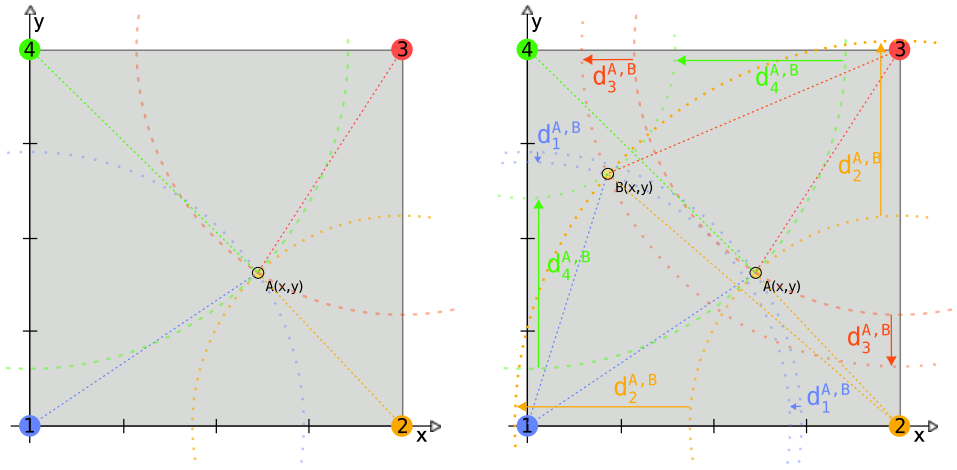
3.3 AUDIO-ANALYSIS

This section covers the analysis of events situated in the performance space. The goal was to get a topology of events, relative to the clients' position, which then can be used for *Audio-Synthesis*.
As mentioned above, for simplification, one design principle was to limit event detection to peak or onset detection of percussive events. The outcome of either of them is described in Technique 1 and 2. Within these two principles the aim was to detect a sample-accurate time difference of percussive events occurring in the signals captured by multiple devices.
This implicates a technical motivation for the following possible calculations:

- Use the first event to determine an initial position and synchronize the clocks between the clients. This allows to calculate the differences of arrival times between the clients for each following event and thus approximate positions of the events in respect to the synchronisation position. Albeit this calculation maybe inaccurate due to the drifts of the clients' clocks as addressed in 3.2.1.

- Use the difference of arrival times between successive events and adjacent clients. This allows to calculate the displacement of the event positions. In this case the drifts can be neglected due the relatively short periods.

- Use a combination of the two methods above by synchronising the first of a short sequence of events. Assuming that within the short timespan of the sequence the clock drifts are negligible.



(a) First event A (b) Second event B

Figure 11: Time differences between successive events

Figure 11 shows the differences of travel time, d_1, d_2, d_3, d_4 of two successive events A and B towards four clients set-up around the events, see also 7. The times $d_i^{A,B}$ can be defined as $T_i^B - T_i^A$, T_i^X being the travel time of the event X to the client i. More precisely the arrival time t_i^X of event x occurring in the input signal of client i and measured in respect to the clock of client i can be defined as follows:

$$t_i^X = t_0^X + T_i^X + L_i + C_{0,i} \tag{1}$$

t_0^X is the events arrival time as it occurs in a hypothetical reference time. T_i^X is the travel time of event X towards client i. L_i equals to the input latency of client i. C_{0i} is equivalent to the clock difference of client i in respect to a hypothetical reference time in which occurs the event t_0^X.
The hypothetical components t_0^X and C_{0i} vanish from the calculated differences d between the arrival times of two events measured by two different clients (1 and 2):

$$d_2^{A,B} - d_1^{A,B} = (T_2^B - T_2^A) - (T_1^B - T_1^A) = (t_2^B - t_2^A) - (t_1^B - t_1^A) \tag{2}$$

More generally the difference of the arrival time difference $D_{i,j}^{A,B}$ between two successive events A and B and two clients i and j can be calculated as follows:

$$D_{i,j}^{A,B} = d_j^{A,B} - d_i^{A,B} = (t_j^B - t_j^A) - (t_i^B - t_i^A) \tag{3}$$

which equates also:

$$D_{i,j}^{A,B} = d_{i,j}^B - d_{i,j}^A = (t_j^B - t_i^B) - (t_j^A - t_i^A) \tag{4}$$

Consequently, the time differences $d_{i,j}^X$ can be obtained as the difference of the occurrence times of an event X measured by two different clients i and j, $t_j^X - t_i^X$. We have implemented two techniques to measure these time differences.

3.3.1 *Technique 1 – Peak Detection*

In our first implementation we focused on identifying the maxima in the audio signal captured by different devices and using them as reference points. An event is triggered in the web-client. When the signal exceeded a certain threshold within a given interval and it is reported to the server.
In addition to clock drift an unreliable timing addressed in 3.2.1 we encountered various other problems such as gaps in the input signal, audio drop-outs, clipping and differences in the signal shape. This section addressed theses problems. For most of them we developed corrective actions that went into further improvements of the system.

Implementation

The peak calculation was done in the audio-processor of the web-client. The incoming samples above the threshold were held in a ring-buffer with a ring-size of 3000 samples. The ring-buffer's highest sample value was considered as the peak and its index or position in time on the web-client was sent to the server. As the different times of each instance of the web-client where synchronized with the first event in order to establish a common clock, as described above. We receive time values on the server based on the common clock plus the time of arrival which then was used to calculate the differences in position.

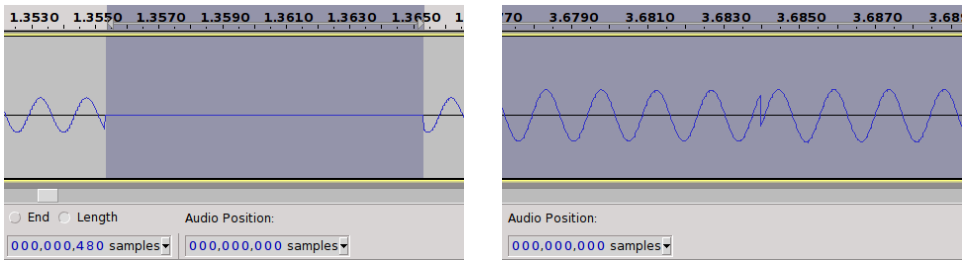
Problems Encountered

Gaps in the scriptprocessor input signal
We encountered unreliable timing of audio frames provided by the ScriptProcessor module in 'onaudioprocessing' events. Often short signal segments and even entire frames were dropped from the signal streams composed by the provided frames.
Workload on the devices, such as other applications running or intensive calculations in JavaScript seem to increase the rate of signal gaps. We traced this behaviour back to the ScriptProcessor module processing on a inappropriate priority.
To improve the performance and to obtain an appropriate threading of the processes we implemented a Web Worker[6]. Web Workers run in an isolated thread and utilize thread-like message passing to achieve parallelism. In this case the analysis of audio events was delegated to a worker.
Even though the worker helped to reduce the signal gaps the unreliable timing did not disappear and still problematic. At the time of writing 'playbackTime', which would provide an accurate time reference, was specified in the W3C-Web Audio API specifications but not implemented yet in available browsers. To bypass this problem we synthesised a reference signal (i.e. timecode) that was sent in parallel to the microphone signal on an additional input channel of the ScriptProcessor module. The reference signal consists of a counter, that starts at 0 and counts samples over one second resulting in a sawtooth signal.

6 Web Worker description

In the event analysis, gaps in the reference signal are reported to the server and taken into account by the calculation of the reference time.

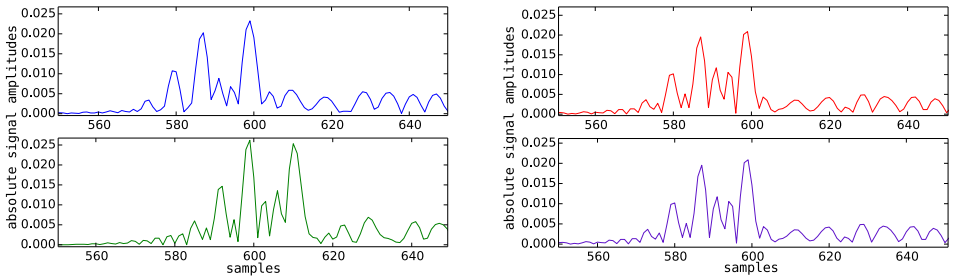
Drop-outs in the audio input
As already addressed above, audio drop-outs where encountered especially on inexpensive mobile phones (Android Staraddict III) or older generation phones (Nexus 3). This resulted in the loss of 480 samples or multiples thereof as shown in Figure 12 and also in Figures 10a, 10b. To circumnavigate the issue and to make use of our cheap phones, we inserted a hack filtering the drop-outs of the device after being reported to the server.
On a device of the latest generation, like the Nexus 4, on the other hand the drop-outs vanished. Admitting that this couldn't be tested thoroughly as there was only part-time access to this device.



(a) Dropouts marked in Audacity (b) Waveform discontinuities

Figure 12: Recorded sine wave signal on Staraddict III phones

Signal Shape
Different signal shapes of the same signal source were encountered on different devices. Especially signals like a finger-snap with more than one successive attack made it difficult to identify the right maximum amplitude.
The issue was examined by analysing an event of a finger-snap stored in the ring-buffers of four different devices. The essential parts of the ring-buffers around the event are illustrated in Figure 13.



(a) Maxima on client 1 & 2 (b) Maxima on client 3 & 4

Figure 13: Anatomy of the signal on different devices

The difficulty in detecting the maximum peak is clearly visible in Figure 13a. Note that client-1 identifies the second amplitude of the signal whereas client-2 reports the first amplitude as its peak. In this

case it resulted in a mismatch between the actual and the reported peak of 12 samples. 12 samples on a sample-rate of 44100 amount to 90cm difference in location, which was not suitable for the target application.

Clipping
In addition to differences in shape, clipping of the incoming signal appeared to be a major problem. On currently available mobile devices the input gain cannot be adjusted. According to the intensity of the signal, the A/D converter already receives a hopelessly distorted signal which makes it useless for peak detection.
Becoming aware of this, we implemented a clip detector. Each time the incoming signal exceeds a certain threshold in the web-client, the number of clipped samples are reported to the server by the clip detector through the websocket connection. This didn't solve the problem but it made us at least aware of the occurrences and its effects on the peak detection.

Results Obtained

What this all amounts to is that we considered peak detection as inappropriate for our purposes. Due to addressed problems like the timing and the signal gaps that we affiliated with the lack of computing power. This will probably be solved with the future development of technology.
However, the encountered variations in the signal shape in combination with the clipping made us to consider the outcome as unsatisfactory for our purposes. Nevertheless, within an ideal set-up with no clipping and a unit-sample signal the results were sample accurate.
A further inconvenience of this technique is due to an absolute threshold used in the peak detection. This makes it very difficult to adapt to the sound events encountered in different environments.

3.3.2 *Technique 2 – Onset Detection and Comparison*

The second implementation goal was to overcome the insufficiencies of Technique 1 (3.3.1), an onset detection was considered as appropriate to circumnavigate the issues we had with peak detection: in particular the clipping and the differences in signal shape. We took into account, that the detection of onsets (i.e. signal transients at the beginning of an event) would be unaffected by clipping and variations of the signal shapes.

Implementation

The onset calculation was done with audio-analysis by the web-client (see also (Bello et al., 2005; Brossier, 2006)). We first calculate an onset detection function as the difference between the logarithm of a short-term and long-term Root-Means-Square (RMS) of the input signal. This is implemented through two ring-buffers, one calculating

a fast varying RMS with a buffer size of 20ms and the other, a slowly varying RMS with a buffer-size of 200ms.
As shown in Figure 14, the onset detection function represents sudden changes of the signal energy. Onsets are detected as the resulting values of the onset detection function exceeding a given threshold and reported to the server.

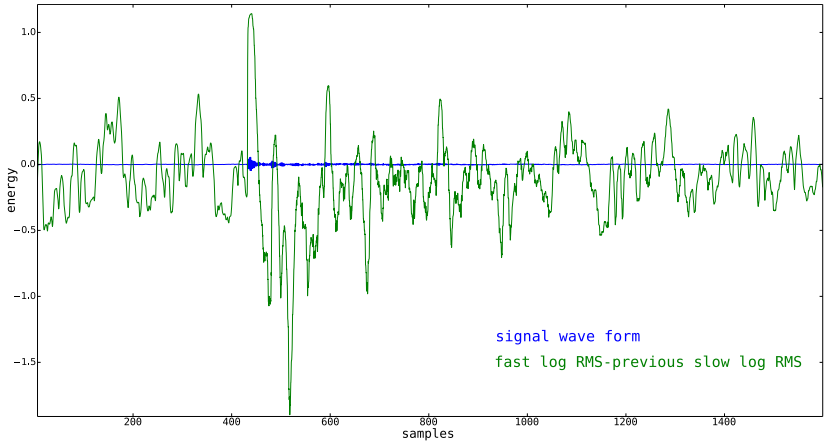


Figure 14: Signal and onset function

Problems Encountered

Signal-to-Noise Ratio
A poor signal-to-noise ratio was encountered. This issue was solved by implementing a low-pass filter in the ScriptProcessor. Filtering at a frequency of 2000Hz brought along a usable signal-to-noise ratio usable for cross-correlation and onset detection.

Cross-correlation
We got up to 10 samples difference in the detection when using the onset to measure the time of an event even with an 'ideal' signal shape like a unit sample. We traced this back to the different signal shapes across the web-client instances on different devices as observed similarly in Technique I (3.3.1). Figure 15 shows the different detected onsets as well as the various shapes of the waveform used for the cross-correlation.

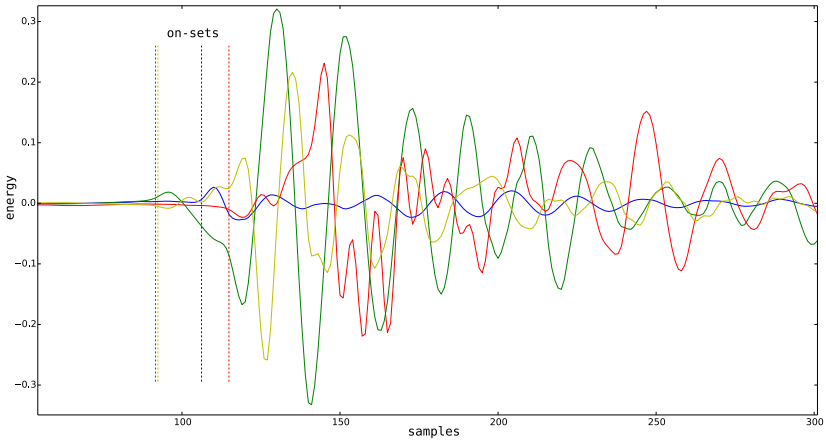


Figure 15: Samples of an event on 4 different web-client instances

As a solution we decided to combine onset detection with cross-correlation. In order to determine an exact time difference between two signal by cross-correlation each web-client sends 10ms of recorded samples (cross-correlation buffer) around a detected onset to the server. This cross-correlation buffer is composed of pre-onset and post-onset samples. It consists of 1/4 buffer length of samples from the pre-onset and 3/4 buffers length of post-onset samples. The cross-correlation buffers of an event recorded independently by each web-client are sent to the server. To calculate the cross-correlation between two web-clients, a window (cc-window) of half the length of the reference signal buffer is taken from its cc-buffer with an offset of 1/4 buffers length to include the relevant event situated at that point. This window then is correlated with each cc-buffer from the remaining clients. The maximum of the cross-correlation can be considered as the time difference of the recorded events.
With this procedure the time difference of an event occurrence on different devices can be calculated as the sum of the onset time difference and the difference resulting from the cross-correlation.

Results Obtained

With the cross-correlation based on the onsets described above, we achieved a sample-accurate time-of-arrival difference of an event on different devices running on different clocks.
Furthermore this combination makes the set-up robust to other signals like noise. Other than Technique I, the threshold doesn't apply to absolute signal amplitudes but to changes in signal energy and automatically adapt to different sound environments.
Ultimately this technique allows for calculating approximate positions and tendencies in movement and directions as well as for reproducing and transforming the topology of inter-onset times measured on the different clients.

3.3.3 *Evaluation*

The system has been constantly evaluated in a set-up consisting of 2-4 identical mobile phones placed 150cm apart in an office space. The phones run instances of the web-client and a further computer provides the server. In particular situations e.g. measurement and debugging, the phones have been replaced with computers. Furthermore at certain points little active speakers have been used for amplification.
An experimental set-up of the system with two similar computers has shown that by applying the method described in Technique II we obtained a difference of ± 100 samples within a series of events – hitting with a pencil a plastic snack-box (see Figure 16) – starting in between the computers and moving towards either side.

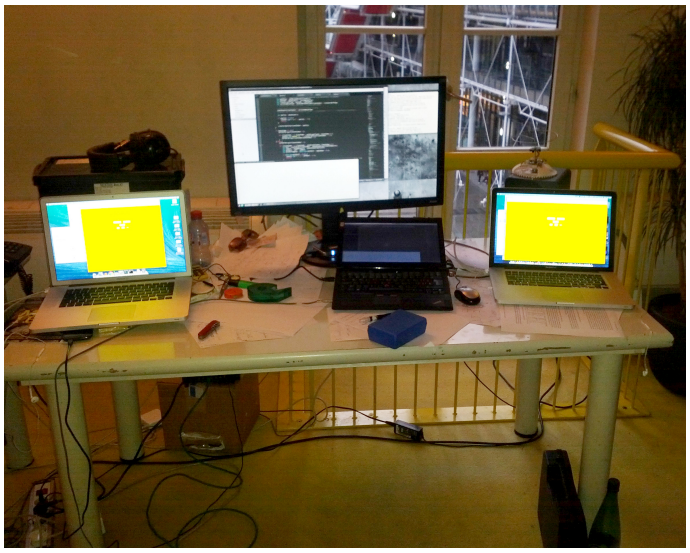


Figure 16: Test arena with the experimental set-up

3.4 AUDIO-SYNTHESIS

In order to create interactive audio applications the parameters of the audio-analysis have been used to control sound synthesis. Multiple devices provide multiple channels hence bring along the capability to spatialise and transform sounds within the environment. Based on this multi-channel set-up we have sketched and partially implemented three different scenarios.

3.4.1 *Scenario I – Sound Synthesis Control*

This scenario is based on a mapping of the movement and direction parameters from the audio-analysis to sound synthesis control. This includes synthesis and modulation of effect parameters like pitch, echo and delay.

For instance, events around an onset triggered by percussive sound like snapping, chirruping or clapping are recorded and sent to the server. On the server the recordings are analysed using the onset and cross-correlation combination described above. The position parameters calculated in the analysis are then used to control the sound-synthesis.
The sound-synthesis e.g. a subtractive synthesis from noise is performed on each client. The sound-synthesis is modulated correspondingly to the relative event position. This results in an ambient tone cluster composed by the synthesis taking place on each device.
The audio-analysis and audio-synthesis can run in parallel thanks to the filter characteristics used for the analysis – extracting event positions from the amplitudes in the signal. Thus a constant tone cluster emerges and events provoke perceptual shifts between tone patterns, which correlate sound structure and activities/behaviour within the arrangement.

This scenario is conceptually related to Pierre Schaeffer's *potentiomètre d'espace* (Schaeffer, 1967). An early electro acoustic experiment presented 1951 in Paris, where performers gestures control the dynamic level of music played from several shellac players. The performer situated within the *potentiomètre d'espace* controls the range of speed of sounds played back on four loudspeakers correspondingly to the body movement.
The scenario is also related to the drone-oriented environment *Dream House* by Marian Zazeela and La Monte Young (1966, - present). This work consists of a "*total environmental set of frequency structures in the media of sound and light*" which maintained drones of single frequencies for up to four years.
Eleh's works are also related. These tend to explore single tones, or combination of them, for example in *Black Mountain 1933* on the Album 'Floating Frequencies / Intuitive Synthesis II' (Eleh, 2007), resulting in an alluring ambient drone.

3.4.2 *Scenario II – Reproduction and Transformation*

This scenario is based on a deliberate reproduction of the topology of events. It allows to repeat a series of events based on the relative times between them. The recorded sequence can be translated (delayed) and transformed in time (rhythm and localisation).

For example a series of events are recorded on four devices via the web-client. After a specified time-out (e.g. one second) following the events, the four series are sent to the server. On the server the recordings are analysed using the onset and cross-correlation combination described above and sent back to the clients.
After a designated time of silence, a spatialised transformation of the topology of the events is synthesised on each device by playing back a sound file, for example a scissor cut spatialized on the four 'channels'/devices in respect to the obtained approximate positions and movements of the preceding series of events. However, the topology of the events are not just reproduced on the four devices. Through the transformation of temporality, i.e. scaling the time of arrivals, the rhythm and spatiality of sounds are modified in relation to the participant's actions.
Note that during the audio-synthesis the audio-analysis has to be muted to not auto-generate events. This particular scenario is been considered separately in the next section

Conceptually close to this scenario are works like *Clocker* by Alvin Lucier (1978). Based on the idea of changing the perception of time "*simply by thinking*", the rhythm of a clock is modified by changes of emotional states measured by a galvanic skin response sensor and played back through a matrix of loudspeakers.
Or $\Sigma = a = b = a + b$ by Eliane Radigue (1969) is another related piece. A double set of 7" vinyl, holding two copies of the same record are intended to be played "*séparément ou simultanément, synchrones*

our asynchrones" at different speeds – 78 RPM, 45 RPM, 33 RPM or 16 RPM to generate distinctive patterns.

3.4.3 *Scenario III – Autopoiesis*

This scenario focuses on the autopoietic[7] generation and organisation of events on the devices. After an initial trigger of events the further development of the interaction between the devices evolves autopoietically through dynamically generated events between the devices themselves, without the need of any further intervention.

In the beginning a series of percussive events are triggered in the environment. Once these events are analysed by the server, they are played back on each device analogous to Scenario II. Although in this case the web-clients are not muted during the synthesis.
As a consequence the events are analysed again already during their synthesis and immediately fed back to the server. The non-muting opens this feedback loop leading into a dynamic process of interaction between the devices themselves.
The way of transforming temporality in the synthesis here determines on the one hand the rhythm and spatiality of the sounds, on the other hand the production and regeneration of the sounds bring along the potential to develop, preserve and produce a certain behaviour over time.

Conceptually close to this scenario are works like the self-generated *Partita for Unattended Computer* by Peter Zinovieff (1968). Performed at the Queen Elizabeth Hall in London in 1968, a computer once initiated by an operator ran on stage without any human intervention, playing an unaccompanied performance of a live-generated computer composition.
Also related is Nicolas Collins (1974) *Pea Soup*. A self-stabilizing network of acoustic feedback evolving over time by nudging the feedbacks pitch to a different resonant frequency every time the feedback builds up, resulting in an "*architectural raga*", influenced by sound and movement.

7 self-creation, self-preservation - Varela et al. (1974)

4

Conclusion

In the arrangement which evolved from the work on this project we have designed and implemented a web-based multi-user system. The system can be put in place instantaneously and is easily scalable in the number of participants. It entails a sample accurate audio clock synchronisation on independent clients in a web based environment and a multi channel sound synthesis system based on the Web Audio API.

Using sound to gather position information is as an essential part of the project. Sound is beneficial on the one hand due to its accuracy in comparison to other sensor modalities available on mobile device (see Figure 5 in Chapter 2) on the other due to its ubiquitousness and simple set-up.
In comparison to other location based systems we accomplished a representative accuracy in localising events in our arrangement. As indicated in the *Audio-Analysis* section, relative position of events could be determined with sample accuracy based on the synchronisation of the audio clock and by using a combination of onset and cross-correlation algorithms. However, these findings have to be considered in regard to the ideal set-up established by the constraints we designed, i.e. the restrictions to static devices and usage of percussive sounds only, which are discussed in the following paragraph.

4.1 DISCUSSION

The constrain of devices movements might be withdrawn by the use of fingerprints (Pourhomayoun and Fowler, 2012) or auditory maps (Martinson and Schultz, 2009). However, this pre-processing would not correspond to our prerequisite of an ad-hoc and easy-to-set-up arrangement. Similarly the usage of additional infrastructure like external microphone arrays (Filonenko et al., 2010) or self made sound receivers (Hoflinger et al., 2012) were not an option.
Nevertheless, we should take into consideration the method of pairwise distance estimation for relative positions as pursued by Kim et al. (2014). It could be a possible method to overcome the restriction of movement within our framework.
Another challenge entailed in mobility could arise from the different angles of arriving signals which result in different waveforms. Whether this difficulty can be smoothed away by the onset cross-correlation combination only practice will show.

Another constraint to be discussed is the percussive sounds. In the current set-up the transients in the percussive sounds are a first approximation of detecting discontinuities in the signal. But as cross-correlation applies to any arbitrary singularity in a signal – except periodic signals – with a modified cross-correlation e.g. a lower threshold on the energy, non-percussive or continuous sounds could be analysed for discontinuities and positions determined. Albeit the sensitivity to noise has to be increased by a higher threshold to the signal in general unless the intention is a previously described autopoietic scenario.

Additionally, working with different sensor modalities and through use of alignment regions (Qiu et al., 2011) it could be possibile to scale the arrangement to larger spaces and/or outdoors. However, without any additional infrastructure the audio signal becomes unobservable at a certain distance and therefore useless for our arrangement.

The capabilities of a multi channel sound synthesis based on the Web Audio API are illustrated in the *Audio-Synthesis* (Section 3.4). They describe a variety of scenarios based on sound synthesis and their modulation on multiple devices or channels.
The different scenarios depicted in the *Audio-Synthesis* reflect on the idea of embodied interaction as a coupling of sound and movement although here not explored to maximize listening capabilities (Berglund et al., 2008) rather than to generate playful performative interactions focusing on the perception of sound through localisation, spatialisation and its lateralisation.

4.2 FUTURE WORK

The developed arrangement on the one hand provides a framework for artistic exploration of interactive composition techniques based on multiple mobile devices. On the other it encompasses the possibility of setting up a low cost localisation system using mobile phones.

Improvements concerning our approach could be made by determining the input latency of the devices. Similar to the outlined chirp, beeps or ping signal method used to localise their 'mates', an audible signal could be used to determine instantly the input latency of each device by emitting a reference sound and analysing the occurrence of this sound in the microphone input of the client. This would allow to use different models of devices in the same set-up.

The mobility is a facet of the work that needs future consideration as it adds another layer of complexity and opens further forms of interaction. Shifting the devices during the performance, engenders – in Alvin Lucier's words – *"an action or process, set into motion and sustained throughout the course of the work, producing unexpected and complex results*" (Lucier, 1998).
With mobile clients not only the topology of the participant events induced between the devices modifies the sounds, additionally each participants dislocation drives the analysis and synthesis. This set-up, modelling sounds in relation to the shift of each individual, could be used to investigate further into perception under dynamic auditory conditions and its effects in a collaborative environment.

Bibliography

Alvin Lucier (1978). *Clocker*. Lovely Music - LCD1019.

Bello, J., Daudet, L., Abdallah, S., Duxbury, C., Davies, M., and Sandler, M. (2005). A tutorial on onset detection in music signals. 13(5):1035–1047.

Berglund, E., Sitte, J., and Wyeth, G. (2008). Active audition using the parameter-less self-organising map. 24(4):401–417.

Blauert, J. (1997). *Spatial Hearing: The Psychophysics of Human Sound Localization*. MIT Press.

Borriello, G., Liu, A., Offer, T., Palistrant, C., and Sharp, R. (2005). WALRUS: Wireless acoustic location with room-level resolution using ultrasound. In *Proceedings of the 3rd International Conference on Mobile Systems, Applications, and Services (MobiSys 2005)*, page 191–203.

Braitenberg, V. (1987). *Vehicles : experiments in synthetic psychology*. Bradford books. MIT Press, 1. aufl. 2 edition.

Brooks, R. (1990). Elephants don't play chess. 6(1):3–15.

Brossier, P. (2006). *Automatic Annotation of Musical Audio for Interactive Applications*. PhD thesis, Queen Mary, University of London, London.

Cariani, P. (1993). To evolve an ear. epistemological implications of gordon pask's electrochemical devices. *Systems research*, 10(3):19–33.

Edward Ihnatowicz (1968). SAM sound activated mobile.

Eleh (2007). *Floating Frequencies/Intuitive Synthesis II*. Important Records-IMPREC158.

Eliane Radigue (1969). $\Sigma = a = b = a + b$. Not On Label.

Filonenko, V., Cullen, C., and Carswell, J. (2010). Investigating ultrasonic positioning on mobile phones. In *Indoor Positioning and Indoor Navigation (IPIN), 2010 International Conference on*, page 1–8. IEEE.

Gerhardt, H. (1995). Phonotaxis in female frogs and toads: Execution and design of experiments. In Klump, G., Dooling, R., Fay, R., and Stebbins, W., editors, *Methods in Comparative Psychoacoustics*, BioMethods, pages 209–220. Birkhäuser Basel.

Girod, L., Lukac, M., Trifa, V., and Estrin, D. (2006). The design and implementation of a self-calibrating distributed acoustic sensing platform. In *Proceedings of the 4th international conference on Embedded networked sensor systems*, page 71–84. ACM.

Handzel, A. A., Andersson, S. B., Gebremichael, M., and Krishnaprasad, P. S. (2003). A biomimetic apparatus for sound-source localization. In *Decision and Control, 2003. Proceedings. 42nd IEEE Conference on*, volume 6, pages 5879–5884. IEEE.

Herrera, J. and Kim, H. S. (2014). Ping-pong: Using smartphones to measure distances and relative positions. In *Proceedings of Meetings on Acoustics*, volume 20, page 055003. Acoustical Society of America.

Hoflinger, F., Zhang, R., Hoppe, J., Bannoura, A., Reindl, L. M., Wendeberg, J., Buhrer, M., and Schindelhauer, C. (2012). Acoustic self-calibrating system for indoor smartphone tracking (ASSIST). In *Indoor Positioning and Indoor Navigation (IPIN), 2012 International Conference on*, page 1–9. IEEE.

Huang, J., Ohnishi, N., and Sugie, N. (1997). Building ears for robots: sound localization and separation. 1(4):157–163.

Janson, T., Schindelhauer, C., and Wendeberg, J. (2010). Self-localization application for iphone using only ambient sound signals. In *Indoor Positioning and Indoor Navigation (IPIN), 2010 International Conference on*, page 1–10. IEEE.

Kaiser, F. (2003). Die mechanismen für das richtungshören bei der fliege ormia ochracea.

Kim, H. S., Herrera, J., and Wang, G. (2014). Ping-pong: Musically discovering locations. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 273–276. CCRMA.

Klump, G. M. (1995). Methods in comparative psychoacoustics.

Kryter, K. (1970). *The effects of noise on man*. Environmental sciences. Academic Press.

Lopes, C. V., Haghighat, A., Mandal, A., Givargis, T., and Baldi, P. (2006). Localization of off-the-shelf mobile devices using audible sound: Architectures, protocols and performance assessment. 10(2):38–50.

Lucier, A. (1998). Origins a form : Acoustical exploration, science and incessancy. *Leonardo Music Journal, Ghosts and Monsters: Technology and Personality in Contemporary Music*, 8(1998):5–11.

Lund, H. H., Webb, B., and Hallam, J. (1997). A robot attracted to the cricket species gryllus bimaculatus. In *4th European Conference on Artificial Life*, page 246–255.

Marian Zazeela and La Monte Young (1966). The dream house sound and light environment.

Martinson, E. and Schultz, A. (2009). Discovery of sound sources by an autonomous mobile robot. 27(3):221–237.

Mason, A. C., Oshinsky, M. L., and Hoy, R. R. (2001). Hyperacute directional hearing in a microscale auditory system. *Nature*, 410(6829):686–690.

McGraw-Hill (2005). *McGraw-Hill Concise Encyclopedia of Bioscience*. Concise Encyclopedia Series. McGraw-Hill.

Miyashita, S., Nagy, Z., Nelson, B. J., and Pfeifer, R. (2009). The influence of shape on parallel self-assembly. 11(4):643–666.

Moore, B. (2012). *An Introduction to the Psychology of Hearing*. Emerald, six edition.

Müller, P. and Robert, D. (2001). A shot in the dark: the silent quest of a free-flying phonotactic fly. 204:1039–52.

Nicolas Collins (1974). *Pea Soup*. Apestraartje.

Noë, A. (2004). *Action in perception*. Representation and mind. MIT Press.

Peng, C., Shen, G., Zhang, Y., Li, Y., and Tan, K. (2007). Beepbeep: a high accuracy acoustic ranging system using cots mobile devices. In *Proceedings of the 5th international conference on Embedded networked sensor systems*, page 1–14. ACM.

Peter Zinovieff (1968). Partita for unattended computer.

Pourhomayoun, M. and Fowler, M. (2012). Improving WLAN-based indoor mobile positioning using sparsity. In *Signals, Systems and Computers (ASILOMAR), 2012 Conference Record of the Forty Sixth Asilomar Conference on*, page 1393–1396. IEEE.

Qiu, J., Chu, D., Meng, X., and Moscibroda, T. (2011). On the feasibility of real-time phone-to-phone 3d localization. In *Proceedings of the 9th ACM Conference on Embedded Networked Sensor Systems*, page 190–203. ACM.

Reeve, R. E. and Webb, B. H. (2003). New neural circuits for robot phonotaxis. 361(1811):2245–66.

Schaeffer, P. (1967). *La musique concrète*. Presses Universitaires de Frances.

Schlienger, D. (2012). Indoors and local positioning systems for interactive and locative audio applications.

Schlienger, D. and Tervo, S. (2014). Acoustic localisation as an alternative to positioning principles in applications presented at NIME 2001-2013. *Nime 2014*.

Schöneich, S. and Berthold, H. (2010). Hyperacute directional hearing and phonotactic steering in the cricket (gryllus bimaculatus deGeer). *PLoS ONE*, 5(12):e15141.

Simon, H. (1996). *The Sciences of the Artificial*. Karl Taylor Compton lectures. MIT Press.

Varela, F., Maturana, H., and Uribe, R. (1974). Autopoiesis: The organization of living systems, its characterization and a model. *Biosystems*, 5(4):187–196.

Wallin, N., Merker, B., and Brown, S. (2001). *The Origins of Music*. A Bradford book. MIT Press.

Webb, B. (1995). Using robots to model animals: a cricket test. 16(2):117–134.