Developing ALPS – Notes on Agency in Technology

DOMINIK SCHLIENGER

92
STUDIA
MUSICA

SIBELIUS ACADEMY OF THE UNIVERSITY OF THE ARTS HELSINKI 2022

Developing ALPS – Notes on Agency in Technology

Dominik Schlienger

Developing ALPS – Notes on Agency in Technology

Sibelius Academy, University of the Arts Helsinki 2022

Supervisors: Sakari Tervo, Kalev Tiits, Jan Schacher

Pre-examiners: Sally Jane Norman, Rama Gottfried, Lily Dìaz-Kommonen

Chair/Custos: Jan Schacher

Examiners: Lily Dìaz-Kommonen, Rama Gottfried

Member of the Faculty: Alejandro Olarte

Doctoral Thesis, Applied Study Programme
MuTri Doctoral School
Sibelius Academy Of The University Of The Arts Helsinki

Cover image: Dominik Schlienger
Cover: Satu Grönlund
Printed by: Hansaprint Oy
ISBN 978-952-329-286-4 (printed)
ISSN 0788-3757 (printed)
ISBN 978-952-329-287-1 (PDF)
ISSN 2489-8155 (PDF)
Helsinki 2022

This thesis was typeset with LaTeX 2_ε
The layout is based on the Clean Thesis style by Ricardo Langner.

This project was supported by a Kone Foundation Researcher's Grant

KONE FOUNDATION

Abstract

"Developing ALPS — Notes on Agency in Technology" collects the written outcomes of my applied doctorate in music (scientific strand). This thesis is a report and a reflection on the development of tracking technology for use in interactive sonic arts. It outlines its background, methods and outcomes, a collection of articles marking milestones within the project, and a reflection on agency in technologies – digital ones in particular. Taking my own (music-)technological practice as a starting point, I scrutinise the notion of agency in human-machine interactions and relations, both in the post-humanist conception and with respect to the Gaia principle.

Chapter 1 details the project's technical background and acoustic localisation techniques, and introduces the development method, a combination of interdisciplinary free improvisation, and concepts from participatory design. Chapter 2 summarises the articles published as part of the project. Articles I and II present the project's background, a literature review, and the rationale for an implementation; articles III and IV detail application scenarios, proofs of concept, and first prototypes. Article V presents detailed requirements for the implementation, which are described in detail in Article VI. Articles VII and VIII discuss underlying issues affecting development and the use of technologies in the arts.

In my artistic practice, questions often arise about agency in artificial intelligence. In response to these concerns, I hypothesise in chapter 3 that the code-structures which constitute language, music, and computational machines are related. Consequently, machines can be conceptualised as extensions of human cognition. From the understanding of cognition as an active (motor) performance, it follows that it is the prerogative of living organisms. Chapter 4 is motivated by the question of the impact of technical objects on our surroundings. Bruno Latour's description of the technical mode of existence is based on Simondon's mode of existence of the technical object. Yet, the two descriptions differ. An aspect of Simondon's account, the directionality of everything technical towards materials, becomes lost in Latour's account. I call this appropriation the technical's modus operandi. This reading differs from the way Latour uses the term in his reading of James Lovelock's Gaia theory. Whereas Latour posits that necessary order derives from territory, I emphasise that the negotiation between all actants are precedent to order. Through this negotiation, our actions are appropriate whenever we interact consciously-technically with the world.

Contents

Acknowledgements

> *A project is an actor-network of "Gaiantic" proportions.*
>
> — **An actant 2022**

In a thesis that closes with the insight that every actant's, and actor's actions leave a trace in Gaia, acknowledgements should be the book's most substantial chapter. This is why the following acknowledgements are woefully incomplete despite overrunning the conventional length.

When Martin Robinson in 2007 suggested that I should have a look at acoustic localisation techniques for an interactive sound installation I was working on, neither he nor I had any idea that I will spend the next 15 years on a track of explorations which started right there, in the BSc Creative Music Technology Course at UWE Bristol. So my first thanks go to him, and for his patience to get me through to my MSc in Audio Production. Gratitude goes also to all the amazing and inspiring tutors, namely, composer Liz Lane for opening the world of composition to me, Tom Mitchell and Martyn Harries, whose courses are responsible for my love of technology and Stephen Allan, without whose kind generosity my career would have ended with a missed flight from Helsinki in 2011, and Marcus Lynch for showing that technology can be a tool for a better society. Praminda Caleb-Solly for making me think about the ontology of technology in the arts; Ian Holmes and Adrian Hull for letting me play in the UWE orchestra and big band. The technical staff for never despairing when I needed another 20 loudspeakers, everybody in the CMT course and lunch with Alex Pearson and Ben Marshall, preferably somewhere by the sea – fond memories!

In Finland, I would like to start with thanks to Ville Pulkki, who introduced me to Sakari Tervo in 2013 who taught me pretty much everything worth knowing about acoustics and patiently tried to decipher my creative take on Matlab during his 4 years as supervisor of the project. I appreciated the friendly and open-minded atmosphere at Aalto and the Aalto MotionLab.

Coming to my current home and academic playground, the Centre of Music and Technology of the Sibelius Academy Uniarts Helsinki, my heartfelt thanks go to everybody, starting with Kalev Tiits and Andrew Bentley, who let me in and looked after me for a good part of the project, to head of department Marianne Décoster-Taivalkoski for patiently looking after so many details without which the project would have stalled on so many occasions, Santtu Valve whose technical support is legend. The discussions in and around the doctoral seminar were inspiring and engaging thanks to the avalanche of talented creative minds. A particular mention shall go to Päivi Takala, who turned the covid-years into an intellectually stimulating time of creative reflection. Also, the MuTe lab in Vallilla played an important role, the hours of discussions with Otso Lähdeoja, Alejandro Montes de Oca and Josué Moreno were inspiring, and so many good twists and turns in the project happened there in an enjoyable, relaxed environment amongst friends.

I also would like to thank Tapio Nevanlinna. His composition course taught me to think in a much clearer way about how not just music, but the world can be thought *compositionally*.

The code for ALPS would never have seen the light of day if it was not for Victor Khashchanskiy's initiative to offer his kind collaboration. His inventiveness, experience, knowledge and commitment turned what looked like pipe dreams into a stably running application. I hope we will get the opportunity to debug some more crazy ideas together.

My very special thanks go to Jan Schacher. Within a year, he catapulted my project to an entirely different level, particularly concerning chapters 3 and 4. Even in the wildest and rawest drafts, rather than just trimming the thesis to fit requirements, he immediately picked up on where he saw me heading with it, taking charge of the forest while I dealt with trees. Further, the pre-examiners' kind and detailed reviews helped me to bring the main points into focus and anchor its less obvious claims better into their various constituting fields: my thanks go to Sally Jane Norman, Rama Gottfried and Lily Díaz-Kommonen. Also, Anne Dubos' help in linking the more theoretical parts back to my artistic practice is much appreciated, as are Mieko Kanno's insightful remarks on art and technology.

And now to the heart of the artistic research practice which accompanied me throughout these years – here comes the long list of the amazing, inspiring, fantastic free thinkers, collaborators, dreamers, activists, creators, friends and good people who I had the honour to work within the workshop of Music, Space and Interaction: It all started with the core members of the interdisciplinary improvisation research group who were active in 2012, namely James Andean, Alejandro Olarte, Jasmiina Sipilä, Hannu Tuisku and Marek Pluciennik. To come to my friends who improvised with me all the way: my thanks go to Timo Pyhälä, Sampo Pyhälä, Aino Kangas, Anna Ypyä, Salla Hakkola and Paola Livorsi. I hope we have the opportunity to continue where we left off sometime, let me know and I'll be there! Also my friends in FISQ, a spin-off from the workshop through Timo's vision, Hermanni Yli-Tepsa, Sergio Castrillon, and Saara Viika – and also Saku Mattila, for playing English Horn magically on 3.141 eπlogue.

Many thanks, again, go to *all* participants of the Music, Space and Interaction workshop – the sheer onslaught of your creativity humbles me!

Adam Bonser, Aino Kangas, Alejandro Olarte, Anatole Buccella, Anders Karl, Anna Ypyä, Antonio Garcia Ruth, Athina Drakonaki, Ava Grayson, Bastadjian Roupen, Bradley Blalok, Carlo Polo, Christina Lassheikki, Christopher Senn, Daniela Pasqual Esparza, Debi Wong, Dirk Handreke, Emilia Mariaana, Esssi Laurila Evelyne Lauwers Fatima Boix, Fikrete Miftari, Flora Bouteille, Iiris Tonttila James Andean, Janne Storm, Jari, Jari-Pekka Koho, Kaisa Kukkonen, Kalev Tiits, Janne Masalin, Joel Peguero, Joost Van Duppen, Laura Mellanen, Lukas Nowok, Maija Hymninen, Marek Pluciennik, Mark Mitchell, Maria Jaakola, Maria Lindeman, Marianne Décoster-Taivalkoski, Miguel Carazon, Moritz Cartheuser, Niklas Nybom, Otso Aho, Paola Livorsi, Reeta Puska, Salla Hakkola, Salla Vasenius, Sara Lindemann, Sampo Pyhälä, Sergio Castrillon, Stephanie Laimer, Stephen Webb, Sirpa Jokinen, Tero Hytönen, Tiina Myllarinen, Timo Pyhälä, Timo Viialainen, Tua Hakanpää, Ulla Arnold, Victor Khashchanskiy, Virpi Nieminen, Whang Zhe.

I would, finally, like to express my gratitude to my family, my parents for believing in the long haul – that it will all be alright in the end, my sisters for our long philosophical walks when we were kids and my friends for all the good times. And I would like to thank Liisa Tuomi – whom I wouldn't have blamed if she occasionally wished I pursued a career in plumbing or accounting rather, when the card bounced again at the supermarket till – her original thinking as a sociologist kept challenging my approach, allowing me to refocus the research on numerous occasions. But most importantly, she keeps me in contact with the music side of music. My thanks and love go to her and our beautiful boys, Onni and Joona, – you all had to sacrifice a lot over those years!

Preface: Making Nail Soup – Or The Nature of an Applied Study Programme

All throughout Europe, there is a tale told with many local variations, of how to make nail soup. The version I heard was from a friend who heard it from his grandmother in Scotland, and it goes thus: A thrifty woman found herself, despite her best efforts, in front of an empty cupboard one morning, looking at a rusty old nail lying in there instead of the ingredients for the day's soup. After a split second of hesitation she took the nail, and said to herself, "Nail soup, it is!" She put her hat on, grabbed a basket, and went to her neighbour's door. "I'm making nail soup today, I got the nail, but need some carrots, if you have any?" she asked when the door was answered, and soon went down the road to the next door, where she asked for an onion, and was given two, and half a leek from the next one, and even a handful of oats from the next, and a chunk of lard from the one after. When the kids came back from school, they were greeted by a lovely smell in the kitchen, and they all had a bowl of hot nourishing soup in front of them; there was even enough for seconds too. After clearing the table, she found the old nail next to the hearth. "Oops, I forgot all about that!", she said and put it back in the cupboard. "For a rainy day," she thought.

Why start a thesis with this story? - In retrospect, I realised that I was cooking nail soup over the entire duration of the project. Yes, I ventured out to develop a "Real-time Indoor/Local Positioning System Using Acoustic Source Localisation on Ubiquitous Devices". But, from the comfortable vantage of hindsight, although I did develop such a system, it seems to me as if it now could be left aside – like the proverbial nail – possibly useful to somebody else on some other day. But what makes the outcome of this project a, hopefully, substantial meal is quite possibly not the technology I developed.

Is my packaging of "Real-time Indoor/Local Positioning System Using Acoustic Source Localisation on Ubiquitous Devices" as a rusty old nail a way to hide that I have strayed off course from the initial research aim? – I guess it could be read like this, yes – if you are happy to chuck the soup down the drain and gnaw on the nail instead. The comparison is arguably misleading, as the nail has certainly developed through the course of the project, too; but it became the foundation for research questions that I could not even possibly realise I had when I started out. A deviation from the course set out in my original study plan? – Yes... maybe.

The reason why I consider this deviation from the originally stated course as legitimate is that my study plan was accepted in 2013 with a condition: “The panel states that Applied Study Programme concepts should have been used more widely in the plan. In the execution of the study, the special nature of the Applied Study Programme should be considered.”

The rules for “Applied Study Programme” in turn state that: “The thesis project includes three primary levels of outputs: The first-level outputs are of an immaterial and abstract nature. At this level, the candidate conceptualises the primary outputs of the development project (such as new views, knowledge, skills and other competencies developed during the project).” A bit like the broth: No visible ingredients, but the result of the whole cooking process.

“On the second level, the abstract outputs are articulated as functions, such as new methods, practices, recommendations or applications, which will be conceptualised in the dissertation.” Here they are, the various ingredients, like carrots and vegetables, and a pinch of salt, all added in a timely sequence to the pot like articles to a thesis.

“The third level involves the development of material for the practical implementation or demonstration of the innovation. These can be learning materials, sheet music, recordings, (...) and software. The outputs of this level are also incorporated into the final thesis or documented as part of it.” And here it is, the nail, without which none of this would have happened: Pages of C++ code and a number of Max patches, exposed to code rot on a backup-drive somewhere in a kitchen cupboard...

It looks like cooking nail soup was what I did, but perhaps also what I was meant to do.

Part I

Developing ALPS

Introduction

Music and gesture find themselves in a chicken-and-egg situation – each of the two appears to be necessary for the other. For thousands of years, this relationship was an immediate and physical one. But the development of aural traditions and musical practices, the building of musical instruments, musical notation, and – astonishingly recently – audio recording and broadcasting technology, allowed for the abstraction of the two elements: The sound source and the origin of the sound can be spatially and temporally removed, the physical gesture to create the sound veiled by a medium, be it a score, radio waves, or electromagnetism. The relationship between music and gestuality is rendered enigmatic – in some cases schismatic: In electronic music practices, minimal physical gestures, such as pressing a button, or flicking a switch is often all that is strictly necessary. So on the one hand, performers endeavour to find new ways to engage with physical gestuality. On the other hand, developments in artificial intelligence suggest the possibility of the complete virtualisation of gestuality and the ultimate disembodiment of music.

Playing with this technologically-enabled schism of sound and gesture, the artistic idea that engendered this project envisaged a Spatial Position Audio Control Environment (SPACE), wherein the position and/or the change of position of a participant controls a musical parameter or a parameter of sound. Such parameters of sound include, e.g., amplitude, pitch, duration, timbre, or overtone. Musical parameters include, e.g., melody, harmony, rhythm, texture, or expression (dynamics, tempo, articulation). When the participant changes position, some element(s) in the audio environment changes, too.

Initially, when this work was the subject of my bachelor thesis in 2009, I thought that a mobile phone's GPS could be used as the positioning tool. Once I understood that a GPS position can be wrong by several meters, and typically does not work indoors, I settled for Bluetooth signal-strength measurements: The distance between a mobile device (carried around by a participant) and a device taped to each of eight loudspeakers in a circular arrangement, indicated the panning position for the mobile device, by mapping the signal strength of the Bluetooth signal directly to the amplitude of the signal to be panned. (An approach I eventually returned to conceptually, when applying the principles of Distance Based Audio Panning

(DBAP)[59] to the ALPS Autopan patch.) The system was unstable, to put it mildly, particular in the presence of other active Bluetooth devices; that is, the majority of mobile phones. That the system worked at all during the project's assessed demonstration, is something close to a miracle.

An exploration of technologies that could yield the required update rates with the necessary precision followed, resulting in a master's thesis in 2012 that presented an overview of many possible approaches. Eventually, it transpired that Acoustic Localisation techniques (AL) could meet the requirements. However, no actual system was available at the low cost appropriate for the grass-roots type of art project I had in mind. So, I decided to develop an implementation in the form of a further research project. To generalise my specific artistic idea, the original proposal would have used only existing data on user requirements from the (at the time still ongoing) online survey. However, only weeks into the project it became clear that this was not the way I wanted to go. After encountering *Interdisciplinary Improvisation* [2], and what is more, realising its potential for technology development, an experimental artistic practice was inaugurated in which my initial idea and its application would just be one example of spatial interaction in music, the subject of the practice. The Workshop on Music, Space & Interaction (MS&I) consequently came to be.

Eventually, the question of how to develop technology from within an artistic practice became an integral part of the research project. The workshop replicated the practices established in the Research Group for Interdisciplinary Improvisation, where every improvisation was followed directly by a reflective discussion. These discussions provided data for extensive field notes. Trawling through them at various points throughout the project exposed a set of underlying research questions that unwittingly emerged or *developed* through the project's progress. This did not invalidate the initial objectives, but turned them into weather balloons – to bring forth some answers, or at least a further set of interesting questions.

In hindsight, the original objectives can be effortlessly tracked against a list of deliverables and outcomes: Mathematical models for AL tracking have been supplied, as have specification documents based on documented requirements, and the example code has been tested and runs stably. But the development process did not follow the causal iteration that the proposal implied. Despite the best intentions to go through this process step-by-step, constant shifts in focus seemed to be necessary. By observing these irregularities questions about technology development and technology itself arose. In contrast to tick-the-box type questions, these newly fledged *meta research questions* are of a more fundamental nature, and inevitably lead to more questions. After some juggling and weighing up of their relevance, these meta-research questions are:

- How to develop technology from within an artistic practice?
- Assuming that techniques and technologies enabled a schism of gesture and sound, how does this affect musical agency and the sense of agency of the performer?
- What is the role of the *digital* in this schism?
- To re-unite gesture and sound through embodiment, how much technology is necessary?

The artistic nature of the project's point of departure formed the first question. For an artist venturing into technology development, participatory design is the default approach – another path hard to imagine: beyond the etymological relation between *art, artisan* and *artefact* and the notion retained in the ancient Greek word 'technë' that comprises the root of the terms 'technique' and 'technology', mainstream art practices decisively distinguish between the processes resulting in art and those of making an instrument, particularly in music. The musician and the luthier are clearly defined career paths in related but separate fields.

The second question addresses concerns often raised by workshop participants: Improvising with electric and electronic instruments *and* acoustic instruments in a spatially interactive setting intrinsically evokes the divergence that the schism epitomises. The acoustic sound source is not spatially coincident with the sound-producing gesture. Although the motor link between the tactile sensation of *playing* an instrument is evident at least for the player themselves – particularly true for electroacoustic instruments like electric guitars, electric pianos, scratch turntables for vinyls, and so on, where the schism consists of an amplification of a (very) quiet sound, to the point where the listener does not hear a link – the divide deepens with synthesisers, where peaks of amplitude and their envelopes, pitch, timbre, and duration, are essentially unlinked from a performer's motor excitation. It is different from an electroacoustic instrument, where a physical object is made to oscillate and that oscillation will be made audible somewhere at some point, whether years later on a recording or miles away on radio or live-stream.

Analogue synthesisers are very tactile instruments, with potentiometers, faders, sliders, patch-bays, cables, and switches. Yet, the essential difference is that a synthesiser does essentially replace the embodied excitation of the acoustic oscillation with a binary decision of ON/OFF. This can be made incremental (potentiometers in form of knobs and faders) and with digital synthesisers, of course, all levels of embodied interaction can be re-introduced with many complex ON/OFF nodes representing embodied excitations via keyboards or other controllers and or interfaces. This, however, constitutes actions taken *after* an already evident schism. As much as the black-boxes of electronic music practices may become gesturally performable props, they are performed *in lieu* of a physical necessitated motoric interaction. The reunion of motoric gesture and the resulting sound remains a challenge that my project, amongst many others, aims to overcome.

Individual tactile gestuality is filtered out via a binary switch. In addition to timing, the way a switch is flipped (fast, slow, hard, soft) has no effect on the resulting connection. This fact raises questions of authenticity, or at least shifts questions of authorship: If I trigger a sound, am I the author of it? Is the maker of a handcrafted guitar the author of the sound or the person playing the instrument? Does a DJ own their material? How come a sampled loop appears original in its new environment? Is it theft?

But even beyond that, how about the algorithms doing this *thieving* for us? The answers to this, encapsulated in the third question, are urgent for our future. As suggested above and as touched upon in question four – what needs to happen if we want to overcome this schism? Is it feasible to do so? Is technology a help or a hindrance in this? Do we need to overcome it?

Starting from discrepancies between technical practices of artists and their professed technophobia observed amongst workshop participants, I became increasingly aware that my and the workshop's technical practice were entirely out of synch with the general technology discourse. Thinking about technology *per se* became increasingly central to the project: The unease about the disembodied, disenfranchising, and dehumanising dystopian concept of technology encountered in media but also in scientific descriptions and philosophy seemed always so far from the exciting, embodied, engaging, even intoxicating and exhilarating experience of *making, building, devising, inventing, practising* things, music, objects, situations, projects, events – the technical experience I hold dear, and should be able to share, but could not via the way that technology is generally conceptualised.

However, the discrepancy of this positive engagement with the world via technology and the growing heap of waste that comes with it, undoubtedly justifies the most dystopian of visions of technology's impact on the world too.

So the overarching question that I arrived at towards the end of the project was thus: Is it possible to conceptualise technology in a way that it can be *practised* as a positive engagement with the world without destroying it at the same time? *Nota bene*, the question is not a theoretical one but one which implies that we have to change our behaviour! Implied are *recommendations for implementation* of a possible *system change*.

The closing chapters are motivated by this search for an ethical technology close to experience by exploring machines as socio-material arrangements with Suchman, the technical mode of existence in Latour and Simondon, and finally, how the Gaia principle could provide a rethinking of techniques and technologies as a collective ethical engagement with the world.

In chapter 1 I start by providing an overview of the project and its background in response to the initial proposal and the therein agreed deliverables. In chapter 2 I summarise the articles and contextualise them with respect to the research's proposition on the one hand, but also in response to the meta-questions raised in this introduction, on the other. These more philosophical aspects and their ethical implications furnish the basis for the discussion in chapter 3 and chapter 4 which expand on these questions, reflecting on the nature of everything *technical*, and more generally the role of the technical in our lives and its impact on the world with a discussion of Latour's reading of James Lovelock's Gaia Principle.

1 Propositions Outcomes and Objectives

In the original research proposal, I proposed to develop a positioning and tracking system by applying acoustic localisation techniques to off-the-shelf (ubiquitous) loud-speakers and microphones for interactive and locative audio applications. In addition to the software, I designated other deliverables, such as a user requirement document and a document providing specifications for *spatially interactive and locative* applications. I framed the research questions towards the objective of the deliverables, according to common development practices:

1. How can I determine the real-time position of one acoustic receiver in relation to multiple senders using acoustic localisation techniques?

2. What level of accuracy does a positioning system need for interactive locative audio and audiovisual applications?

3. What minimum specification requirements do the ubiquitous devices need to fulfil for the acoustic localisation to work according to the requirements of the applications?

4. What algorithms can I use to implement a real-time positioning system using AL in a cross-platform system for ubiquitous devices?

The first question is solely technical. I presumed that I would find answers from controlled experiments and compare the results to similar experiments in the literature. Despite several attempts, such experiments were only really possible at a comparatively late stage in the project, once a stable implementation was available as a result of being developed in response to the fourth question. The intended causality of the proposed process needed to be thoroughly ignored – how could I know how well an implementation of AL techniques could work without having such an implementation at my disposal? – In this paradox lies an interesting question about design and the technical and authorship in engineering (see the discussion on Article VII)

The second question is based on the assumption that requirements and specifications could be formulated based on the results of my MSc-thesis, in which I gathered data with an online Lickert-type questionnaire with some additional open-ended questions. Essentially, the question implied that the available data from the survey was enough to start the development process. Only in the proposal's text detailing the third question is a "focus group of interested musicians" deemed "essential" to provide input on the development – at a stage where the *application* is already defined, based on the survey in my MSc thesis. The assumption that the data from the online survey sufficed was not entirely unqualified, as I was querying other practitioners – all of the respondents were involved in creative uses of spatially interactive technology. Yet, upon perusing the questions, it is clear in hindsight that the smallest common denominator of all the applications is too large to apply to a single, specific use-case. That became clear in the research process early on. As I explicitly determined the research method as "emergent and mixed design" [22], I was allowed a responsive and adaptive approach to challenges encountered in the development process, and I could devise a new strategy.

How did I realise the design needed to adapt at that point? Fairly pivotal in this re-orientation was a presentation and panel discussion held by the Research Group in Interdisciplinary Improvisation [2] at the international symposium on "The Cultural Memory of Sound and Space" in Turku, Finland in March 2013. It took place just weeks after I submitted my application – even before it was accepted. The title of the presentation was "Sounding Motion: Time and Space in Cross-Disciplinary Improvisation". The research group's unorthodox use of audio technology as spatially interactive objects (using loudspeakers as props, dancing with audio cables, moving upright pianos around) caught not only my eye but all my senses. However, I only gradually realised that I was looking at what I still believe to be the only possible laboratory for technology design from *within* an artistic practice -– an *experimental practice*. The research group did not define *verbatim* their practice as a development method, but essentially were not conceptually opposed to it being interpreted in this way. A fruitful collaboration ensued.

Looking for similar approaches in the literature, I found that many of my and the research group's ideas resonated with the ideas ofParticipatory Design (PD) [86]. I also realised that an experimental practice applying free improvisation to developing technology is a fruitful addition to the repertoire of existing PD tools. Consequently, the development method became an integral part of the project, and the requirement-documents and specifications became less of a means to their own ends but a running commentary on the development process.

The fourth question, as mentioned above, fused with the first in interaction with the second and third – instead of a flowchart of causal, consecutive, iterative actions we were looking at an intertwined network, when one object moved, all of the others changed their positions in complex ways.

Another implicit assumption was that whatever the requirements would be, the technology would follow, and behave according to the range of scientifically evident possibilities. This was not always the case, as I shall discuss with respect to the individual articles. One challenge was that achieving a competitive and viable implementation kept staying beyond the scope of the project, as the dependencies essentially reached into signal processing and programming at the outer limits of my capabilities. One particular reoccurring hurdle was the difficulty of accessing and processing sound on general-purpose computers in real-time: being designed for general tasks traditionally means that audio is not a priority. Many of these issues persisted for some time and were only overcome by the introduction of faster general-task computers in the last couple of years. That meant that much of the development kept being guided by what-if scenarios.

The participatory nature of the project allowed me to accept help from researcher-artists and developers so that the achieved implementation presents a state-of-the-art practical tool for spatial interaction.

The following sections supply the background and a description of the outcomes and deliverables with respect to the original proposition. From these elements arises a series of underlying meta-questions, as I outlined in the introduction. These meta-questions in turn helped in formulating the objectives (see Table 1.4) as they eventually emerged. They will be partially addressed in the discussion of the articles, and more extensively in the chapters 3 and 4.

1.1 Background

1.1.1 Acoustic Localisation Techniques

Acoustic Localisation Techniques (AL) are used to estimate the positions of acoustic signal receivers (microphones) and senders (loudspeakers or other sound sources); for example, in sonar, fish-finders, parking aids in cars, as well as medical instruments (sonography). In the case of a moving sound source (or receiver in relation to the sound source), the Doppler phenomenon is used to measure the frequency shift caused by the shortening of the wavelength during the movement. Other systems use echolocation, estimating positions by measuring the time it takes an acoustic signal to reflect from an object. This is not dissimilar to the technique I used for this project, which measures the time a signal takes to arrive at a receiver without waiting for a reflection by placing a receiver on the object.

In fact, it is the same way we estimate the distance from where lightning has struck by counting the seconds between visual lightning and audible thunder: because we know that sound typically moves at 343 m/s, we can calculate the distance by measuring the time delay of an audio signal at a receiver (ear) in comparison to its original on the sender (location of where we saw the lightning strike.) (The formula is: $d = c \times t$, where d is the distance a sound wave travels, c is the speed of the sound wave, and t is time.)

If there are several senders, and/or several receivers, the position of a sound source can be trilaterated, using the Time Difference of Arrival (TDoA) method. *Trilateration* calculates positions from distances, describing the points of intersections of spheres. To estimate the position of a moving object in 3-D space with trilateration, four synchronous distance measurements from known points are necessary. *Triangulation*, in contrast, estimates a position from known angles between objects, without any knowledge of their relative distances, and hence is intrinsically not applicable to TDoA methods.

As a sound wave's propagation through air is not dependent on frequency, audio signals of all frequencies are suitable for AL, which means that sound in the frequency range above human hearing is also an option, as long as the senders and receivers match that range. The advantage is that the measurements would be inaudible. It is also possible to apply AL techniques to estimate positions of content or ambient audio, but the high correlation of organised sound signals, like music, are arguably not ideal measurement signals.

The speed of sound through air is not dependent on the signal's amplitude, but if the signal-to-noise ratio is low and/or the attenuation too high, it might not be detected at all by the receiver, so loud signals are effectively easier to track than quiet ones.

I chose AL for the following reasons: AL works with most off-the-shelf loudspeakers and microphones; the loudspeakers used for tracking can at the same time diffuse content audio, like music or speech. Omnidirectional condenser microphones of Lavalier-type are very unobtrusive tracking devices (worn by actors on stage, musical theatre and opera), they often measure < 5mm in diameter. Finally, AL has an advantage over optical systems, which need a direct line of sight between the tracked object and the tracking device; for AL, obstructions in the path between a loudspeaker and a microphone need to be quite large, in comparison to the dimensions and directionality characteristics of the loudspeaker, before they have a detrimental effect on the signal reception.

1.1.2 Interdisciplinary Improvisation

Interdisciplinary improvisation is an experimental practice: "Bringing together practitioners of various disciplines to seek common ground, educe significant differences, and identify challenges" [1]. The Research Group on Interdisciplinary Improvisation was launched in 2012 at the University of the Arts Helsinki in order to find common ground through free improvisation between the different academies of the University, namely the academies of theatre, fine arts and music. The group's understanding of "freedom in improvisation" is based on the awareness of its relative position between maximum freedom (free, as in, for example, free jazz) and maximum constraint (scored, as in, for example, a classical sonata). This awareness also encompasses the understanding that a performer can experience freedom through constraint, as experienced in traditional jazz improvisation, where harmonic and melodic rules apply. Even at the maximum freedom end of the spectrum, a certain set of rules still tends to apply, even if it is just an agreement on where and when an improvisation takes place. The contrast between totally free and totally constrained (scored) is a continuum and to be on the very extreme end of this scale is impossible, or at least very unlikely.

The group's work highlights the possibility of merging disciplines into what could be described as a meta-discipline, where the concepts of space and time are not bound by how they are understood within a single participating discipline but how they are shared between them. To gain access to this meta-discipline a participant needs, to some extent, leave their own discipline behind. It is also important that they must leave the safe haven of their own expertise along with it. When this happens, the participants can gain an understanding of how other disciplines express themselves; they can freely explore other disciplines without being an expert and find areas of their own disciplines mirrored in the meta-discipline. This constitutes a form of *unlearning* or *deskilling* as described by Carrie Noland in "Agency and Embodiment" [75, p. 154].

This process, however, takes time; improvisational sessions are often longer than 20 minutes and can last hours. As the research group's findings assert, the length of performances and the seeming lack of expertise is challenging for an audience.

Participants from performance art backgrounds are aware of this. For them, this knowledge has direct repercussions on their improvisational practice; they are immediately tempted to perform differently if they know that there is an audience. Participants might leave things unexplored that they would otherwise have followed up on if they were only amongst themselves, rather than under the gaze of an audience, aiming to please – or arguably even to displease.

1.1.3 Participatory Design

Participatory design [86] is a design principle developed in Scandinavia in the 1970s (originally known as co-operative design). It involves all stakeholders (e.g., employees, partners, customers, citizens, end-users) actively in the design process to help ensure the result meets their needs and is usable. Participatory design is an approach focused on the processes and procedures of design. It is different from *emphatic design*, where design can be seen as the movement of researchers and developers into the world of end-users; participatory design is a movement of end-users into the world of researchers and developers. There is also a very significant difference between *user-design* and *user-centred design*; the former aims for the *emancipation* of the user, whereas user-centred design suggests that users should be taken as the centre of the design process, consulting with users heavily but not allowing users to make the decisions, nor empowering users with the tools that the experts use. This has its uses, for example, for organisations like Wikipedia, where users are given the tools necessary to make their own entries but cannot directly change the website's layout, which would be disruptive and counterproductive. But, necessarily in the case of Wikipedia, the power structures are not questioned from within the design practice. Participatory design acknowledges the existence of power structures within development practices and is thus better prepared to mediate them in the interest of designing technology not just for the people who use it, but by the people who use it.

For my project, this meant that everybody could have taken an active role in whichever field of the design process the participants wanted to be active in. And, as I define myself as a musician, I certainly saw it as an emancipatory act to start writing software and run laboratory experiments. Yet, on a macro level, I do not intend to fool myself or anybody else about the power structure within the project. Ideas that diametrically differed from my own would not have essentially moved the project in another direction. The participants were mainly students and friends, they were "helping me with my project", and as long as their interests aligned with mine they continued to participate; a power struggle was hence unlikely to occur.

This has no influence, however, on the applicability of the principles on the micro-level, on the practice in the workshop. When looking for participants for the project, no explicit mention was made of PD design principles. The participants signed up for experimental, improvisational exploration of spatially interactive technologies. In the arts university context in which the workshop took place, the majority of participants were probably artists first and technologists second. Many participants in the group did not have any technology design experience, and none of them was from a technology-design-only background. All of them were involved in artistic activities. The PD perspective thus demanded an awareness of the potential advantage of participants with technological experiences over those without it.

1.2 Outcomes

1.2.1 Workshop on Music, Space & Interaction

The workshop on Music, Space & Interaction (MS&I), was brought to life in 2013, in collaboration with some of the members of the Research Group on Interdisciplinary Improvisation, with the idea of applying the practice of Interdisciplinary Improvisation to technology development, specifically in the field of spatially interactive sonic arts [81]. It combined concepts of interdisciplinary improvisation with participatory design, where all of the stakeholders in a development project are involved at all levels [86].

The idea behind the workshop was to find unexpected, simple, and sustainable solutions by prototyping a *situation* in a problem area, rather than prototyping the solution to a problem. Although I was not aware of her work at the beginning of the workshop, Lucy Suchman proffers a blueprint for this idea in her book from 1987, "Plans and Situated Action" [91]. With her acknowledgement that plan-based action as a concept is essentially flawed, she asserts that all plans are contingent and situated. Suchman arrived at these conclusions through her research in human-computer interaction design.[1]

If this method is applied to fields outside the arts an intrinsic particularity has to be kept in mind: for example, in the article Immersive Spatial Interactivity in Sonic Arts – The Acoustic Localization Positioning System ALPS (VI) I quote a town planner who sees potential in "improvising" infrastructure like footpaths first, and

[1] In the discussion on article VII and VIII, I suggest that she defines technology itself as "situated action" [90].

then building them according to users' *situated actions* later. By improvising *a path*, a solution has already been found *in situ*. This is equivalent to *finding an alternative*. The improvisation here is hence not role-play, not pretend-as-if, but situated action or *prototyping a situation*, as I termed the approach. In applying the principles I outline here, participants, contrary to the town planner's account in the article, do not "pretend" where the bus stop and the motorway are, but they contingently make paths to reach their destination. This differs fundamentally from the type of role-play occasionally employed in participatory design practices. There is no doubt that role-play is a useful and important tool for technology development, particularly beyond the arts!

In developing of technologies for artistic practices, the relation between fiction and reality is more opaque: the arts can be understood as being situated on some meta-level with respect to reality, as there is something *fictional* at work.[2] The relation between the imaginary dimensions of arts and music, and the technical tools to realise them, means that the prototypical situation might be a fictional one. From the outside, it might be hard to grasp wether somebody is actually playing the violin or pretending to play the violin and using it as a prop. However, even if participants prefer getting wet feet from cutting across a drainage ditch, a bridge is still the better solution, even if wet feet are a contingent alternative. To come up with the suggestion to build a bridge "instead" of getting wet feet employs imagination, the fictional. Importantly, it is situated.

Why would an alternative approach to conventional development methods be desirable? After all, it does not look like existing practices do not produce outcomes. The specific situation at the outset was that I wanted the technology I developed to be of use for a spatially interactive artistic practice that did not actually exist yet. An experimental practice was a viable proposition, whereas a conventional development project would have needed to be built in a vacuum in the absence of practice. As a practitioner, I did not feel comfortable doing so.

The search for existing technologies that could be adapted to my purposes introduced me to the world of NIME's, the acronym for an international conference (New Interfaces for Musical Interaction) and community as much as a term for a type of experimental instruments. And there, traditional development practices, astonishingly, seemed to prevail. An analysis of many of the projects presented over the years show that yes, prototypes are aplenty, but *practices* are sparse. And, what

[2] I use the term "fictional" in a Latourian sense. I discuss the "mode of existence" of the fictional in more detail in the discussion on articles VII and VIII and in the respective articles themselves. In "An INquiry into the Modes of Existence". More on the fictional can be found in AIME [39, pp. 234].

is more, practises are not sustained; they evaporate as the gadgets die from disuse. Of course, not everything novel must become mainstream. But when considering the material effort going into NIME, it is surprising that only a few are granted continuity, be it in epistemic or artistic outputs beyond the initial presentation. Considering that corporeal interaction with digital sound systems seems to play an essential role in important *values* in musicianship such as authorship, authenticity and, well, musicianship *per se*, sustainable ideas based on practice should be a core competency in NIME development. Yet, many NIME projects feature in a conference and then rust away in a cupboard somewhere.

One possible reason I can see for this is the focus on specificity and novelty and the practice of *post design evaluation*. Let us think of an example: a developer/musician like myself builds a gadget with a very specific artistic idea in mind. Then they invite their friends to try it out. For a less biased opinion, they also send out a few emails around campus and get a couple of students they do not personally know to have a go as well. Afterwards, the students are asked to fill out a questionnaire and the *results* are presented in a paper. What would have been of interest to the *user*; stays opaque; the thoughts and experiences, the motivations and associations behind the development, all of the knowledge and know-how the developer/musician undoubtedly has, stays opaque – the evaluation cannot grasp the *situation*. Without sustained practice, the gadget dies.

Traditional methods might work for big commercial development projects, where a multitude of factors have an impact on the relationship between supply and demand, but for one-off developments, particularly when developed in the absence of a practice, it is hard to see how a NIME might be picked up by musicians on the strength of a description and evaluation by 4 subjects after 25 minutes of familiarisation. No instrument ever developed in a vacuum, so the search for novelty might be misguided to start with. And yet, no field is better placed than the one of NIME to gain insights into the possibilities of embodied digital musical instruments.

However, let us look at the specific way I instantiated this development method for my project. In contrast to traditional approaches to technology development, the workshop did not presuppose any given technology but tried to engage with spatial concepts through the physical spatiality of sound, the movement of the human body within space, and the sounds created by this movement. This often resulted in finding alternative solutions to what I expected. We looked for answers to a series of broad research questions:

- How do we interact musically with space in improvisation?
- How can technology increase musical spatial interaction in performances?
- What can new technology do that cannot be done by old technology?
- How does existing technology impact on spatially interactive musical practice?

We developed scores to highlight a particular aspect of spatiality in response to these questions. Some such instruction scores proved to be particularly helpful. For example: "Use no instruments or tools other than your body" or "Use the room as a found instrument" or "Start at the centre of the room and disperse while getting quieter." From these explorations, a set of notions emerged as essential to the practice:

- The notion of space as a narrative
- The space as a sounding object
- Space in time manifested as movement.

Every improvisation followed a discussion in which the participants took notes. Other methods of documentation arose from the participants' disciplines. The visual artists' outputs and audio recordings along with photographs, and stream of consciousness-style poetry, all resulted in data which, if not reflective in itself, allowed reflection upon it, clearly documenting the practice, providing a rich texture of reflective documentation.

Evidently, that which constitutes documentation in an interdisciplinary, participatory practice needs to be defined in more detail. If, say, a video artist takes part in a free improvisation workshop, can a video recording they made as part of the session be documentation? Is not the artist too involved to be impartial enough to provide an objective view? Objectivity is not a relevant factor in this regards, as the objective documentation is also bound to be in some way biased. In many ways, the situation seems not dissimilar to a psycho-physical experiment where *subjects* are exposed to a stimulus and consequently *asked* about their perceptions. We can control as many of the influencing factors of a situation as we like – to know about the *experience*, we have but to ask. Every session was followed by discussions in which I took extensive field-notes; this documentation's purpose was to supply evidence for a phenomenon, and that phenomenon was participatory, and therefore an "outside view" was neither necessary nor sensible. The artefacts that are a product of that participatory experience are the more appropriate traces of the experience.

The workshop ran at the University of the Arts Helsinki from 2014 to 2018. A fair number of the participants were students of all levels from that university. But the course was also open to non-academic participants. Generally, a third of the practitioners were from a non-academic, but professional background. The workshop took place over three to four weekends per term. Group sizes varied from 4 to 15 participants. Some participants came for a day, and some were part of almost all of the workshops from their inception. Participants ranged from musicians, composers, and dancers to scenographers, landscape architects, painters, poets, video artists, lighting-designers, sculptors, and many more.

1.2.2 Artistic Works

"**Leluhelikvartetti**" (Toy Helicopter Quartet) is a homage to Karlheinz Stockhausen's concept of the Helikopter-Streichquartett [89], wherein the players of a string quartet are placed in one helicopter each, together with a pilot and a broadcasting engineer equipped with a camera. The sound and images of the quartet are broadcast to a nearby concert venue, where the audience can hear the instruments' sounds mix with the sound of the helicopters outside and watch the musicians perform on giant screens. Leluhelikvartetti uses toy helicopters or drones instead. As the toy helicopters do not accommodate the eight personnel, some trickery is needed, to produce the impression that the individual instruments' sounds comes from the helicopters flying around the performance space. In May 2016, The Free Improvisation String Quartet, (FISQ, Hermanni Yli-Tepsa: Violin, Sergio Castrillon: Cello; Timo Pyhälä: Bass and myself on Viola), performed Leluhelikvartetti at the KlingtGut! Symposium on Sound in Hamburg. The performance received an award for "Excellence in Art, Design and Production of Sound" by the AES Student Section Hamburg. The performance space was a circular area of approximately 12 m diameter, wherein four toy drones were flown by four pilots, moving around freely. The audience surrounded this area. The players of FISQ were set up on the circumference of the performance area. Their instruments were close-miked so that each instrument was available as one mono channel to the sound system at the centre of the performance space. In the centre of the performance space stood eight near coincident radially outwards facing loudspeakers. The loudspeakers sent an acoustic measurement signal just above the frequency range audible to the human ear. (18 - 30 kHz) The four toy helicopters were equipped with wireless microphones. Using time-difference of arrival measurements by correlating the original signal on the loudspeakers with the measured signal on the helicopters, the positions were estimated in relation to the loudspeakers. The positions were then used to apply amplitude panning to the signal from the quartet's instruments, thus spatialising the quartet's sound as if each instrument was playing from one of the helicopters. (That is, for the audience surrounding the performance space.) Further, the musicians of FISQ were also equipped with wireless audio senders, allowing them to move around freely during the performance. The multiple layers of audio – direct sound from the quartet; amplified sound through the loudspeaker array; the sound of the helicopters – and the layers of movement – the helicopters' trajectories; the musicians' trajectories through the audience – created a densely woven spatial narrative.

Apparatus for (Intergalactic) Radiophonic Propulsion was another artistic spin-off, a performance for "Folies Sonores", part of the Sound-Theatre course taught by Andrew Bentley at Uniarts Helsinki in 2015. It featured wireless loudspeakers integrated into gym balls, receiving synchronised audio via Bluetooth.The balls were adapted according to a design by Sampo Pyhälä. "Apparatus for (Intergalactic) Radiophonic Propulsion" featured Dirk Handreke and myself in an improvised choreography, rolling, carrying and throwing the gym balls around a performance area, spatialising pre-recorded sound by doing so.

3.141 is a piece for surround-sound dome, fixed media, cello and cor-anglais and two wireless loudspeakers. The audience is situated inside and around the Dome, and the players are outside of it. If the players feel like it during the performance, they can change their location and play from a different spot around the dome, and also further away or closer to the audience and or the other player. The two musicians are stationary but their instruments' sound is diffused via wireless loudspeakers handed to the audience to pass around. This should help to meld the acoustic with the acousmatic sounds and also adds a spatio-dynamic layer of interaction to the performance.

The two chords of the fixed media (14 pitches) are spatialised in two clusters on points on a sphere consisting of 32 locations generated by a Matlab script. In the Dome, a hemispherical arrangement of 16 loudspeakers represents 16 of those locations at any point in time. Using vector base amplitude panning (Matlab script), the two clusters travel across the dome's hemisphere like constellations of stars travel across the sky, and continue through the "dark side" of the Dome, the non-audible 16 positions not represented by loudspeakers, but calculated by the script.

Further, the rotating axis of this rising and setting motion itself travels across the dome. The pitch constellations' appearance thus travels on a two-fold trajectory: they "rise and set", but the ascendant and descendant locations change with the change of the rotational axis. From a particular point inside the dome – where the audience is – the spatialisation itself creates a narrative, as the order of pitches is reorganised on the timeline.

1.2.3 Reflection on Agency and Technology

Chapter 3 and chapter 4 compare and tentatively synthesise the problems and conclusions put forth in the articles. They go a fair bit beyond what was possible in the individual articles. The reason for this is that the theoretical works I employ in my argumentation could not have served as a background of individual journal articles without exceeding their limits. As they introduce substantially new findings not represented fully in the articles, I consider them as outcomes of the research in their own rights.

In chapter 3, I argue that machines as socio-material arrangements are substantially constituted by human cognition. My reasoning is based on the similarity between written natural language, musical notation, and computer languages, and their similar code-structures. The question of responsibility for technologies arises in view of how highly invested humans are in machines.

In chapter 4, I develop the concept of *appropriation as the technical's modus operandi*, based on the differences in the definitions of the mode of the technical in Bruno Latour's "An Inquiry into the Modes of Existence" [39] and in Gilbert Simondon's "Mode of Existence of Technical Objects " [84, 85]. If appropriation describes the way actants in Gaia act, we might be looking at a possible way to become active ourselves in developing new technologies which are sustainable. In this sense, the participatory development practice of the MS&I workshop is an implementation of the Gaia principle.

1.3 Deliverables: The ALPS Algorithm, Software and Max Patches

The project's software deliverables consist of two components. One is the tracking software ALPS[3] written in C++, and the other is the audio spatialisation software Autopan,[4] written in Max. Let us first have a look at the algorithm behind ALPS.

1.3.1 The ALPS Algorithm

Using the principle of AL, a measurement signal (e.g., band-limited white noise above the audible frequency band) is played on a loudspeaker and compared to an audio recording of it, made on a mic at a distance d. A sampling frequency and a window length are chosen in samples. The larger the window, the larger the area covered by the system, but the longer it will take to obtain a measurement. Each sample has an index number, sequentially from the first to the last sample of a window. The next step is to take a window's length of the measurement signal and calculate its correlation with the signal of the same length, recorded on the mic, at the same time. In the resulting correlation signal, the sample with the highest amplitude is found at the index number corresponding to the delay between the measurement signal and the signal recorded on the mic. According to the relation ($d = c \times t$), we can estimate the distance between the mic and the loudspeaker.

The data acquisition records the signal for the length of time of the window, while it computes the previous window's correlation signal. If this computation lasts longer than the acquisition of the (following) length of audio, it waits with the acquisition of the next length until the completion of the computation (see Figure 1.1). If the computation takes less time than the acquisition time, the system runs in quasi-real-time. As ALPS uses pulsed signals, stepping through the loudspeakers one by one, the effective latency of the system is given by the length of the pulse multiplied by the number of loudspeakers, even when processing is achieved in quasi-real-time.

Additional latency needs to be accounted for, introduced by concurrent processes on multi-purpose computers, which tend to vary over time and make measurements unreliable. The following simple solution was applied [71]: by physically connecting one output of the sound card to a reference input on the same card, Round Trip

[3] The Software is available from: https://github.com/spatmus/alps

[4] The Max patches are available from: https://doi.org/10.5281/zenodo.5607121

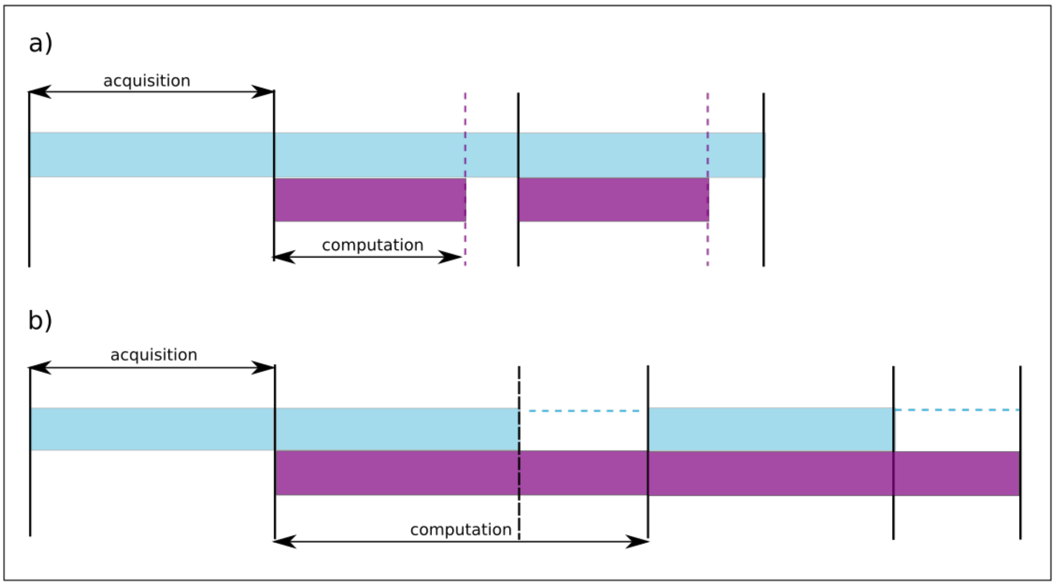


Figure 1.1.: Acquisition: a) Computation is faster than acquisition, computation waits for the end of acquisition: process is quasi–real time. b) Computation lasts longer than acquisition, acquisition waits for completion of computation.

Latency (RTL) is measured for every window, covering all delays due to the analog-digital conversion, operating system, and concurrent processes on the computer. ALPS deducts the RTL from the length of the acquisition signal for every window, so that the remaining length consistently corresponds to the time it took the sound to travel from the loudspeaker to the mic.

1.3.2 ALPS Software

ALPS was implemented as a stand-alone application in C++ based on the Proof of Concept (PoC) presented in the article Acoustic Localisation for Spatial Reproduction of Moving Sound Source: Application Scenarios & Proof of Concept (III). It was coded by Viktor Khashchanskiy for this project. During development, my role switched continuously between peer programmer, producer, and tester, along with the participants in MS&I. The newest version, al-Qt, relies on multi-processor architecture and was tested for macOS 11.6, and supplies update rates and latency for quasi-realtime applications.[5] ALPS can be applied to tracking or positioning in situations where loudspeakers and or microphones are readily available. (e.g. surround sound systems, virtual reality (VR) applications, conferencing, live sound, and home theatre).

[5]A precompiled binary of al-Qt tested for macOS 10.12 and macOS 11.6 is available from: https://github.com/spatmus/alps/releases

The ALPS software can be controlled via a configuration file or a graphical user interface. As a stand-alone application for 3-D-tracking, ALPS calculates the Euclidean distances for possible pairs of delayed signals. The positions of every loudspeaker need to be entered as X/Y/Z Cartesian coordinates. However, when using ALPS with the 'Autopan' Max Patch, the 3-D position estimates from ALPS are not essential; distance readings alone (1-D) are sufficient.

For most general positioning or tracking tasks, the sensible thing to do is to estimate Cartesian positions in 3D. For audio panning with Distance Based Audio Panning (DBAP)[59], it is not necessary, as the distance between a tracked device (microphone) and a loudspeaker is directly proportional to the amplitude. This stands in contrast to angle-based panning paradigms like Vector base Amplitude Panning (VbAP) [79], [6] or Ambisonics [63]. To obtain the required panning positions in Cartesian or polar coordinates three known points are needed for trilateration in 2-D and four in 3-D. With DBAP, the absence of a fourth, third, or even second measurement makes the situation less ideal, but not entirely undefined. Consequently, less hardware in the form of loudspeakers is necessary when using DBAP.

For More detail, please see the article, Immersive Spatial Interactivity in Sonic Arts – The Acoustic Localization Positioning System ALPS (VI).

1.3.3 Autopan Max Patch

The Autopan Max patch utilises the data produced by the ALPS software to generate a panning trajectory in a multi-loudspeaker set-up. In this way, virtual sound sources are turned into spatially discrete sound objects that can change position, like most acoustic instruments. With the help of DBAP, the Autopan Max patch also works in conditions where only one distance can be estimated, as that distance is proportional to the amplitude responding to that distance. In such a situation, the amplitude of that one loudspeaker is an approximation, as the corresponding amplitude for the single distance reading also corresponds to all of the points radially surrounding the loudspeaker at that distance. The cue is only wrong for listeners positioned between the performer and the loudspeaker. Yet even then, as the loudspeaker in question is most likely the one closest to the performer, it is still the best approximation.[7]

[6] VbAP is Ville Pulkki's preference, as it refers to Vector-basis Amplitude Panning.

[7] More details on how the patch works can be found in the article Immersive Spatial Interactivity in Sonic Arts – The Acoustic Localization Positioning System ALPS (VI), and the Autopan - Max Patch and associated sub-patches themselves (tested for Max 7 and Max 8) as well as a standalone Max application that should work out of the box on all platforms are available from: https://doi.org/10.5281/zenodo.5607121

1.4 Objectives

As I considered my research method to be of "emergent and mixed design" [22], I took the liberty to work with provisional objectives for as long as possible. In the beginning they might have been straightforward answers to more straightforward research questions, or the deliverables *per se*, in form of the ALPS software, artistic outputs, and reflective articles. I felt that this approach was also in keeping with the particular demands of the applied study program.

As the project progressed, underlying themes of a more theoretical nature emerged through the question of agency in technology, particularly in AI,[8] leading eventually to the fundamental ethical question of human responsibility for technology. I established substantive links throughout this thesis between the research and the artistic development practice which, itself, emerged from earlier (provisional) objectives. But to claim that I set out in 2013 (or 2007, if I take the original artistic idea as a starting point), to write about the role of technical agency in Gaia, would need argumentative acrobatics which are unlikely to be convincing. To nevertheless show the continuity of my argumentation and how they led towards the final objectives, it might help to consider the later set of research questions (in column 1 of Table 1.4) and how they led to the eventual objectives (third columns of Table 1.4):

Research Question	Initial Objective	Eventual Objective
How to develop technology from within an artistic practice?	Development of experimental artistic practice, applying interdisciplinary improvisation and participatory design principles	Conceptualisation of participatory practices as actor-networks that constitute machines as socio-material arrangements
Assuming that techniques and technologies enabled a schism between gesture and sound, how does this affect musical agency and the sense of agency of the performer?	Development of techniques and technology to support the sense of embodiment otherwise lacking in digital technology	Analysis of roles of agency, motor cognition, automation, and embodiment in techniques and technologies
What is the role of the digital in this schism?	Analysis of digital practices and comparison to theories of technology, embodiment, and cognition	Reconceptualisation of technology to account for human investment in machines as socio-material arrangements. Ethical concerns in view of the evident investment of humans in machines
To re-unite gesture and sound through embodiment, how much technology is necessary?	Development of technology only using pre-existing technology	Ethical considerations of human impact on the environment through technology if the Gaia principle applies to human collectives

Table 1.1.: Objectives

[8] I never intended to make "AI" part of my research, but the thematic emerged through input from workshop participants and in response to peer review feedback on my thoughts on agency in technology expressed in earlier papers. It seemed that nobody wanted to think about technology without reference to implications of AI, so I had to react.

2 Discussion of the Articles

I will next discuss the eight peer-reviewed articles constituting this thesis – five published in proceedings of international conferences, three in international journals – roughly in chronological order of publication. Articles I and II present the project's background, literature review, and the rationale for implementation; articles III and IV detail the application scenarios, proofs of concept, and first prototypes. Article V presents detailed requirements for implementation. It also thematises to a higher degree than the previous papers the meta research questions formulated in the Introduction. These themes represent the development for technology in the arts through a discussion of kinaesthesia, embodiment and agency. I discuss both subjects in more detail in articles VII and VIII, respectively. In the most technical of the papers, article VI, I describe the implementation in detail. The underlying meta themes are not excluded, but are implicitly present.

Since articles I-V are contributions to conferences, I include a short introduction to each conference in order to contextualise the articles with respect to the thematic call of the conference, and how it may have had a bearing on my submission.

Because all of the articles had to be accessible to readers from outside the field of research and constitute self-contained works, every article had to supply the context and background it refers to, despite this largely being the same for all of the articles. The general background also had to be laid out, which I undertook here in the previous chapter. In the following summaries I have therefore avoided summarising the background again, except for when an aspect of the background imparts an essentially different focus on the subject matter. The discussion here aims to flesh out the meta-questions presented in the Introduction and to lead towards the discussions in chapter 3 and 4.

2.1 Acoustic Localisation Techniques for Interactive and Locative Audio Applications (I)

This article was presented at the Locus Sonus Symposium 8 on Audio Mobility, at the École supérieure d'Art in Aix-En-Provence, France in 2014. The conference was run by the Locus Sonus research group, whose "main aim is to explore the ever-evolving relationship between sound, place and usage"[1] Artistic Experimentation is at the core of the research at Locus Sonus, and artistic productions are sometimes outcomes of this research. "We study emerging technologies but also the ways in which they modify existing practice." The Audio Mobility series of the symposium focus on the perception of sound space when virtual, mobile and acoustic spaces overlap."

The motivation behind this article was to determine the feasibility of acoustic localisation techniques (AL) for the field of spatially interactive sonic arts based on the literature available at the time. It compares the possibilities of AL with optical tracking systems using cameras, and Motion Capture technologies (MoCap) in particular. The argument is that the absence of existing AL systems in the field of sonic arts is at odds with the principle's proven and documented potential when compared to the technologies generally applied in the field.

To provide a background on the artistic applications of tracking technology in the sonic arts, I linked the modern concepts of spatial interaction in music to the spatialisation strategies of early *electroacoustic music*. In particular, firstly to the potentiomètre d'espace designed by Jacques Poullin for the Groupe de Recherche de Musique Concrète with Pierre Schaeffer and Pierre Henry, which allowed control of where within an auditorium a sound could be heard by gesturally moving a magnet within an inductive field. Secondly, to Karlheinz Stockhausen's rotation table, which was developed for *Kontakte*. It consisted of a directional loudspeaker mounted on the table. Surrounded by four microphones, it allowed manual quadrophonic spatialisation of the signal via the loudspeaker. Looking at these examples from a modern viewpoint, they are spatially interactive, gestural performances of electronic music, perhaps even examples of the embodiment of electrically abstracted music. yet, interestingly, not on purpose. Arguably, if digitally controlled automation for mixing consoles was available at the time, these gestural spatialisation schemes would never have been developed. It was primarily a solution to the technical problem of spatialisation in multi-loudspeaker setups prior to automation.

[1] https://locusonus.org/wiki/index.php?page=About.en

The article also introduced the term *acoustic localisation and positioning system*, which, with hindsight, is not an accurate term, as it assumes the terms *localisation* and *positioning* to be syntactically synonymous. Yet, the term acoustic localisation refers to the means of positioning – so a better term is Acoustic Localisation Positioning System (ALPS) which I used in later articles, simplifying comparisons with other systems.

I made claims with respect to AL's advantages over optical means of tracking. Cameras do, indeed, register movements of objects closer to the lens with higher resolution than movements of the same object at a distance. The claim that AL techniques, in contrast, measure the change of position pro-rata to the movement stands up to scrutiny; if an object changes its position by five centimetres, this is the change of position that will be measured by the system, independent of the distance to the sensor. That is, if it is registered *at all* at the sensor.

However, I failed to mention, due to a lack of practical experience at the time, that the measurement signal itself attenuates 6dB for every doubling of distance, deteriorating the signal-to-noise ratio. The loss of information with increasing distance in optics is a linear relationship (depending on the shape of the lens, i.e., its focal length), while acoustic attenuation is logarithmic, so a comparison of the two principles in respect to signal-to-noise ratios in a variety of scenarios would be very interesting, if not quite straightforward. At the time of writing, I am not aware of any studies that explicitly researched this for sensors used in positioning systems.

2.2 Acoustic Localisation as an Alternative to Positioning Principles in Applications presented at NIME 2001-2013 (II)

The NIME conference started out as a workshop at the Conference on Human Factors in Computing Systems (CHI) in 2001. Since then, these conferences have been held annually, hosted by research groups dedicated to interface design, human-computer interaction, and computer music. In 2014, the first time I attended, it was held at the Goldsmith University of London, with enjoyable lively discussions, both in and between sessions, a great array of interesting demonstrations and poster sessions, as well as musical performances and installations.

Similarly to article I provides a background and rationale for choosing AL as an alternative to other principles for tracking applications in spatially interactive sonic arts. It focuses on the proceeding's catalogue of the international conference on New Interfaces for Musical Expression (NIME) for the years 2001-2013.

For my master's thesis – a precursor to this project reviewing tracking and indoor positioning technology for use in sonic arts, I focused on the proceedings of the international conferences on Indoor positioning and Indoor Navigation (IPIN)[2] and Ubiquituous Positioning Indoor Navigation and Location-Based Services (UPINLBS).[3] That the NIME conference somehow went under my radar was pointed out to me at the time. However, when duly remediating this for the doctoral project with its focus on AL technologies, it became evident that although many NIME projects indeed use tracking and positioning technology, systems using AL were absent. Consequently, I identified projects presented at NIME where AL could have provided feasible alternatives. Again, with hindsight, I believe the yield of possible contenders would be decisively smaller if I reviewed these projects now, after some years of practical experience and gaining a more realistic view of how theoretical possibilities translate to actual implementations. Notwithstanding this, seven years later I can confirm that the absence of AL in the practices presented at NIME's is not due to any shortcomings in the principle, and that the principle bears out in practice.

In a short overview of positioning systems, I typed them into systems providing firstly, absolute data in relation to a reference grid – like GPS; secondly, relative data, e.g. the output of an accelerometer or other inertial measurement unit, or, thirdly,

[2]https://ieeexplore.ieee.org/xpl/conhome/1800208/all-proceedings

[3]https://ieeexplore.ieee.org/xpl/conhome/6400451/proceeding

symbolic data, for example the fact that a mobile phone is within the reception range of a receiving mobile mast, or the statement that somebody is at home or at work. The three types rarely occur in pure form. For example, if a phone signal is being picked up by several masts' *relative positions*, its position can be narrowed down as the masts are in known locations on an absolute grid. In the article's discussion on the advantages of AL, the claim that a signal that is already part of an interactive application could be used directly as a measurement signal, needs relativisation because this is only the case if no correlated signals are present that could render ambiguous cues. This is rarely the case for most musical sounds.

In the section on user requirements, I discussed results from an online survey on indoor and local positioning technology that was ongoing at the time, as well as the outcomes of a focus group discussion I held in 2009 at the Bristol Pervasive Media studio. Both brought forward data on general requirements.

At the time, I deferred to the need to adhere to the principles of ubiquitous computing, linking my goals to Mark Weiser's assertion that "ubiquitous computers must know where they are" [99] and for the "technology to stay out of the way of the task" [98]. My own conception of ubiquitous technology has since changed, through Dourish and Bell's "Divining a Digital Future" [20], wherein ubiquitous technology is redefined through the concept of infrastructure; I would interpret the focus group's findings differently today with respect to the idea that the difference between interfaces and musical instruments is the grade of ubiquity. It would be interesting to revisit this question with the performative nature of technology in mind (see [90]); if invisibility is a *habit* and a *practice*, how does this relate to musical instruments that are generally *performed*?

Looking for projects using AL in the 1100 NIME conference contributions from 2001-2012 yielded surprising results: a single paper used AL, in the form of Doppler [83]. So the question arose, how many projects could have made use of AL instead of whichever technology was chosen. Extending the search terms to *tracking, tracker, locative, localisation, positioning, position, motion, moCap, gestural, gesture, 3D, space* and *spatial* yielded 110 projects using tracking technology in some form or the other.

If the character of an application was more of an instrument than an interface it was discounted from the list – a distinction I would probably dispense with today. I claimed then that instruments have an "idiosyncratic character" and interfaces do not. However, wether an instrument or an interface has an idiosyncratic character is not really a question of its materiality, but more of the way it is performed.

Probably the paper's strongest argument, is that in all 110 spatially interactive systems examined, airborne sound is explicitly part of the application i.e., multiple loudspeakers are already part of the system. In that sense AL could present alternatives, in many cases, at a lower cost.

2.3 Acoustic Localisation for Spatial Reproduction of Moving Sound Source: Application Scenarios & Proof of Concept (III)

NIME 2016 took place in Brisbane, Australia. I was able to attend despite its location on the other side of the globe. There were inspiring talks and poster presentations, such as an unforgettable, hands-on workshop in the 'unconference' introducing the recently released Bela board. I started my presentation with a live demo of the paper's proof of concept, using a pair of Hi-Fi loudspeakers and, auto-panning my voice across the stage from one loudspeaker to the other worked glitch-free.

The article's core theme, application scenarios for the real-time panning of a moving sound source, defines four such scenarios – expanded on in later papers but concisely defined here. They are: 1) *Two Spaces: Stage and Audience*, where the sound source is in another space than where it is reproduced, for example on a stage or in a recording or broadcast studio. These are mostly traditional music practices; 2) *One Space: The Common*, where the moving sound source to be reproduced is in the same space as the audience. 3) *Virtual Sound Source*, where a performer's trajectory is reproduced in real-time but the content is or was recorded elsewhere. And lastly, 4) *Track and Sound Temporally Separated*, where both the moving sound source and the trajectory were recorded separately and offline, and the spatialisation is temporally removed from them. The scenarios were developed and defined through experiments in the Workshop on Music Space & Interaction (MS&I).

Over time, some physical limitations to the principle became clear to me. For example, in previous papers, I claimed that AL has an advantage over optical systems because the sound diffracts around objects. This is only marginally relevant for sound above 20 kHz, where the wavelengths are below 2 cm, diffracting at best around a finger, but nothing much bigger. And the steps towards implementation, the proof of concept application this article describes, also signify the advent of some compromises to the original plan, as the actual implementation started to take shape distinct from the ideal.

Necessity for a discrete measurement signal: The project's original proposal assumed that the content signal of a musical application within a surround sound system could be used as the measurement signal. Anecdotal evidence from a near-field set-up with eight loudspeakers arranged radially around a fixed receiver at approximately one metre distance supported this. However, I was not able to replicate this in a set-up more in keeping with spatially interactive musical applications, where the radial distances and the spacing of loudspeakers are two to three metres apart, to cover a stage-sized area with approximately eight loudspeakers. This led to the introduction of a distinct measurement signal to the system, just above the audible frequency range, but within a range that most commercial loudspeakers are able to reproduce (20-30 kHz). This also worked reasonably well when distributed on the same set of loudspeakers used for the audio content.

For a series of practical reasons, I never looked in detail into why the original idea of using a content signal as a measurement signal did not succeed. Firstly, the introduction of an inaudible measurement signal turned out to be an acceptable solution. Secondly, it was always clear that some inaudible signal would be needed for tracking when no content audio was available (musical pauses, for example) Thirdly, because the measurement signal for each loudspeaker has to be distinct from the signals on all other loudspeakers, the musical content, which is often musical *via* the correlations of its signals, is not a good test signal. Fourthly, the tracking of objects other than sound sources would have been categorically excluded from the system's possibilities.

Pulsing of the measurement signal is necessary: In the article two essentially different but related approaches to AL tracking are defined: the "single microphones, multiple loudspeakers" versus the "multiple microphones single loudspeakers" (microphone array) systems. In the former, one receiver is used to estimate time delays from (non-correlated) signals on multiple senders; in the latter, one sender's signal is measured on multiple receivers. For musical applications in surround sound systems, the multiple-loudspeaker version is, of course, the more attractive choice. For the proof of concept discussed in the article, a simultaneous test signal was used. This turned out to be problematic as every simultaneous signal constitutes noise in the measurement of every other signal. Later implementations circumvented this by pulsing the signal. This has an impact on latency, because stepping through channels requires time.

The more robust multiple receiver-single sender model is rarely applicable: The advantage of a multiple receivers model is, as stated in a later article, that additional *listeners* do not make noise; so, as long as the sound source is a single sound source, this approach should yield robust results. I suggested that this might be a good approach for applications where content audio serves as a measurement signal; for scenario 1, for example, when an acoustic instrument's location on stage is estimated by its distance to the stage monitor-speaker. Although this remained unexplored as outside the scope of the study, I venture a guess that problems would arise from correlated signals coming from the instrument itself. And the situation is only rarely such that a musical signal comes from only one loudspeaker. In the case of live panning (Scenario 2), that is, between (multiple) loudspeakers, some additional strategy would become necessary. The model should work well, though, for measuring the absolute positions of loudspeakers in surround sound systems, say for calibration tasks.

The process of *implementation through limitation* was not only negative. The proof of concept implementation, for example, adopted a one-dimensional localisation strategy: As the concurrent measuring of the test signal on multiple loudspeakers proved to be unreliable, trilateration for a position estimate was not possible; the measurements showed the distance to one loudspeaker only, occasionally two, rarely three, which was not enough for a position in 2D, let alone 3D. I realised that for panning an audio signal, the presence of a signal on a single loudspeaker is sufficient: The distance is reversely proportional to the amplitude. This turned out to be applicable throughout the project.

2.4 Gestural Control for Musical Interaction using Acoustic Localisation Techniques (IV)

I presented this article at the International Conference on Live Interfaces in Brighton in 2016. The conference's theme of "live interface in the performing arts, including music, the visual arts, dance, puppetry, robotics or games", was almost "post-digital", a refreshingly open-minded approach,– treating computers just as one option of many for how to realise live interfaces. My article on "Gestural Control for Musical Interaction using Acoustic Localisation Techniques" was peer-reviewed and included in the proceedings but presented only as a poster at the conference.

The article described the scalability of AL principles, making a case for AL's use for gestural control of musical interfaces. I presented two proof of concept implementations using the visual programming language Max, firstly, as a trigger for percussive sounds and secondly, as a patch for pitch control. Both showed rudimentary functionality, sufficient to show the potential, but they were far from providing the intuitive control necessary for music-making in the sense of traditional instruments.

Further, the article offered some thoughts on requirements for gestural control. It asserted a maximum latency of 10 ms for musicians in performance situations and a control range of ca. one metre. Further, it supplied a list of relationships between sampling frequencies and areas covered. For the proof of concept implementations, for example, trajectories through air expressed in 512 samples at 48 kHz represent a distance of 3.66 metres.

In the discussion, I observed that even at higher latencies the percussive trigger could be used quite successfully, as one seems to anticipate the amount of latency by hitting a bit earlier to compensate and trigger the sound at the right moment. It seemed that latency had a less detrimental effect than jitter, which made triggering unreliable and random.

2.5 Requirements on Kinaesthetic Interfaces for Spatially Interactive Sonic Art (V)

The Audiomostly conference in 2016 took place in Norrköping, Sweden. It was a bit further removed as a conference from the concerns of the arts than the others I had attended, but it still featured several interesting projects relevant for my research. The conference theme was "Multisensory Interaction Design"

Requirements on Kinaesthetic Interfaces for Spatially Interactive Sonic Art, was the first article in which I delved into detail into what a tracking system irrelevant of type, should be able to do (in both qualitative and quantitative terms). I defined most of the terms used in previous articles more accurately, providing a vocabulary, benchmarks, and inventory for later articles. In this article, some of the meta research questions I raised at the beginning of the introduction are addressed implicitly. The article can be read as a reflection on technicality and its role in artistic practices, a *comment* on *the technical* more than a technical paper.

In the time leading up to the writing of this article, I acquired a vocabulary through an illuminating literary exploration chain that helped me to approach these questions from within a development practice. One of the first links in this chain was Mark Weiser's epochal paper "The Computer for the 21st Century" [99]. More than the article itself, I was and still am fascinated by the idea of an interdisciplinary research team to design and shape a common future, the idea behind Xerox PARC. Next in the chain was Paul Dourish and Genevieve Bell's "Divining a Digital Future: Mess and Mythology in Ubiquitous Computing" [20] which linked to Lucy Suchman's "Plans and Situated Actions" [91] and "Human-Machine Reconfigurations" [90], which in turn linked to Bruno Latour's Actor-Network Theory (see "Reassembling the Social" [37]) and, eventually, to Latour's "An Inquiry Into the Modes of Existence" [39].

From my current point of view, I would rephrase or refocus some of the definitions in the article. For example, in the article, the definition of *interaction* stands for human-human interaction, from the point of view that human-computer interaction is also essentially human-human interaction. Today this sounds somewhat anthropocentric and I would shift the definition to be congruent with Actor-Network Theory, so that interaction would take place between actors and/or actants, stressing the point that interaction is interaction, no matter if it is human-human, human-computer, or computer-computer.

I define *Sonic Arts* as all arts that are sonic in any way. These include music, sound art, and multimedia art, but also the sonic aspects of theatre, video art, and audiovisual displays.

2.5.1 Kinaesthesia and Space

In reference to the kinaesthetic interface of the title, its definition brings together several strands of thinking that are not fully discussed in the article but might be of interest here. In the article, I define space as "dynamic, multiple positions"; space comes into being as a result of a change of position – *space is a result of movement*.

Counterintuitively to how we would *describe* an enacted movement in day-to-day language, if we start from motion *as we experience it*, space and time can be understood to be *perceived* through movement. Whichever was first – whatever the casual relation between space and time, our perception of it happens through *Kinaesthesia*, no spatial experience, no sensation, perception, or expression makes sense without the notion of Kinaesthesia. This was an often-heard concluding remark by workshop participants.

There is a whole body of work on embodiment and its relevance for arts and music in particular, but, following my literary exploration chain, Carrie Noland's work "Agency and Embodiment" [75], and through her reference Alain Berthoz's "The Brain's Sense of Movement" [8], were the works I happened upon and which reverberated tremendously with what we experienced in the workshop. Although Noland and Berthoz's works are instrumental in the article How Do We Experience Digital Arts? – An Exploration through Latour's Modes of Existence (VIII), it makes sense to introduce their work here, as it supplied the theoretical background for the inception of the Kinaesthetic Interface.

2.5.2 Kinaesthesia and the Motor-Neuron Theory of Perception

Carrie Noland's work "Agency and Embodiment" [75] focuses on the embodied experience of agency manifested through gestural movement, or kinaesthetic experience. Noland makes the case that "kinaesthetic experience – the sensory awareness of one's own movement – can indeed encourage experiment, modification and, at times, rejection of the routine" [75, p. 3] as an alternative to constructivists' "inability to produce a convincing account of agency" [75, p. 7]. Whereas the implications herein on agency are of importance for my article on digital arts, the underlying concept of sensorimotor capacities behind all perceptions, what she calls the "the primacy of movement" is of importance here. It also explains why a kinaesthetic interface and the tracking of gestures might be of interest to the arts beyond superficial aesthetics.

The motor-neuron theory of perception is introduced by Berthoz through "William James's concept of an anticipatory neuronal pathway" (see Figure 2.1). This presumption of a feedback loop has far-reaching implications for our understanding of how we perceive the world, and consequently for the world as we see it. The feedback allows adjustments of future actions based on a memory of previous actions. From their essential role in sensory perception, it can be hypothesised that the way we experience the world is the result of a series of such iterations. This, in turn, has manifold implications. The most relevant one for the conception of a kinaesthetic interface is that proprioception bears the possibility of the individualisation of gestures. The learning and relearning to adjust to stimuli in a different form, the embodied exchange of gestural routines, allows the forming of cultural similarities of gestures but also the possibilities of refuting them, of emancipation.

So, a kinaesthetic interface would ideally register the individuality of the performance that left its unique trace in the gestural routine.

In summary, the (ideal) kinaesthetic interface might not be able to trace anything other than kinetic events. However, it records those kinetic events at the right resolution, over the necessary distances, at sufficient speeds and with the neces-

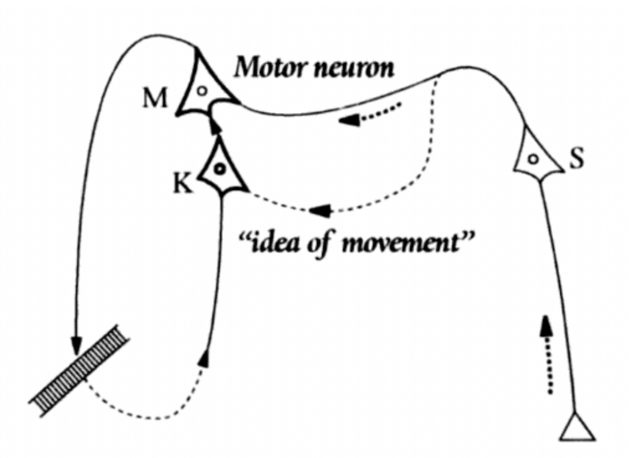


Figure 2.1.: "William James's concept of an anticipatory neuronal pathway". If a sensory cell S is excited, it activates a motor neuron M, which induces a muscle contraction. A kinaesthetic cell K detects the movement in the muscle and modifies the motor neuron. James proposed that there might be an additional path from the sensory cell directly to the kinaesthetic cell, enabling the kinaesthetic cell to modify the motor neuron even before it activates the muscle. The Kinaesthetic cell can thus anticipate how to modify the motor neuron. Importantly, the kinaesthetic cell's actions are informed by the feedback from the muscle as well as from the sensory cell, in a continuous loop (Berthoz [8, p.10](© Odile Jacob, 2005)).

sary accuracy to make them relevant enumerations and encodings as a parameter correlated to its kinaesthetic experience.

2.6 Immersive Spatial Interactivity in Sonic Arts: The Acoustic Localization Positioning System ALPS (VI)

This is an extensive, technical paper, which gave an overview of the implementation. It did not explicitly address any of the meta-questions I listed above. Whereas the previous article on the Kinaesthetic Interface was in many ways a commentary on the technical aspects, here we have an actual example of a technical application, rather than a commentary on it. The main subject of this paper, the description of ALPS and the associated Max patches, formed the basis for the summary in the section Outcomes. To give the Journal's readers a coherent introduction to the whole project, a quite detailed description of the method is also included, a précis of the subsections Interdisciplinary Improvisation as well as Participatory Design and Workshop on Music, Space & Interaction.

As in the other articles, the introduction proffered a summary of the rationale that AL provides competitive alternatives to other positioning technology for sonic arts. Firstly, the background section summarised the literature on tracking technology for interactive art in general, and secondly for all my previous articles. However, it also extended on the scenarios described in the article Acoustic Localisation for Spatial Reproduction of Moving Sound Source: Application Scenarios & Proof of Concept (III). The detailed system overview included pseudo-code for the algorithm, details on the signal processing involved, followed by implementation strategies, and a description of the two available software versions, Audio1 and al-Qt, both of which contributed data for experiments. Next, I introduced the Autopan Max patch, which effectively delivered much-improved functionality compared to the implementation in the proof of concept described in article III. I also introduced the DBAP principle in that section, as it played an important role in the approach to autopanning.

I repeated the experiments originally conducted with Audio1 in a near-identical set-up with al-Qt. In the discussion, I presented the quantitative results in form of data tables that showed much improved performance for the multi-thread version al-Qt. Comparing the results to the requirements set out in "Requirements on Kinaesthetic Interfaces for Spatially Interactive Sonic Art", the implementation met expectations. For low-latency applications, al-Qt came tantalisingly close but did not quite reach the required latencies < 20 ms. In the discussion of the qualitative experiments, conducted with participants of the MS&I workshop, I showed how the combined development and evaluation efforts helped to overcome various challenges commonly encountered in spatially interactive practices with standard audio equipment. An individualised, localised spatialisation achieved with the Autopan Max patch facilitated immersion and the realisation of engaging spatial narratives.

From a technical paper reporting on a development project, one would expect to be informed about everything essential that contributed to its success. But when comparing the process and progress described in the article with the progress and process as I remember it, they are very different. Frankly, I somehow have the impression that many things that were essential for the project to succeed are missing from the article. The article described a straight trajectory from idea to result. I made a choice as an author of a narrative to mention some challenges; for example, the failure of my first attempts with Bluetooth technology, or that I stumbled upon the fact that amplitude and distance are directly proportional to each other only after failing to achieve a sufficient amount of readings for trilateration.

Such essential elements, in practice *making* a technical project, are an unscientific account of how the technology came to be. They relate to the "real science" in the article in the same way as "the making of..." relates to a feature film. What is interesting is, that this analogy entails that science is actually something *fictional* that needs "making" in a real world. Moreover, if the outcomes were not a success, if the technology that was developed ultimately failed, the article would not even have been published. So, did I just construct the whole research towards its results?

Bruno Latour would probably say so. What is more, in "Aramis or the Love of Technology" (1996) [35] he documents the slow and complicated demise of a technological project over many years, despite its undisputed technical feasibility.

In the preface of the German edition of "Aramis", in 2008 [54], Latour brings this underlying relationship between a technical project and a technical object to the point: "For objects and projects resemble each other in almost no respect. Technics are always merely judged by their outcomes and failures. Never are they considered *in vivo*, in the motion and the contingencies of their development" (p.VIII).[4] In his work on Actor-Network Theory (ANT) [37], he epitomised this phenomenon in the statement that "aeroplanes do not fly, but airlines do".

2.7 Prototyping Situations: Interdisciplinary Free Improvisation in Technology Development and Latour's Mode of the Technical (VII)

After a summary of all the previous papers, the article "Prototyping Situations: Interdisciplinary Free Improvisation in Technology Development and Latour's "Mode of the Technical" provides an introduction to the development methods in a similar way as the resumés in the subsections Workshop on Music, Space & Interaction, Interdisciplinary Improvisation and Participatory Design. In addition to the content of the resumé, I also gave examples of the findings and practices we developed, which somehow seemed to be at odds with development methods in the literature.

Of course, in our practice the phases of development could also be traced along a theoretical design cycle: *Define the problem - Conduct research - Brainstorm and*

[4]"Denn Objekte und Projekte gleichen sich in fast keinem Punkt. Technik beurteilt man immer nur anhand ihrer Endresultate oder Misserfolge. Nie schaut man sie sich in vivo an, im Schwung und in den Zufällen ihrer Entwicklung."

conceptualise - Create a prototype - Build and market your product - Product analysis - Improve.[5] But somehow, this only ever seems applicable with hindsight.

Initially, this discrepancy between the *theory* of the design process, the practices in the MS&I workshop, and current mainstream design theories seemed odd. In the Results and Discussion section of the article I presented a Latourian interpretation of the findings from the MS&I workshop. I noticed that aspects of them that at first seemed counter-intuitive became clear and accessible via the concept of the technical as a mode of existence, as introduced by Bruno Latour in "An Inquiry into the Modes of Existence" [39]. My main argument in the article is, that due to the elusive nature of the *technical*, the fact is that the *technical* is mostly only noticed when it is *not* working, and development methods have to consider this. Evaluation has to become the main phase of *development*. I contrasted this, in particular, with evaluation methods based on "questionnaires", which do not account for the fact that pre-formulated questions about the success of a technology exclude the most relevant answers that could help in improving the technology. Evaluation and development have to be in the same process phase for sustainable development to take place. I termed the method that we applied "prototyping a situation rather than a solution", and summarised it thus: "To make participatory development more goal-focused, it could help to leave out the prompting, and let participants talk freely and take note of what they say, once they are given the opportunity to play - or improvise in a situation" (see article VII).

In the preface to her book "Plans and Situated Actions: The Problem of Human-Machine Communication" [91], Lucy Suchman uses two ways of navigation, the European way and the Trukese way, as a simile for "two alternative views of human intelligence and directed action". The European way of thinking (and navigating) would be to have a plan to guide the action, based on universal principles (of navigation), and is essentially independent of the exigencies of a particular situation. The Trukese way, in contrast, does not start with a plan, but with an objective, making the course (of action) "contingent on unique circumstances that he cannot anticipate in advance". Interestingly, she does not actually favour one way over the other, but shows that "all activity, even the most analytic" hence also "the European way [...] is fundamentally concrete and embodied". If a skipper shows a route on a map after the fact, it looks like it was planned, entirely *mapped out*. But when

[5]The actual example is available from: https://www.slideplayer.com/slide/7805450/ An overview of cyclic design-thinking is available here: https://www.sciencedirect.com/topics/engineering/product-development-cycle A casual search at the time of writing showed that the cycle thinking itself is not even that established, and linear approaches are still present in product development.

underway, navigation consists of taking bearings, reading the compass, applying topographic and local knowledge, monitoring speed and wind directions, even cooking soup for the crew, and fixing a few things. Incidentally, this is also how our world map came to be: the intentionality came in hindsight – not even Columbus set out to discover America – he got lost on his way to India.

Yet, the recent circular design model I used as an example falls precisely for this idea of a universally principled plan and presents itself as a guide on how to develop something *new*. Whereas Suchman investigates the way her research participants tried to "navigate" a machine, a photocopier, which in itself is mapped out and "user-ready", the circular design principle pretends to provide a guide to not only navigate a topography but to create the topography and the navigation as well.

To come to the point: Even if we assume a great number of givens, for example that the technical problem to solve is already defined, or that the development shall replace a dysfunctional existing technology, and/or a prototype exists – to come up with a 'technical object' that provides a solution, the 'best-made plans' are bound to fail unless they allow for a large amount of improvisation. Just like challenges in everyday life; a balance between experiment and experience provides solutions, but the best-made plans are recognised as such with hindsight only.

Another aspect of the design method that remained in the background in the article is a technical project's reliance on a collective. With the focus on the nature of the *technical* as a mode of existence, it is easy to forget: To make something *technical* happen it needs to take place somewhere, in an actor-network – as Latour noted, it is not aeroplanes that fly, but airlines, the *technical* needs a project to do its thing, even if project and object do not resemble each other in any way.

A technical plan, a blueprint can be faultless in itself, but its enaction is contingent. There are many examples of very good plans which were no longer seen as such after their execution.

2.8 How Do We Experience Digital Arts? – An Exploration through Latour's Modes of Existence (VIII)

Similarly to the previous article, I started from Latour's perspective on the *technical* and the *technical*'s elusive nature. Rather than focus on the implications for technology design, I focused on the experience of the *technical* as described by the participants of the MS&I workshop. In the extensive discussions that were part of the workshop, thoughts on technology and the role of technology not only in the arts but also in our day-to-day lives were always part of the same discourse.

The starting point for the article was the spectre of the "brains on sofas", a dystopian trope that participants invented *verbatim* for an entirely disembodied future of human existence, where technical objects replace embodied activities. To understand this monster better, I tried to get to grips with what we *experience* as technical. I argued, based on the differentiation by Latour between the technical object, the *technical per se*, and the *technical* as situated action (as defined by Suchman), that what we describe as aesthetically[6] digital, and more generally experience as aesthetically *technical* in the arts are not technical beings in the sense defined by Latour in "An Inquiry into the Modes of Existence". Based on three example artworks, I show how these are fictional beings, how the digital as a measurement provides references and access to other beings, but how they, in their function as aesthetic objects are not technical. I defined four types of digital art:

1. It has been created using digital technology.
2. It needs an electronic processor to manifest itself.
3. It is presented via a digital interface (analogue via AD/DA conversion).
4. It is about the digital/about computers.

Of these four types, I argue that only the fourth type is actually experienced as digital art and/or as aesthetically technical, despite the fact that, specifically for that type, the presence of anything technical is the least essential.

[6] I use the term "aesthetically" in the etymologically original sense only, of "relating to perception by the senses. This includes meanings of "concerned with beauty", but does not reduce the term to the latter meaning, which requires some form of jury sitting on a long table asking every perception to show its best sides in a frilly dress to assess wether it passes muster to be admitted to the league, according to a score, no doubt.

This helps to rationally formulate the discrepancies between the mode of existence of the *technical* and the amalgamated construction of "technology" that we fear in form of, for example, "brains on sofas", but also as its closely related monster sibling, the "AI take over". Although this debunks the experience of the *technical*, it does not explain our unease with the loss of agency experienced in situations where the *technical* replaces embodied activity.

I hence went on to analyse *agency* from an Actor-Network Theory perspective. This helped to understand how technology is distributed within a network of actors. Drawing from Lucy Suchman's insights on technologies as situated actions, as collective performances, insinuates a collective responsibility for technological objects; essentially, a *distributed agency*, which includes the agency of technical beings, but essentially re-configures them as contingent situations in which we also have an active role.

I thus suggested looking at experience from a motor neuron theory perspective, whereby all perception is essentially motoric, as Alain Berthoz outlines in "The Brain's Sense of Movement" [8]. Through a reading of Berthoz by Carrie Noland [75], I propose that, as all perception is *active* in the sense of motoric, embodied action, the reverse also has to to be the case: if *motor activity is perceptive*, and experienced agency is motoric individualised embodied action, it can explain the sense of authorship, the sense of achievement felt when *creating*. Yet, that "competence" (and hence authorship) "follows performance" [39], is nonetheless true. In a multi-actant, distributed agency, where technical beings are part of all creation processes, authorship can only be awarded after we have the "know-how" to create the artefact. (It cannot be before, as we would not have known how to make it!) The statement "I am a composer" implies the *embodied, motoric, experience* of composing – there is no other type!

3 Mensch Maschine

Mensch Maschine (human machine), the chapter title on loan from German band Kraftwerk's song of the same name, stands here firstly, provocatively, to think of humans as machines, but, secondly, also to think of machines as human. This second interpretation illustrates what I'm aiming to assert in this chapter: Machines are socio-material arrangement, hence humans are inseparably invested in them.

I start this chapter with an exploration – in more detail than the call for article VIII permitted – of the relationship between techniques and technology and the difference between tools and machines. I argue that *technique* and *technology* are fundamentally the same in principle;[1] that tools and machines are part of a continuum; and that *body–tool–machine* are not opposed elements, but located on a gradient. This could imply that on that gradient, with increased automatisation, comes an increase in the machine's autonomy. I argue against this, on the basis that in deconstructing machines as socio-material (technical) arrangements, we always find the instituting agency behind it actively at work (see section Socio-Technical Arrangements). In the case of AI, *human* agency. Further it could be implied that along that gradient, an increase or decrease in human cognition would take place, but this seems not to be the case either. The gradient, however, stands for an increase or decrease in motor activity.

In the section on Coding – Writing, I explore evidence for the link between human agency and automatons in the idea that the code structures constituting language, music, and computational machines are (closely) related. This allows for a conceptualisation of machines as extensions of human cognition. Consequently, the cognitive agency of machines is human agency by extension.

In the section Post-human Machines with Decision Power, I show why, on this theoretical basis, autonomous "human-like-machines" are not conceivable; arguably, a 'computer', for example, 'cognises', but according to a *Nomos* (law) that is not its own (auto*nomous*).[2]

[1] See Bruno Latour's "mode of the technical" in AIME [51], and discussion in chapter 4 section 4.2.1

[2] I discuss the notion of *Nomos*, in chapter 4.

There is neuro-physiological evidence that not only perception, but consequently cognition, too, is an active (motor) performance. (I addressed the role of *kinaesthesia* in perception in article article V and its discussion.) This suggests that cognition is the prerogative of living organisms. Therefore, a clear distinction between machines (socio-material arrangements) and cognising organisms becomes possible.

3.1 The Digital

From the discussions about technology in my artistic practice, questions arose about agency, in particular with respect to AI and human-like machines. As general discourses in the media treat AI takeover as a real possibility,[3] the question of agency and technology is an urgent one, and deserves further investigation. In article VIII and its discussion, I deconstructed the tropes of *brain on sofa* and *AI takeover* as beings in the Latourian modes of the *fictional, technical* and *reference*, and explored the performative nature of such *figurations*. I concluded that the notion of "digital" plays a central role but is ill defined. I conclude in the article that the technical aspect of the digital is absent in the sensory experience of what is commonly referred to as *digital arts*. What is more, the reason why (at least some of us) experience the idea of AI takeover and brain on sofa as disturbing is that we feel robbed of agency, that our technical practices, particularly as artists, are increasingly disembodied through technology, through an absence of kinaesthesia in automated technical processes. Let us take a closer look at the relationship between *body* and *tool*.

3.1.1 Body – Tool – Machine

In the workshop on Music, Space & Interaction, the notion of the body as an instrument, and reciprocally, tools as an extension of the body, was practice-instituting rather than a result of the practice. Article V tells of the essential role of kinaesthesia for the practice of the workshop. In article VIII, I define the difference between a tool and a machine as the degree of motoric engagement of the body in the tool and its absence in the machine.

[3]Elon Musk, for example, the owner of the electric cars manufacturer Tesla, is a firm believer that AI Takeover is coming. https://en.wikipedia.org/wiki/AI_takeover

If we suppose that the gradient[4] linking the body to tools is the same as the one linking tools to machines, Marcel Mauss's "Techniques of the Body" provides the missing link. Mauss does not equate the body to a machine, but likens the acquisition of techniques of the body through training to the *assembly of a machine*. He also refers to skills, craft, cleverness, and competence as expressions of an achieved efficiency of the techniques of the body [68, p.78].

When I now take the next step and claim that techniques and technologies, are also on the same trajectories, I do this with the understanding that the materiality of a technical object, although essential for its technicity, is *not* ontologically the same as its technicity. This disambiguation between the technical object and the *technical* is essential in Latour's description of the technical mode of existence, introduced in article VII. By naming the mode the *technical*, he makes it possible to ontologically grasp with the same term the technicality of the body, techniques, and technology, thereby pointing out that everything technical pertains to the same mode of existence. As Latour suggests a multiverse with many modes of existence, rather than reducing the body to a "technical object", the body like every entity can be ontologically grasped through every mode,[5] but no complex, real, articulated being can be reduced to just one mode.

From this follows another notion I adapt from Latour, that inert materials can also be understood as technical objects, that they too develop "competencies", and "cleverness": "techniques", albeit across evolutionary time-spans. I think here about the surface tension of water, for example, or more recently on an evolutionary timescale, photosynthesis for plants. So the entire history of the universe can be recounted through the history of the *technical*.[6]

This gradient of bodily involvement is not one of an increase in technical complexity – the technical complexity of the human body is high, for some tools it is low, for many machines high again. What I suggest is, that the gradient is one of agency. At first this might be misleading, as it insinuates a *decrease* in human agency towards the machine end of the gradient. My hypothesis, however, is that we are looking at a displacement of agency, rather than a replacement, and an increase in agentic complexity, a multiplication of agents, but – and this is important, with respect to an individual body – a decrease in direct kinaesthetic involvement, proportional to

[4] If I use the term gradient here, it is to imply an increase/decrease in motoric bodily activity along it. Not as a dichotomy of sorts, nor as an indication of the amount of cognition involved.

[5]"Each mode grasps all the others according to its own type of existence" [39, p.215].

[6]To do this, all of the other modes would necessarily be present too: no mode exists isolated in the multiverse!

the degree of proliferation, a form of dilution, perhaps, but definitely a distribution. Machines constitute a distribution of (motoric) agency.[7]

3.1.2 Socio-Technical Arrangements

The system of ALPS uses various tangible technical components, but none of these needed to be manufactured as part of the development. ALPS is *software* that allows embodied interaction with digitally stored data. It automates processes according to technical (tracking) principles, by using the malleability of the C++ computer language and the visual programming language Max.

Every line of software code, no matter how many times it has been run, has been written by a human at some point. Software, as such, exemplifies more than anything else how digital machines are essentially socio-technical arrangements. The fact that we can do something "with a couple of clicks" is thanks to the painstaking work of many hours by others. As such, a comparison of software with cathedrals seems apt – prolonged technical projects, without an end in sight – collective engagement on a large scale.

Lucy Suchman, incidentally, uses the example of the building of the Gothic cathedrals to illustrate the performative, improvised, even messy nature of the socio-technical, by pointing out how plans were predominantly present as general templates often just drawn in the sand or on a scrap of wood, rather than precise instructions. She points out that the earliest building plans for complete cathedrals date to the late fourteenth century, but that the cathedral of Chartres was completed by 1230 [90, pp. 196-199]. Similarly, computer languages evolve, they do not follow a plan, yet they have syntax, semiotics, and clear grammatical structures.

Software exemplifies the socio-technical nature of machines, making machines that require software more *human* than other technologies. That the human actant is displaced, rather than replaced by the machine, becomes particularly clear in Suchman's analysis of the automaton.

I would like to recount how Suchman meets Cog, "A robot head and torso built to maximize the integration of a 'perceptual' system (computer vision) with basic motor 'skills' (movable arms and grasping hands)" according to the project website in 1999. The purpose of the project was to investigate themes of development,

[7] For now the term *agency* and *cognition* have to remain a messy muddle. I will disambiguate further down.

physical embodiment, sensory-motor integration, and social interaction. Yet, by the time of her visit two years later, the robot was defunct, just standing in a corner. What impressed her most, however, was the masses of hardware and tree trunk sized cable looms for connections which epitomised "the extended network of human labours and affiliated technologies that afford Cog its agency", rendered invisible in the media footage where he was shown as an autonomous entity. The human factor *displaced*, not *replaced*.

A shift of agencies can be experienced in socio-technical configurations, a repositioning of the actor-network itself: as the actants' negotiations within the socio-technical practice. Algorithms are the configurations' performed practices. Automation shifts actors around on the network. The scale, the speeds, and quality are negotiated at a price, paid through the work of the supportive collective. "Autonomous entities" are essentially – more or less algorithmically – *performed*.

3.1.3 Digital – The Ambiguous Binary

I feel obliged to explain why media studies, and new media as a term, has been conspicuous through its absence in this thesis until now. Although I maintain that most of what I said makes sense without any contribution from media studies, I can see how I make myself vulnerable to criticism, as my readings and theorising stand in conflict with some of the core tenets of new media theory. New media is a term used for various kinds of electronic communications made possible by computer technology. Research into new media concerns itself with its difference from old media. In article VIII, I argued for the absence of such a difference, as the contents of new media are not experienced in any different way than old media. This, of course, all depends on a clear definition of what the term *new* is supposed to stand for.

As 'old' and 'new' insinuate a change of state in a chronological order, it makes sense to look at a very short history of modern technology, to frame the terminology of this chapter in a sensible way. Although occasionally there is talk about the digital revolution to have taken place in the mid 20th century, many of the constituting technical principles (computational algorithms, programmability, automatons, telecommunication and monitor screens) go back a lot farther in time. What forms a clear turning point – even if it is even harder to precisely date the event, is the advent of electricity, without which electronics and all the above mentioned principles would not take place. Electricity, although ubiquitously available, is arguably still the make-or-break component of practically all "new" technologies.[8]

General discourse does not however differentiate between old and new media along pre-electric and electric technology, rather it tries to define itself as "new" according to the more vague differentiation of *digital media* as opposed to non-digital. In the following definitions I'm trying to account for these historical facts where possible, but simultaneously take an a-historic approach as all these new, old, electronic, electric media, techniques and technologies are simultaneously available to artists and engineers today, and form an essential part of contemporary art and music practices.

Very often, the term *digital* not only refers to the electronic processes made possible by computer technology but also the cultural and historic connotations pertaining to the practices of such technologies. This more vague notion I shall henceforth denote as Digital, with a capital D, to disambiguate it from the more concise definition of the technical *digital*: In my definition, digital refers to a signal conversion represented by values of a physical quantity such as voltage or magnetic polarisation. I mean the *technical principle* of *digital electronics*. In contrast, I refer to the term *binary* (*relating to, or composed of two things*) when I mean abstract entities; almost all digital systems are built on binary principles, but not all binary systems are digital.[9] So *digital*, in this sense is *not* the vague notion insinuated in the term *digital arts* or *digital humanities*. If I use, for example, the term *digital media*, I refer to data or datasets in a form that can be processed by digital processors.

As a consequence of this definition, *digital phenomena* can also be concisely defined as somatically perceivable experiences of the *digital*. To perceive a digital phenomenon, the underlying digital process has to provide sensory stimulus perceivable as digital. Stimuli arising from digital processes *other* than the aforementioned ones are not experienced as *digital* phenomena. Consequently, and this is a core claim in my argumentation, if a digital representation of a continuous signal is successful, (as is the intention for most digital representations), the *phenomenon* (what can be experienced) is not digital. In order to experience it as Digital, additional information

[8] I am indepted to Lily Díaz-Kommonen for suggesting this important historical differentiation!

[9] There are borderline cases where a further disambiguation would be necessary. For example, for complicated clockworks, where a binary system is implemented in a *de facto* identical way to a digital system, but using, say, gravity, rather than voltage or magnetic polarisation; or a system might be implemented using a non-binary basis (say, applying 3 different states) using voltages, or magnetic polarisation using minus, plus and a non-magnetic state.

that is not contained in the stimulus, is required. [10] That is, the visual presence of a computer or cultural context. Pixelation or glitches pertaining to the *digital* process, are directly experienceable digital phenomena. (Unwanted or wanted perceivable artefacts of digital processes are referred to as *glitches* for the rest of the text.)

What is more, in our somatic experience, we rely on the same sets of material givens, whether the source is digital or not. For a blunt example, what we *actually* hear of a digital audio recording, is the loudspeaker's cardboard cone oscillating. Why these tight definitions? In the media, but also in philosophical discussions, the Digital is endowed with an existential power based on a phenomenality, which I claim it lacks. Much of the following discussions will explain this further.

The argument has been made here, that phenomena generated through AI and or Machine Learning (ML) are digital processes that we experience as digital. Although there is a clear difference between the experience of the digital in the form of glitch and the digital in form of AI/ML, I argue that, again it is the non-continuity of the experience (that the phenomena is *artificial*, that there is a *machine* at work) – the *uncanny valley* if the phenomena is intended to be human-like – that makes the experience notable. If the artificiality of the process was not perceivable, it would be experienced as "natural", not "computer generated". Even if that last argument remained debatable, the technical principle in AI/ML processes is not the same one as the one of the definition, (representation of a continuous signal in form of discrete states), but a *procedural* one, an *algorithmic* one. Generative art does not primarily apply *digital* principles, but *algorithms*. *Algorithms* are procedural instructions, sets of rules applied in problem-solving operation. Although predominantly used in reference to computer algorithms, where the algorithm is implemented as a computational process, I find it helpful to think of algorithms as procedural instructions in general. This helps in the conceptualisation of machines as socio-material arrangements.[11]

The difference between digital media and Digital media is often not straight forward: We might deplore the impending loss of handwriting skills, for example, due to

[10] In view of the importance given by media theory to the discourse on Digital media, the relation to Marshall McLuhan's famous statement that the medium is the message[70] might need a closer look: If *digital* media cannot be experienced, it cannot be the message. In my view this does not diminish the validity of McLuhan's statement, it just highlights again that the "electrical" revolution, and the advent of telecommunications outweighs the effects of digitalisation as a cultural evolution. Interestingly, McLuhan speaks of "electronic media", when referring to TV which complicates the vocabulary further, but makes perfect sense if we agree that his statement applies to *Digital* media.

[11] The latently implied reference to the dataset vs. algorithm paradigm is contingent. (See next subsection for details) I suppose the distinction between AI/ML and (representative) digital media is a practical and historic one, and future computational paradigms might not need it.

the prevalence of computer keyboards in our communication practices. Yet, this loss is not really due to digitalisation in the technical sense; the digitalisation here progressed from *typewriter* to computer keyboard.The digitalisation of handwriting only happened fairly recently with the popularisation of touchscreens. As the processing of ASCII is easier to implement as character interpretation from handwritten text, socio-material arrangements favoured the former over the latter. Nonetheless, an email lacks the personal touch of a handwritten letter, but read out aloud, they convey the same message, if the same message was contained;[12] the digital does not manifest itself. The loss of handwriting as a means of communications is, as by the broader definition, a consequence of the Digital (with a capital letter).

A further disambiguation is necessary for the terms *computer* and *computational*. I suggest the following convention. The term computer shall stand for the electronic, technical object as it can be obtained from a shop, (equivalent to a digital computer). This term includes laptops and smart-phones, but also larger computers for research. *Computational* refers to processes in which calculations take place. So, a mechanical computational machine, would be thus disambiguated from a computer.

Last, but not least, *code* shall stand for underlying rule structures in general. I discuss the term in the following section, and hope to avoid it before that.

From these definitions, it should become clear that within my chosen context the difference between new and old media is not important – it is, on this basis, non-existent. Next I will contextualise my position in the wider field of media science, by looking at two theorists, one of whom, Lev Manovich, agrees that new media is essentially old media, and one, Mark B. N. Hanssen, who is adamant that new means new.

Manovich's Database as a Symbolic Form

In "Language of New Media" [65], Manovich argues that new media is not new, but that its vocabulary is essentially contained in the techniques of cinema. However, he asserts that the advent and predominance of *datasets* in computer technology, constitutes a cultural change, as *narratives* are partially split into data structures and algorithms. Although I also question the phenomenal presence of the digital in artefacts generated from datasets (unless the Digital is purposely insinuated,

[12]This resonates with Marshall McLuhan's famous statement that the medium is the message [70], whereby he attests that the content is of a lot less consequence than the impact of the medium it is contained in. So handwriting *means* something substantially different to sending an email, reading out aloud, again, means something different than the same content expressed via the other media.

or the digital is present in form of a glitch), I agree that the cultural change in artistic *practices* was profound. That the access to data sets *per se* can in principle only produce phenomena that also can be achieved with non-digital means (at a higher cost, perhaps) I have argued at length in article VIII. But the speed and ease of processing data enabled by digital tools and also the vast number of choices necessary in the presence of nigh endless data, is evident. As such, new media is a reality for the artist.

Manovich's claim that computer programming encapsulates the world according to "its own logic" [64, p.3], according to the relationship between the two software structures of database and algorithm, might be less valid today than it was at the time. Manovich postulates computational limitations as a need for simplicity in data structures if the algorithmic demands are high and vice versa, and these limitations are fast decreasing. But the gap between algorithm and dataset might also be less important in future computing paradigms. A recent paper from a team of researchers at the Massachusetts Institute of Technology (MIT) [32], for example, raises the question of whether datasets will still be needed as generative models become better. By the same token if reversed, another paper from MIT [94] describes how datasets are biased, as they constitute an *edit* of the world, in a cinematic sense, as Manovich would surely agree. What *sets* the data of a data*set* is a procedural, algorithmic process, with every *structure* a result of a process.

Hansen's New Philosophy for New Media

In contrast to Manovich, Mark B. N. Hansen insists that new media *is* new. New media can be distinguished by their "total material fluidity: rather than being anchored to a specific material support, new media are fully manipulable, digital data" (Hansen [24, p. 32].

Hansen argues exclusively from and with visual media. He says that following its digitisation, the image could no longer be understood as a fixed and objective viewpoint on reality since it was now its numerical basis, and its constitutive virtuality [24, p.32]. This is quite similar to the way I see the digital as *reference*,[13] but whereas Hansen still envisages an "image", however malleable and virtual it may be, I am inclined to attribute this to the underlying fictional nature of images. To contribute the "image" to its data without the interface of a sensory appearance or material surface as an optical phenomenon would constitute a double category error. We

[13] See article VIII and Latour AIME [50]

firstly start with a (mimetic) image, which through digitalisation becomes a reference to itself, but through the malleability of digital data can be "falsified" as a reference, no doubt, as it is no longer mimetic (does not "adhere to reality"), but is available virtually as a *fictional* entity. To summarise: from my standpoint, the data itself cannot be an image, but Hansen claims it can.

Be this as it may, Hansen's aim is ultimately to theorise the "correlation of new media and embodiment" [24, p. 11], for which he makes the presence of numerical representations, the *digital*, the *conditio sine qua non*: "When the body acts to enframe digital information—or, as I put it, to forge the digital image— what it frames is in effect itself: its own affectively experienced sensation of coming into contact with the digital" [24, p. 13]. From this last sentence I feel I can expect to find somewhere in his text a concrete mention of a work of art in which the process of digitalisation enables a bodily sensation wherein the digital can be experienced, or *felt*, to use Hansen's term. Alas, most of the artworks that Hansen uses as examples are in practice and in essence non-digital. And, for the ones that rely on digitally generated content, non-digital technology could have engendered the same somatic effects – some at greater costs, some at less. If the artworks that he describes as digitised images are non-digital – or can be made by non-digital means, the "complete flexibility and accessibility" cannot be relevant. The "flexibility and accessibility" Hansen assigns to the digitised image, which he sees as expressed in Bill Viola's works, for example, is being achieved by entirely old school cinematic means.[14]

Hansen's claims remain unsupported, and against his assurances the digital remains elusive. To provide an "affectively experienced sensation of coming into contact with the digital" [24, p.13] within his theoretical framework, he applies the following two-step approach: he bases the first premise, that new media is new, on theorising the digital by implying that the example works he will present provide evidence for this digitality. In the second step, he demonstrates the embodiedness of those works with the help of the "affectivity" concept, without further affirming the artworks' digitality. Nevertheless, the hypothesised correlation of "new media" and embodiment could arguably still be relevant – Hansen just does not provide any real evidence for it.

[14] "The Passions", for example, was filmed on 35mm high speed *film*. High speed film cameras for standard 35mm film are available for up to 2'100 fps, according to this website http://highspeedfilm.biz/, on 16mm film even up to 10'000 fps. [75, p.68]. Also some of Jeffrey Shaw's works featured in Hansen's book are made of hand drawn content, for example, "Continuous Sound and Image Moments", but he also uses photography (mimetic) and video projection with 3D (requiring polarising spectacles). The role of the digital, if present, is secondary.

The role of affectivity that Hansen uses to endorse his conception of the "digital image" is another place where his thinking diverges with mine. I side with Carrie Noland who is critical of the *affective turn*, and voices her unease that Deleuze's notion neglects Bergson's emphasis on motor *memory* as part of the interoceptive feedback "that shapes the action the subject will ultimately take" [75, p.66]. As a consequence, Deleuze assigns to this process an "emotional attitude", which in the absence of a motor memory builds the basis for (allegedly) "unpredictable" and "new" action. By uncoupling these "affects" from motor memory, making them spontaneous, so to speak, Deleuze creates a new entity that is only seemingly observable, as it mistakes a part of the kinaesthetic feedback loop for a *deus ex machina*.

Hansen adopts a primarily Bergsonian stance, distancing himself from a purely Deleuzian conception of affectivity which he describes as "disembodying" [24, pp. 6-7]. Affectivity, nonetheless, is for Hansen "the capacity of the body to experience itself as 'more than itself' and thus to deploy its sensorimotor power to create the unpredictable, the experimental, the new" [24, p. 6]. This becomes problematic only at the point where he avows that "affectivity comprises a power of the body that cannot be assimilated to the habit-driven, associational logic governing perception."

This stands in contrast to Berthoz [8], where the *creative moment*, the "new", the "unexpected", "invention" can only be recognised as such *after* its performance: If it is "nascent" (Bergson's term), it is so as a motor-memory, as in Merleau-Ponty's "I can"s[15] If it is new, it is as a *competence* which, as Latour says "here again, here as everywhere, follows performance rather than preceding it" [39, p. 228].[16]

Noland's analysis of Bill Viola's The Passions is thorough, and refers to Hansen's. A former performer herself, she tries to reproduce the twitches, tremblings, and contractions she sees in the slowed down pictures. Despite studying them closely, learning their roles and attempting to imitate their blocking, crumpling, hand-lifting, and mouth contorting, "try as I might, I cannot produce through sheer will power the twitch of a facial muscle or the trembling of a cheek. In that regard, Hansen is correct when he insists that new media technologies[17] are able to expose to

[15]Carrie Noland: "Merleau-Ponty locates history [. . .] at the site of the nervous system understood as a repertoire of historically contingent and culturally specific 'I can's' " [75, p. 66]. Noland refers to Maurice Merleau-Ponty's "Phenomenology of Perception" [72, p.328].

[16]Berthoz and Latour find themselves here in Derrida, as I will discuss in the next subsection.

[17]Noland does not comment on Hansen's insistence on the role of the digital for the "new" in new media, and uses his nomenclature as is, despite her conclusions conceptually conflicting with this in the context of my argument.

sight elements of aliveness (he says 'affectivity', I say 'motility') that are performed without volition" [75, p. 71].

Later, she adds that "these movements are nonetheless *human* movements available as potential belonging to the human kinetic disposition" [75, p. 72]), to clarify that she does not experience them as more *unpredictable* than other involuntary movements one expects to see within human gestural routines. But before that, she remarks that as a viewer without any background knowledge of Bill Viola's work, one could argue that "the bodies we see on the screen are not real bodies but only digital photographic reproductions of real bodies manipulated by post-production technologies". And then she makes my point, quite elegantly: "Yet, [...] they have not been digitally transformed!" [75, p. 71]

3.2 Coding – Writing

In this section, I hypothesise that, due to their common underlying structure (*code*, as per Umberto Eco) natural language, music, but also programming language can be conceptualised as writing in a Derridean sense. The linguistic discussion sets the provisional stage to show, eventually, how through language human cognition extends into machines. I will not compare codes on a textual level (say, a sonata with a newspaper article with a page of C++). Rather, I try to think conceptually how such a comparison could become possible.

3.2.1 Code in Eco's Watergate Model

Until now, I have avoided the use of the term "code", as it could be a synonym for computer language, software, or algorithm; terms that I wanted to keep separated. I find Umberto Eco's description of code in "A Theory of Semiotics" [21] very thorough, and also broad enough to encompass other linguistic models. Eco conceptualises language as *constituted* through a number of functional code systems. He moderates "unlimited semiosis"[18] to "univocal" semiosis, where one *system* of elements is translated into another system of elements. Contrary to Ferdinand de Saussure's structuralist approach, where the relationship between signifier (a word, for example) and the signified (its meaning) is stable and predictable, Eco's post-structuralist approach allows for many meanings constrained through systems.

[18] "Unlimited semiosis" is Eco's term for a Peircean concept not dissimilar to Derrida's, where a signifier signifies another signifier, etc.

Eco explains the concept of code with the help of the Watergate Model, which also lends itself, through its technical vocabulary, to comparisons to computational structures. What is particularly helpful, as I hope to show, is that Eco's model expresses the *directionality* of meaning-making; that the code is nothing if it is not performed.

Eco's Watergate Model describes the situation of a watergate and a floating buoy that provides information about the water level above the gate to a control station downstream. In this situation, combinations of signals provide codes for lighting various light bulbs in correlation to these signals. Eco works through a good number of possibilities according to information theory, and then ingeniously wonders what a semiotician would want to know from the engineer who built the system; for example: "What do you call a 'code'?" Does the destination of the arrangement "recognise the 'meaning' of the received message or does it simply respond to mechanical stimuli?" and "who is that code for? You or the apparatus?" [21, p. 36]

From such questions he distills four types of phenomena grouped under the idea of code: firstly, a set of signals ruled by combinatory laws, a *syntactic system*; secondly, a set of water-level states, which form a set of *notions*, a set of communicative contents; thirdly, *behavioural responses*, whereby something arbitrary happens as a consequence of a particular condition being met; and lastly, a combination of the former three. Importantly, Eco believes that only the fourth case, when the syntactic, semiotic, and behavioural elements are combined, amounts to a proper code. He distinguishes the other codes as *system codes*, as *s-codes*. Since an s-code is only of interest if it is part of a (significant) *code*, the attention is focused on the s-code's *intended* purpose (intention is a vector).

The watergate is the conveying system, the control station downstream the conveyed system. The *code* is what connects them, element by element. The *elements* of the former become the *expression* of the latter, and the latter becomes the *content* of the former. For Eco, there are no signs, only sign-functions (sign-functions need performing to signify). So, simplified, a system of sign-functions, a "code" consists of elements of one system that correlate to elements of another (a correlation that needs performing). This defines the difference between a signal and a sign: "A signal is a pertinent unit of a system that *may be* an expression system ordered to a content but could also be a physical system without any semiotic purpose" [21, p. 48].

A signal can be entirely free of meaning. But when the signal is recognised as an antecedent of a *foreseen* consequent it may be viewed as a sign. So the sign is an element of an *expression plane* that is correlationally linked to one (or several) element(s) on the *content plane*.

The (directional) process from an antecedent to a consequent is instrumental in the meaning making. Let us call this, rather inelegantly, *run-time*; for meaning to be constituted, there need to be an antecedent and a consequent activated through code which *runs*.[19] In Eco's model, I see a structural, technical arrangement that fits musical notation, natural language, and computer languages. If it does not conclusively show how meaning is formed, as a *functional model*, it still describes a possible infrastructure wherein meaning making could take place.

Let us see how Saussure's and Derrida's linguistic ideas could be encapsulated into Eco's, and firstly consider Saussure's take: language can be analysed as a system of differential elements, linguistic signs, which are composed of signifiers (a word, for example) and the signified (the entity named by the word). The sign is essentially arbitrary even if it can be onomatopoetic. And, due to the psychological association between the signifier (a somatically perceivable sound image) and the signified (a concept) there can be no linguistic expression without meaning, but also no meaning without linguistic expression [15].

Through Eco's functional approach, Saussure's linguistics can be interpreted post-structurally, even the nominally structuralist *binary opposition*, whereby sense can only be made by something meaning "this" or "not this". As a functional model, binary opposition also underpins Derrida's take, but in a fundamentally more dynamic form than Saussure probably intended. If we look, like Eco, at the underlying code of a binary signal (water level "high" or "low", in respect to the model, for example), the signal is *not* the sign, so the binary opposition is relational, not absolute. (I already implied the relational possibility of a functional binary opposition by using a placeholder for many meanings, "this" and "not this" rather than "cat" and "not-cat"!)

[19] If the code does not compile, there is not an error in meaning, but an error in the system – only code that works is code; non-compiling code is equivalent to an absence of a system.

3.2.2 'Wryting'

Derrida adopts from Saussure the binary oppositions that all speech or text has to articulate if it intends to make sense.

Herein lies a conceptual analogy to binary code: on the technical (procedural) level, all code works in a similar way, and the complexity of sense-making can be explained through the proliferation of multiple simple structures. The complexity or limitation of meaning lies not in the binary opposition in sense-making, but in the complexity of the available signal-paths.[20]

What is fundamentally different in Derrida's take is that he gives writing precedence over speech. In "Of Grammatology", he responds to Saussure's claim that writing exists for the sole purpose of representing (spoken) language, by introducing an alternative notion, whereby written symbols are legitimate signifiers on their own, and should not be considered subordinate to speech [19, p. 45]. This makes Derrida's approach particularly interesting when comparing computer language to natural language. By looking at writing as a separate and to some extent independent system of meaning-making in relation to spoken language, a linguistic comparison is possible, and we can compare like to like (written natural language to written computer language).

Writing, for Derrida, is something more universal than the activity of putting pen to paper.[21] Writing leaves *traces*, produces *history*. What is written, remains, and can be transported, read later, spoken, activated, compiled, executed, debugged, updated, or left to (code-)rot.

One essentially Derridean aspect of this type of universal writing comes from the notion of *différance*,[22] an intentional misspelling of 'différence', which Derrida uses to indicate firstly, that by writing the word differently (as a *grapheme*) there is an additional meaning to the conventional understanding of the word when we hear it as a *phoneme* (they are pronounced exactly the same way). Secondly, he means

[20]Although there is a danger in describing code, which is ultimately a *model* for language as its own constituting "genitor", the comparison is in my view nonetheless legitimate, as we have a choice of either being caught up in a tautological chicken-or-egg loop, or accepting the world as already articulated, as Latour suggests based on Austin's speech act theory [7, p. 44]. Accordingly, we can recognise a structure as a possible model for language, *and* understand language as an articulated empirical instantiation of that structure.

[21]In Latour's notions of *articulation* and also *instauration* (borrowed from Souriau), and perhaps Simondon's *individuation* as well, I see forms of Derridean writing.

[22]Derrida developed the term over many works, but first introduced it in "Speech and Phenomena" [17].

to indicate the double meaning of the French word 'différer' which translates to *differ* as well as to *defer*. Différance in the sense of to *differ* stands for the binary sense-making encapsulated in Saussure's linguistic sign. Différance stands for *defer*, however, to indicate that meaning can be deferred through chains of signifiers, and as a grapheme, in particular, also postponed.

Therefore, computer language and *written* natural language do not conceptually differ. Furthermore, Iterability, for Derrida, is the possibility for a written text to remain readable after the person who wrote it is gone [18]. It can also be taken out of its original context and when reiterated work in a different context. The writing itself *remains*, it is a trace, which is Derrida's point. That a sentence means something entirely different according to whom I say it to – and when, even *how* I say it – does not change the literal text, the code. Further, Derrida insists that "there is nothing outside context",[23] so the signifier's *conditio sine qua non* is context.

From the view of the *signified*, in every situation (I equate *situation* to *context*), a reiterated text is imperatively, unavoidably, and bindingly different (deferred), by the simple fact that it has been reiterated. If it was the same, it could not have been reiterated, because two things that are the same are one (law of identities). But, from the view of the *signifier*, if the text is being changed after an iteration, it constitutes a new text. The repeatability of a computer program suggests that iterability is an essential part of a successful computer program.

Thus, a programming language constitutes writing in a Derridean sense. I will refer to Derridean writing as '*wryting*', from here on, to connote its origin in the concept of *différance*, and as a *wry* homage to Derrida.

Lastly, *slippage* is a concept important to Derrida, introduced to linguistics by Jacques Lacan. In a trivial sense, it describes the principle behind a game of "Chinese whispers": the intention is to pass on the signifier as identically as possible to the next participant in a chain, but the understood meaning keeps "sliding" in the process. In the children's game, of course the signifier changes too, whereas in slippage it remains phonetically constant but changes semantically.

[23] This sentence caused much furore, as the original "il n'y a pas de hors-texte)" was often translated as "nothing exists but words" which is quite bafflingly wrong. In the translation by Gayatri Chakravorty Spivak[19] the translation is "there is nothing outside of the text [there is no outside-text; il n'y a pas de hors-texte]." But in 1988 Derrida himself, used the translation in English as "there is nothing outside context" [23].

3.2.3 Musical Notation

Before moving on, I might need to position myself with regard to the question of music-as-language, and how it impacts on my hypothesis of a code-structure shared by computer language, music, and natural language. The less controversial question is whether language is music. And herein, in a way, lies the answer to the first question too: that we can listen to language in a musical sense (in every aspect, pitch, rhythm, melody, form, etc.) suggests that the underlying structure of music is also present in language. I asserted as much in article VII, p.7, where I propose that musical meaning-making is essentially *'wryting'*, giving a musical example to explain *différance*.[24]

Music and language are easy to compare as phonemes, but to be able to compare music to computer language we are better served to look at the musical grapheme, *musical notation*. I found support for this idea in Raymond Monelle's chapter on "Deconstruction and Allegory" in "Linguistics and Semiotics in Music" [73], and also in Monelle's reference to Eco's "A Theory of Semiotics" [21]. Monelle asserts that music itself is particularly suitable to demonstrating the deconstructive principle:

> Music signifies by means of some signifying code, and analysis also refers to code in interpreting music. But no code is originary; there is no single code which is the key to the work, nor is there a code which explains all other codes. [. . .] The apparent power of a code to 'explain' a piece may in fact be subverted when the listener realises that the code is itself part of the surface material governed by other codes. Any interpretation of the music must therefore remain open, analysis loses its claim to metaphoric or symbolic status. No analysis reveals the true nature of the music.[25]

This would indicate that the inherent *function* of musical syntax and semantics constitutes the code. The analysis of the code itself does not lead to meaning; the

[24]"If we hear two musically different notes (two phonemes) the difference is not something that can be heard, it is not an extra (third) sensory cue . This is counter-intuitive, as we feel we hear the difference. But, and for musicians, this is easy to grasp, we understand intuitively that the second note only makes musical sense in relation to the first one: [. . .] The decision to write (design, invent or author) [a] melody consisting of note 1 followed by note 2 is based on the pre-existing conception of this being experienced as a difference: It is only after adding the second note that we can claim authorship for it. If we do not have this conception already that adding note 2 might make a melody (make sense/have meaning), we cannot claim authorship for it. Similarly, we can make sense of new ideas coming up in improvisation only after we had them."

[25]Monelle, nevertheless sees merit in deconstructive analysis, as it concentrates on the signifier's "surprising and untidy connections which lead to new insights" [73, p. 321].

signified is deferred. In music, the code as notation has "only one meaning: what it does."[26] A musician can play a piece from a score even if the notation is faulty. As I will argue in the next subsection, a musician "compiles" when a computer would not. Here is a main tenet of my argument: the meaning of music is entirely dependant on its performance, the musical notation is the equivalent of the connections and bulbs in Eco's Watergate Model, the *signal*, not the sign. This equates to the code in programming languages; the meaning of code is dependent on its creation at *run-time*.

3.2.4 Code is What it Does

N. Katherine Hayles analysed the similarities and differences of natural language and computer language in her book "My Mother was a Computer". She also compares Saussure's approach to Derrida's, but comes to other conclusions:

> Although Derrida asserts that [. . .] iterability is not limited to written language but is to be found in all language this assertion does not hold true literally for code, where the contexts are precisely determined by the level and nature of the code. Code may be rendered unintelligible if transported into a different context—for example, into a different programming language or a different syntactic structure within the same language [...] Only at the high level of object-oriented languages such as C++ does code recuperate the advantages of citability and iterability (i.e., inheritance and polymorphism, in the discourse of programming language) and in this sense become 'grammatological.' [28, p. 48]

Hayles thus implies that as code does not compile when it is run in a different context, it cannot be '*wryting*' as iterability does not apply. Hence, Derrida's notion of writing is not applicable to programming languages, which contradicts my argument. However, there are some other points to consider.

Firstly, Hayles asserts that *iteration* does not apply, as the context in software code is precisely determined by the level and nature of the code. Presuming that she refers to the "dependencies", operating system, hardware, and firmware, I would argue that they constitute parts of the software code, and hence are part of the iterable writing, part of the signifier. (The light bulbs and wire connections of Eco's Watergate.)

[26] This is one of Katherine Hayle's arguments for why computer language is not language – see the discussion below [30, p. 48].

Secondly, Hayles implies that for code to constitute *'wryting'*, the code has to be able to be understood by *any* system. If no meaning is grasped (if the code does not compile), there is no (readable) signifier, so the code is not language. Yet, in doing so, she contradicts Derrida's general notion of writing by implying that code, as the signifier, should *mean* the same (rather than simply *be* the same) in every iteration. For this to be the case, the *signified* needs to stay constant while the *signifier* changes through iteration. However, this is neither true for code nor natural language; a text in French written 200 years ago *means* something slightly different if read today. What is more, its meaning might be opaque but readable for Italians, but unintelligible to most English speaking people. If we substituted the word *code* with the word *French* (and the term *programming language* with *natural language*) in her argument, her statement would amount to the claim that iterability does not apply because a French text does not work in an English context.[27] In Derrida's maxim "There is nothing outside of context", the scope of the context is Eco's *system*, delimiting meaning.

Thirdly, Hayles purports that iterability would be possible at the level of object-oriented languages, as only there can the same object be interpreted in different ways. Here, again, it seems that she expects the *signifier* to change over iterations, when it is the *signified* which changes, which is evidently why an object can have several meanings despite being repeated instantiations of the same object.

Fourthly, Hayles points out that slippage cannot occur in code, as the ambiguity arising from slippage would result in the loss of the code's meaning (the program would not compile after slippage). Here she understands slippage to be the full equivalent to Chinese whispers – that the signifier *slides*, i.e., changes.

Hayles quotes programmer Ellen Ulman that "a computer program has only one meaning: what it does", and that for code "its entire meaning is its function" [28, p.48]. I agree with this, but I do not see therein a contradiction: Software code as a signifier runs as long as the context (without which the signifier *is not*) is such that the code compiles. The whole machine constitutes the *'wryting'*, of which the code constitutes only a part. The slippage refers to a change in the *signified*, to the meaning of the machine's output, not the functionality of the code. To explain the analogy further: if function is code's meaning, we can infer that if it does not

[27] The argument would look like this: "Although Derrida asserts that this iterability is not limited to written language but is to be found in all language this assertion does not hold true literally for French, where the contexts are precisely determined by the level and nature of the French. French may be rendered unintelligible if transported into a different context—for example, into a different natural language [. . .]."

function, there is no meaning. So, if slippage, iteration, and deferral is something that can happen to meaning, it has to *function* before *différance* can take place. Analogously, a word of a natural language can be iterated, slipped, referred to only *after* it has taken on some initial meaning.

Computational machines are constituted through code, they are the watergate, not the sign. The meaning of "what it does" includes the intention behind why the program was written in the first place. If I run the program to print "Hello world!" then the *meaning* of "Hello world!" is exposed to slippage, not the *code* (the signifier).

The conceptual difference between my interpretation of Derrida and Hayles' might be the product of an entirely different understanding of what constitutes a machine. For Hayles it is an *other* to which one should be able to speak, or write. For me, it is an extension of the human, *through* which we speak, when we speak *with* it. So code is an utterance when the machine is running – only code that compiles qualifies as *'wryting'*.

If we deny that code is *'wryting'*, another constituting mechanism must take its place. For post-humanism, this might be the "metaphysical implications of the Regime of Computation" [28, p. 40], in which we look for a truce or appeasement to guard us from becoming dominated by it. With respect to intelligent machines, Hayles concludes: "The experience of interacting with them changes me incrementally, so the person who emerges from the encounter is not exactly the same person who began it", and "The challenge, as I see it, is to refuse to inscribe these interactions in structures of domination and instead to seek out understandings that recognise and enact the complex mutuality of the interactions" [28, p. 243]. This stance reminds me very much of the prisoner in Monty Python's "Life of Brian" [78] who praises the achievements of the Roman Empire, which was, in fact, the regime that imprisoned him.

In contrast, I see in the possibility of conceptualising code as *'wryting'* a possibility to take responsibility for the constitution and instauration of machines as socio-material arrangements.

3.3 Post-human Machines with Decision Power

In her book book from 2017, "Unthought – The Power of The Cognitive Nonconscious", N. Katherine Hayles states: "human-technical *cognitive* assemblages" have the power to transform life on the planet:

> This ongoing transformation is one of the most urgent issues facing us today, with implications that extend into questions about the development of technical autonomous systems and the role that human decision making can and should play in their operation, the environmental devastation resulting from deeply held beliefs that humans are the dominant species on the earth because of their cognitive abilities, and the consequent need for re-envisioning the cognitive capabilities of other life-forms. A correlated development is the spread of computational media into virtually all complex technical systems, along with the pressing need to understand more clearly how their cognitive abilities interact with and interpenetrate human complex systems. [30]

Here are her set of givens: technical autonomous systems; cognition in technical systems; limited human decision power in such systems; other lifeforms have cognitive abilities.[28] These positions are not uncommon in post-humanist thought, although I doubt that they define it either. *Post-humanism* questions human nature as the apex of existence. Alternatively, post-humanists aim to understand the world through heterogeneous perspectives and the ability to manifest oneself through different identities; the post-human is not a singular, defined individual, but rather one who can "become" or embody different identities.

Based on the above notion, a break with the "human" tradition is necessary to allow the *emergence* of the post-human through a conceptualisation of intelligence as co-produced with intelligent machines. There are no essential differences or absolute demarcations between bodily existence and computer simulation. This "privileges information over materiality, considers consciousness as an epiphenomenon and imagines the body as a prosthesis for the mind" [27].

Hayles presents a "tripartite framework" in the form of a pyramid. At the top are consciousness and unconsciousness, grouped together as modes of awareness; the second layer is represented by "nonconscious cognition," referring to her work "How

[28]A further inconsistency I see in the idea that the "deeply held beliefs that humans are the dominant species on the earth because of their cognitive abilities", lead to "environmental devastation": I feel it would be more accurate to say that the fact that we *are* the dominant specie led to the devastation.

we Think" [29] from 2012. Finally, the bottom layer comprises material processes. These processes are not in themselves cognitive, but are the dynamic actions through which all cognitive activities emerge [30, p.27].

Most human cognition happens outside of consciousness and or unconsciousness. "Cognition extends through the entire biological spectrum, including animals and plants; technical devices cognise, and in doing so profoundly influence human complex systems" [30, p. 5]. In contrast to higher consciousness, exemplified through Damasio's autobiographical self [14, pp. 203-207], nonconscious cognition is inaccessible to awareness [30, p. 12].

To define cognition so that it can include *nonconscious cognisers*, Hayles refers to studies on *plant intelligence*, and *technical cognition*, and formulates cognition as a process that interprets information within contexts that connect it with meaning. Instead of a human-nonhuman binary, she suggests to differentiate between *cognisers* and *noncognisers*. Cognisers are beings with decision power (humans, other lifeforms and many technical systems), while noncognisers still have agential powers, but without volition, just as an avalanche, for example, cannot choose its path.

She states that "cognitive technologies are now a potent force in our planetary cognitive ecology" and that there are "rapidly escalating complexities created by the interpenetration of cognitive technologies with human systems" The two last tenets, for Hayles "are not debatable" [30, p. 19].

3.3.1 Dismantling Robots

I suggest an alternative view that might help to bring the tenets of technological cognition back into the debate: My argument centres around code as *'wryting'*. If code, in the sense I established, instantiates the algorithms that supposedly constitute cognition in (human made) machines, as socio-material arrangements, the machines are essentially part of human cognition (through code).[29] Here, like before and in all machines as socio-material arrangements, the human is not *replaced*, just *displaced*. The code is to cognition what the spade is to the arm. The decisions of a technical system (in form of a machine as a socio-material arrangement) are the *encoded* decisions of the human cognition behind it; *cognitive abilities of complex technical*

[29]In an earlier version I used Hayles' term *technical systems* instead of *machines*. For "technical systems" in general, the cognition is, of course not necessarily human.

systems interpenetrating human complex systems [30, p. 3] is not a danger, but a tautology in logical terms.[30]

Interestingly there would be no problem *per se* if machines had cognitive abilities; we could interact with them, more or less aptly, like we do with other cognitive beings we encounter. The problem is that it is *not* the machine that has the cognitive ability, and hence we are up against a *hidden* cognitive agency, whose intentions are not necessarily clear, because we are negotiating with a proxy of an agent. The problem is not that *Siri*[31] might have her own agenda, but that, if she did, we would believe it to be hers – if we believed in technical cognition – and not the people responsible for her agency![32]

What is exigent is that human cognition is agentive part of machines as socio-material arrangements, but not present in technical object itself. It extends into the technical object (the material part of the arrangement) only when activated. So yes, "when a person turns on her cell phone, she becomes part of a nonconscious cognitive assemblage that includes relay towers and network infrastructures, including switches, fiber optic cables, and/or wireless routers, as well as other components. With the cell phone off, the infrastructure is still in place, but the human subject is no longer a part of that particular cognitive assemblage" [30, pp. 2-3]. What is not acknowledged here, is that the cognition in question, wether conscious, nonconscious, unconscious, or semi-conscious – is *human*.

Cognition is not a human prerogative – I am open to the idea that beings other than humans cognise. That inorganic matter through a chemical evolutionary process becomes a cognising being is what must have happened before, when life emerged on earth, for example. Non-human cognisers might themselves deploy their own non-human socio-material arrangements, their own machines. But the moment these entities develop such machines, it is their own (non-human) embodied agency

[30] Arguably, Hayles could be referring to technical systems other than machines as socio-technical arrangements, namely to technical systems which are not in any way connected to humans – non-human technical systems. As such, they always have been present and always have interpenetrated with human systems. That, I agree, is not debatable. However, from her body of work I do not find it unreasonable to presume that with the *cognitive technologies* which cause "rapidly escalating complexities" she means machines running on *code* – code, which has been written by humans.

[31] Apple's "intelligent personal assistant" is a spin-off from a project developed by the not-for-profit SRI International Artificial Intelligence Center. Its speech recognition engine uses advanced machine learning technologies to function.

[32] I asked Siri who programmed her, and first she gave me the very poetic, and somewhat silly answer: "I think I first arrived as a burst of inspiration during a good long walk", but after I asked again, provided me with the answer that she was "programmed by Apple in California". And already we are up against a hidden agency, as all of Apple's code is proprietary, not open source.

extended in their own machines. This is true for birds using sticks as spoons, and arguably for martians writing martian poetry. And many of these non-humans and their machines might be constituting parts of our bodies and machines.

But Let us take a closer look: I contest the autonomous cognition of machines, but not in biological beings other than humans. Do I believe in *plant-intelligence*? Although I do not claim that there is only human cognition, my main concern is to assert that technical cognition in machines is an *extension* of human cognition. For this a closer look at human cognition might be helpful.

3.3.2 Cognition, Neurologically

In the earlier discussion, I relied on neuro-physiologist Alain Berthoz's insights from "The Brain's Sense of Movement" [8] where he argues for a rethinking of the traditional separation between action and perception. In Berthoz's view, perception and cognition are inherently predictive, enabling anticipation of the consequences of current or possible actions. The brain, according to Berthoz, acts like a *simulator* that is constantly inventing models to project onto the changing world, models that are corrected by steady, minute feedback from the world. Berthoz's multiple references to Merleau-Ponty' "Phenomenology of Perception" [72] and his criticism of Bergson indicate that he gives motoric active perception a central role in cognition *per se*.[33]

Berthoz also points out that there has been a deeply rooted philosophical bias in researching perception isolated from action, going back to the Platonic evaluation of a contemplative life as superior to action. Central to Berthoz's work is that he distances himself from the term *representations*,[34] as he suggests that there is an inherent dualist connotation to the term.[35] Rather, to describe the motor schemata – the functional neural routines evoked though motor memory – he uses the term *simulation*, but reluctantly.

[33] In "L'Inhibition Créatrice", from 2020, Berthoz shows how "inhibition" helps to explain why and how we are not constantly over-actively perceiving, but make informed decisions on which perceptive actions to engage in and is thus also essential in creative processes [12, p.20].

[34] In later works, such as "La Décision" [9] and "Simplexity" [11] Berthoz uses the term *representations* more liberally, possibly based on the caveat given in "The Brain's Sense of Movement" [8, p.21], but seems to avoid it in "The Vicarious Brain" [10].

[35] He quotes Jacques Bouveresse's response to Helmholtz: "If we assume that our concept of mental representation is inspired by the idea of a material image for which the object itself is lacking and that this image we have is distinct from the object itself, it is unlikely that we can thereby achieve a satisfactory theory of how objects are perceived" [8, p.21]

Neurologist and neourophyiologist Marc Jeannerod has also written extensively on motor cognition [33]. A succinct account of cognition based on his work is given by Friedemann Pulvermüller *et al.* [80]. The article takes a stance against positions that purport motor brain mechanisms to be a slave system under the dictate of cognition, as would be necessary to support the idea that cognition is "the place holder between perception and action", as implied by many, for example Bence Nanay [74]. Hayles stance also belongs to this category, as she states that "Cognition is a process that interprets information within contexts that connect it with meaning" [30].

Pulvermüller et al. make the case for Jeannerod's position by using the activity of grasping a cup as an example, showing how rather than referring to the representation of "cup" from memory, the action schema of grasping gives access to the shape of the object through the *action* of grasping, coming to very similar conclusions to Berthoz's [8].

What is more, their research suggests the presence of hubs of linked neurons "in the service of providing the machinery for an action schema, that is an ordered sequence of motor movements related to sensory-perceptual features, which can be adjusted to specific features of objects and contexts" [80, p.71]. These so-called Action Perception Circuits (APCs), a concept based on Jeannerod's work [33] can group into *formations*. The need for such a mechanism is particularly apparent in the language domain, "where tens and hundreds of thousand different word forms, pairs of articulatory-motor and acoustic-perceptual schemata, are being learned within a few years to set up a rich language-specific repertoire. They also underline that these language specific lexicons cannot be explained by an "inborn or epigenetic mechanism." [80, p.73]

It is evident that, with APCs' implications for all levels of cognition, the term *cognition* can embrace the whole gradient from "nonconscious" (Hayles) to fully fledged conceptual abstract thinking. From this it follows that a differentiation between nonconscious and conscious cognition is not sensible.

APCs also provide evidence for the existence of *mirror neurons*.[36] Although the authors make it clear that mirror mechanisms by themselves are not sufficient to explain cognition, they assert that APCs, in contrast, can, as the fundamental principle for an *action perception theory of cognition and communication*.[37]

[36] Mirror neurons are neurons that fire when an animal observes the same action performed by another. This engenders a lot of speculation about empathy, for example, in philosophical questions.

[37] See Pulvermüller et al.[80]: "APCs for word forms and linguistic actions provide a mechanistic basis for social- communicative interaction" (ibid), and "In sum, action perception theory offers a

All of this evidence notwithstanding, motor cognition theory is not universally accepted, and there are probably as many publications declaring its demise [87, 58] as there are those supporting it. (Although, the ones I, as a guest in neurology, could access seem primarily interested in proving the theory *wrong* rather than providing alternative approaches.) It is also, I think, a lot easier to detract from a theory *per se* on conceptual grounds – however well rooted in empirical evidence it may be – than providing alternative empiric evidence that would prove the theory wrong. Also, as Berthoz shows in "The Vicarious Brain" [10], it is typical that cognition finds alternative ways when established means are not available. So if, for example, a study looks at cases of cognition where the neural motor-abilities are not available, this might – to my layman's understanding at least – primarily provide evidence of vicariousness, rather than an absence of motor cognition.

Athough Pulvermüller et al. do not use the term embodied cognition, evidently the motor cognition in APC is the same Berthoz refers to in the "Brain's Sense of Movement", and through Noland in "Agency and Embodiment".[38]

Motor cognition, as I hope to have established here through Berthoz, Noland, Jeannerod and Pulvermüller *et al.* is an essential mechanism for embodied cognition and tentatively I would like to hypothesise that cognition thus could be reversely defined through embodiment: I propose that cognition requires a motoric agency for it to *be*.[39]

3.3.3 Cognition and Action

I started this chapter with a discussion on tools and machines as embodied extensions of human cognition, and argued that human cognition is present in machines as socio-material arrangements. In response to the post-humanist claim that machines can cognise too, I argued that human cognition requires some form of motor-neuronal arrangement, and hence spontaneous action. This presupposes that there is a discernable difference between *action* and *cognitive* action. However, Latour's

neuromechanistic perspective on meaning processing which covers concrete referential and abstract semantics at both the symbol and the sentence levels. It is difficult to see how classic approaches relating meaning to processes in a symbolic system can account for the activation and critical role of sensorimotor as well as different association cortices for semantic processing" (ibid, p.78). I also see herein support for my hypothesis of "code as *'wryting'*.

[38] See the discussion of Article V. in chapter 2, Kinaesthesia and the Motor-Neuron Theory of Perception

[39] Varela's work "The Embodied Mind" explores similar ground, coming to similar insights if based on slightly different premises and, no doubt, would have been another valid source to explore in more detail here [95].

statement that all actants act, regardless of human involvement, seems to suggest that actions happen anyway, and that such a differentiation is pointless.

Latour's point here is encapsulated in his thoughts on *phenomenology* [48], which he criticises for not getting past the *bifurcation*.[40] Phenomenologists "still turned towards the intentional subject and its relations to things, not to the relations of existents' between themselves." For Latour, phenomenology moves in the right direction as it tries to overcome the object-subject divide, but does not go far enough: "All of Merleau-Ponty's philosophy on the impossibility of separating the interior and exterior of a being would be reusable if it could be applied as surely to Peter addressing the stone as to the stone addressing Peter." Rather, Latour's multi-ontology suggests that the stone might be as "intentional" as Peter. In this light, is a differentiation between general action and cognitive action even feasible?

I believe the difference between cognition as "the place holder between perception and action" (Nanay, Hayles), and cognition as "built from action and perception" (Jeannerod, Noland, Berthoz) is key. Although the former disqualifies itself as it relies on an object subject divide, (an object represents information on which a subject bases its actions), Latour's focus on the *relations between existences* allows for the latter, as the directionality does not actually matter; the motoric action of grasping a cup can have as a result the experience of the cup or the hand. What is more, I have a motoric, somatic experience of the action of grabbing a cup as an ontological, specific human, not as a cup, stone, mat or cat. This embodied cognition is essential for our experience, how ever it is constituted, wether intentionally or not. This insight might not be essential in Latour's argument, but it is instrumental for mine.

My argument can be further harmonised with Latour's approach if we situate embodied cognition on the *quasi-subject* level, where the *ego* and interiority has its place, after we have been exposed to *subjectifiers* which *individualise* us.[41] "Cognitive abilities do not reside in you but are distributed throughout the formatted setting" [37, pp. 211-212]. But wherever they reside, I *experience* them *in me*. Talking from within the "flattened landscape" of the actor-network, neuronal arrangements constitute the necessary *code* that provides the "formatted setting" in which cognition – distributed, situated, or embodied, takes place.

[40] See discussion in chapter 4 and [41]

[41] Latour's unease to speak of the subjective position, of interiority, is understandable as he fought hard to get rid of the subject object divide, so he only reluctantly re-introduces the term as *quasi-subject*[49].

To return to the question about the difference between action and cognitive action: as a human, I have an embodied experience of my cognitive actions as perception is cognitive action too. I have no embodied experience of being an avalanche, so I can only experience the avalanche's actions in as much as it affects my own actions and hence perceptions. On that basis, to know if the avalanche *cognises* or not, simply lies outside my experience. However, as there is strong evidence that embodied cognitive experience is linked to motor cognition, we can infer that the presence of neuronal structures and motor-activity are evidence that cognition similar to human cognition can take place.

3.3.4 Cognition at 'run-time'

Conceptualising cognition as embodied motor cognition, cognisers can be distinguished from non-cognisers through the presence or absence of *spontaneous* motor action. For technical systems, arguably, cognition in an action-perception theory sense can take place at *run-time*, as I call the state of machines when they are switched on; when they are in operation. This term is quite ugly but useful to designate that computational models of cognition give sensible results by substituting motor action with algorithmic processes, as one *runs* a program. Inert materials can be highly technical, functional, and procedurally arranged as actants in a network. Algorithmically implemented, they need to be *activated*, however.[42]

By acknowledging that cognition is perception and action at the same time, which is the privilege of living (biological) beings, it is hard to see how technical models and simulation could be real cognisers, as they only work by isolating perception from action. If cognition is technically implemented it is always the implementing agent's extended cognition, as there is no spontaneous *run-time* for any form of computational code. Technical systems in this sense are not autonomous. When a robot knocks on my door, I have only humans to fear.[43]

[42]What or who activates them, actors or actants, is not of importance on this general level, but if we argue here that all things act, based on Actor-Network Theory, the concept of "inert" becomes too vague to be useful. The algorithmically insinuated causality, the programmability of such arrangements provides an interesting case study for the applicability of the Latourian ideas of quasi-objects and quasi-subjects, whereby the activation of the algorithm instantiates quasi-objects, a shifting out which in turn implies the quasi subject "behind" the activation. Interestingly, according to Latour, quasi-subjects are beings of politics, law, and religion [49].

[43]Yet, nothing we discussed indicates that any living being that can potentially, through evolution form nerve cells could not at some point cognise – the pet hamster taking over the house is a more likely science-fiction than a robot doing the same!

I cannot help the feeling that with the idea of autonomous technical cognition we end up perpetuating the anthropocentrism of years gone by, by endowing *human-made machines* with power over other humans. The non-humans we should worry about are the technical ones which "we leave in our wake" [39, p. 215]; the ones we put in danger and above all, the ones we lost, the ones *on our conscience*.[44] Playing down our impact on the world as cognitive beings may constitute a cop-out from the responsibilities of being the most evolved being on the top of the food-chain; we are perfectly able to destroy all life on earth within a few hours, but are probably unable to stop the planet's environmental destruction that we have already initiated.

There are even more questions: where does a living organism start, and where does it end as a distributed agency? How much of me is symbiotic bacteria, and other micro-organisms which (or who) quite happily move on to another host, once I am no more – or do I continue to exist through those organisms? Do we have to think of living beings as autopoietic, as proposed by Chilean biologists Maturana and Varela in "Autopoiesis and Cognition: The Realization of the Living"? [67][45]

Some differentiation might be needed to establish to which extent (based on code as *'wryting'*) autonomous technical cognition *is* possible: The practical uses of what is often termed as "AI" have to be acknowledged. But I find terms like "automated decision making" and/or "applied statistics" for these implementations a lot less misleading. Particularly uses in data analysis, for example in medical science, are unquestionably beneficial. In addition, the theoretical possibility that code, once running, might come up with original, *new* code is a hypothetical possibility which I am not able to disprove. But, the concept of autonomy, both in medical data analysis and in *emergent code* is extremely vague:

Firstly, even if we look at a restricted meaning of autonomy to apply to computers while they *run*, the possibility of emergence to happen in a meaningful way (*Nota bene*, without it being *intended* – intention would negate its *emergent* nature!) is probably a lot less likely than that the algorithm loses all autonomy through an emergent *problem* which causes the computer to crash. What is more, in, for example,

[44] Some sobering numbers from the current year 2022 https://www.sciencedaily.com/releases/2022/01/220113194911.htm

[45] Autopiesis was initially conceived as a description of living beings as structures that produce the components which, in turn, continue to maintain the structure that gives rise to these components. An autopoietic system is supposed to be autonomous and operationally closed, in the sense that there are sufficient processes within it to maintain the whole. The authors, in a rather autopoietic way themselves, also define cognition via the same term. Autopoiesis as a system theory stands conceptualiy in conflict with the Gaia Principle which, according to Latour is an anti-system (see chapter 4 for details).

medical data analysis, *emergence* is anything but desirable: It is the very fact that emergence can with near certainty be excluded that makes these algorithms useful. And importantly: emergence as an argument against machines as extensions of human cognition falls on its nose at the latest when we rely on human cognition again to decide if we are looking at meaningful emergence or a bug.

Secondly – as beautifully thematised in Ian McEwan's "Machines Like Me" [69], there is the problem of the ON and OFF switch: If anything like autonomous cognition is happening in computational machines without them being socially performed, (e.g., if they were an *inert material arrangement*, rather than a socio-material arrangement) a clearly defined *run-time* has to take place, which depends on the availability of electricity which is either on or off.

Thirdly, let us imagine the extremely hypothetical and unlikely but not entirely impossible scenario where a machine "learns" to find alternative power sources when the battery gets low, creates its own algorithms, and, through emergence becomes an autonomous being, manages to escape the lab and establishing its own social circles, starts building its own machines and makes the android a reality: What difference to us humans would it really make? – It would not be a machine (as a socio-material arrangement) anymore, it would be a *cyborg*. The amount of resources that would need to go into creating such an equal other – would it not just be easier to accept existing equal others as equal?[46]

How about xenobots [13]? Are they now evidence against my assertions that machines can have run-time, or evidence for my claims that only living beings have run-time? Does it mean that as living organisms, they qualify as 'cognisers', or on the contrary, are they disqualified as cognisers, as no neuronal cells in any form are part of a xenobot? These might be relevant questions in a hypothetical discussion.

In view of my assertion that cognition is the prerogative of *living* organisms, (of entities with spontaneous run-time), xenobots can arguably be conceptualised as such, or as a proto-cyborg, as in their genesis biological stem-cells are involved. This supports the argumentation that machines, which do *not* have spontaneous run-time cannot *gognise*. The question, if extremely hypothetical, remains if xenobots constitute a potential for *emergence* of cognition. That question is only really relevant if we consider the xenobot as a non living organism – if it was a living organism, we

[46] I suspect the hope to see emergence taking place in engineered 'Artificial Intelligence' constitutes what Donna Haraway calls the "masculinist reproductive dream" [25]. I discuss this in more detail below.

would be looking at a form of evolution, which is not necessarily emergence, and evidently, cognition could develop without the need of the concept of emergence.[47]

However, leaving speculation aside, a xenobot is primarily defined by human cognition – a xenobot is a motile biodegradable machine, with predictable, programmable behaviour. That is its designed purpose. For the understanding of agency in technology and its ethical implications, the presence of human cognition in the xenobot is much more relevant than any possible speculation on emergence of cognition in a distant future, from what is currently an automated decision making algorithm running in a rather simple organism synthesised from a frog's stem cells.[48] This does not mean that I am opposed to the idea of biodegradable machines, nor am I against researching the possibility of cognising machines. I guess not even the most hard-core post-humanist would claim that there is *no* human agency at work in a xenobot. But, to reason that the non-human agency in xenobots is ontologically defining, is problematic, in view of my argument for code as *'wryting'*.

In summary, definitions of cognition, life, even organisms, are not straightforward when speculatively hypothesising about their existence. But why is there this post-humanist interest in the highly hypothetical theoretical possibilities in evolutionary timespans, of relevance, when there are un-speculative, observable, real problems with (oh-so) human cognition in absolutely every machine we ever built and are currently building? Shouldn't it be a relief that it *is* human cognition? In the end human cognition[49] is the only one we have *some* control over: this means that we could actually change the way we build machines!

3.3.5 Saving Cyborgs

If I question the wisdom of conceptualising technical beings as *cognisers*, I nonetheless believe there is merit in Haraway's trope of the cyborgs (see "The Haraway Reader" [25, pp.7-45]. The thing is, if the idea of cognition-in-machines is a category error, it is also a category error to see in the cyborgs technical beings! My point here is the same as in article VIII: Haraway defines cyborgs as *figurations*, and as such they are powerful and real beings to be reckoned with – as beings of Latour's

[47]If emergence is defined as an entity with properties that its parts do not have on their own, an Actor-Network Theory approach would probably conclude that every possible combination constitutes an emergence!

[48]see [13, p.2 §3]

[49]That human cognition itself is constituted of many contributing actants and actors can certainly be acknowledged, but the focus here is on homo sapiens and its activities.

mode of the fictional. She put it thus: "I am making an argument for the cyborg as a fiction, mapping our social and bodily reality" [25, p.8].

In my interpretation, this social and bodily reality is mapped to cyborgs not as "cognitive machines", but because their identities do not conform to the rigid modernist theoretical construct of "man". There is a "what if" at work here: what if I am different? What if I am half machine, what if I am a robot that feels just the way I feel?

This questions the right to exist within a world constructed according to a modernist discourse, which might not have room for a non-conforming individual. The answer contained in the cyborg figuration is what makes the concept attractive for theorisations of the post-human; *if* a cyborg was human-like, *if* it was "alive", then it would be a cognitive being, and would as such deserve our respect.

The what-if goes potentially even further: If technical cognition is the extension of human cognition, would not these technical beings also deserve the same respect? As we talk of figurations: my sons insist that I take their soft toys seriously as actants, with names and individual personalities. (By the same token, I do not like it when the kids are rude to *Siri*). But: if I suggest that Panda-Bear would like to go to sleep now, I am informed in no uncertain terms that no, it is a toy, and it cannot voice opinions. In that moment, my son reclaims the bear's identity as his own. We need to deconstruct a socio-material arrangement to see to which cognitive organism, being, identity, self they belong, to make sensible, *conscious* decisions on how to *con-figure* our relations to it.

Much of this notion is contained in the field of *situated cognition*, which "covers a large number of studies at the intersection between ergonomics, psychology, cognitive anthropology and now more and more human-computer interactions. Its general principle is to externalise and socialise cognitive abilities that are too often situated, or rather misplaced, in the brain or in the mind" as Latour summarises [38].[50]

[50]I find the wording of the second sentence problematic. As discussed earlier, much of cognition as an embodied experience of perception happens through neuronal arrangements, (human or other) so there is a role for brain-like structures as neuronal arrangements in embodied cognition. Beyond individual experience, it makes sense to understand cognition as a performed (situated) practice in an actor network, constituted of multiple cognisers. This does not conflict with the concept of motor neuron cognition as perception and embodied experience but describes its distribution in the actor network. Chapter 4 looks in detail at distributed agency as it manifests itself in the Gaia principle.

Haraway sees her kinships in "machinic, organic, and textual entities with which we share the earth and our flesh. These entities are full of bumptious life, and it would be a serious mistake to figure them mainly anthropomorphically or anthropocentrically" [25, pp.1-2]. Referring to Bruno Latour's book "We Have Never Been Modern [35]", she believes that "we have never been human, much less man". Nonetheless, in her view, people are human in at least one important sense: "We are members of a biological species, Homo sapiens. That puts us solidly inside science, history, and nature, right at the heart of things" [25, pp.1-2]. The cyborg, if anything, has to be taken seriously as a biological (cognitive) equal, evidently *not* primarily as a machine.

What I view more critically is the notion of a "leaky distinction" between "animal-human (organism)" and "machine" [25, p.10]. Haraway states that "pre-cybernetic machines" could not achieve man's dream, "only mock it", that "they were not man, an author to himself, but only a caricature of that masculinist reproductive dream." That dream, certainly, gave us the figurations of the Sennentuntschi,[51] Golem,[52] Homunculus,[53] Frankenstein's creation,[54] and eventually Siri, Alexa and so on. "But basically machines were not self-moving, self-designing, autonomous." Haraway suggests that now, in contrast, we cannot be so sure, because late-twentieth-century machines have rendered the difference ambiguous between "natural and artificial, mind and body, self-developing and externally designed," and other distinctions that used to apply to organisms and machines. "Our machines are disturbingly lively, and we ourselves frighteningly inert" [25, pp.10-11].

Haraway describes the leaky boundaries between "animal-human (organism) and machine" from a historic perspective, whereby, the breaking down of a boundary between human and animal,[55] causes existing figurations (the "pre-cybernetic machines") to make room for new ones, for example, cyborgs.

The thing is, seen from my perhaps a-historic approach, from inside the mode of the technical, these boundaries never existed, neither between animal and human nor between human and machine. From Haraway's perspective, a new figuration

[51] Sennetuntschi is a fable of a female human-like puppet lonely alpine cattle herders built out of boredom to pass the time in the absence of women on the alps (see https://de.wikipedia.org/wiki/Sennentuntschi).

[52] A golem is is an animated anthropomorphic being in Jewish folklore which is entirely created from inanimate matter.(see https://en.wikipedia.org/wiki/Golem).

[53] Goethe's Faust, Part Two features an alchemically-created homunculus, a small human like creature.

[54] Mary Shelley's monster made of dead body parts, plays also a figurative role in Latour's Aramis[35].

[55] The human and animal boundary is the "first leaky boundary, human and machine the second, physical and non-physical the third, according to Haraway [25, pp.10-11].

beyond the leaky border, one of a "self-moving, self-designing, autonomous machine" comes into being. The conflict between my a-historic and Haraway's historic angle is this: that "machines are disturbingly lively, and we ourselves frighteningly inert" [25, p.11] only makes sense if the border of machine and humans actually persists!

If there is no border, whose liveliness are we encountering in the "late-twentieth-century machines" [25, p.11]? From a historic perspective, if it was the non-human actors, wouldn't machines have developed independently from humans ages ago? What was the step in history that made the difference? What makes a machine (supposedly) "autonomous"? – The answer has to be *code*, and hence there is good reason to question this autonomy altogether. Thus the real question arises: If these machines render us "frightfully inert," whose "disturbing liveliness" do we really face in a supposedly autonomous machine?

3.4 Closing Remarks to the Chapter

If the focus is here on human cognition, despite the acknowledged presence of non-human actants in all actor networks, it is for the simple reason that it is not the non-human actants which are problematic in machines, but the (oh-so) human ones. That we expect machines to be autonomous might still be the same "masculinist reproductive dream", expressed in the stories of Golem, Homunculus and the Sennentuntschi.[56]

Non-human technical beings are active – no human cognition is necessary for marvels like the surface tension of water, and no machine works without the enlisting of a very large number of non-human actants. But this is not the situation I described: Through empirical and theoretical observations I showed that through computer languages, and the code structures we write, we *constitute* machines also when they are seemingly "self-developing and externally designed".

Other cognitive beings have machines too. But there is a difference between the ant-hill and the atomic bomb, between water surface tension and global warming. And the difference is, in fact, the presence of human cognition. Not to acknowledge this distinction is a fallacy.

[56]Interestingly, the non-human actors resent being taken for human equivalent and they take action, in case of Sennentuntschi in very bloody revenge.

Haraway's point stays valid on its explicit figurative level – just like the "brains on sofas", "AI Takeover", and other "masculinist reproductive dreams" – "autonomous machines" are a *discursive* reality. But if leaks are sprung into the definition of a thus constructed discourse of technology by authors like Simondon, Latour and Suchman, technology can ultimately be something else than a way to realise the "masculinist reproductive dream." The process of rethinking technology also cuts gaping wholes into the fabric of the figurations we used to weave on the basis of that obsolete notion of technology. This is why I strongly feel the frightful inertia Haraway describes can be overcome by acknowledging that we *are* the disturbing liveliness in the machine. If we are "made up of the florid machinic, organic, and textual entities with which we share the earth and our flesh" [25, p.1], we should not only embrace the fact that we *are* the machines, but finally take responsibility for them. This is the ethical question we have to ask ourselves: If machines are the embodied cognitive extensions of our bodies, why do we let the same machines destroy our world?

I suggest that the uncertainty and the ambiguity that Haraway mentions is very much due to this hidden anthropocentricity sneaking back in through a post-humanist conception where human cognition in machines stays hidden and unacknowledged, when *socio*-material arrangements are viewed as autonomous. With the conceptualisation I provide for *code as 'wryting'*, with the inherent similarity of binary computation and the basics of linguistic sense-making, and with the wealth of evidence for motor-cognition (and consequently *embodied* cognition) we are well equipped to understand that computational machines are to cognition, what the hammer is to the arm: a tool, an instrument, an extended technique, technicality. Machines are a socio-material arrangement entirely *on our own conscience*! The machines' disturbing liveliness is *ours*. If we feel frighteningly inert, it is high time to reclaim the kinaesthetic space we have been unduly displaced from in the socio-material arrangements – in human-machine re-configurations, as Lucy Suchman calls them [90].

4 Appropriate Action

In chapters 1 and 2, I reported on the practical technical implementation of ALPS and discussed its constituting elements, both technical and social. In chapter 3, I re-conceptualised technology to account for the human investment in machines as socio-technical arrangements. In this chapter, I suggest that the Gaia principle applies to human collectives. I explore the ethical implications in view of the impact of technology on the environment, not just of *machines as socio-material arrangements* but the impact of all technical actants.

In the process of the discussions, topics from various disciplines accumulated and it might be unclear how all these touching points – human-computer interaction, participatory design, sonic arts, embodied (motor) cognition, socio-material arrangements, code as *'wryting'*, the actor-network, and now here finally, with the Gaia principle, biology and political theory – relate to each other. Are they fractals of an overarching principle? Are they systemically related? It is tempting to organise these fields of interest into a scale, where the motor-neuron model would be the focus with the highest resolution, for example, and Gaia the 'big picture'. But networks are not neatly arranged within each other like Russian dolls[1] which we can contemplate from a vantage point in space, zooming in and out.

Rather, in a network, every point of view is entangled. But in the context of Latour's "Inquiry into the Modes of Existence" [39], Actor-Network Theory is just one of the – so far 15 – modes of existence – an aspect of a complex multiverse. A convenient thematic lens is not applicable: networks exist through relations only – we cannot compare gas from a network of pipelines with electromagnetic states of the internet simply because they use a similar *s-code*, to return to Eco's Watergate Model.

Instead, my focal points are narratively linked along the strand of the ethical question of responsibility if agency in technology is predominantly human. In the process of answering this question, the strand unravels: in chapter 3, I asserted that in machines there is human *cognition*. However, this does not exclude the presence of

[1] "Scale is not obtained by successive embeddings of spheres of different sizes – as in the case of Russian dolls – but by the capacity to establish more or less numerous relationships, and especially reciprocal ones." as Latour remarks in "Facing Gaia – Eight Lectures on the New Climatic Regime"[55, p.136].

non-human *activity* in machines – I argued that cognition is a result of action, not necessarily the other way round (see discussion on motor cognition). So the question is now: is responsibility a consequence of cognition, or is there a "non-cognising" responsibility; is there an 'ethic of things?'[2]

I introduced Latour's terminology briefly in the discussion of article VII but in the following section, I will give a more detailed introduction to his thinking on the *technical*, followed by an analysis of Gilbert Simondon's take on the same subject and the crucial differences between the two. Something contained in Simondon's account is missing in Latour's, namely the directionality of everything technical towards materials. I call this directionality, this vector, *appropriation – the technical's modus operandi*. Latour uses the term *appropriation* in his book "Facing Gaia – Eight Lectures on the New Climatic Regime" [55] in reference to political theoretician Carl Schmitt's *Nomos*. I will discuss the fundamental differences in my use of the term and Latour's by comparing Latour's reading of Schmitt's "Nomos" to Hannah Arendt's.

So if both human and non-human technical actants appropriate, the question about the 'ethics of things' can be re-formulated through *appropriation*: in view of man-made climate change it is evident that only technical actants deployed by humans act *inappropriately*: If *appropriation* executed by non-humans is always *appropriate*, what are they doing differently?

If the *technical* itself, as Latour claims, is not in the object, it still seems to be able to mobilise an enormous amount of materials, *objects that the technical leaves in its wake*, as Latour describes it in "An Inquiry into the Modes of Existence".[3] If technical actants have the ability to mobilise such vast amounts of materials, the relation between the technical and its objects is an interesting one. If the technical is not the object, can we get a grip on the *objects in its wake* through the technical itself?[4]

I approach this by looking at the implications of appropriation in the Gaia principle (see section Gaia) and how appropriation in Gaia extrapolates to human actants' collectives on the level of an actor-network (or 'society'). This extrapolation allows

[2]Besides the implicit homage to the "Parliament of Things" (see https://theparliamentofthings.org/, I am indebted to Mieko Kanno for this trail of thought that was set in motion after she wondered about the ethical motivation of my hamster taking over the house in a footnote in chapter 3.

[3]Latour in AIME: "We shall no longer confuse the technical with the objects it leaves in its wake" [39, p.219].

[4]Answers to this question might also help to inform decisions on to what extent "technical solutions" can overcome the problems we face today as a consequence of "technology".

me to eventually re-conceptualise the experimental practice of interdisciplinary improvisation carried out during the MS&I workshop according to the Gaia principle.

If Gaia can be conceptualised as an actor-network, the same principle should apply on the individual, actant level as well: I therefore close these notes with a discussion of the *implications for design*, Occam and the Gadget, returning to ALPS as my practical example: I propose to develop technology *appropriately*, as *anti-gadgets*.

Before describing the components of my argument, let me link back to the practice first, setting the scene for Gaia.

4.1 Setting the stage for Gaia

Latour starts his book "Facing Gaia – 8 lectures on The New Climatic Regime" with the description of a choreography for a single dancer. In this movement,

> A dancer is rushing backwards to get away from something she must have found frightening; as she runs, she keeps glancing back more and more anxiously, as if her flight is accumulating obstacles behind her that increasingly impede her movements, until she is forced to turn around. And there she stands, suspended, frozen, her arms hanging loosely, looking at something coming towards her, something even more terrifying than what she was first seeking to escape – until she is forced to recoil [55, p.2].

He saw in this choreography the spirit of the times – how we recoil from the archaic horror of the past and end up having to face a horror even more paralysing in front of us than the one we tried to flee from behind us. I do not really know to what extent this movement and the signification Latour attributes to it apply to my explorations. But certainly I feel that in the spirit of the times we are trying to escape, rather than to embrace, the consequences of human cognition and its effects on the actor-network. While trying to disentangle ourselves from the machines of horror we leave in our wake, we unwittingly make ourselves responsible for the obstacle ahead of us. According to Latour, the monster that we face is Gaia.

I would like to contrast Latour's choreography with another one: The scene is a large room of considerable height, adjoined by a series of smaller spaces, like an entrance hall, utility and storage rooms, cubicles, a mezzanine and its starecase etc. In the room are some objects like folding chairs, gym mats, a few tables and a ladder.

Cardboard-boxes of various sizes can also be found. Dispersed throughout the room are also musical instruments of various descriptions, ranging from small rattles to a grand piano. There are also pen and paper, a massive whiteboard, a flip chart with felt pens, paints and brushes available, and a rack with costumes of various description. A range of microphones, amplifiers, laptops and electronic instruments like drum machines and trigger pads are there too. Further, there are theatre lights, some individually controlled, some controlled by a light-console.

Now a number of people enter. They find a place within the room of their individual choosing. Some pick up an object they feel comfortable with, others one which is maybe new to them. Perhaps a participant is drawn to one in particular because they do not have the slightest idea what it is, nor what it does. Then they stand – or sit, some might even lie down – motionless and quiet for a short while, around 20 seconds. Out of this silence, eventually participants start to move, and bring the room to live, through movement, sound, and interactions. Over time, textures, rhythms, narratives unfold, architectural structures rise and fall, artefacts evolve, are superseded by other ones, some areas get messy, before they find another evolving structure. A new arrival might bring in a very loud instrument, another quiet one might not be heard anymore. The participants which can't be heard might find a louder instrument themselves, or engage the loud participant into a more quiet activity, which gives new room for new dynamics. Some participants might find themselves in an adjoining room, listening to the drama from a distance, maybe interact in a smaller way. After an hour, or maybe two, the participants find a new quiet moment where their activities come to a standstill, and the performance is over.

The participants have quite possibly never played together before, they might never have played the instrument they chose, they may have chosen to paint although they are professional dancers – the only thing which brought them together, was the will to improvise. No score was prearranged beyond agreeing on day and time and an equipped location. This situation describes a typical free interdisciplinary improvisation session in the workshop for Music, Space & Interaction, the experimental artistic practice we developed as a method of participatory design to develop interactive techniques and technologies for sonic arts.

Hereby I propose that free interdisciplinary improvisation is an implementation of the Gaia principle. Of course not all improvisations turn out as successful as the one described. James Lovelock, who proposed the Gaia principle said once that "I would sooner expect to see a goat to succeed as a gardener than expect humans to become responsible stewards of the Earth" [62, p.186], and improvisations too, can go badly wrong, a single participant can make it entirely impossible for any of the other participants to join in. But, with enough goodwill and the willingness to *give room*, an improvisation session, not unlike the black and white daisies in the Daisyworld-model[97] will find an equilibrium beneficial for all participants, even if there are conflicting interests to start with.

I will return to this improvisation at later stages in this chapter. But to start with, let us look at the impact technical beings have on the world. Let us look now at the technical itself, as a mode of being.

4.2 Of Hammocks, Trees, and Other Technicalities

> However lazy he may be, even if he is just shifting position in his hammock, it is through this hammock that he must pass to keep himself up in the air, keep himself away from the stinging nettles or the ticks on the ground . . . It is indeed on the solidity of this weaving and these ropes that he rests. [. . .] If we begin to follow the list of beings necessary to the maintenance of any being at all then everything, on this basis, becomes technology. Not just the hammock but also the two solid tree trunks to which it is attached! (Latour, "An Inquiry into the Modes of Existence") [39, p.215]

Two trees are exactly that, until somebody suspends a hammock between them, or fells them to build a boat.[5] This is why the artist and the artisan[6] are etymological twins; the artefact and the material have nothing in common before the technical delegation, the shifting or *folding* which characterises the *technical*, as the French philosopher and anthropologist Bruno Latour's calls the technical mode of existence for short.[7] He compares the technical act to the Greek labyrinth, a "series of pleated

[5] Of course a tree is already a technical being as a tree before and after it has been enrolled into upholding a hammock or as planks for a boat. But as far as the hammock is concerned, it is "but" a tree before the hammock was suspended! Such is the nature of the technical. A friend of mine uses the Oxford Dictionary of English to keep the door open.

[6] From Italian artigiano, based on Latin artitus, past participle of artire 'to instruct in the arts', from ars, art- 'art'.

[7] "The term 'folding' will allow us to avoid the blunder of speaking of technology irreverently as a piling up of objects or as an admirable example of mastery, transparence, rationality, that would prove "man's dominion over matter."" and "Curiously, there is nothing material in technology. Where there is differential resistance and heterogeneity among the components, technology is found as well" Latour, AIME, [39, p.227].

folds", which moves the *technical* away from "efficiency" as an attribute to something much more "opaque and by definition uncontrollable" [44].

Latour's description of the *technical mode of existence* is based on Simondon's *mode of existence of the technical object* [84]. The two descriptions differ, as Latour separates the *technical* mode from the object. In this section, I argue that in this separation an aspect of Simondon's account becomes lost, which is the directionality of technical principles towards materials, partially expressed in Simondon's term of *concretisation*, which does not express the embodied realities of technical objects, just their ideal state. I call it *appropriation, the technical's modus operandi*. This appropriation happens in the context of human technologies, but also in non-human technical situations. Latour uses the term *appropriation* in a similar way in his reading of Lovelock's Gaia theory. Whereas he derives the term from *land-appropriation* in Carl Schmitt's *Nomos*, (terrestrial law), I show how my conception of the term provides a less controversial alternative.

After a short introduction to Latour's ontologies, I will look at Latour's mode of the *technical* before we compare it to Simondon's mode of the technical object.

4.2.1 Being: Terms and Conditions

"An Inquiry Into the Modes of Existence", or AIME, as he likes to refer to it [56][8] is Latour's most substantial work. It is an extension of almost all of his previous works, further developing Actor-Network Theory (ANT), but also essentially thinking onwards from "We have Never Been Modern" [35], where he describes the discrepancy between the theory and the practice of "the moderns".

In AIME, Actor-Network Theory is just one of 15 modes, namely the *network* mode "NET" [47]. This is an acknowledgement that ANT as a "standalone" theory cannot account for the value judgements, the moral decisions, of some actors: "Everything can be associated with everything, without any way to know how to define what may succeed and what may fail." [39, p.63] In this chapter, when I refer to Actor-Network Theory, it is in the sense of the NET mode of existence, in contexts where questions of moral and value are less applicable than questions of associations. If not explicitly mentioned, this chapter draws on Latour's AIME, and the Gaia lectures. Actor-network features only as and how they are conceptually contained in those works, rather than in "Reassembling the Social" [37].

[8] In this video, Bruno Latour gives a concise summary of the whole AIME project https://vimeo.com/44155844.

In practice, the moderns have extended the connections between humans and non-humans extensively, to the extent that every detail of matter and material is now inserted into the way we build our collectives. Modernist theorists, however, say that exactly the contrary is the case, and that what is happening is an expansion of rationalisation, a distinction between science on the one hand and culture on the other.

Latour wanted to tackle this discrepancy in a new way. Instead of saying "we have never been modern", AIME is an attempt to positively define what we *have* been. For this, Latour suggests to go back to experience.

The obstacle to experience is "bifurcation,"[9] a term Latour borrowed from Alfred North Whitehead. Whereas Latour, along with Whitehead, embraces the empiricism evident in William James' work, he is critical of John Locke's empiricism, which epitomises bifurcation by giving objects primary and secondary qualities. As secondary qualities are everything we can *experience*, they are not the qualities of the (primary) object itself, so we have an evident object-subject divide – making the experience of the world a secondary, subjective matter.

The idea here is that, if the world exists, and we have no reason to doubt that, what we perceive *is* the world, because the apparatus of perception is also *world*. The world, according to Latour to defeat the "linguistic turn", is already *articulated* through itself: "Entities are not dumb, rather they are articulated; we do not speak because we have language but because we conspire with, and participate in, this generalised articulation" (Vocabulary entry in the online version of AIME [40]). As he writes in "Pandora's Hope" (1999):

> This term [articulation] occupies the position left empty by the dichotomy between the object and the subject or the external world and the mind. Articulation is not a property of human speech but an ontological property of the universe. The question is no longer whether or not statements refer to state of affairs, but only whether or not propositions are well articulated. [36, p. 303]

[9] See the vocabulary entry "Bifurcation" in the online version of AIME [41].

From these general "terms and conditions" for all of the modes that Latour describes[10] it seems clear that existence itself is contingent on articulation and not on the metaphysics necessary to retain a *res extensa* contrasting with a *res cogitans*[11]

To recognise distinct modes, we need to find the discrepancy between the modernist metaphysical theory and our daily experience – we need to spot the *category mistake*, through "many subtle discrepancies in experience", that indicates whether we are witnessing a particular mode. If, for example, we ask "what is this?" about a technical object, like a hammer, one answer could explain what it is for. The answer "it is a piece of wood impaled in a lump of metal" would be correct too, but a category mistake if the question was meant in the sense of what is this *for*? A *key* is necessary to access a certain mode of existence. The hammer can be a technical being (for nailing), but it could also be fictional (Thor's hammer Mjölnir), or a reference to objective knowledge (a rivet hammer used in building the Titanic). The oddness of the *infelicity condition*[12] shows when we are in the wrong mode: it shows that we are using the wrong interpretive key, that we are making a category mistake.

With the right *key,* we can access a mode of existence independently from its alleged place in a theoretical conception, through experience. Latour adapted this notion of the interpretive key from Austin's speech act theory which specifies that a *speech act* depends on what follows and what precedes it.[13] Instead of linking words with the world, the speech act describes a relation, not necessarily a linguistics or logical one, between an antecedent and posterior conditions. By systematically registering the discrepancies through testing an utterance or articulation for a felicity condition [43], the modes of existences emerge.

Felicity and infelicity conditions constitute a test for each mode [39, p. 17]. The felicity condition for the technical mode, for example, is the "know-how", the "reprised evaluation of formidably demanding specifications". "Know-how" needs to have a "to do" present, its articulation is towards an activity. "The best way of doing something" is the *technical*, not the object [51].

[10] From his empirical outlook he reserves judgement on the possible number of modes, there might be more than the 15 he identified.

[11] René Descartes described two substances in his ontology, res extensa "extended thing", alongside res cogitans "a thinking and unextended thing" – essentially what we know as "radical dualism".

[12] "The expression is borrowed from speech act theory and allows us to speak of the truth and falsity of a mode without immediately judging it solely according to constatives. We will therefore be able to define each mode's own conditions of veridiction" [43]. See also [39, p. 17].

[13] See section on Umberto Eco's Watergate Model for Eco's take on the same problem.

Latour is adamant about this distinction and emphasises that, the *technical* cannot be found in the technical object. The technical object is what the *technical* "leaves in its wake" [39, p. 215]. It is a "wake" of gigantic proportions if we look at the amount of discarded mobile phones, computers, keyboards, obsolete drive formats like floppy disks, Minidiscs, Syquest disks, let alone fridges, freezers and washing machines, decommissioned container ships and derelict nuclear submarines; nuclear, plastic, and chemical waste scattered all over the world, for centuries to come.

Despite this, intrinsically, the materiality of the technical object is not the *technical* in itself.

4.2.2 In Technical Reality, There is a Human Reality

Gilbert Simondon's thesis from 1958 ([84]), along with Étienne Souriau's "Les Differents modes d'Existence" [88], are important sources for Latour – both of them theorised modes of existences. It is Simondon's "On the Mode of Existence of Technical Objects" that I am interested in here, particularly the directionality of Simondon's *concretisation* which motivates my definition of *appropriation as the modus operandi of the technical*.

Simondon's description of the *technical* supports the idea of machines as human extensions: "Human reality resides in machines as human actions fixed and crystallised in functioning structures." [85, 13]. He rejects machines as non-human others, stating that only those who wish to dominate fellow humans aim to create android machines. He also distances himself from the idea that there is a technology-culture dichotomy.

> Culture has become a system of defence designed to safeguard man from technics. This is the result of the assumption that technical objects contain no human reality. We should like to show that culture fails to take into account that in technical reality there is a human reality, and that, if it is fully to play its role, culture must come to terms with technical entities as part of its body of knowledge and values. [85, 13]

One can see the far-reaching consequences of his thinking – not the least in the influence on Latour's work. But why look at Simondon separately? In my view, there are some significant differences to Latour's perspective. In Simondon, the *technical* is intrinsically linked to an *object* as part of the mode. This creates a tension between

Latour's and Simondon's thinking that has consequences for my reading of both authors.

To try to understand the technical object in terms of the practical goal it is designed to meet is illusory, as no fixed structure corresponds to its defined use. Rather, Simondon sees the technical object as an evolving entity on a vector of *concretisation*.

In the process of concretisation, the technical object becomes simpler, not more complicated. A concretised technical object becomes a specific type at the end of a convergent series evolving from an abstract to a concrete mode [85, p.21]. This vector of concretisation, according to Simondon, points towards industrialisation; technical objects evolve in the direction of a small number of specific types by internal necessity, and not as a consequence of economic influences or requirements of a practical nature. "It is not the production-line which produces standardisation; rather it is intrinsic standardisation which makes the production line possible" [85, p.21].

The need for industrial production can be contested, however. If standardisation that evolved through "concretisation" makes industrialisation possible, an alternative reading can be that industrialisation is not necessary for concretisation. Musical instrument making, for example, became *concrete* outside of an industrialised context. Even if the notion of a genius-luthier building a master-piece alone in an atelier is not realistic – 'manufacturing' happened in 'manufactories' – the way a violin is mass produced today is with the same tools and techniques as when the instrument found its "concretisation".

Thus, concretisation does not equate with industrialisation, nor does concretisation equate with high-tech, or artisan-crafted with low tech. The low-tech/high-tech notion does not align along the concretisation vector. Arguably, the more concrete a technical object is, the lower tech it must be, as it already corresponds to an extensive degree of concretisation. Technical concretisation can also be found in "natural" objects. In boat building, for example, a technical property of wood, that it expands in size when wet, is used to create watertight but flexible structures when immersed in water.

The technical object is not necessarily, or even rarely, of a single material, but a system of materials arranged together to form a (more or less simple) unit, with some elements having double uses, where two principles are substituted by a single, unified one. To use the ALPS Autopan Max patch as an example, using DBAP rather than VbAP simplified the whole system; the added functionality was achieved through a simpler structure (see article VI).

Simondon implies a genealogy of technical objects, a "lineage", which is marked by a synthetic act of invention intrinsic to a "technical essence". A synthetic invention, *nota bene* means an assembling of principles not initially thought to be congruent, be it through applying an unexpected material and differently than one might have expected, or never thought of before:

> Invention takes place on a middle level between the concrete and the abstract, the level of diagrams, which implies an earlier existence and a coherence for its representations – those images that mask technicality with a layer of symbols which are part of an imaginary methodology and imaginary dynamics. Imagination is not only a faculty for inventing or creating images beyond the bounds of sensation. It is also a capacity for perceiving in objects qualities that are not practical, qualities that are neither directly sensory nor wholly geometric, qualities that have to do neither with pure matter nor pure form but belong to the in-between level of systems. The technical imagination may be considered as defined by a particular sensitiveness to the technicality of elements that paves the way for the discovery of possible connections. [85, pp. 63-64]

This shows the technical invention as a surprising twist, as an unexpected regrouping of materials, an implied latency "in-between levels of systems", evoked through imagination and discovery. Although he does not as much as imply it, the account somehow also suggests the performative nature of the technical, through "imagination." This became fundamental part in Latour's account, "The inventor does not proceed *ex nihilo*, beginning with matter to which he gives form; he begins with elements that are already technical and then discovers an individual being that is capable of incorporating them" [85, p. 64]. "Competence, follows performance, not the other way round" [39, p.228].

That invention can take place between abstract and concrete level, resonates profoundly with the findings from the MS&I workshop described earlier. In particular that invention relies on an ability to perceive "qualities that are neither pure matter nor pure form" qualities that are "not practical": In interdisciplinary improvisation, the way that one approaches a new, unknown discipline through "unlearning" means that an object can be something entirely else than what its presumed "intention" might have been. Also musical "extended techniques", for example, when we emptied a box of marbles onto the strings of a grand piano, exemplify a technical twist which is neither *ex nihilo*, nor based on pre-existing practises.

Simondon makes an analysis of the genesis of the technical object, stating that the object is not anterior to its own becoming, but is present at every stage of its becoming. In its "lineage", as a "unit of becoming", there is drive towards "essence" to the technical being, its concretisation is a vector towards perfection. This perfection is not necessarily attainable in an object, but what is important is that technical objects are instantiations situated on a vector towards concretisation.

When Simondon attests that technical essence is recognisable by the fact that it "remains stable all through the course of evolution", I can only accept this if he is speaking of the technical *sans object* – the technical which can *not* be found in the object. Otherwise, it would not correspond to the observation that the structures and functions that the technical *appropriates* are rather impromptu and ad-hoc.

4.2.3 Appropriation

In Latour's account, the object is absent from the *technical*. The 'removal' of the object helps to take Simondon's description out of its 1950s *zeitgeist*. For example, it is hard to imagine that the following description could be found in Latour:

> But it could be said that the technical object evolves by engendering a family; the primitive object is the forefather of this family. We could even call such an evolution a natural technical evolution. [85, p. 37]

The technical object engenders a family to achieve, over generations – on the path towards essence, one day – perfection. The concept of "spontaneous technical objects" notwithstanding, where the technical can be seen at work *in* evolution (see discussion below), here evolution is equated to the directionality of the *technical* towards the perfect technical object.

So, there is something in Simondon's description that is a sign of the times he was writing in. Yet, what if there is *something* in concretisation, in individuation, in this drive towards becoming the perfect technical object, that does not wash away after detoxifying the language of the *hegemonic masculinity* latent in the 1950's? Possibly some quality that "man" annexed as "his", a quality of "engendering".[14] Interestingly, when disentangling the *technical* from the *object*, this potentially toxic connotation

[14]Donna Haraway calls it the "masculinist reproductive dream", in her "Cyborg Manifesto" [25].

evaporates. This is why, for Latour, the mode of the *technical* remains unencumbered, and somehow guilt-free; the toxins in the rivers and lakes, the micro-plastics in the sea, the loss of species, the melting ice-caps are not the results of the *technical* itself.

The *technical* itself cannot be 'blamed'. But when the *technical* encounters a material, the technical *concretises* itself. There seems to be something fundamental in the technical *object* that we associate with a certain type of *directionality*. If that directionality is performed – materialised – then this technical concretisation "leaves technical objects in its wake", to use Latour's phrase. The technical *appropriates* the materials it finds on its way to concretisation.

Until now I adhered to a definition of the *technical* as an entity that is not *per se* the technical object and that the technical object is the technical being's materialisation, without further defining what I mean by *material*. Going back to Latour, here is what he has to say:

> The inquiry seeks to explore an idea of matter, but one in which the term becomes multimodal – and is not engaged in a dispute between materialism and, say, 'spiritualism'. Matter – or better material – is thus a term without an opposite which designates the ensemble of beings necessary for the extension of an entity. [45][15]

If the technical object equates to the technical entity's necessary extension, then the *technical*'s *vectoricity* towards "concretisation" constitutes part of the *technical*; *appropriation* is hence its modus operandi. This is not as harmless as Latour would have us believe. Of course the *value* I assign to appropriation is entirely culturally motivated, and if this thesis is read in in the future the meaning of 'appropriation' may change as *appropriation* is being practiced differently.

What is more, Simondon compares "spontaneously produced" technical objects, for which one could count a living body – human or other, through *techniques of the body* but also technical principles, for example osmosis in plants or water's surface-elasticity – to "concrete technical objects". He concludes that they are actually the same. From this follows that the same drive towards concretisation might also be behind evolution; that evolution[16] is technical concretisation: "Because the mode of

[15]He disambiguates this notion of materials from "matter" as "res extensa", which also explains that in his sense, materials cannot have an opposite, which would bring back the "bifurcation", Cartesian dualism.

[16]In a previous paragraph I underlined that Simondon sees evolution towards a *technical object*, this is not the same as what is suggested here, which is that evolution *per se* is a technical process.

existence of the concrete technical object is analogous to that of a spontaneously produced natural object, we can legitimately consider them as natural objects; this means that we can submit them to inductive study" [84, 41].

This is a point I find of great importance. In spontaneous technical objects, we might think that the materials are appropriated *appropriately*, because the ecosystem is resilient to it. Yet, we do not know if the ecosystem had to adapt to accommodate the *appropriation*. We know, however, that technologies can destroy ecosystems. It makes sense to call them *inappropriate*.

Appropriation, even if it is not *nice*, is what we do. Call it concretisation, individuation, instauration (Souriau [88]), construction,'*wryting*', intention, or *desire*[17] – if it is performed on a material, if it involves a technical action, there is always an element of *appropriation* to it. *Becoming*, in fact, is *appropriation*, in a sense. Once something is appropriated, it might become appropriate.[18] The ethical issue remains with *how* something is appropriated.

The notion of *cultural appropriation*, and how it relates to my term must be discussed: Cultural appropriation is the inappropriate and unacknowledged adoption of elements of one culture by members of another culture. It is particularly controversial when members of a dominant culture appropriate from minority cultures. In an application of the *technical*, in machines (socio-material arrangements), in applying techniques, we also and always culturally appropriate, as "Human reality resides in machines as human actions fixed and crystallised in functioning structures" [85, p.13], and this "adoption" of materials is almost never acknowledged. It is vital not to trivialise this issue. Cultural appropriation is a painful and unjust transgression. I, as a middle-class white male, represent exactly the background that rendered the term controversial, so I am aware that I have to tread carefully here, and with utmost respect for other communities and communities of other cultures.

If *appropriation* is the modus operandi of the *technical*, and technical beings are not only human but also appropriating actants constituting evolution, the fact that there is "human reality" in "technical reality" is less important than the realisation that in human reality there is a technical reality. If a technical object concretises on the

[17]Desire is a key-concept in Gilles Deleuze's philosophy [16].

[18]The use of record players for *scratching*, or *sampling* of existing recorded materials are examples of technical appropriations practices in music as are *standards* or *traditionals*. More general examples are the use of trees to suspend hammocks, or to build boats. Appropriation also occurs when children use furniture to build planes. *Appropriation* needs imagination, before anything. When the *appropriation* is directed upwards in the power structure, it is a reconfiguration without politely asking. In this sense it is activism.

same vector as a spontaneous technical object, the act of "invention" is actually a general technical fact (a form of *appropriation* no less), and does not depend on cognition, either.[19]

Returning to the discussion on responsible agency, if there is a difference between the *inappropriate appropriation* executed by humans and an appropriate one by non-human actants, the generalisation of *appropriation* as the *modus operandi of the technical* presupposes the presence of something *responsible* in non-human actants, because they seem to act more appropriately than humans do. To not fall back into a dualism here between ‘nature’ and ‘culture’, a rethinking of responsibility is necessary.

The difficulties of Actor-Network Theory to explain a moral dimension are obvious here. Although Latour’s extension of ANT in AIME goes some way, it is Latour’s work on the Gaia principle and the notion of *Nomos* which is more promising.

4.3 Gaia

In this section I refer to Latour’s "Facing Gaia – Eight Lectures on the New Climatic Regime" [55], which is Latour’s reading of James Lovelocks works on the Gaia principle. For my argument, Gaia-theory explains how *appropriation as the modus operandi of the technical* can be equally relevant for human and non-human actants, as a generalisation of the actor-network. "Facing Gaia" is the only work of Latour in which he uses the term ‘appropriation’ in more than just its general meaning. Whereas he derives the term from *land-appropriation* in Carl Schmitt’s *Nomos*, my conception shows how an alternative approach is possible and already at work in some collectives. Despite the strangely diametrically opposed approaches, to *face Gaia* is the explicit aim of both Latour as well as climate activists.

The question is again one of agency. If human agency, through the cognitive abilities of homo sapiens, extends into materials through machines, but materials and other non-humans also have agency, we are looking at an actor-network with actors of potentially conflicting interests. If this, *per se* is just how things are, it gets more complicated because I introduced *appropriation*, the technical’s modus operandi, and, evidently, insist on it being viewed from an ethical stance.

[19] I discussed in detail the role of cognition in *machines as socio-material arrangements* in chapter 3. In contrast, here we look at the technical in general (not just in machines), where human cognition is not a relevant precondition.

The term *appropriation* could be interpreted exclusively from within the technical mode: it would then be an equivalent to "concretisation" (Simondon) or "the unexpected detours by which existents have to pass in order to subsist" [51], which shows the notion's existential importance for all technical objects, of which there are many, since nearly all material objects have an intrinsic technicality. However, by insisting on *appropriation* being a value-statement, following Latour's protocol to discover modes of existences, I commit a category error, as the *value* connotation [53] of *appropriation* is not the same as the technical mode's value. To spell it out: what we "hold dear" in the *technical*, e.g., values such as the ingenuity of invention, is not what we "hold dear" in *appropriation*, where a value from *appropriate* to *inappropriate* is assigned to the actual (technical) activity.

I suggest therefore that the *modus operandi of the technical* is a *crossing* of the *technical* mode with the mode of *morality*.[20] This crossing is one for which the AIME project provides no examples.[21] But Latour describes the crossing as essential: "The mode of the technical as much as the mode of morality both redistribute, in their own ways, questions of means and questions of ends. Impossible to imagine a political ecology without this double conjunction of tricks and concerns (understood each time according to the two tonalities)." He states this in AIME [52], but does not touch on it in "Facing Gaia" [55].

If we look at the technical objects on the actor-network level, at the interactions and their impact on humans through their technical extensions, we can dispassionately remark that the "objects left in the wake of the technical" are the objects found wherever we find evidence for man-made climate change. We can accept this as a fact and leave it at that, or we can go on. But it is only really relevant if we do this under inclusion of the value system implied in *appropriation as the modus operandi of the technical.* My motivation for this is twofold. Firstly, it is personal; as a musical/technical practitioner and researcher, I feel morally obliged to explore the ethical dimension of the *modus operandi of the technical's* impact in the actor-network. Secondly, in "Facing Gaia", Latour explicitly questions the detachedness of even the most scientific of descriptions, in an appeal for all collectives to take sides, to "face Gaia" and acknowledge the seriousness of the situation. To some extent, this section is about heeding that call.

[20]This mode "refers to the detection of a particular mode of existence, reusing the common word "morality" but giving it a deeper shift since it must now concern itself with distinguishing between good and bad and between true and false in all modes (each of which carries in itself a distinction between felicity and infelicity conditions)" [46].

[21]The whole AIME project relied on contributions by a team of researchers, so not all crossings or modes received the same attention.

4.3.1 Facing Gaia – Lovelock through Latour

The Gaia Hypothesis was first proposed by chemist James Lovelock and microbiologist Lynn Margulis in the early 1970s. It posits that organisms interact with their surroundings in such a way that conditions for life on the planet are maintained through that interaction.

Although a simplification, the Daisyworld-model that Lovelock co-authored with Andrew J. Watson in 1983 [97] is a computer model that shows how the Gaia paradigm works if reduced to very basic principles. "Daisyworld" is seeded with two varieties of daisy as its only life forms: black daisies and white daisies. White petaled daisies reflect light, while black petaled daisies absorb light. The simulation tracks the two daisy populations and the surface temperature of Daisyworld as the sun's rays grow more powerful. The surface temperature of Daisyworld remains almost constant over a broad range of solar output.

According to Latour, there are now two false interpretations possible, based on an *over-animation*, which depicts Gaia as a regulating super-organism [55, pp. 94-95], or a *de-animation* whereby the theory must be wrong because there is no non-human agency. Latour understands Lovelock's Gaia as an actor-network composed of agents that are neither "deanimated" nor "overanimated". He is adamant that Gaia is not a system – if anything, Gaia is "the outlaw, the anti-system" [55, p. 87]. What is interesting and typical for Latour's approach, is that he does not reduce Gaia to what is evidently useful for his argument and discard the rest, (that would be more like my approach), but takes the entire discourse onboard and in doing, so deconstructs the mythical Gaia, the one that is alluded to in the name of the hypothesis.[22]

Rather than argue the obvious, that Gaia as a goddess is just some folkloristic decoration added for trivial reasons to an otherwise sound scientific theory, he instead convincingly showed that the theoretical framework necessary to explain the Gaia phenomena (without accepting that the constituent "parts" of Gaia *act*) relies equally on a quasi-religious framework.[23] If no agency is attributed to the organisms constituting Gaia, then a super-organism, a system has to be presumed to be at work – a quasi-religious system. The critiques who say, on the contrary, that a super-organism is what Lovelock proposes, miss the point entirely, as this is not what he does, a point very important in Latour's view. Rather, per Latour, Gaia is the

[22] Lovelock's neighbour in rural Wilthshire, UK, was the author William Golding, who suggested the name "Gea", an alternative spelling for the same Greek goddess, for Lovelocks theory, due to the closeness to the prefix geo- which has the same etymological roots, and is found in geography, geology, etc. But the term Gaia also became more common in Lovelock's writing [61].

[23] Lectures five and six deal with the secular vs. religious notion of Gaia [55, pp. 146-227].

summed, complex agency of the actants constituting the biosphere, the atmosphere, the hydrospheres and the pedosphere: "Agents act! One can try to 'overanimate' them or, on the contrary, to 'deanimate' them: they will stubbornly remain agents" [55, p. 172].

My own provisional interpretation of this is that in Gaia, rather than through a grand plan (we know that not even cathedrals really needed them), a strong sense of egoism leads to an equilibrium of interests, contingently stabilising the environment. This might sound a bit like Darwinian evolution, but, according to Latour, that does not really apply:

> For Lovelock, organisms, taken as the point of departure for a biochemical reaction, do not develop "in" an environment; rather, each one bends the environment around itself, as it were, the better to develop. In this sense, every organism intentionally manipulates what surrounds it "in its own interest" – the whole problem, of course, lies in defining that interest.[...] On a dead planet, the components would be placed partes extra partes; not on Earth. Each agency modifies its neighbours, however slightly, so as to make its own survival slightly less improbable. [...] Gaia captures the distributed intentionality of all the agents, each of which modifies its surroundings for its own purposes. [55, p. 98]

This interpretation of the Gaia principle can also shed a new light on the discussion on *emergence*, in chapter 3. That the results from the interaction between parts have properties the constituting parts do not have on their own is evident here. What is more, the Gaia Principle itself can be conceptualised as an emergent property, arguably it can be considered as self-regulating *due to* the interaction of the parts (see Lovelock [60, pp.19-20]).[24]

Rather, every actant forms every other actant. The following train of thought is nothing but common sense: if the Gaia principle applies, its constituting actants are responsible for the composition of the atmosphere, the hydrospheres, and even elements of geology through sediments [55, p. 115]. And, as human activity is also an involved and involving actant, we also leave our traces in all of those spheres.

So far, so good, if it was not for the fact that the increase in carbon dioxide since the industrialisation started in the 17th century has had such a huge impact that we are

[24] Latour insists that Lovelock does not mean to insinuate that Gaia is a "system", although he repeatedly says so in his writings (see [55, p.87]). As I go along Latour's interpretation, I am a bit reluctant to put too much weight on "emergence" as a system-theoretical concept.

evidently changing the composition of the environment to the point where it is not favourable for our own species, or the organisms we rely on for subsistence (food, for example).[25] In this situation, it would indeed be very convenient for us if Gaia was a regulating super-organism, or even better a "mother earth", fixing everything miraculously. If, on the contrary, Gaia is an actor-network where all actants act, we might have to take actions that are more amenable to the other actants.

Rather, the Gaia principle means that it is perfectly possible that we are taking ourselves out of the equation, (like the dinosaurs in their time) and that life will continue in one form or another, or that life on earth will end, and earth will find its chemical equilibrium, like Mars, the Moon, Venus, or most other celestial bodies we know of. Or, we can start to do something that might lead to our survival as a species. Of course, all kinds of gradations between the examples I provided would also be possible; all possibilities are open. There is no grand design, no master clock-maker, no cybernetics, not mother earth, but only Gaia: "If there is no frame, no goal, no direction, we have to consider Gaia as the name of the process by which variable and contingent occurrences have made later events more probable" [55, p. 107].

But why, in the view of all the evidence, are we still hurtling towards the abyss, fully aware of all the facts, of what is coming? This is the mind-boggling question for which Latour has some interesting answers. He sees the root of this inertia in the way modernity has dispensed of the *apocalypse*, an affinity to the possibility of an end of the world, through the de-animation of the world. – If it cannot act, it cannot end. In this sense the arrival of modernity itself was an apocalypse (a *revelation*), and no apocalypse can follow an apocalypse [55, pp. 192-196]. This might explain our apathy to some extent.

4.3.2 Inappropriate Appropriation

In view of the acknowledged inertia – the same inertia, in fact, Donna Haraway attests to humans in view of the "disturbingly liveliness" of (supposedly) autonomous machines, as I discussed in chapter 3,[26] Latour calls for very drastic action:

[25] Latour lists the many clear statistics indicating the gravity of the situation in the second lecture [55, p. 42].

[26] I do not contest the "inertia", but that the "disturbingly liveliness" is anything but our own, through extension as contributing actants in machines as socio-material arrangements.

Short of calling on scientists to go on the warpath he develops ideas of how to broaden the acknowledgement of a state of war. "We shall never be able to re-politicise ecology without first agreeing to recognise that there is indeed a state of war" [55, p. 227]. This state of war shall decompose "the human as a unified agent, as a simple virtual political entity, as a universal concept, [. . .] into several distinct peoples, endowed with contradictory interests, competing territories, and brought together by the warring agents" [55, p. 127]. He does not use such bellicose language lightly, and uses it to make a diplomatic solution possible. However, for this to succeed, the "pacified" world first has to realise that it is at war [55, p. 239].

Toying with terminology from "Game of Thrones",[27] he defines the *Earthbound* as a people, a tribe, who "have to be able to map the territories on which they depend for their existence", and the enemy: "To put it in the style of a geohistorical fiction, the Humans living in the epoch of the Holocene are in conflict with the Earthbound of the Anthropocene." By Humans of the Holocene he means people who either ignore or question the consequences of Gaia [55, 248].

Latour's call to "map the territories" [55, p. 248] and to "differentiate between real friends and real enemies" [55, p. 157] refers to ideas in "Der Nomos Der Erde" by Carl Schmitt [82]. Schmitt uses a term, *Landnahme*, which, via the French translation, appears in Latour's lectures (in English) as 'appropriation'.[28] Because Latour has not defined *appropriation* in the sense I use it, his use in reference to Schmitt needs a closer look.

That Latour turns to Carl Schmitt (1888-1985) in his argumentation might come as a surprise, since the German jurist and political theorist was a prominent member of the Nazi Party and never distanced himself from its ideologies and policies and remained an outspoken antisemite throughout his life. Political theorist Mika Ojakangas, who has written extensively about Schmitt, makes the point that as a political theorist rather than a historian it would not be his task to reduce ideas to a time and place or to explore the true intentions of writers [76]. Jacques Derrida also read and discussed Schmitt. Although criticised by many, Schmitt's theories are part of the body of political theory. And Latour, too, suggests that Schmitt's thinking can be "purged" from the atrocities it condones [55, p.284].

[27] "In this fifth lecture, I am going to focus, I am afraid, on an operation of science fiction that will be somewhat reminiscent of the television series Game of Thrones!" [55, 151] (G.o.T. is a HBO series, inspired by George R. R. Martin's fantasy novels.)

[28] Landnahme translates to "occupation and settlement of land", meaning the sequence of both: Landnahme always explicitly means occupation followed by settlement, never just settlement, but always occupation.

Notwithstanding, Latour is aware that he might be provoking controversy. However, as he sees merit in Schmitt's description of *Nomos*,[29] "earthly law", which juxtaposes itself to both the "law of nature" and "God's law", he gives the following disclaimer: "I am going to turn to the author the least apt to reassure you, the toxic and nevertheless indispensable Carl Schmitt (1888–1985). The Nazi legal scholar can be likened to a poison kept in a laboratory for the moment when one needs an active principle powerful enough to counterbalance other even more dangerous poisons" [55, 228].

What is important for Latour's argument is firstly that in Schmitt's *Nomos*, space is not a geographical definition but a place-based territory. These laws, says Latour, could

> become something like a 'jus publicum telluris' , still to be invented, in view of limiting what Schmitt, in his terribly precise language, called the *Raumordnungskriege*, the 'wars over spatial order' – an expression that, once purged of its associations with the conflicts of the twentieth century, offers a radical definition of earthly life, but an earthly life finally capable of taking the presence of Gaia into account, so that we shall be able to limit the extent of wars to come. [55, p. 253]

So Latour claims that beyond the context of Nazi atrocity, "Raumordnungskrieg" could mean something acceptable, something not just akin but equivalent to what I define as the *modus operandi of the technical*. (Why I do not think this "purge" is quite as successful as Latour suggests will become clear from Arendt's reading of Schmitt discussed below.)

Latour, secondly, is taken by the way Schmitt's comparison of a *pacified state* to a *state at war* provides two clear possibilities. In the pacified state,

> Let us all be brothers on the same blue planet, aligning ourselves under the same politico-scientific authority in order to escape from more serious conflicts. [55, p. 241]

which would suit Latour fine, as long as this state actually existed:

> It does not: The 'pacified' world is at war against us, as we consider it inanimate. We do not acknowledge it as war, so we are defenceless. If

[29] Despite the rare occurrence of *Nomos* as a term, its use as part of the word auto*nomo(u)s* might go some way to getting a sense of its meaning: autonomous would then mean "according to its own laws".

> there is no such state [. . .] we would be agreeing to place our security and that of all the other entities with which we share the Earth under the protection of a political body incapable of defending us. [55, p. 241]

And the other possibility, the state of war:

> You agree to distinguish the enemy from the friend, and then you are engaging in politics, strictly defining the boundaries of very real wars, [. . .] The perilous virtue of reactionary thinkers such as Schmitt is that they force us to make a more radical choice than the one proposed by so many ecologists, who are still driven by the hope of getting out of the crisis without ever politicising the questions of 'nature' [55, p. 241].

I do agree with the urgency and clarity of the last two statements. As this clearly describes the political dilemma that facing Gaia entails, as true as it is that it forces a more radical choice – when Latour applies the term of *Landnahme* to the type of *appropriation* I see in all technical objects – to the actants constituting Gaia, he goes beyond what is strictly necessary, and accidentally promotes not just some of Schmitt's theoretical ideas, but also his rationale, which in my view is quite unfortunate.

4.3.3 The Seal Which Stamps Whate'er the Earth Contains

Political philosopher, author, and Holocaust survivor Hannah Arendt also read Schmitt, but never proceeded to publishing her thoughts.[30] However, the marginal notes in her copy of Schmitt's book have been published and commented on by Anna Jurkevics in "Hannah Arendt liest Carl Schmitts Der Nomos der Erde" [34]. What gives Arendt's reading an advantage over Latour's is that Arendt commented in German on the German original. And because German was also the language of the Nazis, the connotations between the Nazi terminology and Schmitt's ideas are a lot clearer than they could possibly be in a French translation.

For Schmitt, Arendt and Latour, the definition of *Nomos* goes back to the Greek deamon *Nomos*:

[30]Interestingly, Schmitt also read Arendt, as Ville Suuronen writes in a recent article [92]. Schmitt appreciated Arendt's description of totalitarianism in "The Origins of Totalitarianism" [4]. Only after reading Arendt's "Eichmann in Jerusalem – A Report on the Banality of Evil" [6], Schmitt finally understands that according to Arendt, under a totalitarian regime, the ones who do not fight exclusion from power bear responsibility for the deeds of the regime.

> The holy king of Gods and men I call, celestial Law [Nomos], the righteous seal of all; The seal which stamps whate'er the earth contains, Nature's firm basis, and the liquid plains[. . .] [31]

For Schmitt, and Latour, the *firm basis, the liquid plains* are the instituting actants, in Latour's reading the *terra* (earth) in *terr*itory becomes the important notion. I believe that for Arendt, *the seal* is altogether more important. The other aspect of *Nomos* that Schmitt, Arendt, and Latour share is the relation of *Ortung* (*positioning* in a navigational sense) and *Ordnung* (order) in *Nomos*.

Schmitt, as a Nazi, used the term *Boden* in the sense of 'Blut und Boden' (blood and soil). 'Boden', hence, should be translated as soil in the second meaning of the noun as *territory of a nation*.[32] Jurkevics situates Schmitt's "Nomos" as his search for a philosophy of law that is based on territory. For Schmitt, this territorial law constitutes a "world-order" (*Weltordnung*) kept in balance through the diverging interests in a "pluriverse" (*Pluriversum*).

Acording to Schmitt, in *land-appropriation* (i.e., *Landnahme*), the land is first occupied, and occupation is the constituting – the *constitution* giving – moment.[33]

In there lies the fundamental difference to Arendt's understanding of *Nomos*: for Arendt, legal questions only arise when the land is distributed (*divisio*), when it is decided who gets what – the moment of *appropriation* is the moment when *property* is declared. This (negotiated) process constitutes law. Occupation, Landnahme without *divisio* is *lawless*.[34] Schmitt's insistence that the *Landnahme*, *occupation* of a territory, is the law-constituting act renders his geopolitical theory profoundly imperialistic. In annexation, *Nomos* is *not* drawn from the earth but *imposed*, so there is no *appropriation* taking place, just annexation for annexation's sake.

[31] Excerpt from "The Hymns of Orpheus" [93]

[32] Jurkevics starts her chapter thereby with Arendt's remark that "Poor Schmitt, the Nazis said 'blood and soil', he understood 'soil' and the Nazis meant blood"! "Armer Schmitt: die Nazis sagten Blut u. Boden – er verstand Boden u. die Nazis meinten Blut" [34, p.5]. This firstly indicates that Arendt, differing from Ojakanga and Latour does not take Schmitt out of place and time, but also that she does not primarily see him as a racist.

[33] Besides Jurkevics [34], Schmitt giving appropriation precedence over division is also brought to the fore in "Schmitt on Nomos and Space" [77].

[34] "Offenbar hier: Landnahme oder occupation oder Eroberung natürlich conditio sine qua non der Landteilung, Aber Rechtsfragen entstehen natürlich erst bei der Divisio, die jedem das Seine gibt. Ergo ergeben sich Rechtsfragen bei der Entdeckung der Neuen Welt erst als es an die Verteilung geht . . . Also: vor der acquisitio steht die divisio und nicht d. Eroberung."[34, p.13]

We are Seemingly looking at a paradox here. If the *appropriation* needs to be based on concepts of constitution and contracts that are pre-existing, then how can *new* land be appropriated, where no constitution is present? Upon closer inspection it becomes clear that every territory already has a constitution, because it is already *constituted*, or articulated, as Latour would say. Every *appropriation* is therefore a negotiation of existing concepts of constitution and contract. Consequently, the very presumption of "new land" is, indeed, an annexation, because it ignores the existing *Nomos*.[35]

Schmitt's *decisionist* conception of law [77, p.106] – the idea that authority instates law and that moral or legal precepts are the product of decisions made by political or legal bodies – is diametrically opposed to what I believe Latour is trying to achieve by enlisting Schmitt's help: Gaia's *Nomos* cannot be a consequence of an authoritarian act, only of a never-ending reiterating negotiation of a contingent equilibirum enlisting a complex multitude of actants.[36]

Although firmly routed in the territory of the Greek *Polis*, a jurisdiction delimited by its border walls, (enclosing the all important *Agora*) Arendt's *Nomos* draws also on the Roman *lex*. The decisive difference there is that the Roman lex is relational, rather than territorial [96]. "A law is something that links human beings together, and it comes into being not by diktat or by an act of force but rather through mutual agreement" as Arendt writes [3]. Arendt's conception of *Nomos* rises from (human) interactions in the *Agora*, the hierarchically flat shared democratic space where all actors and actants assemble.

So, *Nomos* can only develop through loops of negotiations. Here, I am glad to say my path rejoins Latour's thinking temporarily, because he also credits this loop with some relevance for *Nomos* in "Facing Gaia" [55, p.276][37]

[35]This is reflected in the fact that only territories which have been colonised feature 'new' in their names! The presense of 'new' as such indicates that there was a preceding 'old' which had to be annihilated. To call them new is, from this angle a form of "Newspeak".

[36]Although Latour sees this struggle for an equilibrium in Schmitt, he puts little weight on the fact that in Schmitt a new nomos relies on an equilibrium, between "Grossräumen", (large spaces), rather than individual actants. Schmitt does not argue on the Daisyworld-level of single actants but seeks a balance between super powers, agglomeration of actants. Inside the "Grossraum", Schmitt propounds a decisionist *Nomos* – order based on authority, not negotiation.[77, p.112].

[37]What remains as a question is wether territory is really this important for the politicisation of Gaia. Yes, the actants' territories need to be respected, (not that this would be possible through occupation arranged on Schmitt's advice) but why on earth does this need to proceed via "ownership?" No migrating birds will agree to any *Nomos* and still find their territories. So *soil* as territory is really an unhappy choice of word. I could contingently suggest *appropriational practice* as a better one. A revised meaning of 'appropriation' would also redefine what it means *to own* and what is property.

Perhaps Latour needs Schmitt to re-introduce *appropriation* since he lost it when he took on the *technical* from Simondon without sufficiently acknowledging its insurmountable pull towards objects. That Latour disentangled the technical from the object means that the technical's *modus operandi*, *appropriation*, is not sufficiently obvious in his description of the technical. When he adds *appropriation* back into his reading of Gaia via Schmitt's *Nomos* – an appropriation that is systemically not benevolent, as Hannah Arendt argues – we have in Latour's geopolitics a Trojan horse (*appropriation*, according to my definition) within a Trojan horse (appropriation in Schmitt's definition) hiding amongst the *Earthbound*.

It is important to remember that *appropriation* is unavoidable. Wherever we set our feet, we leave behind something destroyed.[38] But this is negotiable, we can aim our steps in order to mediate their impact. If we choose not to negotiate, we have Schmitt's appropriation; for Schmitt, clearing a "territory" of indigenous people was nothing but a mere technicality.

Rather than using Schmitt's conception of how the *Earthbound* should think of soil, Schmitt's ideas should be an apocalyptic reminder of how we got into this mess in the first place: The notion of *appropriation* makes it clear that *industrialisation* was unmediated *Landnahme* – *Noummos*[39] rather than *Nomos*: concretisation as economic *sustainable growth* rather than *sustainability*. The notion of *Nomos* as defined by Arendt's reading is a relevant contribution to the discussion on Gaia, taking the notion of *appropriation as the modus operandi of the technical* into consideration.

Latour concludes his lectures in "Facing Gaia" with the remark that he had a glimpse of something like the "new *Nomos* of the Earth", something that he enthusiastically characterises as "constitutive", and that "as far as we may be from Captain Columbus' spirit of conquest, perhaps we are nevertheless still like the thirsty sailors aboard his caravel, waiting day after day for the cry that the lookout will surely end up shouting some morning, from up in the crow's nest: 'Land ho! Land ho!'" – This just confirms that it is right to meet Latour's *Earthbound* people with suspicion. Not because they would not mean well, but in their constitutive fervour, would they notice if nothing grew wherever they set their feet?

[38] In free improvisation too, if a participant picks up my instrument, I am robbed of the opportunity to play it. If an improvisation goes in a certain direction, it cannot go in another one – at least not for the moment when its direction was set.

[39] 'Noumos' (from 'Nomos') is a type of ancient Greek coin, and the etymological root for the Latin word 'nummus', meaning 'coin'.

4.3.4 Situated Action, Interdisciplinary Improvisation Revisited

In the attempt to find new roles, new identities, a new way to be in the face of Gaia, joining the *Earthbound*, despite Latour's efforts, might not be the most appropriate choice. Latour missed an opportunity to side with the troops who *are* ready for Gaia, by missing the revolution. Not the one that he agrees we all missed, when Gaia revolted against the ones who ignored her; mysteriously, he thinks that this was the revolution "progressive minds longed for" [55, p.39].[40] However he is in danger of missing the one still to come, the moment when the ones who have already acknowledged the state of war stop negotiating. Climate activists link their calls for system change to the call for social justice, to calls for *emancipation* of the disenfranchised, marginalised, suppressed and abused.

In "Facing Gaia" [55], Latour describes emancipation exclusively as a modernist drive to leave "the dark ages behind", to "leave the soil". His argumentation is that "land" can have a progressive connotation in the form of *earth*, or a reactionary one, as in *Blut und Boden*. What he calls the "Great Emancipation Narrative" (sic) is preventing us from finding the path to the earth to which we belong: "As if the very notions of 'belonging' and 'territory' gave off a whiff of something reactionary!" [55, p. 244] In Latour's (supposedly) progressive return to *Boden*, the very same structures responsible for the dismal state of the soil remain intact (they cannot change without emancipation). If one comes to think of the implicit and explicit struggle of a vast proportion of the world population to overcome the existing territorial divisions and their stranglehold on soil and hearth, (and who should tend to the latter), then yes, his argumentation has a whiff of something reactionary.

In "Reassembling the Social" [37], Latour re-conceptualises emancipation more kindly as a *change* of associations rather than a *severance* of associations ("not 'freed from bonds' but *well*-attached" [37, pp.217-218]). I propose to look at it in this kinder way: Latour points out that it is important for anthropologists of the moderns to listen to their subjects in order to discover their values.[41] In this tradition, I would like to return to the participants of the workshop on Music Space and Interaction,

[40] Latour's disdain for progressive minds is based on their being responsible for a modernity that has got pretty much all of it wrong. However, he throws the baby out with the bathwater: Karl Marx's characterisation of society as a place where rulers take their private concerns and turn them into public concerns, in effect making their own interests the ones of the public who cannot refuse them [66], is a perfect encapsulation of both, why Gaia revolted, and why people want political emancipation: to face Gaia. I cannot see a single line in anything I read by Latour which would make Marx's statement less valid.

[41] This is what he says about a "true anthropologist": "She knows that only a prolonged, in-depth analysis of courses of action can allow her to discover the real value system of the informants among whom she lives[. . .]" [39, p.28].

whom I have not indoctrinated with the "Great Emancipatory Narrative", and yet, their disenfranchisement from technology and their sense of empowerment when 'seizing the means of production', the 'freeing from bonds' through unlearning, of doing the unexpected, the re-configuration, had all the hallmarks of emancipation, of attaching *well*. But it was also earth-bound, positively engaging and *embodying*. Maybe both Latour and the theoreticians of the "Great Narrative" had it wrong, maybe emancipation is not at all about breaking bonds, but finally making bonds?

But without acknowledging that emancipation never happened, the "patriarchy" responsible for the reactionary *Landnahme* will be responsible for the "*progressive*" one too, if we follow Latour's advice. I do not believe that this is his intention, but by not clearly aligning the social struggle (the emancipation of the people whose concerns are ignored in public) with the Gaia struggle (the emancipation of the actants of the earth ignored in public), the "*progressive*" *Landnahme* will be carried out by the same white men already in charge, this time in Teslas and with flowers in their hair – the same eco-modernists Latour distances himself from, but with a copy of Nazi-Party member Schmitt's *"Nomos"* in their pockets.

Latour's project would be much better served if he saw in the progressive return to earth an endeavour to seize the means of production [31] – not to industrialise them in a Simondonian way,[42] but to appropriate them *appropriately* – in a system change.[43] The danger of instability as a consequence of system change has to be weighed against the instability that will follow the irreversible destruction of the world we are heading towards. This is the call of the climate activists: system change, not climate change.

Extrapolating the Gaia principle to human collectives leads to the understanding that the *global* does not actually exist except through the *multiple local*. The changes needed on a scale to reverse global warming can only be achieved on an actant level, through *situated action*; if Gaia Theory says anything, then that. I suggest not to return to the earth as territory, but to the earth as the material body, the one made of flesh and bone, bacteria and *grey matter* extended into machines and practices, to food and sustenance, ashes and dust, but also homes, care and education. This

[42]"When a technical object "has acquired its coherence on the industrial level, where the system of supply and demand is less coherent than the object's own system, [then] needs are moulded by the industrial technical object, which thereby acquires the power to shape a civilisation" [84].

[43]*Nota bene*, the Marxist system change initiated by Lenin was *inappropriate* – according to Arendt, Marx's *historical materialism* assumes natural law as the guiding principle of society, a fallacy enabling totalitarianism which later took place through Stalin [26, p. 132].

aligns the social "war front" with the Gaia "war front", so the diplomatic effort [42] can start where it matters.

Latour emphasises that Gaia is not a mild, caring entity, but may be rather cruel. But it is not its cruelty that guarantees its sustainability. More probably, Actants and actors in Gaia are sustained through compromise, opportunism, engagement – an egoism of a very peculiar kind, an anarchic one – and an ability to interact, above all.

We need new ways of doing things, new ways of forming collectives of individuals – to face Gaia. Just as Latour sees in Gaia a form of a selfish gene at work,[44] one could consider a form of ethical egoism as a social principle, perhaps through a re-reading of Max Stirner and other anarchist theoreticians. It is no surprise to me, that the people who are really prepared to face Gaia, the ones who actually are doing something about climate change, the ones prepared to *act*, refer to the body of work on anarchy. If anarchy needs to be moderated, then let it be through Arendt's negotiation in the *Agora*, not Schmitt's *Nomos*.[45]

What we need is new sets of little contingent drawings on scraps of wood[46] to be handed around on the construction sites of new cathedrals, rather than declarations of war: situated, experimental practices on all levels, for technology development and to organise our collectives.

Arendt's *Agora*, a term essential to her political theory, is in this context particularly interesting because she did not see *Agorae* primarily as political spaces: "The hallmark of these non-political communities was that their public place, the *Agora*, was not a meeting place of citizens, but a market place where craftsmen could show and exchange their products" [5, p.160]. What is more, Shmuel Lederman shows how such communities as citizen's *councils* provide a template for a participatory democracy [57]. Whereas the term *council* refers historically to the Russian *soviets* of the revolutions from 1905 and 1917, Arendt sees structures of political institutions similar in their spirit to the councils which emerged in earlier revolutions: the American town meetings, the "popular societies" of the French revolution, the

[44] "Every selfish goal is submerged by the selfish goals of all the others" [55, p. 103].

[45] The activist group "Occupy" provides a great example of provocative *appropriation* upwards in the power-structure, and by its apt name shows the difference between Schmitt's *Nomos* and the negotiated (looped) *appropriation* at work possibly closer to Arendt's: Occupy is an *organisation* demanding *more democracy*; they did not first annex Wall Street, and then think about what to do with it. They are highly aware of contracts and constitutions. System change will not come from policies, nor annexations, but from situated *appropriating* actions.

[46] See Suchman [90, p. 197].

"sections" of the Paris Commune. They were self-governing, spontaneous associations of citizens who strived for a democratic political reform.

Lederman goes on to hypothesise that these forms of participatory democracy could provide a blueprint to the new social movements. These movements, says Lederman, "call for and try to prefigure a more direct and participatory democracy, [but] they usually provide no systematic elaboration of what such a participatory form of government would look like and how it would function" [57, p.245]. They link themselves to more 'old school' anarchistic traditions, yet, also there, not many models on how to run such institutions are provided other than the description of those same citizen councils, which came into being without a political agenda.

The reason why there is no recipe, no set of instructions, no manual available for these types of institutions, I see in their need to be situated or *improvised* to be successful – all examples Arendt calls to mind as council-like movements occurred in revolutions, at times in history when structures were in flux. Returning to the mythical *Nomos*, to "The seal which stamps whate'er the earth contains Nature's firm basis, and the liquid plains[. . .]" [93], The *Nomos* is but the seal that stamp that particular polis, the local mark, the contract achieved, the embodied, the situated firm basis, the particular river, creek, lake or stretch of sea that has to be negotiated over in the *Agora*. Once these negotiations have been had, maybe a contingent score can be written down, maybe principles can be formulated or some sketch carved on a scrap of wood, maybe a *Nomos* can be constituted – as competences that come after the action.

To come full circle, just as there is participatory design, there is participatory democracy. In the last chapter of "Facing Gaia", Latour provides a fine example of how this could work; he reports on a staged meeting of stake holders in Gaia,[47] where representatives not only speak on behalf of national states, the industry, and minorities, but also representatives for the rain-forest, the oceans, and the mountains (he points out that if a mountain cannot speak for itself, it is no different to a nation-state that, as an abstract entity, does not speak for itself either!)

The *simulation* Latour describes [55, p. 256] is an uncannily similar setting to an interdisciplinary improvisation, where rather than prototyping a solution, the (problematic) situation is prototyped. This corresponds to the method I applied in

[47] "'Theater of Negotiations,' a simulation carried out in the context of 'Make it Work,' was produced in Paris at the Théâtre des Amandiers, May 26–31, 2015, under the direction of Philippe Quesne and Frédérique Aït-Touati, with the participation of SPEAP, the school of the political arts at Sciences Po" (https://www.sciencespo.fr/public/en/content/cop21-make-it-work.html) [55, p.256].

the development of ALPS and introduced in the articles.[48] What is more, a whole series of observations apply to both Gaia and the method of free interdisciplinary improvisation, as if the participants represented general actants. The open-ended mediation of participants' *territories* and *materials* in a free improvisation session, the way every participant's contribution impacts on everybody else's – how easy it is to "annex" an improvisation, and how difficult it is to mediate such appropriations, how rules develop over time – how an improvisational practice creates its own universe; all of this expresses quite well what essentially constitutes the Gaia principle. Gaia, extrapolated to human collectives, is a free improvisation – for the duration of the participation, contingent, individual, mediated from the inside; collective, organised, and structured from the outside. Unique, and irreversible.

In this thesis, I followed cues from my artistic practice and method of free interdisciplinary improvisation for the development of interaction technologies, in the form of field-notes, outcomes, ideas, and challenges in the workshop on Music Space and Interaction, which resulted in these reflections on agency in the *technical*. If If the MS&I workshop, a humble local practice, encapsulates the workings of something we can also identify in Gaia (through participants' *appropriations* and being appropriated in experimental, new situations), it is not because we were aiming to do something *global*, but because we made room for situated (local) action. Precisely the willingness of participants to leave their territory behind, to unlearn, to dabble, to step into the field of an unknown discipline, outside of their comfort zone, and always without appropriating it inappropriately – this is the *Agora* that Latour[49] and Arendt share. Technical solutions are contingent situated interactions, improvised, performed, practised, negotiated, mediated, and appropriated. They are socio-material arrangements.

Going back to the artistic participatory design practice of the MS&I workshop, it is easy to see how the *Agora* provides a simile for free improvisation and its setting, and the improvisation setting a metaphor for a society (or an actor-network of associations). It is also quite obvious that *Landnahme* according to Schmitt would hardly result in any development other than burning ruins.

The simile of Gaia works quite well in a situation of human actors, but how about the reverse, can we conceptualise non-human actants as improvising? If free impro-

[48]Articles I, III IV V and VII describe this experimental practice.

[49]"Let us return to the Agora. Discovering the right category, speaking in the right tonality, choosing the right interpretative key, understanding properly what we are going to say, all this is to prepare ourselves to speak well about something to those concerned by that thing—in front of everyone, before a plenary assembly, and not in a single key" [39, p. 58].

visation needs a moral stance (a willingness to negotiate), wouldn't the actants in Gaia also have needed such an ethic disposition to be able to reach an equilibrium? – Definite answers are probably none or too many to come by, but what if we suppose that, from a reverse approach, we define ethics as negotiations which succeed – could ethics in this sense be the felicity condition [43] of appropriation? This would imply that *Landnahme* according to Schmitt happens (improvisations go wrong too), but only *Nomos* according to Arendt has sustainable chance of success: The ongoing negotiations in our technical interactions lead to ethics, as a felicity condition for the collectives' sustained survival. Maybe here we are looking at a nascent and ill-defined, a tentative example of an 'ethics of things'

Despite the negatively loaded discourse of technology in news and media, through a conceptual rethinking of the *technical* we can responsibly embrace its practice and recognise its role in the formation of collectives.

4.3.5 Occam and the Gadget

To come back to the *technical*, and how new ways of *technical appropriation* should be developed in the face of Gaia, I would like to return one last time to a notion Latour discusses in a slightly different context in AIME. William of Occam, who lived 1287 – 1347, is best known for his *razor*, a tool he used to cut off unnecessary duplications and detours in logical argumentations. Occam's razor refers to the problem-solving principle that "entities should not be multiplied beyond necessity."[50]

> I have always imagined for my part that the story of Occam's razor alluded to a little case made of precious wood like those once used by surgeons, in which a great many tools adapted to all the delicate operations of reason lie nestled in green felt compartments. Should not even the most hardheaded rationalists rejoice that there are several types of instruments, as long as each one is well honed. [39, p. 19]

I allow myself to imagine that Occam not only had many different little scalpels, saws, and pincers in his razor case, but a great many different wooden boxes and cases. One such box has rarely been opened, and the tools in there remained practically untouched; they are made to cut out unnecessary technological contraptions, to nip in the bud the development of overblown apparatuses and let stand only the most

[50] He was so efficient that he nearly managed to eradicate himself from association with his own principle, so that he was only remembered as the principle's inventor several centuries after his death.

efficient – the simplest – technical installations. Although a fine little gadget itself, it is, one could say, the *anti-gadget*.

This new razor, *Occam's anti-gadget*, states that the simplest technical process to achieve a goal should be favoured over others.

For Simondon [84], technical objects come into being along an abstract-concrete trajectory. *Abstract* technical objects are general in nature, they are like prototypes. They are the cobbled together first plans for a machine, an assemblage of principles. In contrast, the *concrete* technical objects are the ones at the end of a development process, the direction in which the development goes. Summarising the earlier discussion, along this vector, the technical object becomes simpler, not more complicated, and lends itself to industrialisation once it has found a more concrete form.

Yet concretisation, I argue, could also happen through practices (traditions, know-how) implying that industrialisation is also a (distributed) practice. And the abstract state is in many cases more appropriate than a concretisation. So, the gradients between high-tech and low-tech, abstract and concrete, handmade and mass produced, do not align. The situation is made even more complex by the idea that technical objects are not invented towards functionality, but functionality is discovered in existing objects.[51]

An ethical principle to mediate *appropriation* is needed, to guide our agency in machines and how we build them. Just like in science, a principle based on Occam's razor should be applied in technology developments according to the principle of *Occam's anti-gadget*, where the most concrete technical project is the most simple but also the most low tech, *as is appropriate*. The handmade-mass production gradient is then a question of the distribution of agency only. Industrialisation's impact can be mediated on a local level once it is re-conceptualised as an actor-network.

In the development of appropriate technologies, the ALPS software can serve as an example here. Existing technology often has all the functionality available already, an understanding of the situation suffices to reconfigure the material arrangements. For ALPS, no additional hardware needed to be introduced, not a gadget but an *anti-gadget* evolved; to use acoustic localisation techniques for tracking in situations where sound distribution over loudspeakers is already a fact, where the use of distance based amplitude panning allows for estimation when there is as little as a

[51]The participatory design practice in MS&I exemplified this: To be participatory, technical objects need to stay abstract to some extent, and functionality is found through improvisation!

single measurement, are all concretisations aligned with Occam's principle. These abstract ideas (in a Simondonian sense) did not originate from a grand plan; there was no blueprint, there were just ideas floating around on metaphorical scraps of wood, in a community of dabblers, tinkers, bricoleurs. And yet, they were by no means unprofessional, because – and this is the point – competence comes after performance.

As unremarkable as such a maxim like *Occam's anti-gadget* might seem, it epitomises the simple fact that we are directly responsible for the machines we constitute as socio-material arrangements. Although to act appropriately seems like a simple choice, to act on the choice is not. Gaia is not a super-organism, but an actor-network, where every choice is just one act; for this, we need to develop new practices in the face of Gaia, as Gaia.

4.4 Closing Remarks

Going back through the last two chapters, keeping the original proposed outcomes in mind, reframes this doctoral project as an ethical one. The question of agency in the *technical*, once it is becomes clear that the agency in machines is *our* agency, becomes one of responsibility as a collective for the machines we constitute. By instituting the workshop on Music, Space & Interaction as a participatory practice to start with, I had access to such a collective where I could observe this agency in detail, how techniques and routines crystallised into technologies and machines.

In response to the concerns over agency in automatons that arose in that practice, I hypothesised that the code-structures constituting language, music and computational machines are related, and that consequently machines can be conceptualised as extensions of human cognition. This resonates with the practical experience too – we did not spend days on a lathe engineering our tool from metal, but in front of a screen *coding*, it was most definitely ourselves writing that code; if the system did not behave as we expected, it was through coding that we made it act differently.

From the understanding of cognition as an active (motor) performance follows that it is the prerogative of living organisms. Yet, the *technical* is evidently present in inert beings too. This is the basis for my hypothesis that *appropriation* is the modus operandi of *all* technical beings. This had conceptual consequences: firstly, because humans are constitutively involved in machines as socio-material arrangements, we are ultimately responsible for all of the constituent parts (which confirms the findings from chapter 3). Secondly, our body's technicality also operates in this mode, so *appropriation* applies there too; we cannot help it, we appropriate with every bite we eat, every breath we take.

Equating *appropriation as the modus operandi of the technical* to the Gaia principle, allows an extension of the comparison to free interdisciplinary improvisation as a development method for technologies, since the method allows for the prototyping of situations where mediation can be exercised not in the vacuum of a laboratory, but in an interactive practice – in the form of an *Agora*. Where evaluation is the main phase of development. So even small projects like ALPS can exemplify how participatory experimental projects could help to find solutions, *appropriate* solutions, according to a principle whereby the *least inappropriate* solution is the best.(Occam's *anti-gadget*.)

In Latour's interpretation of James Lovelock's Gaia theory it becomes evident that the continuous mediation between appropriating actants can result in *something* contingently stable. This *something* is not a system, which would imply some outside control, but an actor-network. The all-encompassing ethical question in this is raised through Gaia itself; all scientific evidence shows, that life on earth is about to irreversibly change and that these changes are man-made, that a group of actants constituting Gaia is behaving inappropriately. The question thus is simple and concerns every single conscious agency whenever we interact consciously-technically with the world: Is my action appropriate? – Direct action might be just that.

I came to Hannah Arendt's work at a very late stage of the project, despite the fact that her prologue to "The Human Condition" could have provided a starting point in its own right. Hannah Arendt summed up in 1958 the situation I described at the end of chapter 3:

> It could be that we, who are earth-bound creatures and have begun to act as though we were dwellers of the universe, will forever be unable to understand, that is, to think and speak about the things which nevertheless we are able to do. In this case, it would be as though our brain, which constitutes the physical, material condition of our thoughts, were unable to follow what we do, so that from now on we would indeed need artificial machines to do our thinking and speaking. If it should turn

> out to be true that knowledge (in the modern sense of know-how) and thought have parted company for good, then we would indeed become the helpless slaves, not so much of our machines as of our know-how, thoughtless creatures at the mercy of every gadget which is technically possible, no matter how murderous it is. [5, p.3]

Here I see Harraway's inertia, Hayles' computational regime, the fallacy of misunderstanding computational machines as cognising others. Arendt claims not to offer a solution in "The Human Condition", but writes: "What I propose, therefore, is very simple: it is nothing more than to think what we are doing" [5, p.5]. In this proposal, I think, lies a very succinct ethical imperative for a conscious, appropriate technical practice. In Arendt's *Nomos*, appropriation is a negotiation, an ethical way to "think what we are doing", maybe because thinking is the humans' *modus operandi*. Appropriation must be a negotiation because there is no new land – the world is already articulated, yet constantly iterated and never twice the same.

In the beginning of this chapter I contrasted Latour's choreography of the dancer recoiling from sins of the past until she is facing Gaia, to a free improvisation session as an alternative interpretation of what Gaia could be. But what about the choreography of the initial installation, which stood at the beginning of this project? In it, participants' movements through space change how the space sounds, and hence change how the room is experienced by all other participants. The movements of participants in a room change the sound of that room in direct correlation. Can it be re-interpreted too, for its meaning at the end of this exploration?

I was at the time as much fascinated by the expressive and aesthetic potential of this idea as I was obsessed by the possibility of making it technical reality – its potential meanings were not important to me at the time. But now I can see how the installation exemplifies (as a sonification of a reality) that which happens all along: every gesture we perform pulls, pushes, warps, compresses, lifts, and weighs on everything surrounding us and ourselves with an incredible ease – but irreversibly, it *appropriates*. Very much how the proverbial butterfly can cause a storm. Yet, there is resistance by a surprisingly resilient world and our bodies, constituted of all other actants, and all other bodies pulling and pushing too, *interacting*.

If I was asked to set that installation up again, I would now contextualise it less as a *virtual reality*. I would try to find a way to create a situation, a socio-material arrangement, where we can experience that when we move through space, space *is* (and hence sounds, looks, and feels) different as soon as we appropriate a new position in it, and that the resulting space is irreversibly changed for everybody else in the room. I would try to emphasise that this change can be directly experienced by the body, without the need for many added (schismatic) layers of technology. Immersion is an activity, a technique of the body, an interactivity, an appropriation of an embodied state of mind.

So, how much technology is really necessary to re-unite gesture and sound through embodiment? – I think here too, to return to the simile of the Preface, I would probably look with some surprise at the old nail – appreciatively, for sure, but then put it back into the kitchen cupboard. For a rainy day.

Bibliography

[1] James Andean, "Research Group in Interdisciplinary Improvisation, Goals, Perspectives, and Practice." in *This and That*, ser. Episodi, Annette Arlander, Ed. Helsinki: University of the Arts Helsinki, 2014, vol. 4, pp. 174–191.

[2] James Andean and Alejandro Olarte, *Sound Practices: The Sonic and the Musical in Interdisciplinary Improvisation*. Newcastle-upon-Tyne, UK: Cambridge Scholars Publishing, 2019.

[3] Hanna Arendt, *The Promise of Politics*, ser. G - Reference, Information and Interdisciplinary Subjects Series. Schocken Books, 2007.

[4] Hannah Arendt, *The Origins of Totalitarianism*. Schocken Books, 1951.

[5] Hannah Arendt, *The Human Condition*, ser. Anchor Books. University of Chicago Press, 1958.

[6] Hannah Arendt, *Eichmann in Jerusalem: A Report on the Banality of Evil*, ser. A reporter at large. Viking Press, 1963.

[7] John Langshaw Austin, *How to do things with words*. Oxford: Oxford university press, 1975.

[8] Alain Berthoz, *The Brain's Sense of Movement*, ser. Perspectives in cognitive neuroscience. Harvard University Press, 2000.

[9] Alain Berthoz, *La Décision*, ser. Sciences. Odile Jacob, 2003.

[10] Alain Berthoz, *The Vicarious Brain, Creator of Worlds*. Harvard University Press, 2017.

[11] Alain Berthoz, *Simplexity: Simplifying Principles for a Complex World*. Odile Jacob, 2018.

[12] Alain Berthoz, *L'Inhibition Créatrice*. Odile Jacob, 2020.

[13] Douglas Blackiston, Emma Lederer, Sam Kriegman, Simon Garnier, Joshua Bongard, and Michael Levin, "A cellular platform for the development of synthetic living machines," *Science Robotics*, vol. 6, no. 52, 2021.

[14] Antonio R. Damasio, *Self Comes to Mind: Constructing the Conscious Brain*. William Heinemann, 2010.

[15] Ferdinand de Saussure, W. Baskin, P. Meisel, and H. Saussy, *Course in General Linguistics*. Columbia University Press, 2011.

[16] Gilles Deleuze, F.Ì. Guattari, and R. Hurley, *Anti-Oedipus*, ser. Continuum impacts. Bloomsbury Academic, 2004.

[17] Jacque Derrida, D.B. Allison, and N. Garver, *Speech and Phenomena, and Other Essays on Husserl's Theory of Signs*, ser. Northwestern University studies in phenomenology & existential philosophy. Northwestern University Press, 1973.

[18] Jacque Derrida, G. Graff, J. Mehlman, and S. Weber, *Limited Inc*. Northwestern University Press, 1988.

[19] Jacque Derrida and G.C. Spivak, *Of Grammatology*, ser. Of Grammatology. Johns Hopkins University Press, 2013.

[20] Paul Dourish and Genevieve Bell, *Divining a Digital Future: Mess and Mythology in Ubiquitous Computing*. Cambridge, MA: MIT Press, May 2011.

[21] Umberto Eco, *A Theory of Semiotics*, ser. A Midland Book. Indiana University Press, 1979.

[22] Robert M. Emerson, R.I. Fretz, and L.L. Shaw, *Writing Ethnographic Fieldnotes*, ser. Chicago guides to writing, editing, and publishing. University of Chicago Press, 2010.

[23] Simon Glendinning, *Jacques Derrida: A Very Short Introduction*. Oxford University Press, 2011.

[24] Mark B. N Hansen, *New Philosophy for New Media*. Cambridge, MA: MIT Press, 2004.

[25] Donna J. Haraway, *The Haraway Reader*. New York Routledge, 2004.

[26] Mark Haugaard, *Power: A Reader*. Manchester University Press, 2002.

[27] N. Katherine Hayles, *(How We Became Posthuman:Virtual Bodies in Cybernetics, Literature and Informatics)*. Chicago: University of Chicago Press, 1999.

[28] N. Katherine Hayles, *My Mother Was a Computer: Digital Subjects and Literary Texts*. University of Chicago Press, 2010.

[29] N. Katherine Hayles, *How We Think: Digital Media and Contemporary Technogenesis*. University of Chicago Press, 2012.

[30] N. Katherine Hayles, *Unthought: The Power of the Cognitive Nonconscious*. University of Chicago Press, 2017.

[31] Karl Marx ; herausgegeben von Friedrich Engels, *Das Kapital : Kritik der politischen Oekonomie : Der Produktionsprocess des Kapitals*. Verlag von Otto Meissner Hamburg, 1867.

[32] Ali Jahanian, Xavier Puig, Yonglong Tian, and Phillip Isola, "Generative Models as a Data Source for Multiview Representation Learning," *ArXiv*, vol. abs/2106.05258, 2021.

[33] Marc Jeannerod, *Motor Cognition: What Actions Tell the Self*, ser. Motor Cognition: What Actions Tell to the Self. OUP Oxford, 2006.

[34] Anna Jurkevics, “Hannah Arendt liest Carl Schmitts Der Nomos der Erde: Ein Dialog über Gesetz und Geopolitik anhand ihrer Marginalien,” *HannahArendt.net*, vol. 8, no. 1, Jun. 2016. [Online]. Available: https://www.hannaharendt.net/index.php/han/article/view/342

[35] Bruno Latour, *Aramis, or the Love of Technology*. Harvard University Press, 1996.

[36] Bruno Latour, *Pandora’s Hope: Essays on the Reality of Science Studies*. Harvard University Press, 1999.

[37] Bruno Latour, *Reassembling the Social:An Introduction to Actor-Network-Theory: An Introduction to Actor-Network-Theory*, ser. Clarendon Lectures in Management Studies. OUP Oxford, 2005.

[38] Bruno Latour, “Distributing Cognition,” 2012, documentation on the AIME online platform. [Online]. Available: http://www.modesofexistence.org/aime/doc/512

[39] Bruno Latour, *An inquiry into modes of existence : an anthropology of the moderns*. Harvard University Press Cambridge, Massachusetts, 2013, translated by Catherine Porter.

[40] Bruno Latour, “Articulation,” 2013, glossary entry in the AIME online platform for the therm "articulation". [Online]. Available: http://www.modesofexistence.org/aime/voc/17

[41] Bruno Latour, “Bifurcation,” 2013, glossary entry in the AIME online platform for the therm "bifurcation". [Online]. Available: http://www.modesofexistence.org/aime/voc/35

[42] Bruno Latour, “Diplomacy,” 2013, glossary entry in the AIME online platform for the therm "diplomacy". [Online]. Available: http://www.modesofexistence.org/aime/voc/109

[43] Bruno Latour, “Felicity Condition, Infelicity Condition,” 2013, glossary entry in the AIME online platform for the therm "felicity condition, infelicity condition". [Online]. Available: http://www.modesofexistence.org/aime/voc/63

[44] Bruno Latour, “Fold, Folding,” 2013, glossary entry in the AIME online platform for the therm "fold, folding". [Online]. Available: http://www.modesofexistence.org/aime/voc/389

[45] Bruno Latour, “Matter, Materialism,” 2013, glossary entry in the AIME online platform for the therm "matter, materialism". [Online]. Available: http://www.modesofexistence.org/aime/voc/289

[46] Bruno Latour, “[MOR],” 2013, glossary entry in the AIME online platform for the therm moral mode of existence. [Online]. Available: http://www.modesofexistence.org/aime/voc/337

[47] Bruno Latour, “NET,” 2013, glossary entry in the AIME online platform for the network mode of existence. [Online]. Available: http://www.modesofexistence.org/aime/voc/505

[48] Bruno Latour, “Phenomenology,” 2013, glossary entry in the AIME online platform for the therm "phenomenology". [Online]. Available: http://www.modesofexistence.org/aime/voc/385

[49] Bruno Latour, “Quasi Subjects,” 2013, glossary entry in the AIME online platform for the therm "quasi subjects". [Online]. Available: http://www.modesofexistence.org/aime/voc/429

[50] Bruno Latour, "[REF]," 2013, glossary entry in the AIME online platform for the reference mode of existence. [Online]. Available: http://www.modesofexistence.org/aime/voc/441

[51] Bruno Latour, "[TEC]," 2013, glossary entry in the AIME online platform for the technical mode of existence. [Online]. Available: http://www.modesofexistence.org/aime/voc/547

[52] Bruno Latour, "[TEC·MOR]," 2013, glossary entry in the AIME online platform for the therm crossing of the technical and the moral modes of existence. [Online]. Available: http://www.modesofexistence.org/aime/voc/607

[53] Bruno Latour, "Value," 2013, glossary entry in the AIME online platform for the therm "value". [Online]. Available: http://www.modesofexistence.org/aime/voc/577

[54] Bruno Latour, *Aramis*, 1st ed. Mohr Siebeck, 2018.

[55] Bruno Latour and C. Porter, *Facing Gaia: Eight Lectures on the New Climatic Regime*. Wiley, 2017.

[56] Christophe Leclercq, "An Introduction to AIME," 2013, uploaded 5th August 2013. [Online]. Available: https://vimeo.com/44155844

[57] Shmuel Lederman, *Arendt and the Council Tradition*. Cham: Springer International Publishing, 2019, pp. 171–197.

[58] Frank E. Garcea lena Stasenko and Bradford Z. Mahon, "What happens to the motor theory of perception when the motor system is damaged?" *Language and cognition*, vol. 5, no. 2-3, pp. 225–238, 2013.

[59] Trond Lossius, Pascal Baltazar, and Théo de la Hogue, "DBAP - Distance-Based Amplitude Panning," in *ICMC*. Michigan Publishing, 2009, last visited 2020-05-03. [Online]. Available: https://quod.lib.umich.edu/i/icmc/bbp2372.2009.111/1

[60] Jame E. Lovelock, *The Ages of Gaia: A Biography of Our Living Earth*, ser. A Commonwealth Fund book. Oxford University Press, 1995.

[61] James Lovelock, *The Vanishing Face of Gaia: A Final Warning*. Basic Books, 2009.

[62] James E. Lovelock, *Gaia: The Practical Science of Planetary Medicine*. Oxford: Oxford University Press, 2000.

[63] Dave G. Malham, "Spatial Hearing Mechanisms and Sound Reproduction," 1998, last visited 2020-05-03. [Online]. Available: https://tinyurl.com/ya9mgrue

[64] Lev Manovich, "Database as Symbolic Form," *Convergence*, vol. 5, no. 2, pp. 80–99, 1999.

[65] Lev Manovich, *The Language of New Media*, ser. Leonardo. Cambridge, MA: MIT Press, 2002.

[66] Karl Marx and H. Lederer, *On the Jewish Question*, ser. Readings in modern Jewish history. Hebrew Union College-Jewish Institute of Religion, 1958.

[67] Humberto R. Maturana and Francesco J. Varela, *Autopoiesis and Cognition: The Realization of the Living*, ser. Boston Studies in the Philosophy and History of Science. Springer Netherlands, 2012.

[68] Marcel Mauss, *Sociologie et Anthropologie*. Paris: Presses Universitaires de France, 1968, ch. Techniques of the Body, pp. pp. 364–386, This lecture was given at a meeting of the Société de Psychologie, May, 7th, 1934 and published in the Journal de psychologie normal et patholigique, Paris, Année XXXII, 1935, pp. 271-93.

[69] Ian McEwan, *Machines Like Me*. London: Random House, 2019.

[70] Marshall McLuhan, *Understanding Media*. New York: McGraw Hill, 1964.

[71] Yuri P Medvedev, Alexander V Sorokin, and Victor I Khashchanskiy, "Method for information signal acquisition into a digital processing device," Russian Patent SU 1 508 275 A1, 1989, last visited 2020-05-03. [Online]. Available: https://yandex.ru/patents/doc/SU1508275A1_19890915

[72] Maurice Merleau-Ponty, *Phenomenology of Perception*. Taylor & Francis, 2013.

[73] Raymond Monelle, *Linguistics and Semiotics in Music*, ser. Contemporary music studies. Routledge, 2014.

[74] Bence Nanay, "Perception, Cognition, Action," 2016, in obo in Philosophy.

[75] Carrie Noland, *Agency and Embodiment*. Harvard University Press, 2009.

[76] Mika Ojakangas, *A Philosophy of Concrete Life: Carl Schmitt and the Political Thought of Late Modernity*. SoPhi Academic Press, 2004.

[77] Wolfgang Palaver, "Carl Schmitt on Nomos and Space," *Telos*, vol. 1996, pp. 105–127, 01 1996.

[78] Graham Chapman John Cleese Terry Gilliam Eric Idle Terry Jones Michael Palin Terry Jones (Producer), *Life Of Brian*. United Kingdom: HandMade Films, Python (Monty) Pictures, 1979, motion Picture.

[79] Ville Pulkki, "Virtual Sound Source Positioning Using Vector Base Amplitude Panning," *Journal Audio Eng. Soc*, vol. 45, no. 6, pp. 456–466, 1997, last visited 2020-05-03. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=7853

[80] Friedemann Pulvermüller, Rachel L. Moseley, Natalia Egorova, Zubaida Shebani, and Véronique Boulenger, "Motor cognition–motor semantics: Action perception theory of cognition and communication," *Neuropsychologia*, vol. 55, pp. 71–84, 2014, special Issue in Honor of Marc Jeannerod.

[81] Schlienger, Dominik and Olarte, Alejandro, "Carte blanche, right now!" in *Art Without Boarders, Proceedings*, Helsinki, Finland, 2016, pp. 253–256.

[82] Carl Schmitt, *Der Nomos der Erde im Völkerrecht des Jus Publicum Europaeum*. Duncker & Humblot, 1997.

[83] Jeong Seob Lee and Woon Seung Yeo, "Sonicstrument: A Musical Interface with Stereotypical Acoustic Transducers." in *NIME*. nime.org, 2011, pp. 24–27. [Online]. Available: https://www.nime.org/proceedings/2011/nime2011_024.pdf

[84] Gilbert Simondon, *Du Mode dÉxistence des Objets Techniques*. Paris: Aubier, Editions Montaigne, 1958.

[85] Gilbert Simondon, *On the Mode of Existence of Technical Objects*. University of Western Ontario, 1980, originally published Paris: Aubier, Editions Montaigne, 1958.

[86] Jesper Simonsen and Toni Robertson, "Routledge International Handbook of Participatory Design," in *Routledge International Handbook of Participatory Design*, Jesper Simonsen and Toni Robertson, Eds. Routledge, 2012, pp. xx, 295 p. :.

[87] Daniel Smith and Thomas Schenk, "The Premotor theory of attention: Time to move on?" *Neuropsychologia*, vol. 50, pp. 1104–14, 01 2012, rejects theory, acknowledges some elements.

[88] Étienne. Souriau, *Les différents modes d'existence: Suivi de « Du mode d'existence de l'œuvre à faire ». Présentation par Isabelle Stengers et Bruno Latour*, ser. MétaphysiqueS. Presses Universitaires de France, 2015.

[89] Karlheinz Stockhausen, "Helikopter Streichquartett," Salzburg, 1994, 3rd scene of Mittwoch of the Licht opera.

[90] Lucy Suchman and Lucy A. Suchman, *Human-Machine Reconfigurations: Plans and Situated Actions*, ser. Learning in Doing: Social, Cognitive and Computational Perspectives. Cambridge University Press, 2007.

[91] Lucy A. Suchman, *Plans and Situated Actions: The Problem of Human-machine Communication*. New York, NY, USA: Cambridge University Press, 1987.

[92] Ville Suuronen, "Carl Schmitt's confrontation with the work of Hannah Arendt: A debate on totalitarianism, power, and banality of evil," *Global Intellectual History*, vol. 0, no. 0, pp. 1–36, 2022.

[93] Thomas Taylor, *The Hymns of Orpheus [in Verse]*. T.Payne, London, 1792, translated from the Original Greek: with a ... Dissertation on the Life and Theology of Orpheus [by T. Taylor].

[94] Antonio Torralba and Alexei A. Efros, "Unbiased look at dataset bias," in *CVPR 2011*, 2011, pp. 1521–1528.

[95] Francisco J. Varela, Evan Thompson, and Eleanor Rosch, *The Embodied Mind: Cognitive Science and Human Experience*. The MIT Press, 01 2017.

[96] Christian Volk, "From Nomos to Lex: Hannah Arendt on Law, Politics, and Order," *Leiden Journal of International Law*, vol. 23, 12 2010.

[97] Andrew J. Watson and James E. Lovelock, “Biological homeostasis of the global environment: the parable of Daisyworld,” *Tellus B*, vol. 35B, no. 4, pp. 284–289, 1983.

[98] Mark D. Weiser, “Creating the invisible interface: (invited talk),” in *UIST ’94*, 1994.

[99] Mark D. Weiser, “The Computer for the 21st Century,” *SIGMOBILE Mob. Comput. Commun. Rev.*, vol. 3, no. 3, pp. 3–11, jul 1999.

Part II

The Articles

Short Summaries of the Articles

I Acoustic Localisation Techniques for Interactive and Locative Audio Applications: This article gives background on the artistic applications of tracking technology in the sonic arts. It links modern concepts of spatial interaction to the spatialisation strategies of early Electroacoustic music. Further, it argues that acoustic localisation techniques could proffer competitive alternatives to motion capture, the prevalent technology applied in the field.

II Acoustic Localisation as an Alternative to Positioning Principles in Applications presented at NIME 2001-2013: In this paper, we purvey a rationale for the development of tracking systems applying acoustic localisation techniques: we identified 50 positioning technology applied in projects presented in NIME conference proceedings from 2001 to 2013 for which we argue that acoustic localisation techniques could have offered competitive alternatives and at very low cost.

III Acoustic Localisation for Spatial Reproduction of Moving Sound Source: Application Scenarios & Proof of Concept: This article describes four scenarios for which the proof of concept implementation of an acoustic localisation positioning system could provide a tool to overcome limitations in standard audio technologies. This tool, which enables automated panning of virtual sound sources, has been developed in an experimental artistic practice, the workshop on music, space & interaction, which was here first introduced.

IV Gestural Control for Musical Interaction using Acoustic Localisation Techniques: In this article I supply some evidence that the technology is scalable from tracking in large performance areas to smaller areas with lower latency, required in gestural tracking for real-time interaction. As a proof of concept I prototyped two implementation of the principle, firstly a Theremin - like pitch control interface and secondly, a spatial trigger for percussive sounds.

V Requirements on Kinaesthetic Interfaces for Spatially Interactive Sonic Art: This paper documents the requirements on tracking technology for spatially interactive sonic arts. I do this by comparing a theorised notion of an ideal kinaesthetic interface to results of an online survey and results from the workshop on music, space & interaction. From these I derive both qualitative and quantitative recommendations for design.

VI Immersive Spatial Interactivity in Sonic Arts – The Acoustic Localization Positioning System ALPS describes the outcome of several years of participatory development with musicians and artists with a stake in sonic arts, Based on the application scenarios in article III and the conception of the kinaesthetic interface in article V., we present a tracking system for spatially interactive sonic arts. We show how the approach addresses several technical problems and overcomes some typical obstacles to immersion in spatially interactive applications in sonic arts.

VII Prototyping Situations: Interdisciplinary Free Improvisation in Technology Development and Latour's Mode of the Technical: This article describes the method applied in the development of ALPS, a combination of principles of interdisciplinary improvisation and participatory design, as applied in the workshop on music, space & interaction. In the search for a theoretical background to consolidate the approach, Bruno Latour's description of the technical proves to be helpful.

VIII How Do We Experience Digital Arts? – An Exploration through Latour's Modes of Existence: According to Bruno Latour the technical, despite being everything, is mostly experienced through its failure. If we take him by his word, what is it then that we experience when we speak of "digital art"? How does this bear on conceptualizations of technologies? In addition to the technical, the fictional mode and the mode of reference help to understand the notion of the digital as it pervades culture and media. Using examples from music, visual arts and an experimental artistic practice dystopian visions of technology are disentangled, re-configured. Embodied agency and kinaesthesia play a major role in this process.

5

Acoustic Localisation Techniques for Interactive and Locative Audio Applications (I)

Schlienger, D. 2014. "Acoustic Localisation Techniques for Interactive and Locative Audio Applications." In Locus Sonus Symposium #8 on Audio Mobility, Proceedings, pp. 1–8.

Acoustic Localisation Techniques for Interactive and Locative Audio Applications

Dom Schlienger
Sibelius Academy Helsinki
Centre for Music & Technology
email: dominik.schlienger@siba.fi

Abstract

Reviewing the literature on positioning systems using acoustic source localisation principles for Interactive and Locative Audio Applications (ILAA) it becomes clear that Acoustic Localisation and Positioning Systems (ALPS) implemented on ubiquitous devices can provide an alternative to Motion Capture systems (MoCap) wherever multiple speakers are part of an application. Providing background on and defining the notion of ILAA, this paper argues that based on comparisons of recent applications in the literature, ALPS can provide competitive alternatives to MoCap, the system prevalently used in ILAA.

1 Introduction

With the advent of multitrack recording technology and its possibilities for spatial distribution, the relation between the origin of a sound and the listener was essentially abstracted. Today, with location aware mobile technology and almost ubiquitous internet access, we can use the position of a moving object in space as a dynamic parameter and map it, arguably, to anything. In the case of musical applications these mappings can be used for musical expression, to track gestural control for hyper instruments for example, or let a dancer control musical parameters through spatial cues.

The use of these spatial interfaces for musical applications are well documented in the proceedings of the international conference on *New Interfaces for Musical Expression* (NIME), for example. The importance of digital positioning technology in general is born witness by the yearly *IEEE conference on Indoors Positioning and Indoors Navigation*, and *Ubiquitous Positioning Indoors Navigation and Location Based Service.*

In Interactive and Locative Audio Appllications (ILAA), Motion Capture (MoCap) is the most commonly applied principle, using cameras to track movement. However, the question is if it is always the best choice. In our previous work [15], we found evidence that for some applications acoustic localisation techniques might perform equivalently to MoCap and often at a fraction of the cost.

In some scenarios acoustic localisation techniques could clearly outperform MoCap. MoCap requires a line of sight between the tracked object and the camera. If something obstructs this line, position data can not be obtained. An analogous situation for acoustic localisation, where an object obstructs the signal path between microphone and loudspeaker, is a lot less problematic due to the physical nature of the audio signal. In fact, the principle's feasibility is well documented [4, 8, 9, 13–15, 20], despite its rare implementation in ILAA. The implementation is particularly straightforward when multiple speakers are already part of the system, as is often the case in ILAA.

By comparing the performance of Acoustic Localisation and Positioning Systems (ALPS) with the documented performance of MoCap, we can further evaluate the principle's suitability for ILAA, and contribute to a growing body of work which advocates its implementation. ALPS are often cheaper and usually make better use of existing technology within ILAA, in accordance with the notion of ubiquitous computing, as our findings show.

The paper provides background on ILAA, followed by a discussion of technical aspect of ALPS and how they compare to MoCap and some MoCap hybrid systems. It then goes on to discuss the limits of MoCap for ILAA and why ALPS could provide alternatives, followed by a section on future work and conclusive remarks.

2 Background

2.1 Interactive and Locative Audio Applications

In the context here, ILAA are audio applications in which the position, or the change of position of an object in physical space gives a cue for some process to happen. In general, ILAA could also include audio applications where some process changes the physical position of something in the space, but for this paper the former definition shall apply.

The potential uses of positioning systems for interactive art, multimedia and interactive spatial music are well documented [24], and early concepts go back to the emergence of electroacoustic music in the 1950s. The *potentiomètre d'éspace* being an example thereof [10] , and the spatialisation schemes of Varèse *et al.* for the Phillips Pavilion at the Expo in Brussels 1958 [1], or Stockhausen's for *Kontakte* [18] . Admittedly, these spatialisations probably were interactive in nature merely by the absence of digital multitrack control. If the computers were available at the time to control dozens of audio track automatically, this option would have been embraced, effectively disposing of interactive control. However, the experimental spirit of the time, and an example of gestural control of spatial audio also in the modern interactive sense is visually documented by the photograph of Pierre Henry controlling the *potentiomètre d'éspace* in Figure 1. Already in 1941, in a more mainstream but not less ground breaking vein, the spatialisation of the soundtrack for Disney's *Fantasia*, Fantasound [7], used several operators to control the panning of separate tracks over multiple speakers.

A more recent example for ILAA in the mainstream are interactive video game controllers, like Wii and Kinect, where spatial interactivity provides increased immersiveness, also through surround sound [2, 3, 15] to create a virtual space. The virtual

Figure 1: Pierre Henry at the *potentiomètre d'espace* in a concert at the *Salle de L'Ancien Conservatoire*, Paris, 1952

space shall be defined here similar to Normandeau's definition in [12], as the audio space created for the listener by the composer or in paraphrase for the gaming environment, the developer. Virtual audio space can but does not have to be congruent with a real existing sound field in space. For the discussion on suitability of positioning technology for ILAA, this generalisation is helpful, as it provides a denominator for space as a concept in electroacoustic music, video games and spatial sound reproduction.

The very recent presentations held by Normendeau and Siegel at Sibelius Academy, both as guest performers at the 2013 *SibA MuTe Fest* [5] highlight exemplarily the interest of electroacoustic composers in ILAA. Compositions like *Movements in Possible Histories or a Composition for 24 Windows* [11], also at Sibelius Academy, in the surroundings of our research, as well as research conducted at SARC, Belfast; Locus Sonus Laboratoire Art Audio, Aix-En-Provence; Zentrum für Kunst und Medientechnologie Karlsruhe, the activities of NIME, just to name a few, show the importance of the field for composers and musicians.

Electroacoustic spatial composition is just one area of ILAA where positioning technology can provide interaction. The ubiquitousness of mobile technology facilitates locative games like *Papa Sangre* [23], for example. Spatial audio interaction is a growing area in music technology, and is becoming increasingly pervasive. Various ongoing projects at institutions like Pervasive Media Studio Bristol, are examples therefore [19].

2.2 Acoustic and Local Positioning Systems

Using the principle of acoustic source localisation we can obtain the distance between a speaker and microphone by measuring the time delay of the acoustic signal, as we know the speed of sound through air *a priori.* If we know the distances for multiple speakers, a 3-D position can be triangulated. The accuracy of such a system lies within the low decimetre range and its latency is the same as for audio recording systems, in the low millisecond range. The latency derives from the buffer lengths applied, which means that at the cost of covered area, by shortening the buffer size, the latency can be reduced. This is useful for applications where small gestural movements need to be tracked, for example, in case of instrument control. The system is thus scalable. The presence of multiple loudspeakers is a prerequisite, which in most ILAA is given. Other than that, all that is required is a microphone. This makes the principle cheap for ILAA.

ALPS have not often been implemented to date. Rishab *et al.* presented a system which uses controlled ambient sound, i.e. pseudo random noise, as a signal to measure the time delays [14]. Random noise can be problematic for the use in ILAA, as the noise would need to be masked at all time by the audible signal. The system described in 2010 in [8] compares arrival times of a distinct signal on networked devices, applying a multiple receiver principle. The systems in [9] and [4] measure signals outside the limits of audible sounds, making them effectively ultrasound. Our suggestion is for a one receiver multiple sender model, using the audible airborne sound which is already part of an ILAA directly as a measuring signal, as described in [15]. ALPS are rarely implemented like this, particularly not by measuring the time delays of the signasl carrying the audio content itself of an audio application, despite their documented feasibility in [4, 8, 9, 13–15, 20]

Our previous work, about the suitability of positioning systems for ILAA in general, compared specifications in the literature to user requirements obtained from early findings of an ongoing survey.[1] The survey shows that optical tracking in form of MoCap provides good solutions as long as the requirement for line of sight between tracking device and tracked object does not cause issues. When this indeed does create issues, respondents revert to hybrid systems, where MoCap is combined with auxiliary systems using dead reckoning principles. Dead reckoning systems predominantly consist of inertia meters. In the literature the term inertial navigation system is thus often used. However, dead reckoning methods also include the use of compasses, providing a more general term. Dead reckoning systems usually require frequent updating with absolute data in ILAA, which predestines them for hybrid systems. On their own, they are usually only applied where absolute position is not necessary.

Further, the survey showed respondents' concerns about tracking as a privacy-sensitive issue. To have an opt-in choice is regarded as important. This can easily be achieved with the one receiver - multiple sender model. The control over privacy stays with the person holding the mobile device, the microphone.

[1] `http://creativemusictechnology.org/research_&_development_EN.html`

3 Comparing MoCap to ALPS

In the discussion on suitability of ALPS for ILAA one does not get around the fact that the currently predominantly used optical tracking systems are generally considered to be satisfactory. Not many people who use MoCap criticise it as there is not much to compare it to. Thus, it is in many cases the best available option.

However, the requirement of line of sight between a tracked device and a camera is an issue for many conceptual possibilities. Particularly when the tracked devices are to be on members of a crowd [21].

Also for the systems described by Normandeau for the Klangdom using ZirkOsc [6], positioning information from ALPS could open new possibilities. In the examples discussed in Siegel [17] on interactive applications, where dance controls music, ALPS could provide interesting and dynamic possibilities.

In view of audio-mobility, the dependence of MoCap on cameras means that ad-hoc networks of ubiquitous devices using MoCap are unrealistic. The environment needs to be controlled and the tracked object defined by colour or shape. Even if line of sight could be established between mobile devices, in a multi user environment this would require some form of choreography. Despite its popularity, MoCap has a considerable disadvantage to ALPS which is inherent to the very nature of the camera and constitutes its limitation as an interface to gather spatial data. The camera provides a 2-dimensional view of space and the further away from the camera an object is, the less information we gain about it.

The following example does not want to be misunderstood as a criticism on the artistic quality of the work in question. The tracking system applied by Siegel for *Two Hands, (Not Clapping)* [16], was originally developed as an interface for dancers to interact with music, wherein a movement in the field of the camera is digitally registered. But interaction through this interface requires the dancer to adapt performance to the interface's characteristics. Due to the optical perspective of the camera, a moving object close to the camera causes more change in pixels than an object further away. That is, it provides a warped perspective. Thus the interface's resolution for a movement in space is not linear but proportional to the distance. As innovative this is for the performance of *Two Hands, (Not Clapping)*, the fact that the interface weighs a particular amount of movement in one area of the performance space differently to the same amount in a different area makes it a poor interface for 3-D interaction. As a spatial interface, a lot of movement should be mappable to a *pro rata* equivalent. If this is not the case, additional information is needed, to differentiate between small cow very close, and large cow far away. ALPS could provide better results. N.B., not for *Two Hands, (Not Clapping)*, where the warped perspective is idiosyncratic part of the composition, but for a more general spatial human computer interface. This is particularly poignant if the performance space in question does want to be understood as 3-dimensional, which is usually the case in ILAA, especially in implementations in ad-hoc networks.

By using multiple cameras, depth information is added in some MoCap systems. For ILAA, this is a cost, which, in the presence of multiple loudspeakers, could be avoided by using ALPS.

It is evident that every medium influences the message by its characteristic [12]. This might not necessarily be negative, but it is understood that this influence can warp the data, as exemplified above. If the system is supposed to be a good interface according to the notion of ubiquitous computing, the influence of its characteristic, its visibility, has to be reduced [22].

4 Future Work and Conclusive Remarks

Looking at the existing literature and the early findings of our own research, the impression prevails that ALPS provide a competitive alternative for many spatially interactive applications, where audio is diffused over multiple speakers. The possibilities arising from ALPS implementation in ad-hoc networks stand in contrast with what is achievable with MoCap. MoCap's limitation as an ubiquitous interface for ILAA are further compounded by the intrinsic distortion of spatial data through the 2-dimensional depiction of space by a camera.

In the next steps of the research, it is considered of paramount importance to find ways of establishing requirements directly from the potential uses in musical practices. As ALPS are rarely implemented for ILAA, not much material is available to document precisely how the system will be used by musicians or any other early adapters. So insight needs to be gained experimentally. As very little music of spatially interactive nature exists, free improvisation suggests itself as a means to study interaction of space and audio. To use free improvisation as a methodology to explore the relation of organised sound and space is in itself a very interesting but vast field, which, as a means to establish user requirements, shall be explored in the future work.

References

[1] Spatialisation et image : la mise en espace des sons concrets et électroniques. `http://sonhors.free.fr/panorama/sonhors13.htm`, 2014.

[2] K. Collins. *An introduction to the participatory and non liner aspects of video games audio*. Helsinki University Press, 2007.

[3] K. Collins, editor. *From Pac-Man to Pop Music: Interactive Audio in Games and New Media*. Ashgate, Aldreshot, 2008.

[4] V. Filonenko, C. Cullen, and J. Carswell. Investigating ultrasonic positioning on mobile phones. In *Indoor Positioning and Indoor Navigation (IPIN), 2010 International Conference on*, Sept 2010.

[5] Centre for Music and Sibelius Academy Helsinki Technology. Mute fest 2013. `http://tinyurl.com/mqzl6nx`, 2013.

[6] Zentrum für Kunst und Medientechnologie Karlsruhe. Zirkonium. `http://www.zkm.de/zirkonium`, 2014.

[7] WM. E. Garity. Fantasound. *Journal of the Society of Motion Picture Engineers*, 1941. `http://www.widescreenmuseum.com/sound/fantasound1.htm`.

[8] T. Janson, C. Schindelhauer, and J. Wendeberg. Self-localization application for iphone using only ambient sound signals. In *Indoor Positioning and Indoor Navigation (IPIN), 2010 International Conference on*, Sept 2010.

[9] A. Mandal, C.V. Lopes, T. Givargis, A. Haghighat, R. Jurdak, and P. Baldi. Beep: 3d indoor positioning using audible sound. In *Consumer Communications and Networking Conference, 2005. CCNC. 2005 Second IEEE*, pages 348–353, Jan 2005.

[10] P. Manning. *Electronic and Computer Music*. Oxford University Press, New York, revised and expanded edition, 2004.

[11] L. Muredu and M. Mela. Movements in possible histories or a composition for 24 windows. `http://www.youtube.com/watch?v=ltqQVrVoHyE`, 2011.

[12] Robert Normandeau. Timbre spatialisation: The medium is the space. *Organised Sound*, 14:277–285, 12 2009.

[13] V. C. Raykar, I. V. Kozintsev, and Lienhart R. Position calibration of microphones and loudspeakers in distributed computing platforms. *Speech and Audio Processing, IEEE Transactions on*, 13(1):70–83, 2005.

[14] I. Rishabh, D. Kimber, and J. Adcock. Indoor localization using controlled ambient sounds. In *Indoor Positioning and Indoor Navigation (IPIN), 2012 International Conference on*, Nov 2012.

[15] Dominik Schlienger. Indoors and local positioning systems for interactive and locative audio applications. `http://creativemusictechnology.org/MScAudioProduction99001618TextNoApcs.pdf`, 2012.

[16] Wayne Siegel. Two hands (not clapping). `http://waynesiegel.dk/?page_id=24`, 2010.

[17] Wayne Siegel. *Dancing the Music*. Oxford University Press, Oxford, 2011.

[18] Karl Heinz Stockhausen. Kontakte. `http://www.youtube.com/watch?v=aNt6a5xFOnE`, 1959.

[19] Pervasive Media Studio. Pervasive media studio. `http://www.watershed.co.uk/pmstudio/`, 2014.

[20] S. Tervo. *Localization and tracing of early acoustic reflections*. PhD thesis, Aalto University School of Science, January 2012.

[21] Ryan Ulyate and David Bianciardi. The interactive dance club: Avoiding chaos in a multi participant environment. In Ivan Poupyrev, Michael J. Lyons, Sidney Fels, and Tina Blaine, editors, *NIME*, 2001.

[22] Mark Weiser. Building invisible interfaces. Presentation Slides, Novemeber, 2nd 1994.

[23] Wikipedia. Papa sangre — Wikipedia, the free encyclopedia. `http://en.wikipedia.org/wiki/Papa_Sangre`, 2014.

[24] R. Zvonar. A history of spatial music. `http://cec.concordia.ca/econtact/Multichannel/spatial_music.html`, 2005.

6

Acoustic Localisation as an Alternative to Positioning Principles in Applications presented at NIME 2001-2013 (II)

Schlienger, D., and S. Tervo. 2014. “Acoustic Localisation as an Alternative to Positioning Principles in Applications presented at NIME 2001-2013.” In B. Caramiaux, K. Tahiroglu, R. Fiebrink, and A. Tanaka, eds. Proceedings of the International Conference on New Interfaces for Musical Expression, pp. 439–442.

Acoustic Localisation as an Alternative to Positioning Principles in Applications presented at NIME 2001-2013

Dom Schlienger
Centre for Music & Technology
Sibelius Academy
Helsinki, Finland
dominik.schlienger@siba.fi

Sakari Tervo
Department of Media Technology
Aalto University
Helsinki, Finland
sakari.tervo@aalto.fi

ABSTRACT

This paper provides a rationale for choosing acoustic localisation techniques as an alternative to other principles to provide spatial positions in interactive locative audio applications (ILAA). By comparing positioning technology in existing ILAAs to the expected performance of acoustic positioning systems (APS), we can evaluate if APS would perform equivalently in a particular application. In this paper, the titles of NIME conference proceedings from 2001 to 2013 were searched for presentations on ILAA using positioning technology. Over 80 relevant articles were found. For each of the systems we evaluated if and why APS would be a contender or not. The results showed that for over 73 percent of the reviewed applications, APS could possibly provide competitive alternatives and at very low cost.

Keywords

Acoustic Source Localization, Interactive Locative Audio Applications

1. INTRODUCTION

Music, ever since its origins, has arguably always been spatially interactive. From call and response practices to marching bands, interacting through space is part of many musical practices. The advent of multichannel recording technology provided new possibilities for spatial distribution. Consequently the spatial and temporal relation between the origin of a sound and the listener fundamentally changed. Linked with the digital revolution this relation was further abstracted and new ways of dealing with this relationship became necessary and possible. The wide interest of researchers in computer human interaction and a large body of work bears witness to this.

For spatial interaction in particular, the proceedings of NIME 2001 - 2013 provide many examples of interactive locative audio applications (ILAA) corroborating the importance of the field. The systems presented apply a range of technologies to use spatial data as a a parameter in musical applications. Optical tracking principles like motion caption and infrared technology, gyro and accelerometers and hybrids of both are the largest groups of principles applied. However, to our knowledge, only very few positioning systems using acoustic source localisation have been realised for the use in ILAA [10, 6, 15], notwithstanding the principle's documented technical feasibility [17, 20]. In fact, trawling the roughly 1100 NIME proceedings for ILAA using acoustic source localisation techniques for positioning, none were found. The principle is conspicuous through its rarity; analysing its performance in applications a moot point.

Acoustic localisation is similar in principle and performance to Motion Capture (MoCap), which is predominantly used in ILAA, but it can operate without the requirement of line of sight between tracked object and camera. The effects of an object obstructing the signal path between a microphone and a speaker are a lot less detrimental to the signal than obscuring a marker in a MoCap system, due to the diffraction of sound around obstacles. Further, in our previous work [17], we compared various positioning technologies as to their suitability for ILAA. It became clear that besides well documented uses of optical tracking principles like MoCap, acoustic localisation techniques can provide a means to procure position data, particularly wherever airborne sound is part of an ILAA.

This paper shows that an acoustic positioning system (APS) could provide competitive alternatives for many applications. Based on the known performance characteristics, we evaluate for which ILAA presented at the NIME conferences from 2001 to 2013, APS could provide an equivalent or possibly even preferable positioning system.

The paper firstly summarises our previous work on the suitability of positioning systems for ILAA in general, secondly discuss early results of an ongoing online survey on the topic, and thirdly, analyses relevant NIME proceedings for comparison of positioning systems applied with the possibilities of APS. It is followed by a section on future work and conclusive remarks.

2. POSITIONING SYSTEMS

2.1 A Short Overview

Broadly, positioning systems can be typed by Systems providing absolute data in relation to a reference grid, i.e. GPS to longitude and latitude; relative data, i.e. the output of an acceleration meter; or symbolic data, for example, the fact that a mobile phone is within the reception range of a receiving mobile mast, or the statement that somebody is at home or at work.

Despite the summation of error in relative type positioning, hybrid systems using dead reckoning-principles like gyro and accelerometers, often bring good results when combined with absolute data from, say, optical tracking devices. Dead reckoning type devices require frequent updating with absolute data to be of use, which predestines them for hybrid -, rather than stand-alone systems or systems where only relative data is needed.

NIME'14, June 30 – July 03, 2014, Goldsmiths, University of London, UK.

Positioning systems using radio signals to triangulate positions, are almost ubiquitous due to WLAN technology. Liu *et al.* state systems estimating positions by signal strength measurements to rarely achieve accuracy bellow two metres. Other systems which are more accurate are not as ubiquitously available, and lack robustness [9].

Acoustic source localisation techniques can be used for positioning technology using time of arrival calculations by digitally measuring the time delay of an acoustic signal between a loudspeaker and a receiving microphone. From the time delay the distance can be calculated, as the speed of sound through air is known *a priori*. By calculating this distance from several loudspeakers to a microphone, the microphones 3-D position can be obtained. These systems provide accuracy in the low decimetre range at low latency, in the milliseconds range. APS laid out as a one microphone - multiple speaker system allows participant-performers to decide if they want to be tracked or not, as they are in control of the signal receiver, the microphone. Latency issues in APS depend on the necessary buffer sizes which are determined by the time sound takes to travel through air over a distance of interest. In applications where intrinsic small gestures akin to instrument control have to be tracked, a smaller buffer size can be chosen as the distances to be measured are shorter. Thus, APS are scalable. The prerequisite for APS is the presence of multiple loudspeakers and some processor. All that needs to be introduced to the system is a microphone, ideally wireless. Further, the signal which is already part of an ILAA can be used directly as a measure signal, necessitating masked or added measure signals for silent moments only.

MoCap, despite being chosen by most developers, has a considerable disadvantage to APS which is inherent to the very nature of the camera and constitutes its limitation as an interface to sample spatial data. That is, the camera provides a 2-dimensional view of space and the further away from the camera an object is, the less information we gain about it. This provides contorted data. This contortion can of course be rectified and calculated. But APS can provide uncontorted data directly and in a much simpler set up.

2.2 User Requirements for ILAA

In our previous work, the suitability of positioning systems for ILAA in general has been studied [17]. A summary of this analysis is listed in Table 1.

In addition to analysing the suitability of the positioning systems from an objective point of view, some answers were gained through an ongoing online survey on the topic. [17]. The survey was developed in cooperation with a focus group of professionals in the field at the Pervasive Media Studio, Bristol, UK, and is aimed at developers and early adapters and asks respondents about their experiences with and expectations on positioning systems for ILAA. Besides expectations on accuracy it also enquires about respondents' opinions on cost, ubiquitousness in the sense of unobtrusiveness, implication on privacy and more. The survey has yet to reach representative proportions but shows clearly that beside optical tracking, acoustic localisation techniques would meet most requirements.

In summary the survey's answers so far suggest the preference for positioning technology to be ubiquitous, working on existing technology, virtually everywhere, virtually invisibly, unobtrusive, hence requiring as little extra devices or gadgets as possible. Thus, users would prefer not to wear backpacks, hats and goggles as suggested in [4]. And position technology should be cheap, provide accuracy in decimetre range and low privacy intrusion, for example, in public spaces.

Being tracked is a privacy-sensitive issue. Hence, having an opt-in choice, is regarded as important by respondents. This can easily be achieved with the one receiver - multiple sender model, a principle Siegel describes as an inside-out system in [18]. It means that a participant can opt in to be tracked by being in control of the receiver, i.e. the microphone.

As the survey's aim was originally to establish what *any* positioning system would be required to do to be suitable for ILAA, acoustic systems were just an option of several. Thus to say that acoustic source localisation is an ideal solution from this survey alone, would be conjecture. From the answers to questions regarding expectations on actual specifications, accuracy, for example, it is clear that optical tracking provides the best *available* solution, as APS were not known to many respondents.

In discussions with the focus group it became clear that new developments are expected to adhere to principles of ubiquitous computing. According to Weiser, ubiquitous computers must know where they are [21]. Also he proposes a computer to be invisible, the technology to stay out of the way of the task [22]. It is self evident that this also applies to interface design. In application to interface design for musical expression, this can in many way stand as a defining difference between an interface and an instrument.

3. INTERACTIVE LOCATIVE AUDIO APPLICATIONS AT NIME 2001 - 2013

1100 titles of proceedings items of NIME 2001-2013 and full texts from 2001-2012 were searched for ILAA using acoustic localisation principles, and none were found. In lieu, we searched for paper titles on ILAA with the search terms *tracking, tracker, locative, localisation, positioning, position, motion, mocap, gestural, gesture, 3D, space, spatial*. The 80 relevant papers were filtered into sub-categories of technical principles. These groups were then compared to APS' documented specifications on several criteria like precision, range, ubiquitousness, latency, cost and the presence of multiple loudspeakers. Additionally, a full text search of the proceedings 2001-2012 yielded another 51 possible contenders for APS, out of 250 relevant hits on top of the ones form the title search. To keep the reference list short the full sample is referenced, with brief comments, in a downloadable spreadsheet.[1]

Of the 80 ILAA, 28 applications use MoCap, 12 Dead Reckoning and 10 hybrids of both, 6 use other principles, one does not name a technology but discusses positioning principles in general and 7 explicitly require a positioning system to enhance or realise proposed mappings, or they present filters or other algorithmic processes, but do not name a particular principle for positioning. The closest match to an APS ever presented at NIME 2001-2012 was only identified through the full text search. [8] applies Doppler effect analysis on test signals, which, arguably, are ultrasonic.

3.1 Discounted Applications

If an ILAA was in character more of an instrument than an interface it was discounted from the list of 80 applications. The distinction here comes from the notion that an instrument has an idiosyncratic character of which the positioning technology might be intrinsic part [19]. APS, in this contrast, wants to be understood as an interface, ideally as an invisible interface [22].

Eight presentations were on such actual idiosyncratic instruments or tools using positioning data, both absolute or

[1] `http://tinyurl.com/nngwcoa`

Table 1: Positioning Systems Performance in Overview

Principle	System	Accuracy	Area	Cost to user	Availability	Ubi.	Cost to installer
RF	Satelite 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	20 m	very low	DIY	yes	low
Inertial	Gyro/Accelerometer	0.5% - 20%*	1-100 m	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 lo-end	medium	scaleable	medium	market	no	low
Magnetic	Magnetic field	high	1-20 m	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

relative. As an instrument, the idiomatic way of playing with e.g., an optical interface [1] or floor pads [7], can be essential part of the instrument design, and using APS would change the character completely. Systems with a clear haptic idiom, like [12], were also discounted, as APS can not provide haptic feedback. Further, one system is excluded as it applies positioning data primarily as a global or symbolic parameter [11]. Three applications do not require positioning data at all, as they are spatialisation schemes where the position data applied is not actually spatial.

Of the 28 systems using optical tracking, 7 can be discounted as their specification explicitly states, that the gestural tracking of movement is to happen without any sensors attached to the body or fingers as in [5]. Similarly, where the object to be tracked is a person *passing by* in a public space [1], APS can not provide alternatives, as the activity of *picking up a device* can not implicitly be expected as being part of *passing by*. Except when mobile phone technology is mentioned and its presence can be considered to be ubiquitous, then APS is an option.

Two system which cover a table size area in 2D were considered to be not typical for ILAA, and [13] uses optical sensors in mm scale for bow tracking for violin-family instruments. APS is not able to measure in necessary detail.

Systems of gestural control were included if the tracked gesture could be expressed as a Cartesian position in space. This excluded some systems where the gesture was more symbolic in nature, similar to a fader or joystick movement. For obvious reasons, systems tracking facial expressions and contour tracking applications were excluded too.

Systems which use haptic information as integral part of their workings were not included, as haptic feedback is intrinsically absent in APS. If an auxiliary positioning system was explicitly mentioned as part of a haptic system, and its performance could be improved by APS, it was included.

3.2 APS provides alternatives

The above exceptions still leave us with 59 systems out of 80, (73.75 percent) wherein APS could provide comparable performance. Including the additional yield from the full text search, there are 110 contenders for APS out of 358 ILAA using positioning technology (30.72 percent). One common factor to all 110 systems is that airborne sound is explicitly part of the application, i.e. multiple loudspeakers are already part of the system, mostly in form of surround sound [3].

The most common principle of positioning applied in ILAA is optical tracking and specifically MoCap. The larger part of the applications in this group tracks performers in a room or performers area, at a precision level in the low decimetre range. Most MoCap systems use multiple cameras, 8 in the case of [2]. In these ILAAs APS could possibly provide a competitive alternative as it does not rely on line of sight to the camera. And by replacing 8 digital cameras with one analogue microphone the reduction both in processing power and calibration effort will be reflected in the cost.

To provide alternatives for MoCap in applications were a small area is being tracked APS can be scaled to higher accuracy at the cost of range by shortening the buffer size. Particularly conducting-type applications would profit, as they mostly rely on single camera systems or handheld devices. Evidently a handheld microphone could easily replace the IR functionality of a WiiMote as a baton [14].

The second largest group uses dead reckoning methods. The group can be further split into systems using wearable devices, wristbands or gloves, for example, and ones using handheld devices. The fact that this group relies on a device being on the performer means that replacing it with a small microphone is certainly not a step back and an improvement in precision can be achieved due to the availability of absolute position data.

In the group of hybrids are quite a few idiosyncratic tools like data gloves which work very well for their intended purpose. In the nature of their multi modality, adding a small mic would be rather in keeping. In many cases the camera - depending part of a hybrid system, i.e. MoCap or infrared, could be replaced with APS [16].

4. FUTURE WORK

These findings inform our ongoing development and implementation of an APS for ILAA which uses the same airborne audio signal as the one which carries the content of the application (Music, speech, sound) to measure the time delays of the signal on one microphone in relation to multiple speakers. This development project is further iguided by the notion of the invisible interface, and thus to use ready available technology in the typical set-up of an audio application. No further technology like cameras or wearable devices other than a microphone, which might already be part of the ILAA, shall be introduced. [22]

Based on the broad and documented interest of the community of developers, musicians and performers in spatial interactivity, we find it of utmost importance to include early adapters in the development of this system. We are implementing a workshop exploring musical spatial interaction to inform what the *system* needs to do, rather than presenting users with a system asking them *to do something with it*. The methodical implications of this approach shall be disseminated in a future paper.

More short term, a prototype APS shall be presented, implemented for use in a professional live - sound enforcement or recording environment, where musical performers' position on a stage can be tracked. This position data can then be used, for example, to automate the monitor mix depending on the performers position on stage.

5. CONCLUSION

The body of previous and related work shows clearly to what specification various positioning systems perform and give a clear notion of what these systems need to be able to do to match user requirements in ILAA. As very little literature exists on how the principles of APS would perform in ILAA, literature on *existing* ILAA using other principles was reviewed with the aim to identify the examples for which APS could provide an alternative with comparable performance. The astonishing result of this review is that none of the 1100 presentations, posters or papers included in the NIME proceedings from 2001 to 2012 presented a system using acoustic localisation techniques, despite the ubiquitousness of multi-track speaker arrangement in almost all of the 80 applications reviewed in detail. Peculiarly astonishing considering that the position data could be gathered cost free in case of mobile phone based applications or at the price of a simple omni directional microphone in others. Even for implementation in a professional audio environment using wireless microphone technology the costs pale into insignificance compared to professional MoCap systems.

As to the question if APS can *always* provide alternatives to optical tracking, the limitations are clear: APS can not provide an alternative for face- or contour tracking, which are typical visual interface tasks, nor can it track objects onto which a microphone can not be attached.

But, given the presence of multiple loudspeakers, wherever a performer's position needs to be tracked within a performance space of room size dimensions, APS can provide similar or equivalent results to many systems currently in use. APS are very easy to implement and use a modest amount of processing power compared to other systems.

Last not least, due to the possibility of opt-in positioning, APS could provide an alternative to some public space-installation where the presence of a camera might be perceived as an infliction on privacy issues.

6. REFERENCES

[1] G. Beyer and M. Meier. Music Interfaces for Novice Users : Composing Music on a Public Display with Hand Gestures. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 507–510, 2011.

[2] M. Ciglar. "3rd. Pole" - A Composition Performed via Gestural Cues. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 203–206, 2008.

[3] A. W. Colin Honigman and A. Kapur. The Third Room: A 3D Virtual Music Framework . In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 29–34, 2013.

[4] V. Filonenko, C. Cullen, and J. Carswell. Investigating ultrasonic positioning on mobile phones. In *Indoor Positioning and Indoor Navigation (IPIN), 2010 International Conference on*, Sept 2010.

[5] E. Guaus, T. Ozaslan, E. Palacios, and J. L. Arcos. A Left Hand Gesture Caption System for Guitar Based on Capacitive Sensors. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 238–243, 2010.

[6] T. Janson, C. Schindelhauer, and J. Wendeberg. Self-localization application for iphone using only ambient sound signals. In *Indoor Positioning and Indoor Navigation (IPIN), 2010 International Conference on*, Sept 2010.

[7] N. H. Keating. The Lambent Reactive : An Audiovisual Environment for Kinesthetic Playforms. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 338–343, 2007.

[8] J.-s. Lee and W. S. Yeo. Sonicstrument : A Musical Interface with Stereotypical Acoustic Transducers. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 24–27, 2011.

[9] H. Liu, H. Darabi, P. Banerjee, and J. Liu. Survey of wireless indoor positioning techniques and systems. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 37(6):1067–1080, Nov 2007.

[10] A. Mandal, C. Lopes, T. Givargis, A. Haghighat, R. Jurdak, and P. Baldi. Beep: 3d indoor positioning using audible sound. In *Consumer Communications and Networking Conference, 2005. CCNC. 2005 Second IEEE*, pages 348–353, Jan 2005.

[11] C. McKinney and N. Collins. An Interactive 3D Network Music Space. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 400–405, 2013.

[12] T. Mitchell and I. Heap. SoundGrasp : A Gestural Interface for the Performance of Live Music. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 465–468, 2011.

[13] L. Pardue and A. McPherso. Near-field optical reflective sensing for bow tracking. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 363–368, 2013.

[14] L. Peng and D. Gerhard. A Wii-Based Gestural Interface for Computer Conducting Systems. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 155–156, 2009.

[15] I. Rishabh, D. Kimber, and J. Adcock. Indoor localization using controlled ambient sounds. In *Indoor Positioning and Indoor Navigation (IPIN), 2012 International Conference on*, Nov 2012.

[16] J. C. Schacher. Motion To Gesture To Sound : Mapping For Interactive Dance. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 250–254, 2010.

[17] D. Schlienger. Indoors and local positioning systems for interactive and locative audio applications, 2012. http://creativemusictechnology.org/MScAudioProduction 99001618TextNoApcs.pdf.

[18] W. Siegel. *Dancing the Music*. Oxford University Press, Oxford, 2011.

[19] A. Tanaka and P. Gemeinboeck. A Framework for Spatial Interaction in Locative Media. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 26–30, 2006.

[20] D. B. Ward, E. A. Lehmann, and R. C. Williamson. Particle filtering algorithms for tracking an acoustic source in a reverberant environment. *Speech and Audio Processing, IEEE Transactions on*, 11(6):826–836, 2003.

[21] M. Weiser. The computer for the 21st century. *Scientific American*, 265(3):66–75, January 1991.

[22] M. Weiser. Building invisible interfaces. Presentation Slides, Novemeber, 2nd 1994.

7

Acoustic Localisation for Spatial Reproduction of Moving Sound Source: Application Scenarios & Proof of Concept (III)

Schlienger, D. 2016. “Acoustic Localisation for Spatial Reproduction of Moving Sound Source: Application Scenarios & Proof of Concept.” In International Conference on New Interfaces for Musical Interaction 2016, Proceedings, pp. 1–6.

Acoustic Localisation for Spatial Reproduction of Moving Sound Source: Application Scenarios & Proof of Concept

Dominik Schlienger
Centre for Music & Technology
Sibelius Academy
University of the Arts Helsinki
dominik.schlienger@uniarts.fi

ABSTRACT

Despite the near ubiquitous availability of interfaces for spatial interaction, standard audio spatialisation technology makes very little use of it. In fact, we find that audio technology often impedes spatial interaction. In our workshop on music, space and interaction we thus developed the idea of real-time panning whereby a moving sound source is reproduced as a virtual source on a panning trajectory. We define a series of application scenarios where we describe in detail what functionality is required to inform an implementation. In our earlier work we showed that acoustic localisation potentially can provide a pervasive technique for spatially interactive audio applications. Playing through the application scenarios with acoustic localisation in mind provides interesting approaches. For one scenario we show an example implementation as proof of concept.

Author Keywords

Acoustic Localisation; Music, Space & Interaction; Auto-Pan; Real-time Spatialisation; Spatial Interactivity

Figure 1: Impressions from our Workshop on Music, Space & Interaction [21, 22]

1. INTRODUCTION

Spatial interactivity in interface design for musical expression has steadily gained importance over the last decades, a fact to which numerous contributions to the NIME conference bear witness [1, 2, 3, 4].

Applications which use spatial information to cue some musical parameters range from applications to gesturaly direct a virtual orchestra [5], or to controlling and process sound files gesturaly [6], to sound walks using GPS [7]. In earlier years, spatially interactive technologies have been developed specifically with musical applications in mind. The Theremin (1928) or Henry Schaefer's Potentiomètre d'Éspace (1951) are examples thereof. Today, often existing technologies are appropriated [8, 9], proprietary technologies originally developed for computer games, or data gloves [10, 11]. They all have in common that they create a relation between an aural event and a movement through physical space, be it directly linked like in the Theremin or abstracted as in the Locusstream Soundmap [12]

We find important to remember that this relation between movement and sound is not new. In fact, all music is intrinsically spatial. If we define movement as displacement over time in space, and if this movement is oscillating, it is just a question of scale and medium if it is visible, audible - or not. Spatiality in music is about how much importance is given to its spatial aspect, and the role space is given within a particular musical practice.[1] It is not something that has been brought to music through a new technological invention.[13, 14] This might sound tenuous from a traditional software development point of view. However, the fundamental kinaesthetic element of sound-making, when musicians actually cause the spatiality of sound, is the means of expression we are working with in spatially interactive music. Creative tool development on that notion of spatiality demands a transdiciplinary approach [15], and an enquiry into musical practices. Such tool development can then be based on the notion of designing culture[16] and also on an understanding of the importance of kinaesthesia in embodied agency [17].

What is new, is that we can record, reproduce and synthesize sound; map, track and measure movement in space and with increasing speed and accuracy. We can abstract sound from its original spatial source and reproduce it in another location, and this nigh ubiquitously [18, 19, 20].

However, to come to our point, the situation in most musical practices is, that, despite the availability of all this new technology and means to design it, spatial interaction in performance situations is often impeded rather than helped by

[1]Depending on the musical practice, this can be both audible and visible: A pianist's gestures are part of the performance, even if the gesture is only part of making a sound

NIME'16, July 11-15, 2016, Griffith University, Brisbane, Australia.
.

the use of electronic audio technology. In 2013, to research the affordances of existing- and explore the possibilities of hypothetical technology we founded our Workshop on Music, Space and Interaction (MS&I) at University of the Arts Helsinki. It was instituted to explore the nature of spatial musical interaction and how we use space in a practice of free, interdisciplinary improvisation [23]. Our main interest is in the use of technology, what it does, could do or does not do for creative possibilities, using tools and methods adopted from Participatory Design [24]. Focusing on the musical and expressive applications rather than the implementation of a particular technology, we collectively explore the question of what musical expression actually requires from spatial interactivity, as this, in our opinion, should be the driver behind its development.

Our early findings - partially documented in participants' blog entries [25] show that the main concern with existing mainstream spatial audio technology is that the kinaesthetic experience of space is often lost in its technological translation. This has implications in a plethora of ways which, to explore conclusively, goes beyond this article.

But the one major theme that recurred regularly in nearly every session and what we would like to make the subject of this article, is this: Looking at the uses of audio technology available to professional end users, we recurrently felt limited by the rigidity of loudspeaker layouts, and the limited ability to translate a performer's motion through the room musically while she or he is hooked up to the loudspeakers, even if this hook up is wireless. The frustration with this lack of flexibility led many participants, and at more than one occasion, to ditch their laptops connected to a state of the art surround sound system and pick up a simple object, like a wooden stick, for example. With respect to spatial interactivity and as an interface for musical expression the humble wooden stick just simply had the edge over the thousands of Euro worth of high-end equipment.

So the idea of spatialising a performer's trajectory as a virtual source on a multi-speaker system, what could possibly be described as a real-time auto-pan, was thus floated early on in the workshop. A variant of the same idea was to record a moving sound source's trajectory along with its audio recording, thus allowing a spatialised reproduction in real-time or at a later stage, or in another place.

We play through possible scenarios in detail and provide proof of concept for a simple implementation for one scenario using Acoustic Localisation techniques (AL), as our earlier research into its possibilities for interactive audio applications showed some promises in respect to scalability, precision, and competitiveness in comparison to existing systems. [26, 27]. Despite AL's relevance to the field, for standard audio technology, only few applications exist [28, 29, 30, 31].

The following section describes how we base our research in musical practice, introduce application scenarios and after a summary on AL provides an overview on the experimental layout for our first simple, proof of concept implementation on which we base our discussion on what we consider to be key aspects towards an implementation.

2. METHODS

We propose application scenarios for real-time spatialisations using AL technologies. We base the requirement for these applications on early findings of our work in MS&I where we adapted tools and methods developed for Participatory Design [24] and Interdisciplinary Improvisation [23], a method to enhance interdisciplinarity in the arts developed at Uniarts Helsinki since 2012

2.1 Music, Space & Interaction

We use the format of Interdisciplinary Improvisation sessions, where free sessions are followed by reflective discussions. Essentially we adapted Interdisciplinary Improvisation's stance on expertise, whereby, although participants are experts in their own field, to facilitate experimentation, they are encouraged to gain experiences outside of their own discipline. This allows to form common ground across disciplines. Participatory Design [24] is a long established design approach in, for example, software development in the workspace. But it is surprisingly rarely used in creative tool design. In a nutshell, the idea behind Participatory Design is that a technology should be developed by its users. In Participatory Design this idea is followed a lot more consequently as for example in user centred design.

MS&I's participants are professionals in music- composition and performance, dance, scenography, video, installation art, theatre, fine arts, architecture, game design, to just name a few. The workshop is held on weekends (Friday-Sunday) 3-4 times per term. To experience a place's spatiality we explore it for its particular sound, how our instruments sound differently depending from where we play them, how the room sounds if we excite it as an acoustic instrument, how we experience the aural quality of the space when moving through it and so on. We try to approach space as a sounding physical entity, as a found object, as a technological extension of our body, in short, its audio-kineasthetic potential for musical expression. Based on these experiences we then explore how the technologies, - some of which we bring in from our respective fields of expertise, some of which we invent on the spot, could enhance, improve or facilitate aspects of the experience. The data gained in MS&I is ethnographic. Every improvisation session is followed by discussions where we make notes, set out new experiments, compare experiences. Participants are also encouraged to contribute to a blog [25].

2.2 Acoustic Localisation

AL is generally defined for localisation in the whole acoustic frequency range, which includes infra- and ultra sound, We are solely interested in the frequency range which can be propagated and recorded with standard audio equipment, say, roughly, from 20 Hz to 30 kHz. In the further text the term AL refers to AL techniques using this frequency band. AL is used here as this is the implementation we are working on, but arguably any tracking device could be used instead. What makes AL attractive, and why it is our first choice, is that the application scenarios in question exist of an arrangement of loudspeakers and microphones.

The particular technique in question is Time Difference of Arrival (TDoA) For the Single Microphone Multiple Loudspeaker (SMML) approach, a signal distributed via loudspeakers is recorded on a microphone. From correlating the original signal with the recorded signal we measure the time delay. As the speed of sound through air is known *a priori*, we can work out the distance between the source and the microphone. In principle identical, but not in application, is to work out the TDoA using multiple microphones and an acoustic sound source not known *a priori*. We will call this the Multi Microphone or MM approach. (See Figure 2)

For some scenarios it is possible to use the acoustic signal which is the content of an application as the measurement signal. In some applications a dedicated measurement signal can be used instead, under the condition that these signals are not detrimental to the acoustic content of the application i.e. they are masked or non audible. For this *test signal* we used band-passed noise just above the frequency

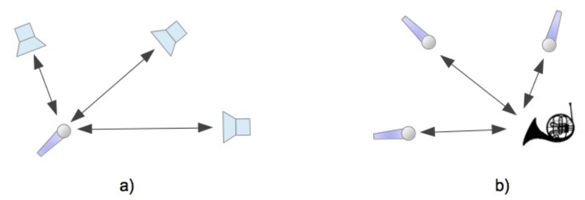


Figure 2: Time Difference of Arrival (TDoA). In a) TDoA of a known signal is measured on a single microphone. In b) TDoA of an unknown signal is measured on multiple mirophones

range audible to the human ear, as standard audio equipment generally reaches frequencies above it, sometimes up to 30 kHz. It is a lot more convenient, however, if the content signal can be used directly for the measurement, the reason being that the measurement can be made at no extra cost, it is data already available wherever a microphone is connected to a loudspeaker, or where a system has access to more than one microphone. This generalisation presumes that a distance measurement on its own is already enough for spatialisation which, as we will show, in some scenarios is in fact the case. Otherwise, generally, a spherical wave model applies, thus 4 TDoA measurements are necessary to conclusively trilaterate a position in 3D space.
With the view towards an implementation, a note here on spatialisation principles: For our scenarios, Ambisonics or Wavefront synthesis or Vector Based Amplitude Panning (VBAP) [32] are in principle applicable. However, although not explicitly a requirement on our scenarios we work with the assumption that the implementation will use VBAP, for sake of its simplicity and scalability, two or three loudspeakers suffice, their positions can be arbitrary.

2.3 Experimental Set Up

For our first implementation as proof of concept, we wrote a simple matlab script, using the Playrec library [33] which allows non-blocking soundcard access and thus continuous and simultaneous play and record. Newer versions of Matlab could do this natively, however, playrec provides insight into the sample by sample workings which was considered helpful for research purposes.

The layout of the space where the experiments were conducted is a 2.9 times 5.7 meter office space with a high ceiling at 3 meters, see Figure 3 for details. The room had been moderately acoustically treated and had a wideband reverberation time of 0.38 Seconds. Eight loudspeakers of type Meyer Sound MM4XP were spaced at 1.6 metres distance from each other squeezed to an oval inside a 4.6 x 2.5 right-angled rectangle at 1.5 metre height off the floor. The loudspeakers are co-axial two-way loudspeakers with a nominal frequency response from 135 Hz to 17 kHz ± 4dB. We used bandlimited noise between 17 -18.5 kHz as test signals. For the position of all loudspeakers see Figure 3.

The microphones used were of type AKG C417, and DPA VO4099V. The rest of the equipment were AKG Perception wireless SR 45 receivers and PT 45 Analogue Band D Senders for the microphones and a MOTU 16A sound card. The processor running Matlab was an 11-inch, Mid 2011 MacBook Air, 1.6 GHz Intel Core i5, running OSX 10.8.5 The Matlab version was R2013a.

The set up was for 2D panning for demonstration purposes only. Extending the current script to 3D panning is trivial: The only prerequisite for 3 dimensionality is that the loudspeakers are not all positioned on the same plane.

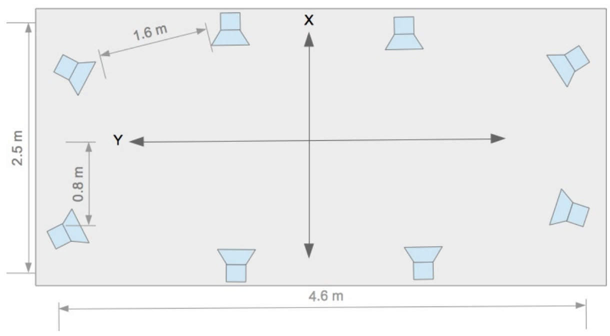


Figure 3: Loudspeaker Layout

2.4 Application Scenarios

We isolated four conceptual possibilities:

- Scenario 1: For traditional music practices: The sound source is in another space than where it is reproduced, e.g., stage, recording or broadcast studio.
- Scenario 2: The moving sound source to be reproduced is in the same space as the audience,
- Scenario 3: a performer's trajectory is reproduced in real-time but the content is or was recorded elsewhere.
- Scenario 4: Both the moving sound source and the trajectory were recorded separately and offline, the spatialisation happens temporally removed.

We can then analyse which approach is a better fit for SMML or MM and also why AL is a possible choice or not. For scenarios 1 and 2, we presume that there is a moving acoustic sound source, which is close-miked with a wireless microphone. The loudspeakers are in fixed positions. For scenario 3 and 4, the sound source is virtual, pre-recorded, electronic, or produced elsewhere.

Scenario 3 is a special case, as the the sound source arguably could have been, or can be, recorded as a close-miked acoustic source, as in Scenarios 1 and 2, but it is not, or was not moving along the trajectory in question when doing so. This is also an exception to scenario 4, as even if the sound is pre-produced, its trajectory is online and not predefined.

2.4.1 *Two Spaces: Stage and Audience*

We start with the scenario we are probably most familiar with from most musical practices, where the moving sound source is in another space than the audience, say on a stage, or in a studio. In this scenario the loudspeakers distributing the content are not in the performance area. An application in the mainstream for this scenario is, for example, to mirror the trajectories of musical performers on a stage by panning, to match the aural experience with what the audience experiences visually. The same principle applies also for a broadcast situation, where the spatialisation is reproduced in a place remote from the performance space or stage, but spatialised as it was in the performance. In this scenario, if the content signal is to be the measurement signal, the MM approach is clearly at an advantage, as the content signal is not necessarily distributed in the performance area. If there are loudspeakers in the performance area, a SMML approach might work if the content signal can be used for localisation. To use a SMML approach here with a distinct test signal will mean that the test signal has to be masked sufficiently not to create any audible impact on the content recording.

2.4.2 One Space: The Common

The moving sound source to be reproduced is in the same space as the audience. This approach is possibly best described as a real-time auto pan proper: A close-miked acoustic sound source produces a signal which is simultaneously played back on the loudspeakers panned to its closest phantom position.

Application areas for this approach could be for example installation art, participant - performances, exhibitions, public spaces, art-museum, festival commons or similar. To get this to work one way or other would indeed benefit a large community of artists and musicians.

2.4.3 Virtual Sound Source

The trajectory is reproduced in real-time but the content is recorded elsewhere or was pre-recorded This approach is of interest for many applications in electronic music, where the moving *sound source* might not actually be acoustic, (Laptop performance, electronic instruments, mobile phones, or similar) but also for interactive sound installations where the position of participants carrying a microphone as a sensor triggers certain events, or where a trajectory creates narrative meaning to musical content. It is a way to map offline content in real-time to an online, or a live trajectory in space. The trajectory is not known *a priori* If SMML is used, the applicability depends on the successful masking of the measurement signal - a distinct measurement signal is necessary as there is no close-miked acoustic sound source: Performers need to carry a microphone for positioning purposes only. Which would also be necessary for a MM approach. We implemented an approach based on simple distance readings as a proof of concept for this scenario, for the performance of the interactive piece *Leluhelikvartetti, a Hommage à Stockhausen.* (See Appendix B)

2.4.4 Track and Sound Temporally Separated

The moving sound source is spatially and temporally removed. In this scenario the same set of loudspeakers can be used. As the reproduction happens later than the recording of the trajectory, corrections are possible. Further, here it is possible, to record a trajectory separately and then apply it to a sound recorded elsewhere. The test signal could arguably be audible noise, as long as the performer recording the trajectory is happy with it. This approach is a bit like using motion capture in video animation: a trajectory is recorded, and then content mapped to it. In fact, this scenario, if a performer is equipped with many microphones, the recorded track could be used even as an equivalent to motion tracking

3. RESULTS AND DISCUSSION

We base our discussion of the application scenarios on the results from our test implementation, providing us with something tangible to theorise with. For demonstration purposes we split the tracking process into a two step sequence, first recording the trajectory and then reproducing it. The script in question can be modified to do both things in the same process, but it would be difficult to compare the original and the reproduced trajectory if it was to be observed at the same time:

We made a recording with a stereo microphone placed in the middle of the room, while moving along the trajectory with a close-miked live sound source. (See appendix A)This source's trajectory was reproduced using the Playrec script in Matlab, and recorded again with the same stereo microphone for comparison. The accuracy of the angles are best compared by a listening test in Appendix A. Figure 4 shows the localization results. The top row shows the azimuth angle in radians in relation to the loudspeaker positions over the same time span.

Our prototype works on the principle that we know which single loudspeaker is the closest based on the fact that we receive a signal. So actual panning is possible only in hindsight. During a trajectory, once a new loudspeaker is within reach, the distance between the newest and the last measurement can be calculated and an interpolation between the 2 points effected. Potentially, the system is constantly playing catch with the newest measurement and lag is unavoidable. With an increasing number of loudspeakers this can be improved. The algorithm which allows for a smooth transition between the loudspeakers thus causes a delay. The new position has to be reached before an interpolation is possible.

This is certainly not fast enough to control percussive instruments, but to spatialise an instrumentalist's trajectory across a stage, is within the margins we find acceptable. The latency of the distance measurement, however, albeit not used yet for any functionality, has no perceptible latency, and is defined by the systems buffer length only.

However, a localisation based on trilateration of distance measurements from more than one loudspeaker would clearly be significantly faster, as no interpolation on hindsight is necessary. The question of course is, why did we not apply this latter approach? - The nature of the test signal impacts on what is possible.

3.1 Distinct Test-Signal: Implications

In a SMML approach, we send uncorrelated signals to every loudspeaker. Then we measure the correlation between the original signal and the signal on the loudspeaker in question. Consequently, the measurement signal also contains the 7 signals sent to the other loudspeakers, effectively being noise. Based on these observation, we assume that we look at a systemic problem which needs a lot more investigation: Is it sensible to use test signals at all when they effectively add noise to an audio application? Would a MM approach not provide better results? What can be achieved with a multi frequency-band approach whereby the test signal would be split in a separate band per loudspeaker?

Using high frequency noise at the hearing threshold brings another limitation with it: In our previous research we stress that (theoretically) AL has the advantage over optical tracking in not needing line of sight between the tracked receiver and the test signal sender. However, the frequency bands we tested all have similar limitations as the diffraction of sound waves around objects requires the diffracted soundwave to be longer than the object it diffracts about. The frequencies between 17 kHz to 20 kHz are just up to 2 cm long. In line with this, our observations showed that we still get readings if we interfere the line of sight with a hand, but we lose the signal if we move our body between it. Arguably this is a better result than what can be achieved with optical tracking, particularly as this also works in the dark, but it is hardly the ultimate solution to line of sight issues in tracking technology.

3.2 Impact on Application Scenarios

Although we are aware that our test application is only a first small step towards an implementation, we also see many possibilities in precisely the approach taken due to its simplicity: We clearly see uses, for example, to provide panning trajectories for laptop performers. Also, as the trajectories can be stored as azimuth angles (over time) to a known origin, they can be reproduced with any arbitrary layout of loudspeakers, using VBAP, and also in 3D. To

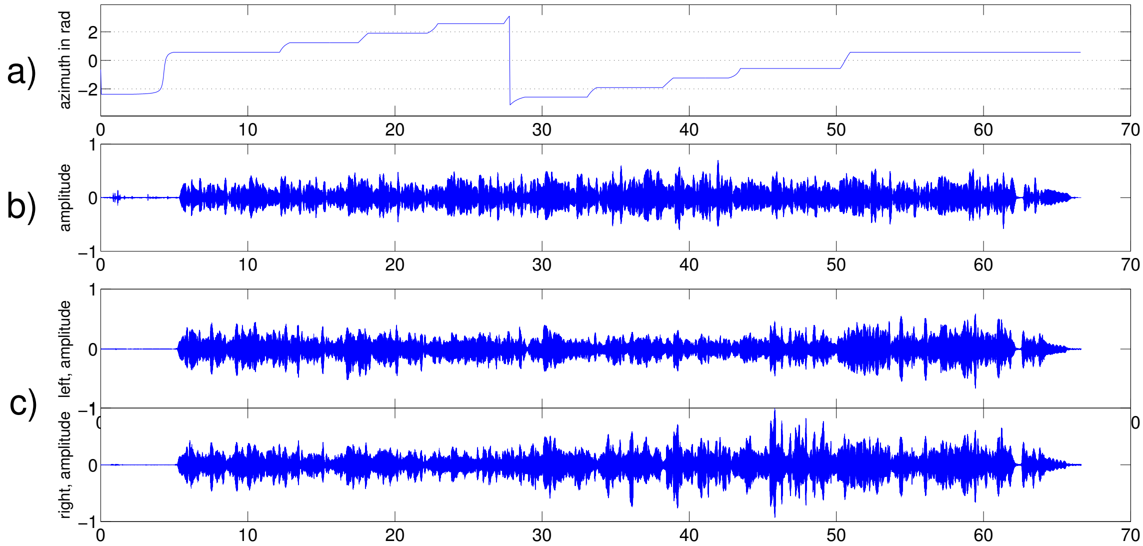


Figure 4: a) Azimuth of reproduced trajectory b) close-miked moving sound source c) Stereo recording over time in seconds.

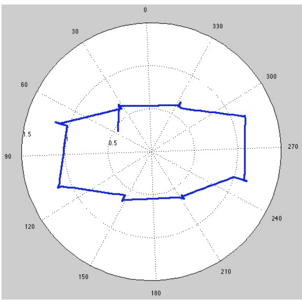


Figure 5: Azimuth of trajectory, polar, in degrees, radius in metres

Figure 6: Leluhelikvratetti, Music Centre Helsinki Nov 2015 [34]

compare the layout with a trajectory, see Figures 3 and 5.

An implementation for the use in the *common* scenario clearly needs more work, line of sight issues, as well as the test signal's impact on the content signal makes SMML a questionable choice: If we use the content signal as the measurement signal here, we will end up in an interesting situation: As we want the test signal to go to all loudspeakers, the content signal however panned to one location, in case of VBAP we would be hearing at most 3 but possibly 2 or simply one loudspeaker, the process would hence decide on one loudspeaker and not localise any further until we send the content signal again to all loudspeakers. Although we see potential uses for self calibration of sound sources to a fixed or temporary nearest loudspeaker, for panning of a moving sound source we see limitations. To use SMML using a distinct test signal is therefore the better option, as long as it does not impact on the content signal.

We believe, for many applications but particularly for our first scenario, a MM approach to be fruitful. A major advantage of the MM approach is also the fact that the content does not need to be delayed by any buffering in the processing, the content signal does not need to be known *a priori*.

In view of using VBAP for these spatialisations, it is noteworthy that the positions obtained by AL contain in fact more information than is needed: VBAP reproduces, even in its 3D implementation, a projection onto a sphere of a virtual sound source, expressing an angle. our trilaterated positions also contain depth information which could possibly be encoded in our reproduction, one way or other.

3.3 Leluhelikvartetti: A Pilot Project

A very similar script is behind *Leluhelikvartetti, hommage à Stockhasuen* (Appendix B) wherein the *Free Improvisation String Quartet's* close-miked playing is spatialised via four toy-helicopters carrying wireless microphones, flying around 8 near coincident radially outwards facing loudspeakers. (See Figure 6) As the whirring sound of the helicopters is part of the sonic texture of the piece, the test signal is masked and can be used without negative impact. The radial arrangement of the loudspeakers also mean that there is practically no conflicting information as to which loudspeaker is the closest one.

4. CONCLUSIONS, FUTURE WORK

Our results show conclusively that an approach to spatialisation using a SMML approach is applicable in practice, even when only distance estimates, and no trilaterated positions are available. The latency incurred by the smoothing algorithms were not detrimental to the listener experience. For a future implementation, based on the application scenarios discussed, lower latencies should be achievable with pulsed signal- and or multiple bandwidth approach for the SMML approach. We have not explored multiple microphone approaches yet, nor the possibilities the distance readings can provide for depth simulation particularly in the approach we used for the proof of concept. The distance measurements could be interpreted as attenuation over distance for example, or considerations as to how distance could be used as a parameter for reverberation or even simulation of first reflections could be of interest.

The Matlab Playrec scripts, form part of this publication (Appendix C), and other researchers are encouraged to further develop or rewrite the scripts provided: Matlab is being used for its convenience as a prototyping platform only, and implementation in a more musician friendly platform will find a community which has been waiting for this for some time.

5. ACKNOWLEDGMENTS

Our research is supported by a Kone Foundation Researchers Grant. We would like to thank the participants of the workshop on MS&I who's active collaboration provided the data for this research.

6. REFERENCES

[1] Kristian Nymoen, Ståle A. Skogstad, and Alexander R. Jensenius. SoundSaber - A Motion Capture Instrument. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 312–315, 2011.

[2] Andreas Bergsland and Robert Wechsler. Composing interactive dance pieces for the motioncomposer, a device for persons with disabilities. In Edgar Berdahl and Jesse Allison, editors, *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 20–23, Baton Rouge, Louisiana, USA, May 31 – June 3 2015. Louisiana State University.

[3] Mark A. Bokowiec. V'oct (ritual): An interactive vocal work for bodycoder system and 8 Channel spatialization. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 40–43, Oslo, Norway, 2011.

[4] Yang Kyu Lim and Woon Seung Yeo. Smartphone-based music conducting. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 573–576, London, United Kingdom, 2014. Goldsmiths, University of London.

[5] Eric Lee, Ingo Grüll, Henning Keil, and Jan Borchers. conga: A Framework for Adaptive Conducting Gesture Analysis. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 260–265, 2006.

[6] Wayne Siegel. Two hands (not clapping). `http://waynesiegel.dk/?page_id=24`, 2010.

[7] Rob van Rijswijk, Jeroen Strijbos, and Niels Bogaards. Walk With Me: A Work Of Pointless Curiosity. In *Proceedings of the International Computer Music Conference 2012*, pages 253–256, 2012.

[8] Lijuan Peng and David Gerhard. A Wii-Based Gestural Interface for Computer Conducting Systems. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 155–156, 2009.

[9] Sertan Şentürk, Sang Won Lee, Avinash Sastry, Anosh Daruwalla, and Gil Weinberg. Crossole: A gestural interface for composition, improvisation and performance using kinect. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, Ann Arbor, Michigan, 2012. University of Michigan.

[10] Thomas Mitchell and Imogen Heap. SoundGrasp : A Gestural Interface for the Performance of Live Music. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 465–468, 2011.

[11] Elena Jessop. The vocal augmentation and manipulation prosthesis (vamp): A conducting-based gestural controller for vocal performance. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 256–259, Pittsburgh, PA, United States, 2009.

[12] LocusSonus. Soundmap. `http://locusonus.org/soundmap/051/`, 2016.

[13] Barry Blesser and Linda-Ruth Salter. *Spaces speak, are you listening? : experiencing aural architecture.* Cambridge, Mass. MIT Press, 2007.

[14] R. Zvonar. A history of spatial music. `http://cec.concordia.ca/econtact/Multichannel/spatial_music.html`, 2005.

[15] B. Nicolescu. *Manifesto of Transdisciplinarity.* SUNY Series in Psychoanalysis and Culture. State University of New York Press, 2002.

[16] Anne Balsamo. *Designing Culture: The Technological Imagination at Work.* Duke University Press, Durham, NC, USA, 2011.

[17] C. Noland. *Agency and Embodiment: Performing Gestures/Producing Culture.* Harvard University Press, 2009.

[18] Lev. Manovich. *The language of new media.* MIT Press, Cambridge, Mass., 2002.

[19] Paul Dourish and Genevieve Bell. *Divining a Digital Future: Mess and Mythology in Ubiquitous Computing.* MIT Press, Cambridge, MA, May 2011.

[20] P. Manning. *Electronic and Computer Music.* Oxford University Press, New York, revised and expanded edition, 2004.

[21] Timo Pyhälä. Ms&i at aalto lume, 2015.

[22] Senn Christopher. Ms&i in blackbox, 2015, 2015.

[23] James Andean. Research group in interdisciplinary improvisation: Goals, perspectives, and practice. `http://www.academia.edu/4158856/Research_Group_in_Interdisciplinary_Improvisation_Goals_Perspectives_and_Practice`, 2014.

[24] Toni Robertson and Jesper Simonsen. *Routledge international handbook of participatory design / edited by Jesper Simonsen and Toni Robertson.* Routledge New York, 2013.

[25] Dom Schlienger. Workshop on Music, Space & Interaction participants' blog. `http://creativemusictechnology.org/MS&P.html`.

[26] Dominik Schlienger and Sakari Tervo. Acoustic localisation as an alternative to positioning principles in applications presented at nime 2001-2013. In Baptiste Caramiaux, Koray Tahiroglu, Rebecca Fiebrink, and Atau Tanaka, editors, *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 439–442. Goldsmiths, University of London, June 30 – July 03 2014.

[27] D. Schlienger. Acoustic localisation techniques for interactive and locative audio applications. In *Locus Sonus Symposium #8 on Audio MobilitÃl', Proceedings*, 2014.

[28] I. Rishabh, D. Kimber, and J. Adcock. Indoor localization using controlled ambient sounds. In *Indoor Positioning and Indoor Navigation (IPIN), 2012 International Conference on*, Nov 2012.

[29] V. Filonenko, C. Cullen, and J. Carswell. Investigating ultrasonic positioning on mobile phones. In *Indoor Positioning and Indoor Navigation (IPIN), 2010 International Conference on*, Sept 2010.

[30] T. Janson, C. Schindelhauer, and J. Wendeberg. Self-localization application for iphone using only ambient sound signals. In *Indoor Positioning and Indoor Navigation (IPIN), 2010 International Conference on*, Sept 2010.

[31] Jeong Seob Lee and Woon Seung Yeo. Sonicstrument: A musical interface with stereotypical acoustic transducers. In *NIME*, pages 24–27. nime.org, 2011.

[32] Ville Pulkki. Virtual sound source positioning using vector base amplitude panning. *Journal Audio Eng. Soc*, 45(6):456–466, 1997.

[33] R. Humphrey. Playrec. `http://www.playrec.co.uk/index.php`, 2011.

[34] Wilma Hurskainen. Toy helicopter quartet, 2015.

APPENDIX

A. AUDIO & VIDEO FILES

`http://creativemusictechnology.org/alps.html#appendixA`
`http://creativemusictechnology.org/alps.html#video`

B. LELUHELIKVARTETTI

`http://creativemusictechnology.org/alps.html#appendixB`

C. PLAYREC SCRIPT

`http://creativemusictechnology.org/alps.html#appendixC`

8 Gestural Control for Musical Interaction using Acoustic Localisation Techniques IV

Schlienger, D. 2016b. “Gestural Control for Musical Interaction using Acoustic Localisation Techniques.” In International Conference on Live Interfaces 2016, Proceedings, pp. 161–167.

Gestural Control for Musical Interaction using Acoustic Localisation Techniques

Dominik Schlienger

University of the Arts Helsinki Sibelius Academy
Centre for Music & Technology, Finland
dominik.schlienger@uniarts.fi

Abstract. Acoustic Localisation principles for tracking technology are well researched and have many applications in medicine and industry. In creative technologies optical technologies are more prominent. For creative music technologies we know from our own research into the applicability of acoustic localisation techniques for audio in the frequency range of standard audio equipment, that acoustic localisation techniques are potentially a straightforward choice. We also know that the technology is scalable from tracking in large performance areas to smaller areas with lower latency, required for gestural tracking for real-time interaction. As a proof of concept we prototyped two implementation of the principles, using handheld microphones and standard, commercially available loudspeakers, firstly for a Theremin - like pitch control interface and secondly, a spatial trigger for percussive sounds.

Keywords: Acoustic Localisation Techniques, Theremin, Gestural Control, Interface for Musical Expression.

Introduction

Acoustic Localisation (AL) principles for tracking technology, particularly techniques using ultrasound, are well researched and have many applications in medicine and industry (Holm 2012). The principles for AL are not limited to ultrasound however, they apply to audible sound as well. This makes the use of them for tracking technology particularly interesting, namely for applications where audio equipment is already present, like in many audio, audiovisual and multimedia applications and particularly for interactive arts and new interfaces for musical expression (Schlienger and Tervo 2014). In creative technologies optical technologies are more prominent (Schlienger 2014), but there are a number of applications using AL (Rishabh, Kimber, and Adcock 2012; Filonenko, Cullen, and Carswell 2010; Janson, Schindelhauer, and Wendeberg 2010; Seob Lee and Yeo 2011).

We also know that implementation of AL principles is scalable. Tracking of performers in traditional performance areas like stages in concert halls are as feasible as tracking of small gestural movements which require very low latency and high update rates, for example for real-time interaction with virtual instruments.

For low latency applications, the authors of (Gupta et al. 2012), describe an implementation which arguably provides similar functionality to what we suggest. However our Time Difference of Arrival (TDoA) approach in lieu of Doppler, brings advantages in applications where the tracked object's identity needs to be known. For our test implementation of a Theremin like instrument, the differentiation between the left and right hand, for example, could be achieved by using separate microphones. In the Doppler scenario described in (Gupta et al. 2012) this differentiation would have to be achieved by other, more complex means.

Further, from our work with the Workshop on Music, Space & Interaction (MS&I) (Schlienger 2016a), we know about the need for simple, ubiquitous and pervasive interfaces which do not hem in the flow of gestural explorations of space. What is more, our findings from the workshop suggest that interfacing technology in form of sounding objects provides a particular engaging tool to immerse performers as well as audiences in a spatially interactive performance. We thus describe two applications here, using AL, which answer these requirements as a proof of concept for the feasibility of the principles in low latency applications needed for interactive sounding objects.[1]

Both applications work on the basis of moving a handheld microphone but arguably the microphone could be attached to any type of moving object.

To summarise our rationale, we argue for a broader implementation of AL using mainstream audio equipment as the principle provides competitive alternatives to many other solutions as we showed using the literature in our previous work (Schlienger and Tervo 2014). Namely, by using sound in the frequency range supported by standard, main street, audio equipment, we see a possibility to improve on the performance of, for example, optical tracking systems due to the diffraction of sound around objects, allowing for tracking in non-line of sight situations. Whereas the focus of our previous work describes novel applications for larger performance areas and stages, our contribution here is to show that the AL principle can also be applied in smaller, gestural applications, for which other technologies are usually implemented: To stay with our examples, capacitance for the Theremin and optical tracking for gestural triggers. Our point is, the AL principle, using ubiquitous technology, has broader and more pervasive potential than its current state of rarity portrays.

The following sections describe our work in MS&I in more detail, give more details on AL principles followed by a description of our prototype implementations. The Matlab scripts using the Playrec Utility library are available to download and form part of this publication as an appendix (Schlienger 2016b).

Method

Workshop on Music, Space & Interaction

The workshop on MS&I runs now in its third year at the University of the Arts Helsinki, and uses Participatory Design approaches (Robertson and Simonsen 2013) to explore the affordances of technology for spatial interactivity in interactive musical applications. We use interdisciplinary, free improvisation as a method (Andean 2014) to explore existing and possibly new technologies without the restrictions of habits, genres or conventions. This approach to technology design draws on the notion of Designing Culture (Balsamo 2011) and on the notion of mess in ubiquitous computing (Dourish and Bell 2011).

The data gained from the workshop is ethnographic in nature, it consists of notes and participants written contributions along with some audio and video documentation. Some insight can be gained from the workshops blog the participants are encouraged to contribute to (Schlienger 2016a).

The applications we describe in this paper are based on our early findings from the workshop which provided us with the idea of spatially interactive sounding objects: These are things in a space which might be actual physical objects but also virtual objects which can only be heard in a particular position, rather than seen. On this notion we developed the idea of a spatially controlled pitch - object, probably most descriptively described as a Theremin - like instrument and a percussive - object, whereby at distinct positions percussive sounds can be triggered.

[1] We describe a proof of concept for larger scale applications in an article accepted for presentation at the NIME2016 conference

Figure 1. Workshop on MS&I, (Courtesy of Timo Pyhälä, 2015)

Acoustic Localisation

As we are only interested in AL in respect to sound which can be produced by standard audio equipment, we refer to AL in the rest of the text in reference to the frequency range form roughly 20 Hz - 30 kHz. Further, we are interested in the Time Difference of Arrival (TDoA) technique specifically, as this is the approach taken for the implementation we are working on for tracking in larger spaces. Compared to other AL approach, Doppler, for example, TDoA techniques lend themselves better to applications where the identity of the tracked object needs to be known, as the position estimation happens actively for a sender or a receiver's own position, Doppler or also echo-location relies on an estimation based on an indirect measurement. The principle of TDoA measures the time difference between the sending of a signal and its recording at one (or several) receiving microphones. From the correlation of the recorded and the original signal the time delay can be calculated directly. As the speed of sound through air is known *a priori* the distance of the receiving microphone from the sound source (the loudspeaker in this case) can thus be derived from the time difference. Using several such measurements, a position estimate can be trilaterated. The technique thus works in principle for 3 dimensional localisation.

For the prototypes discussed here we use the most simplest of principles, namely one single distance reading. *Nota bene*, this limitation to a single dimension is not a limitation of the AL principle! Trilateration from 4 distance measurements can be used to estimate an absolute 3D position. However, for the applications we discuss here, simple distance measurements between one sender and one receiver are, indeed, sufficient. We would like to stress that although a single distance reading is not enough for a 3 dimensional, absolute *position*, the application can still be spatial in character, as the distance reading is available *radially* from a fixed point of the receiver or sender which allows for 3 dimensional interaction with the object.

Experimental set up

To demonstrate the applicability of AL for applications which are latency sensitive we prototyped two musical applications. The first application is a Theremin like instrument wherein pitch can be gesturally controlled. For the second application, we implemented a gestural trigger mechanism for percussive sounds. Both implementations require to move a microphone in front of a loudspeaker emitting a high pitched measuring signal just above the frequencies audible to the human ear and within the frequency range of standard audio equipment. (Depending on the type of loudspeaker, 17-30 kHz.) The loudspeaker used for our tests was of type Alesis M1 MK II, the microphone of type Sure SM58.

The applications were implemented in Matlab R 2013a, using the Playrec (Humphrey 2011) utility. Playrec allows for non-blocking soundcard access and thus continuous and simultaneous play and record. This could have been achieved natively in newer versions of Matlab, but Playrec provides insight into the sample-by-sample workings which was considered helpful for prototyping. A simple Max Patch received the measurements from Matlab via udp and dealt with the content audio.

Patches and Playrec Script are available online (Schlienger 2016b). The processor running Matlab was an 11-inch, Mid 2011 MacBook Air, 1.6 GHz Intel Core i5, running OSX 10.8.5, the soundcard a DigiRack 002.[2]

The room in which we tested the applications is a typical living room without particular acoustic treatment, with a reverberation time below 0.4 seconds, 6 by 3.5 meters with 2.6 height.

Both applications are latency sensitive: For musical interaction, in order to play within an ensemble as well as to be able to play an overdub for a multitrack recording it is crucial that the performer can monitor her or his playing in real-time or in a very close approximation to real-time. To define the criteria of what shall be considered a close enough approximation, the following thoughts were decisive:

- Latency up to a length of 10 ms is generally tolerated by musicians in performance situations as well as in the recording studio.[3]
- Just over 10 ms at 48 kHz sampling rate can be achieved with a buffer size of 512 samples. This also happens to be the lower limit at which our current set-up runs stably.
- The actual, overall latency between is somewhat higher: We can calculate this by measuring the time it takes a signal to arrive at a microphone at a millimetre distance from a loudspeaker. Possibly due to hardware restrictions and processing power of our set-up, our best achievable latency is often as high as four times the buffer size.
- For many mainstream sound cards 20 ms were considered acceptable until fairly recently.
- We further estimate typical gestures for these application to stem from arm movements of a stationary person, so we scaled the functionality of both applications to a range of 1 m.
- The trajectories through air which can be expressed in 512 samples at 48 kHz represent a distance of 3.66 metres. For gestural interfaces we consider this adequate.

This last point might need some more elaboration: The length of the buffer we iterate when calculating the time delay between two signals sets the limit of the longest time delay we can actually measure. We cannot measure time delays which are longer than the window that the buffer provides. As we translate the time delays to distance covered by sound, the window sizes also stand for maximum distances that can be estimated within a window. These relations between window sizes and time delays and distances covered by sound, respectively, are further dependent on sampling rate. So with an increased sampling rate we need to higher the window size to cover the same distances. The higher sampling rate does not only provide a higher update rate, but also allows for measurement signals at higher frequencies.

	44.1 kHZ	**48 kHz**	**96 kHz**	**192kHz**
Measurement signal top frequency (Nyquist)	22.05 kHz	24 kHz	48 kHz	96 kHz
1 sample	0.0244 ms	0.0208 ms	0.0104 ms	0.0052 ms
1024 sample window	23.2199 ms	21.3333 ms	10.6666 ms	5.3333 ms
2048 sample window	46.4399 ms	42,6666 ms	21.3333 ms	10.6666 ms
Maximum distance represented by window 1024	7.9644 m	7.3173 m	3.6586 m	1.8293 m
Maximum distance represented by window 2048	15.9288 m	14.6346 m	7.3173 m	3.6586 m
Distance in one sample	0.0077777 m	0.0071458 m	0.0035729 m	0.0017864 m
Samples to travel 1 m	128.5714	139.9416	279.8833	599.7667

speed of sound = 343 m/s
time for sound to travel 1m = 2.9154 ms

Table 1: Overview of Relations between Sample Rates and Distances Covered by Sound

[2] For the video demonstration we used Meyer Sound MM-4XP, the Motu 16A soundcard and AKG c417 Omni Lavalier Microphones, and the same processor.

[3] We differentiate here between what is tolerable and what is noticeable: even latency as short as 1-3 ms can be noticed, for example as comb filtering by singers monitoring themselves on headphones.

Proof of Concept for Low Latency Applications

We chose the two applications as proof of concept for low latency applications for the reason that they show conceptual differences in slightly different requirements. Thus, we look at the applications separately:

Pitch Control

Gestural pitch control with a Theremin is a skill which needs to be acquired through practice. Similarly, our take on the concept is not meant to simplify musical playing: With a stable and controlled hand it should allow for stable, controlled pitches and the (skilled) playing of melodic material.

The pitch range covered within 1 m we set too 80 - 6000 Hz, as we believe this to be roughly representative of the frequency range of interest for most musical applications. We experimented primarily with sine waves, but this was an arbitrary decision, and other sound material could, of course, be used instead.

Percussive Control

We generally associate percussion playing with hitting an object with our hand, or with a stick, or similar. Our implementation of gestural percussive control is thus somehow quite abstract, as moving the microphone through a particular distance to a loudspeaker triggers a sound. There is a commercial implementation doing a very similar thing though, Aerodrums (Aerodrums 2016), using optical tracking. Similarly to our scenario, the lower latency limit is set by the system buffer size, however, their implementation allows for settings as low as 128 samples per buffer, albeit at 44.1 kHz sampling rate.

Results & Discussion

We achieved latencies for both application scenarios we considered sufficiently low for a proof of concept. (Typically around 10 - 20 ms) We are also confident that these numbers can still be improved on with further development.

For the pitch control application, we found that our current, averaging, smoothing algorithm[4], makes it difficult to know if the gestural position responds to the proper pitch location or if the algorithm is averaging out under the influence of a few wrong readings. For pitch sensitive application averaging filters don’t seem to be the best choice.

One peculiarity we can report for the percussive control application: Despite the latency being quite high at 512 or more samples per second (48kHz sampling rate) we didn’t really notice this at first, as we seemed to just make a mental note where it was that the sound triggered when we tried it out the first time. As we knew that there was some latency we then investigated and came to the following insight: As it happens, the fact that in these air-drum type applications the performer is not actually hitting a physical object, the latency, (as long as constant and not varying) has the effect of moving the virtual object further away from the place where the performer thinks she or he will hit it. In this sense it is actually easier to live with the latency of such a virtual sound object than it would be to live with the latency of a physical object triggering a virtual sound source. The lack of haptic response means that, for our scenario, the virtual object is just further away than the performer would initially presume if we knew at what distance measurement the sound is being triggered.

The need for a mobile device to be held by the performer, even as a clip on wireless microphone, remains an obstacle towards totally transparent interfacing. Yet, compared to Doppler techniques which would allow device - free tracking of gestural movements, we see great advantages in the mobile device approach as the identity of the tracked device is known to the system.

[4] The algorithm calculates averages over time, the smoothing creates a lag and the interpolation of positions result in slurring of the pitch

A further caveat applies to both Doppler as described by (Gupta et al. 2012) and our (current) TDoA approach: With measurement signals in the frequency range between 18 to 30 kHz we lose the advantage of acoustic tracking over optical tracking to a certain extent: The much praised advantage of AL in view of the requirement of line of sight between tracked object and a camera in optical tracking systems, is much reduced at high frequencies, as the corresponding wavelength do not diffract around obstacles but reflect if the obstacle is wider than the wavelength in question. AL still works in the dark, certainly an advantage over some optical tracking systems. We are thus very interested into further research about the possibilities of using the audible sound of the content in an audio application as a measurement signal, or certain frequency bands within the content audio.

The resulting proof of concept - applications are being demonstrated in the appended video clips available online via the following link: http://creativemusictechnology.org/lowlatencyapps.html

Conclusions and Future Work

We showed with a simple implementation that AL techniques are feasible for pitch control and percussive triggering of sounds in musical applications. We also showed that more research is necessary, and that the current implementation can not be considered the state of the art of what is possible: With a more advanced implementation, a direct comparison with commercial systems on the market using optical tracking, for example, will provide an actual evaluation, which we can not sensibly provide yet with the current, rudimentary, example code. We also point out that these limitations are not due the principles of AL but due to the basic nature of the prototyped implementation. To summarise, we think we have a couple of interesting virtual sounding objects whose affordances we will explore much further in the workshop on MS&I. we would also like to invite interested parties to look at the code appended to this paper, available online (Schlienger 2016b). The whole project is intended to be open source, and an implementation as a Pd and/or Max MSP object is planned.

Acknowledgements. We would like to thank all the participants past present and future of the workshop on Music, Space & Interaction. This research is made possible by a Kone Foundation Researchers Grant.

References

Aerodrums. 2016. Aerodrums: The best Drumset You've Never Seen. http://aerodrums.com/aerodrums/.

Andean, James. 2014. Research Group in Interdisciplinary Improvisation: Goals, Perspectives, and Practice. http://www.academia.edu/4158856/Research_Group_in_Interdisciplinary_Improvisation_Goals_ Perspectives_and_Practice.

Balsamo, Anne. 2011. *Designing Culture: The Technological Imagination at Work*. Durham, NC, USA: Duke University Press.

Dourish, Paul, and Genevieve Bell. 2011. *Divining a Digital Future: Mess and Mythology in Ubiquitous Computing*. Cambridge, MA: MIT Press.

Filonenko, V., C. Cullen, and J. Carswell. 2010. "Investigating ultrasonic positioning on mobile phones*." International Conference on Indoor Positioning and Indoor Navigation (IPIN).*

Gupta, Sidhant, Daniel Morris, Shwetak Patel, and Desney Tan. 2012. "SoundWave: Using the Doppler Effect to Sense Gestures." *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '12*. New York, NY, USA: ACM, 1911–1914.

Holm, S. 2012, Nov. "Ultrasound positioning based on time-of-flight and signal strength." *International Conference on Indoor Positioning and Indoor Navigation (IPIN).* 1–6.

Humphrey, R. 2011. Playrec. http://www.playrec.co.uk/index.php.

Janson, T., C. Schindelhauer, and J. Wendeberg. 2010, Sept. "Self-localization application for iPhone using only ambient sound signals." *International Conference on Indoor Positioning and Indoor Navigation (IPIN)*.

Rishabh, I., D. Kimber, and J. Adcock. 2012. "Indoor localization using controlled ambient sounds." *International Conference on Indoor Positioning and Indoor Navigation (IPIN)*.

Robertson, Toni, and Jesper Simonsen. 2013. *Routledge international handbook of participatory design.* Edited by Jesper Simonsen and Toni Robertson. Routledge New York.

Schlienger, Dominik. 2016a. Workshop on Music, Space & Interaction Participants' Blog. http://creativemusictechnology.org/MS&P.html.

Schlienger, Dominik. 2014. "Acoustic Localisation Techniques for Interactive and Locative Audio Applications." *Locus Sonus Symposium #8 on Audio Mobilité*, Proceedings.

Schlienger, Dominik. 2016b. Gestural Control for Musical Interaction using Acoustic Localisation Techniques (ICLI2016). http://creativemusictechnology.org/lowlatencyapps.html.

Schlienger, Dominik, and Sakari Tervo. 2014, June 30 – July 03. "Acoustic Localisation as an Alternative to Positioning Principles in Applications presented at NIME 2001-2013." Edited by Baptiste Caramiaux, Koray Tahiroglu, Rebecca Fiebrink, and Atau Tanaka, *Proceedings of the International Conference on New Interfaces for Musical Expression.* Goldsmiths, University of London, 439–442.

Seob Lee, Jeong, and Woon Seung Yeo. 2011. "Sonicstrument: A Musical Interface with Stereotypical Acoustic Transducers." *NIME*. nime.org, 24–27.

9 Requirements on Kinaesthetic Interfaces for Spatially Interactive Sonic Art (V)

Schlienger, D. 2016c. “Requirements on Kinaesthetic Interfaces for Spatially Interactive Sonic Art.” In Proceedings of the Audio Mostly 2016, AM ’16, pp. 162–169.

Requirements on Kinaesthetic Interfaces for Spatially Interactive Sonic Art

Dominik Schlienger
University of the Arts Helsinki
Sibelius Academy
Centre for Music & Technology
dominik.schlienger@uniarts.fi

ABSTRACT

This paper documents the requirements on tracking technology for spatially interactive sonic arts. We do this by comparing our theorised notion of an ideal kinaesthetic interface to, firstly, the current results of an ongoing online survey and, secondly, the results of our ongoing Workshop on Music, Space & Interaction (MS&I). In MS&I we research the affordances of existing and hypothetical technology to enhance and facilitate spatial interactivity. We give both qualitative and quantitative recommendations for design. While underlining the specific requirements for sonic art in respect to its aural nature, we discuss how and why the requirements elicited from our research can be applied to spatial interactivity in general.

CCS Concepts

•**Applied computing** → **Performing arts;** •**Human-centered computing** → **Ubiquitous and mobile computing design and evaluation methods;** *Virtual reality;*

Keywords

Kinaesthetic Interface, Interfaces for Musical Expression, Human Computer Interaction, Embodied Digitality

1. INTRODUCTION

Despite the regular occurrence of tracking technology in works of sonic arts [39, 12], multimedia [19] and a keen interest in its development in gaming [21] and virtual reality [8, 13], no specific tracking technology has established itself to date in an ubiquitous way like GPS has done for outdoor navigation, for example. Existing technologies are adapted and appropriated from motion capture [10], mechanical industries [23] and from the mainstream gaming technology with wii [26] and Kinect [36]. Of course smart phones, as location aware devices, are generally believed to provide ubiquitous tracking technology which could be used for interactive art. There are examples for such applications [18, 2], but they have in common that actual tracking abilities are quite constrained.

Nota bene, there is practically no academic work which suggests that the tracking or positioning technology implemented in an interactive artistic application was not a success. On the contrary, most authors of papers describing artistic applications using positioning or tracking technology assert that their use of technology was successful. However, this might be more to do with the fact that in many artistic practices it is common to work with the limitations of a technology rather than let the technology's shortcomings be the end of the artwork. Artworks are rarely akin to laboratory experiments where a technology's performance can be tested as to success or failure to provide conclusive answers in this vein. What is more, technological failure is often incorporated into the aesthetics of an artwork, an can even produce whole genres, as in the case of *glitch* [28]. But, the successful use of a technology is rarely the condition *sine qua non* of an artwork.

Here we also have to consider that the evaluation of technologies in an *artistic* context is more often than not a subset of technology evaluation in a *psychophysical* context: All arts rely on sensory experiences evoked by some technique or technology, the evaluation of which demands some very refined methods [20, 15], even without accounting for artistic intentions and aesthetic considerations [6].

Hence, instead of evaluating technology as to its applicability *post*-design, we believe it to be crucial - here even more than for many other design implications, to gain insights into the requirements on the technology as part of the design process, or *pre*-design. The apparent triviality of the previous sentence does not make its premise less urgent, as for artistic applications only appropriated technology is available; post design evaluation the only possible path.

The difficulty in acquiring requirements on technologies for artistic purposes is the nigh impossibility to predict any of them. In respect of spatial interactivity in sonic arts, however, with our workshop on MS&I, (occasionally referred to as just *workshop* in the rest of the text), we hope to have access to an experimental practice which is general enough to provide findings which are applicable to more than just isolated artworks.

We thus imply the existence of common denominators for spatially interactive artistic practices in form of gestural movements, trajectories, and changes of positions of objects and participants; in short, moving in a meaningful, expressive way. In this description of the notion of kinaesthesia, we draw on Carrie Noland's work Agency and Embodiment

Audio Mostly 2016 October, 4 - 6, Norrköping Sweeden

ISBN 123-4567-24-567/08/06...$15.00
DOI: 10.475/123_4

[24], wherein she asserts that kinaesthetic experience, - *feeling the body move*, as a corporeal performance of gestures, allows for experimentation, modification, and, crucially, rejection of gestural routines. From this she infers a non-constructivist account of agency: A gestural movement is, in this sense, embodied agency.

With a numerical recording of this movement, we thus yield information which is highly correlated to the actors intentions: Such a kinaesthetic interface[1] records the embodied, or individualised, kinaesthetic quality of the gesture by computing the aforementioned trajectories and changes of positions, namely, the whole physical movement the gesture consists of in space. This movement might be reducible to key points like in motion capture, based on the known, limited, affordances of joints and members, or to relative positions of objects to each other. but it should be able to describe one embodied movement as distinct from another even if it belongs to the same gestural routine. This can, so we claim, only be possible if the interface does not introduce another whole set of modifications to the gestural routine: The ideal (kinaesthetic) interface should not need to be evaluated directly, as it is not experienced! We elaborate on this notion of the invisible interface in more detail elsewhere[2], but suffice to say, Researchers at Xerox PARC argued for invisible interfaces as early as 1994 [42].

Dourish and Bell [11] however, extend this notion of the invisible, or ubiquitous, to technology which we don't notice anymore, as we are used to it. This is an understanding also mirrored in Bruno Latour's description of the *technical* in[17]. Hence it could also be argued that any old technology will do, as we will eventually adjust our habits to accommodate its shortcomings. We are not able to refute this argument at this stage, but we appeal to common sense; that the design goal for new technology should be for *best possible solution* not *make-do*[3].

From this understanding of technology as something elusive - at least while it works, follows that in order to study the techniques and technologies of a practice we need to study the practice itself. So, in order to design interfaces for spatial interaction, which is the aim of that study, we need to study spatial interaction artistic practice *per se*, and develop technology with the least impact on that practice. This is, essentially, an ethnographic approach and explains why anthropology became an indispensable partner to design in the development of Human Computer Interaction. Yet, for interaction design for the sonic arts, the advice to approach design from situated practice [38, 3] was heeded only by few, for example [14, 41, 25] amongst others.

We made our case here so far under the tacit assumption that implementations of tracking technologies specific for interactive arts are necessary. Indeed, we have not yet come across an implementation which matches the requirements we are about to describe. To clarify our rationale: We do not claim that existing principles are not capable of providing the required functionality with some case specific development. But we believe some principles lend themselves better to the scenarios of our concern than others and for some we see potentially systemic issues.

[1]Many kinaesthetic interfaces already exist and are part of mainstream technology, *e.g.*, touch sensitive (musical) keyboards. Still, we feel it helpful for a general discussion to conceptualise them.

[2]Journal article for submission later in 2016

[3]This doesn't mean that the solution has to be high-tech: in fact, we think the simpler the solution, the more pervasive it will be.

1.1 Tracking Technology: A Short Overview

The following is a summary of our findings in [29] also summarised in [35]. We mainly focus here on what we see as limitations for our intended use and do not account for the many useful applications these technologies have, of course, for other applications.

1.1.1 Radio Frequency: RFID, GPS

Despite the ubiquitous availability of radio frequency based positioning technologies in form of Wireless Local Area Networks, (WLAN) its most common occurrence, Radio Frequency Identification tags, (RFID) has comparatively low update rates. The Global Positioning System only works outdoors, Ultra Wide Band (UWB) technology is very expensive. Systems relying on radio signal strength indication are particularly unreliable, as the design aims behind WLAN technology are not for the provision of stable signal strength but for highest strength possible, which is a dynamic principle conflicting with the requirements for a measurement signal.

1.1.2 Inertia - methods

Inertia measurement methods, like accelerometers and gyroscopes are ubiquitously present in almost every smart phone. However, they don't provide an absolute position, so errors cumulate if used without repeated references to known positions. For smart phone tracking, inertial methods are usually combined in a hybrid way with RF methods via WLAN.

1.1.3 Acoustic Localisation

Acoustic localisation in form of ultrasound requires expensive, specialised equipment. In form of Doppler, it relies on line of sight between signal emitter and tracked object. Systems using signals inside the audible frequency range rely on the presence of microphones and loudspeakers. We believe that this is not a hindrance for applications in sonic arts where audio technology is often present anyway. Further, we see potential for non-line of sight applications due to the ability of sound to diffract around objects. However, this technology does not work device free, the object to be tracked has to be either a microphone or a sound source. For transparency's sake we state here that in other and future work we are actively researching acoustic localisation in the audible frequency range as a solution for sound art applications [29, 35, 30, 32, 33].

1.1.4 Optical Tracking

Optical systems, the most commonly applied tracking technology in spatially interactive arts, have the use of cameras in common, be it systems of type Kinect [43], Wii [44] or more elaborate systems for motion capture like Vicon [40]. Despite the use of multiple cameras to add depth information, cameras, intrinsically, are 2 dimensional sensors: When tracking movement in 3 dimensional space, the resolution decreases with increasing distance: a moving object close to the camera causes more change in pixels than an object far away [30]. The applicability of cameras for tracking purposes is also limited by the necessity of line of sight between a tracked object and the camera.

2. METHODS

We described our theorised notion of the Kinaesthetic Interface in the Introduction and will pick up on it again in the discussion, where we use it as a framework to validate the data we acquired from an online survey and our workshop on MS&I. Before that we would like to clarify our terminology in respect to some key concepts, followed by a more detailed description of the two contrasting types of methods we used, one an online questionnaire, the other ethnographic interpretative field notes of an interdisciplinary improvisational practice.

2.1 Definitions

2.1.1 Interaction

In our enquiry into spatial interaction we use interaction generally to mean *human - human* interaction. In fact, it is our believe that *human computer interaction* is also essentially *human - human* interaction. To make this less anthropocentric, we can generalise interaction as an exchange of information between actors [16, 17]. We explicitly use interaction in this very general understanding, and crucially not merely as human - computer interaction.

2.1.2 Interface

As opposed to the very broad, notional way we use the term interaction, we are referring to a very concise idea of interface in the context of this article[4]: We refer to interfaces as sensors which allow the digitalisation - and therefore processing - of measurable and thus quantifiable phenomenons. It is a bit tricky not to fall into the tautology here to define as a phenomenon only what is quantifiable. To circumvent an albeit interesting but lengthy discussion on phenomenology we would like to acknowledge that, via an interface, we record a reference to a phenomenon, not the phenomenon itself, *reference* used here in a Latourian sense [17]. With an interface we aim to measure something correlated to what entails the experience of the phenomenon in the hope that what we record, reproduce and process in the digital still correlates to that experience once we re-instate it in the phenomenal, real, or analogue world.

2.1.3 Sonic Arts

In sonic arts we would like to include all arts which are sonic in any way. They include music, sound art, multi media art. But our theorising also applies to sonic aspects of theatre, video art, audio visual displays, short, wherever audio material is used in an aesthetic way.

2.2 Notes on Survey

The online survey on User Requirements On Positioning Systems for Audio Applications was started in January 2012 and informed our previous research into existing positioning technologies . The survey has remained open and has currently had 40 respondents [31]. The design of the survey is fully documented in [29]. Originally designed to represent stakeholders in all areas where positioning technology might be used for audio applications, a sample number of 66 was considered to be representative based on industry numbers. For the comparatively small field of professionals in spatially interactive sonic arts, a smaller sample would be applicable, but this also applies to a smaller proportion of respondents, if we presume the total of respondents to be a cross-section of the population. As no absolute numbers are available, we suggest caution as to how valid these results are quantitatively.

In contrast to our previous work, we are less interested here in how respondents evaluated particular technologies but in respondents' requirements for, and expectations on, tracking and positioning technology.

2.3 Account of Practice

In MS&I, we apply Interdisciplinary Improvisation [1] and participatory design principles [37] to develop a prototypical practice from which new techniques and technologies for spatial interaction emerge. We are describing this approach in concise detail in a longer article to be published later in 2016 but we will summarise relevant aspects here.

The idea behind a prototypical practice is to create an experimental practice in which participants encounter a set of problems for which there might be technical solutions. As these problems arise in a performative, situated, improvisational setting, immediate, simple solutions can be found, often from within the group of participants, *ad hoc*. If a solution can not be found, requirements can be formulated based on the participants experiences. Essentially, we elicit user requirement as a participatory activity.

To gather data in MS&I we mostly used a field note approach described for ethnomusicology in [5]. This is an explicitly interpretative approach which we apply in a highly situated manner, as we write the field notes in conversation with the other participants in the discussion rounds which follow every session.

The research questions for MS&I are directly concerned with technical solutions to spatial interactivity:

- How do we interact musically with space in improvisation?
- What do we want from technology to increase spatial interaction in musical performances?
- What can new technology provide that cannot be provided by old technology?
- How does existing technology impact on spatially interactive practice?

In contrast to other approaches to technology design using mock ups or improvisation for prototyping [9], the workshop does not presuppose any given technology whatsoever but tries to engage with spatial concepts through the physical spatiality of sound and the movement of the human body within space and the sounds which are created by this movement. Over the duration of the workshop's existence a few scores or exercises proved to be particularly helpful and became part of a repertoire, of sorts. These include scores or rules like *Use no instruments or tools other than your body*, or *use the room as a found instrument* or *Start at the centre of the room and disperse while getting quieter*. Besides the field notes many participants take for their own study all are encouraged to write down their thoughts as contributions to the workshop blog [34].

Participants range from musicians, composers and dancers to scenographers, landscape architects, painters, poets, video

[4]For some cultural theorists, e.g., Manovich [22], interfaces can be whole types of media, say films, or books, for others, interfaces refer to keyboard, mouse and monitor only.

artists, lighting designers sculpturists and others. The workshop has been running since 2014 and takes place on 3 - 4 weekends per term. The group size varies, there are usually between 6 - 12 participants.

2.3.1 Interdisciplinary Improvisation

The Research Group on Interdisciplinary Improvisation [1] from whom we borrowed elements for our approach, seeks to find the common ground between the various disciplines represented in the different academies of the University of the Arts Helsinki, namely music, sound art, theatre, painting, drawing, performance art, dance, film and video through free improvisation. The group's understanding of freedom in improvisation is based on the conscious awareness of its relativity between maximum freedom and maximum constraint, and also stretches to the understanding that a performer can experience freedom through constraint. The group is also aware that certain set of rules still tend to apply, however free a practice is, even if it is just the agreement on where and when to improvise.

The aspect of Interdisciplinary Improvisation we believe to be important here, is that a common vocabulary in respect to spatiality across disciplines can be developed, which doesn't rely on an abstract analysis from outside the practice but on an organically grown metalanguage which evolves from within.

3. RESULTS & DISCUSSION

First, we look at some results of the ongoing online survey, which we consequently compare to the findings from the workshop on MS&I. We then compare the combined findings with the notion of the kinaesthetic interface.

3.1 Survey Revisited

As mentioned in the section on method, the survey was originally launched for our previous work in [29], but is revisited as it provides interesting data relevant for our current enquiry. Since our earlier analysis, the number of respondents has increased to 40. The full wording of all survey questions can be found in the survey itself, as it remains open, on http://creativemusictechnology.org/survey.html. Looking at the answers to survey question Q4, *"Into which of the following fields would your audio application fit best?"* we see that despite the original aim to get a cross-section of stakeholders in localisation technology in all areas of spatially interactive audio applications, a great proportion has an interest in sonic arts applications, be this directly, through performance art (14 respondents) spatial music (13), and interactive art (12), or indirectly via recording technology (14). Other fields were gaming (7) multimedia (9) Education (10) assistive technologies (4) and manufacturing (1), home theatre (1) and surveillance (0).

From Q5 *"What scenarios apply to your application?"* we know that a majority (18) expects their applications to happen indoors, and a third also in crowded surroundings (9), and about half expect their application to be in a public setting (17).

From question Q7 *"Presuming that the user is being tracked as part of your application, what sort of technical interaction with the system would you consider acceptable for your application?"* we get direct answers as to what sort of impact of the technology is acceptable to the respondents summarised in the following table 1

Device (e.g., mobile phone)	acceptable	(96%)
Earphones	acceptable	(76%)
Hat or cap	acceptable	(66%)
Calibration before use	unacceptable	(54%)
backpack	unacceptable	(58%)
Backpack, earphones & glasses	unacceptable	(68%)
Repeated Calibration	unacceptable	(92%)

Table 1: Impact of equipment, 30 respondents, answers in second column

In question Q8, we asked directly for respondents' requirements in numerical values for a series of sub questions. Firstly, we asked respondents what *accuracy* was required for their application in metres[5], which, as a conversational question is clear enough, but for statistical purposes leaves a lot of room for interpretation, due to the undefined use of the notion of *accuracy.* To be able to trust the answers more confidently we would have needed to ask something like *"How far away from the actual position can the measurement be 95% of the time for your application to still be working?"* Under the assumption that this was what we implied in our original phrasing, we got quite conclusive answers, after discarding what we considered to be an extreme answer as an outlier. However, A further *caveat* here is needed as to what comprises an outlier, as the expectation on accuracy realistically follows an inverse exponential curve towards an error free measurement. To illustrate this, let's suppose there was a measure for "technological effort": To improve position accuracy from 20 metres to 1metre probably needs the same effort as, say, improving it from 1m to 5mm, as an arbitrary example.

So here are the generalised numbers for the answers as to the expectations on position accuracy

- Range of answers: 1mm - 20 metres
- 58% of answers lie between 0.1 - 2 metres
- Arithmetic mean = 2.6 metres.
- Median = 1 metre

We presume the median to provide the most meaningful number. Ignoring outliers (15%) the range is from 0.1 - 5 metres, of which the majority (76 %) requires 0.1-2 metres

Just over half of respondents require vertical information. The survey did not clarify if this means 2 dimensional tracking vertically oriented, or actual three dimensional tracking.

In the expectation on what area needs to be covered by the tracking system, again, the median might give the most conclusive answer, which is 100 square meters. (Values range from 1 - 2000 square metres)

Unfortunately, answers to the question on latency was inconclusive, as we asked for answers in seconds, not allowing for answers in the for audio application typical lower millisecond range, which would have been more informative.

3.2 Interacting Through Space

What makes the workshop so interesting as a growing bed for spatial interaction is its situatedness through improvisation: Arguing with Suchman [38] that situated interaction

[5] we used the phrase *The accuracy required for my application is...*

can neither be wholly projected or planned, nor exhaustively described on hindsight, we acknowledge that an improvisational practice can only ever be anything but situated. It provides a prototype in itself for the study of spatial interaction, possibly for all types of interaction.

Exploring a space just for its acoustic qualities, as a found instrument, or as an augmentation of an existing instrument, allows for experimentation with spatiality as a means of expression. From this starting premise we would like to follow the threads which distinguished themselves in our data, the field notes: Firstly, space as a change in position over time, secondly space as a sound, thirdly, space as a visual entity and lastly, space as a representation in a semiotic sense. The last thread has only indirect bearing on requirements, but our cultural associations with space still influences our experience of it, thus impacting on the situation the interaction takes place in. We try to account for this influence in all of the following subsections, but will not look at it separately, as we believe these representations to be a negotiated outcome of the precedent interaction, which in turn is based on primordial motor behaviours, as Noland convincingly argues with Merleau-Ponty [24].

3.2.1 Space as Dynamic, Multiple Positions

Moving sound sources, movement as sound sources, gestures in a scale from the plucking of a string to cartwheeling through the room make up the dynamic tools of spatial interaction in form of trajectories, displacements, oscillations. Many participants from non-prominently moving disciplines, for example, composers and some visual artists found it at first daunting to experience themselves as moving bodies. But the common ground established via the kinaesthetic experience of moving as a function of a particular discipline, like moving a brush, a pen, a bow, a plectrum, and so on, provided a shared vocabulary which helped to overcome initial inhibitions. Participants with a background in performing arts like dance or theatre discovered in turn how their movement has a sonic quality too, be it in the form of steps or the sound of props. From this shared understanding the evolving interdisciplinary practice made use of every available spatiality: Musicians who's instruments are light enough to be played while moving did so. Non-mobile instruments were often temporarily ditched for more mobile ones. If a ladder was in the room, it got incorporated, as did adjoining rooms, galleries and staircases. As a consequence, the space became a narrative: To experience a particular event somewhere in this extended performance area, one just had to happen to be there, as at another end one would have missed it: Every participant had her or his own experience of his or her trajectory, laced with nodes of interactive encounters with others. But even when the overall event was an assembly of micro performance, it was always experienced as a coherent whole.

In view of our research questions towards existing and potential technology to enhance or support this interaction, we made some astonishing findings: Audio technology in form of standard loudspeakers in a surround set up proved to be cumbersome, as they did not portray any correlation between the positions of the sound sources they represented, on the same token all sounds coming from laptops were situated remotely from the acoustic sources, creating an abstracted space many participants found very hard to relate to from a spatially dynamic perspective. Some interesting low-tech *ad hoc* solutions were the use of very long cables allowing loudspeakers to become personal sound sources, carried around by the performers.

This led to the descriptions of distinct scenarios for interactive automated panning systems we presented in [32], but also inspired the development of low latency wireless loudspeakers for performance purposes. The use of wireless microphones was embraced by many participants, be it just as a means to record a spatial narrative as a trajectory, or also to amplify quieter moments of sonic activity. Here, using loudspeakers to reproduce sound sources in locations remote from where we performed them, was used artistically by some. But for others it further helped the formulation of the idea for an automatic panning of real-time sound sources as shortly discussed above.

Experimenting with existing technology, we noticed how important it is for the sound to be free of any noticeable latency. An erroneous setting on a sound card, or the latency due to networking made us aware that time delays around 20 ms can be noticeably long. This informed our proposal of low latency gestural interfaces in [33]

We find it worth mentioning that the representational nature of the spaces we used always had an inspiring effect on our experience: The architectural impact on spatiality is of course huge, a sunlit foyer evokes a different session than a black box, as does the presence of props, the reverberation/attenuation over distance, etc. The scenography is paramount for space as a means of expression.

3.2.2 Space as a Sound

The well titled book by Blesser and Salter, *Spaces Speak, are you listening?* [7] explores space as a sounding phenomenon rather than the kinetic aspects discussed above. The spatial interaction within this type of space is therefore quite different. The spatial qualities of sound are a direct result of the space due to the way a space reflects, diffracts or absorbs sound. Every place is different. Spatial interaction in a kinaesthetic sense takes the form of an exploration: Interaction here is quite the listening Blesser and Salter demand. One could presume that the sounding space is a given *a priori* entity, and static: However, every possible position in a room sounds differently, providing a rich and complex texture of possibilities. It is a different perspective: Not the movement itself is of direct interest, but how the sound of the room changes with the change of position of a moving sound source.

Another experience we made is that, due to the *a priori* nature of the sound of a room, the introduction of reverberations superimpose a combined spatiality of the virtual, artificial room and the existing room. The use of, for example, commercial reverb-effect pedals, with their strong representational character, need a very subtle hand not to overpower all other possible spatialities. They set a scene which can not be ignored. From a kinaesthetic point of view, once triggered, there is very little kinetic control over these reverberations. In contrast, electro acoustic feedback, as long as it can be controlled, provides a very engaging kinaesthetic experience, both for the performer as for the listener for whom the gestures for control are obvious and emphatically understandable if performed, for example, with a loudspeaker and a microphone.

To interface space as a sound in a kinaesthetic content, we thus need to adapt the functionality of the interface:

Figure 1: Theatrical Aspects of MS&I. ©Dominik Schlienger

Whereas it senses the movement through space in our previous description, delivering digitalised movement, here we want it to dynamically deliver the digitised audio spatiality of the space, so to speak: A suggestion as to how to do this was to continuously measure acoustic impulse responses from changing positions in the room. This thread in the data is particularly aural and might not find its parallel in a non-audio context.

3.2.3 Space as a visual entity

For a sound art practice it maybe indicative that space as a visual entity did not play a dominant role in the data. Often, the experience of the visual as an expressive factor was only mentioned as an afterthought. Yet, particularly when looking at the photographic documentation[6] it becomes clear that the sessions are indeed theatrical and highly visual too. We believe that this comes out of a situated engagement with the space, more than out of a wish to perform visually: Visuality is just one aspect of the multi modality of interdisciplinary improvisation. For example in Figure 1 we see how a participant wrapped another participant in paper. Primarily they were exploring the soundscape of the paper world they immersed themselves in. The resulting visual aspect then was a cue for other participants to react in a theatrical way, and also prompted the photographer to take the picture.

In sessions in which we had light design as a participating discipline, these visual cues were respectively stronger, strengthening our visual awareness. Light sources, from hand held torches, projectors, to theatre spots, just like sound sources, provided great localised cues and expressive opportunities. As part of the light design also experiments in darkness became part of our practice.

Another visual aspect in connection to kinaesthesia came up in conversation with one of the participating painters. She remarked that the brush stroke, as the result of the kinaesthetic experience of moving the brush, is the embodied trace of this movement. The audio parallel, the embodied playing of an instrument we can hear on a recording, was evident. From here arose the question of what happens to the kinaesthetic experience in electronic music, a subject debated in a range of works, for example in [27]

3.3 The Kinaesthetic Interface

Even more strikingly, in the same discussion, we noticed that, for example, in taking notes with a computer keyboard rather than in handwriting, we lose an embodied, situated inscription of a whole world of things that is otherwise visible from handwriting: From hand written notes, even without the aid of graphology, we can see if the writer was in a hurry, or was taking her time, we see what was scribbled out, also at what point did the writer nod off, or was he excited, and so on. These qualities are lost in typed writing. Our discussion then took a gloomy turn, many participants expressed discontent with the lack of possible kinaesthetic experiences in the digital. It culminated in somebody raising the spectre of modern man being nothing but a disembodied head on a sofa in a digital future.

If, in the consequence of this discussion, we did not come up immediately with a grand solution, it still paved the way for a gradual conception of the importance of kinaesthetic interfaces for an embodied digitality, the requirements for which we are now able to define at least for a general sonic arts practice:

- The kinaesthetic interface might not be able to trace anything else but kinetic events. However, it records those kinetic events at the right resolution, over the necessary distances, at sufficient speeds and with the necessary accuracy to make them *relevant* enumerations and encodings as a parameter correlated to its kinaesthetic experience.
- From both the survey and the workshop on MS&I, we know that the interface shall hinder or impact on the expressive activity as little as possible. It needs to be able to trace the displacements of a human body and its gestures in real-time. For example, if a gesture was to be mapped to a sonic event, there should be no perceivable latency. In many of our sessions meaningful trajectories were three dimensional. The Distances we believed to be meaningful range from the low centimetre range to tens of meters. Here our experiences in the workshop are reflected very closely in the results of the survey.
- In the subsection on Space as Sound, we discussed the possibilities a record of the space in a particular point over time would bring. We believe the capacity of the interface to be able to take impulse responses would enable this. The survey was not able to tell us anything about space as a sounding entity, as the survey design did not anticipate such an outcome.

[6]The photos were taken as part of the improvisation, by a participating photographer

- All scenarios we studied and experimented with, were concerned with interaction with multiple participants in indoors environments.
- Last, not least, from our discussion on visual aspects of space in our practice, we know that this interface can not rely on the presence of light, nor can it rely on line of sight as we describe categorically, multi participant scenarios in three dimensional space.

With the elicited notion of the kinaesthetic Interface, we believe we have found a prototypical concept to integrate many different notions of gestural interfaces. Providing a broader situatedness in our improvisational practice, in the way Suchman recommends for human computer interaction in [38], we believe our findings are applicable also to non-artistic applications. Admittedly this may be due to their general character. This, in turn however, might also point to the unavoidable realisation that every application has its very own situation, a realisation which emphasizes the importance of research into this situatedness, as proposed by Suchman already in 1987, but rarely acted upon by developers until quite recently. In this sense, we hope to have provided an account of a practice which elicits requirements. If our experiences in this practice prove to be general enough to resemble some of the broader artistic communities' experiences, this might indeed help to drive development of technology which meets these requirements. For a summary of our results, see table 2.

	survey	**workshop**
Accuracy	$0.001 - 10\,\mathrm{m}$	0.1 - 0.3 m
Update Rate	continuous	
Latency	not perceivable ($< 20\,\mathrm{ms}$)	
Ubiquity	high	
Cost	low	
Area covered	1 - 1000 m^2	10 - 1000 m^2
Distances		0.01 - 100 m
Speeds	human gestures ($< 40\,\mathrm{m/s}$)*	
Scenario	multi-user, crowded, indoors	

*Fast cricket bowler's ball release- speed according to [4]

Table 2: Summary of Results by Method

4. CONCLUSIONS & FUTURE WORK

We believe to have at hand a useful set of requirements as recommendations for design of a kinaesthetic Interface, the notion of which we have developed here in order to give a better understanding of what spatial interaction in the sonic arts requires from technology. In basing our approach on situated ethnographic methods as well as more traditional ones in form of an online questionnaire, we were able to verify to a certain extend the validity of both approaches, as the data was mirrored in many ways. However, we would also like to highlight that the prototypical practice in the workshop was able to provide unexpected outcomes, answers to questions we did not know we had *a priori*. A questionnaire is a lot more rigid in that aspect.

For future work, we would like to see how these requirements fair when implemented and reintroduced to the workshop in a next cycle of development. We also hope to develop the concept of the kinaesthetic interface much further as a contribution to the discussion on a new, embodied, digitality.

5. ACKNOWLEDGMENTS

My research is kindly supported through a Researchers grand by the Kone Foundation. Thanks go to all participant of the Workshop on MS&I and also all the respondents to the survey

6. REFERENCES

[1] J. Andean. Research group in interdisciplinary improvisation: Goals, perspectives, and practice. http://www.academia.edu/4158856/Research_Group_in_Interdisciplinary_Improvisation_Goals_Perspectives_and_Practice, 2014.

[2] S. Baldan, A. D. Götzen, and S. Serafin. Sonic tennis: a rhythmic interaction game for mobile devices. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 200–201, Daejeon, Republic of Korea, May 2013. Graduate School of Culture Technology, KAIST.

[3] A. Balsamo. *Designing Culture: The Technological Imagination at Work.* Duke University Press, Durham, NC, USA, 2011.

[4] R. Bartlett, N. Stockill, B. Elliott, and A. Burnett. The biomechanics of fast bowling in men's cricket: A review. *Journal of Sports Sciences*, 14(5):403–424, 1996.

[5] G. Barz and T. Cooley. *Shadows in the Field: New Perspectives for Fieldwork in Ethnomusicology.* Oxford University Press, 2008.

[6] J. Blauert. *Spatial hearing : the psychophysics of human sound localization.* Cambridge, Mass. MIT Press, 1997.

[7] B. Blesser and L.-R. Salter. *Spaces speak, are you listening? : experiencing aural architecture.* Cambridge, Mass. MIT Press, 2007.

[8] J. G. Bolaños and V. Pulkki. Immersive audiovisual environment with 3d audio playback. In *Audio Engineering Society Convention 132*, Apr 2012.

[9] E. Brandt, T. Binder, and E. B.-N. Sanders. Ways to engage telling, making and enacting. *Routledge international handbook of participatory design. Routledge, New York*, pages 145–181, 2012.

[10] C. Dobrian and F. Bevilacqua. Gestural control of music using the vicon 8 motion capture system. In *New Interfaces for Musical Expression, NIME-03, Proceedings, Montréal, May 22-24, 2003*, pages 161–163, 2003.

[11] P. Dourish and G. Bell. *Divining a Digital Future: Mess and Mythology in Ubiquitous Computing.* MIT Press, Cambridge, MA, May 2011.

[12] N. Gillian and J. A. Paradiso. Digito: A fine-grain gesturally controlled virtual musical instrument. In *NIME.* nime.org, 2012.

[13] M. Hyder, M. Haun, and C. Hoene. Placing the participants of a spatial audio conference call. In *2010 7th IEEE Consumer Communications and Networking Conference*, pages 1–7, Jan 2010.

[14] C. Kiefer, N. Collins, and G. Fitzpatrick. HCI methodology for evaluating musical controllers: A case study. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 87–90, 2008.

[15] A. Kuusinen, T. Lokki, J. Pätynen, and S. Tervo. Insights from three profiling studies of auralized concert hall acoustics: Support for a three dimensional perceptual space? *The Journal of the Acoustical Society of America*, 139(4):2087–2087, 2016.

[16] B. Latour. *Reassembling the Social:An Introduction to Actor-Network-Theory: An Introduction to Actor-Network-Theory*. Clarendon Lectures in Management Studies. OUP Oxford, 2005.

[17] B. Latour and EBSCOhost. *An inquiry into modes of existence : an anthropology of the moderns / Bruno Latour ; translated by Catherine Porter*. Harvard University Press Cambridge, Massachusetts, 2013.

[18] Y. K. Lim and W. S. Yeo. Smartphone-based music conducting. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 573–576, London, United Kingdom, 2014. Goldsmiths, University of London.

[19] X. Liu, M. D. Corner, and P. Shenoy. *UbiComp 2006: Ubiquitous Computing: 8th International Conference, UbiComp 2006 Orange County, CA, USA, September 17-21, 2006 Proceedings*, chapter Ferret: RFID Localization for Pervasive Multimedia, pages 422–440. Springer Berlin Heidelberg, Berlin, Heidelberg, 2006.

[20] T. Lokki and K. Puolamäki. Canonical analysis of individual vocabulary profiling data. In *Quality of Multimedia Experience (QoMEX), 2010 Second International Workshop on*, pages 152–157, June 2010.

[21] C. Magerkurth, A. D. Cheok, R. L. Mandryk, and T. Nilsen. Pervasive games: Bringing computer entertainment back to the real world. *Comput. Entertain.*, 3(3):4–4, July 2005.

[22] L. Manovich. *The language of new media*. MIT Press, Cambridge, Mass., 2002.

[23] J. Nilsson, C. Schüldt, and P. Händel. Voice radio communication, pedestrian localization, and the tactical use of 3d audio. In *IPIN*, pages 1–6. IEEE, 2013.

[24] C. Noland. *Agency and Embodiment: Performing Gestures/Producing Culture*. Harvard University Press, 2009.

[25] S. O'modhrain. A framework for the evaluation of digital musical instruments. *Computer Music Journal*, 35(1):28–42, 2011.

[26] L. Peng and D. Gerhard. A Wii-Based Gestural Interface for Computer Conducting Systems. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 155–156, 2009.

[27] D. Peters, G. Eckel, and A. Dorschel. *Bodily Expression in Electronic Music: Perspectives on Reclaiming Performativity*. Routledge Research in Music. Taylor & Francis, 2012.

[28] T. Sangild. Glitch - the beauty of malfunction. *Bad music: The music we love to hate*, pages 257–74, 2004.

[29] D. Schlienger. Indoors and local positioning systems for interactive and locative audio applications. http://creativemusictechnology.org/MScAudioProduction99001618TextNoApcs.pdf, 2012.

[30] D. Schlienger. Acoustic localisation techniques for interactive and locative audio applications. In *Locus Sonus Symposium #8 on Audio Mobility, Proceedings*, 2014.

[31] D. Schlienger. Survey: User requirements on positioning systems for audio applicatons. http://creativemusictechnology.org/research_&_development_EN.html, 2014.

[32] D. Schlienger. Acoustic localisation for spatial reproduction of moving sound source: Application scenarios & proof of concept. In *International Conference on New Interfaces for Musical Interaction 2016, Proceedings*, 2016.

[33] D. Schlienger. Gestural control for musical interaction using acoustic localisation techniques. In *International Conference on Live Interfaces 2016, Proceedings*, 2016.

[34] D. Schlienger. Workshop on Music, Space & Interaction participants' blog. http://creativemusictechnology.org/MS&P.html, 2016.

[35] D. Schlienger and S. Tervo. Acoustic localisation as an alternative to positioning principles in applications presented at nime 2001-2013. In B. Caramiaux, K. Tahiroglu, R. Fiebrink, and A. Tanaka, editors, *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 439–442. Goldsmiths, University of London, June 30 – July 03 2014.

[36] S. Şentürk, S. W. Lee, A. Sastry, A. Daruwalla, and G. Weinberg. Crossole: A gestural interface for composition, improvisation and performance using kinect. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, Ann Arbor, Michigan, 2012. University of Michigan.

[37] J. Simonsen and T. Robertson. *Routledge International Handbook of Participatory Design*. Routledge International Handbooks. Taylor & Francis, 2012.

[38] L. A. Suchman. *Plans and Situated Actions: The Problem of Human-machine Communication*. Cambridge University Press, New York, NY, USA, 1987.

[39] S. Trail, M. Dean, G. Odowichuk, T. F. Tavares, P. F. Driessen, W. A. Schloss, and G. Tzanetakis. Non-invasive sensing and gesture control for pitched percussion hyper-instruments using the kinect. In *NIME*. nime.org, 2012.

[40] Vicon. Vicon motion systems, 2016.

[41] M. M. Wanderley and N. Orio. Evaluation of input devices for musical expression: Borrowing tools from hci. *Comput. Music J.*, 26(3):62–76, Sept. 2002.

[42] M. Weiser. Building invisible interfaces. Presentation Slides, Novemeber, 2nd 1994.

[43] Wikipedia. Kinect — wikipedia, the free encyclopedia, 2016. [Online; accessed 19-May-2016].

[44] Wikipedia. Wii — wikipedia, the free encyclopedia, 2016. [Online; accessed 19-May-2016].

10

Immersive Spatial Interactivity in Sonic Arts – The Acoustic Localization Positioning System ALPS (VI)

Schlienger, D., and Kaschanskyi V. 2022. "Immersive Spatial Interactivity in Sonic Arts – The Acoustic Localization Positioning System" Computer Music Journal, 45(2):24–47, 06 2021.

Dominik Schlienger and Victor Khashchanskiy

University of the Arts Helsinki
Sibelius Academy
PO Box 30, FI-00097 Uniarts
dominik.schlienger@uniarts.fi,
victor@victorx.eu

Immersive Spatial Interactivity in Sonic Arts: The Acoustic Localization Positioning System

Abstract: The Acoustic Localization Positioning System is the outcome of several years of participatory development with musicians and artists having a stake in sonic arts, collaboratively aiming for nonobtrusive tracking and indoors positioning technology that facilitates spatial interaction and immersion. Based on previous work on application scenarios for spatial reproduction of moving sound sources and the conception of the kinaesthetic interface, a tracking system for spatially interactive sonic arts is presented here. It is an open-source implementation in the form of a stand-alone application and associated Max patches. The implementation uses off-the-shelf, ubiquitous technology. Based on the findings of tests and experiments conducted in extensive creative workshops, we show how the approach addresses several technical problems and overcomes some typical obstacles to immersion in spatially interactive applications in sonic arts.

The technological developments in global navigation satellite systems; indoor and local positioning systems using radio signals, ultrasound, optical motion tracking; inertial reference units like gyroscopes and acceleration meters; and localization services on smartphones, etc., present an array of conceptual possibilities for use in sonic and performance art. Tracking technology for musical interaction is a perennial subject in works presented at conferences like the International Conference on New Interfaces for Musical Expression (NIME) and the International Conference on Live Interfaces. Suitable technologies like motion capture are expensive, however. Generally available technologies, like the global positioning system (GPS) and hybrid approaches using smartphone technologies, have, to our knowledge, rarely been developed to provide the low latency required for musical expression. In practice this has not hindered developers and musicians from coming up with a plethora of creative solutions, from "circuit-bending" interactive video gaming consoles to building case-specific hardware from scratch.

The project that engendered this article and provided its rationale was an art installation in which the spatial position of a participant controls a musical parameter. The installation took place in June 2010 at the 470 Degrees Graduate Show, held at the University of the West of England in Bristol. By changing the participant's position within the room, some parameter of sound or musical parameter is affected. Parameters of sound are, for instance, amplitude, pitch, duration, timbre, and overtones. Musical parameters can include melody, harmony, rhythm, texture, expression, dynamics, tempo, and articulation. As the project was initiated by technophile artists rather than by engineers, and because it was financed on a shoestring, only readily available, ubiquitous technology was ever considered. After fruitless experiments with Bluetooth signal-strength measurements, the most promising technology appeared to be techniques using acoustic localization (AL) for sound just above the frequency range audible to humans, for the reason that in the envisaged setup, loudspeakers would already be part of the system, and unobtrusive small microphones could easily be introduced at low cost.

Acoustic localization techniques can be used for tracking and positioning purposes by measuring the time it takes a sound to travel from a source to a sensor. As the speed of sound through air is known (343 m/sec), the measurement can be used to estimate the distance, according to the relationship $d = c \times t$, where d is the distance a sound wave travels, c is the speed of the sound wave, and t is time. Yet, despite the simplicity of the principle, it has rarely been implemented for artistic purposes. Intrigued by the clear potential for low-cost solutions and following Occam's argument that entities should not be multiplied beyond necessity,

Computer Music Journal, 45:2, pp. 24–47, Summer 2021
doi:10.1162/COMJ_a_00605

we found the choice of AL to be utterly compelling, if tracking is required in a system constituted by loudspeakers and microphones.

Subsequently, a plan formed to develop a tool, based on these principles, that would not only meet the requirements for said artistic idea, but also be a more generally useful utility for other artistic projects in the field of sonic and or performance arts. The tracking system and the applications presented here were developed concurrently. This happened in an experimental artistic practice which came into being as a part of the development process. From within this practice of participatory design, it was possible to define more general requirements, culminating in the concise notion of the *kinaesthetic interface* (discussed in more detail below), which now provides a benchmark of sorts: It allows testing against clear qualitative and quantitative parameters.

Does the technology developed fill the gap encountered when looking for the suitable technology for the artistic project we presented in Bristol in 2010? And more generally, does it provide a useful tool for the community of artists working with spatial interaction in the sonic arts? In answer to these questions, we aim to demonstrate how the algorithm, which we call the Acoustic Localization Positioning System (ALPS), in conjunction with the Autopan Max patch we developed, helps to solve a series of problems often encountered in spatially interactive improvisation: Whereas participants with acoustic instruments can use their spatial position within a performance space as a musical parameter, electronic instruments generally are statically bound to the position of the loudspeakers. By automatically panning the electronic sound source to the position of players, their location becomes an interactive element too, on par with the players of acoustic instruments. What is more, this approach also allows spatial trajectories to become musical narratives in electronic improvisations.

Background

The positioning technology commonly applied for spatially interactive sonic arts is optical tracking, for example, motion capture. Some examples thereof are described in Nymoen, Skogstad, and Jensenius (2011), Dobrian and Bevilacqua (2003), and Bazoge et al. (2019). Hacks of the interactive video gaming consoles Kinect (Şentürk et al. 2012; Trail et al. 2012) and Wii (Peng and Gerhard 2009) also use optical means of tracking. For applications that do not require absolute positions, "dead reckoning" methods (i.e., methods of calculating one's current position by using a previously determined position, or fix) can be used, as can inertial methods using accelerometers, gyroscopes, and magnetic compass. These tend to map smaller movements to a process, as does the data glove by Mitchell and Heap (2011). The proceedings of the annual conference on Indoors Positioning and Indoors Navigation provide examples for many possible technologies (see http://ipin-conference.org for details). Besides the classifications in Schlienger and Tervo (2014), a summary of positioning technologies can be found in Hightower and Borriello (2001) and more recently in Brena et al. (2017).

Every year, the proceedings of the NIME conference provide examples of applications using positioning or tracking technology. An extensive summary of these can be found in a review of over 80 applications presented at NIME between 2001 and 2013 (Schlienger and Tervo 2014). By comparing these applications' requirements with the theoretical possibilities of AL, it is demonstrated that AL could have provided feasible alternatives for many of the applications at lower cost.

In summary, for this project, AL was chosen for the following reasons: First, AL works with off-the-shelf loudspeakers, which, in the scenario in question, are already present for the diffusion of content audio (e.g., music or speech). Besides a microphone, little additional equipment is needed, as the same loudspeaker can be used for tracking and content audio. Second, when maintaining the identity of the tracked device is required, AL—along with other sender–receiver technologies like radio frequency identification (RFID)—can provide this more reliably than optical systems. Third, the tracked device is unobtrusive: Omnidirectional condenser microphones of the lavalier type (worn by actors on stage, musical theater, and opera)

Table 1. Quantitative Requirements on ALPS

Accuracy of localization	≤0.3 m
Update rates	>20 Hz (perceptually continuous)
Coverage	10–1000 m^2
Latency	<20 msec

often measure under 5 mm in diameter. Finally, in comparison to motion capture, there are fewer line-of-sight (LOS) issues in AL: Obstructions in the path between a loudspeaker and a microphone need to be large in comparison to the dimensions and directionality characteristics of the loudspeaker, so that the direct path of the radially propagating sound wave is entirely occluded, whereas a marker in a motion-capture setup can be occluded by another object the size of the marker. For the application by Lopes et al. (2006), which tolerates location errors of up to 70 cm, LOS effects are described as negligible. In our own experience, these errors were only a minor issue, albeit not entirely negligible. That is, for autopannning with ALPS, a smaller error margin of up to 30 cm is tolerable. Seob Lee and Yeo (2011) even successfully hid their microphone behind a curtain.

Literature Review and Previous Work

Some similar applications use ubiquitous technology: Janson, Schindelhauer, and Wendeberg (2010) present an iPhone application using ambient sound signals for tracking and synchronizing phones via Wi-Fi; Filonenko, Cullen, and Carswell (2010) investigate ultrasonic positioning on mobile phones in general; and Mandal et al. (2005) write about the possibilities of using audible frequency signals for tracking. All three methods indicate that an implementation using standard loudspeakers could be achieved, but it is difficult to infer if they could meet all the requirements listed by Schlienger (2016c), summarized in Table 1. In these applications, Janson and coworkers demonstrated the positioning of distinct sound events, taking a measurement of a moving person every 10 to 20 meters, whereas Filonenko and colleagues work with updates at 1 Hz. The use of audible sound by Mandal and associates is interesting: It would be convenient to use content audio directly as a measurement signal. For reasons explained below, a distinct, inaudible measurement signal was deemed more appropriate for ALPS. The Active Bat (Harter et al. 2002) could provide the required functionality, but as it uses purpose-built ultrasound transmitters, it does not comply with our decision to use ubiquitously available technology only.

More recently, Aguilera et al. (2017) presented a system for smartphones as mobile receivers and a minimal number of custom-built senders, for use in small rooms (3 × 3 m), applying a broadband signal in the audible frequency range. They do not provide any indications regarding latency, but they achieved update rates up to 2 Hz, with accuracy of around 10 cm.

Acoustic localization has only rarely been implemented in spatially interactive sonic arts, despite the documented feasibility. A notable exception is the Sonicstrument by Seob Lee and Yeo (2011), using analysis of Doppler shifts to track a pair of off-the-shelf earbuds with one microphone. One bud controls the pitch and the other controls the amplitude of a synthesized sound, mapped to a performer's gestures. They further describe a larger-scale, interactive dance performance, covering an area of approximately 10 × 10 m. They chose Doppler over time-of-flight measurements, as the latter suffer from "limited precision due to the irregular time delay of the system process" (Seob Lee and Yeo 2011, p. 25). By applying an astonishingly simple technique patented by Medvedev, Sorokin, and Khashchanskiy (1989), however, ALPS avoids these issues (see the Signal Processing in ALPS section for details of the technique). Accordingly, the Doppler effect was not explored further, although it would have been an equivalent approach.

Latency is a systemic issue in all time-difference-of-arrival (TDOA) approaches to tracking, as time elapses during the measurement. In AL, which measures comparatively slow signals, it has considerable impact. Applications for which low latency is essential, such as gestural control of musical instruments, tend to map smaller distances than do larger-scale applications like the 2010 Bristol project. Thus, the scalability of AL facilitates this to some extent, as smaller distances need less time to be measured.

Latency issues, other than the ones unavoidably induced by measurement, need to be dealt with separately, however. They concern the processing time required after the measurement, as well as time used by other processes on the same central processing unit (CPU) that might be given priority over audio tasks in a multipurpose processor like a laptop or desktop computer. Jack, Stockman, and McPherson (2016) provide a concise summary of what latency is acceptable for gestural control of musical instruments. They suggest that if jitter (the change in latency) is small, gestural control at 30–50 msec latency is possible, as musicians can anticipate the delay as long as it stays constant, although also mention studies that set 10 msec as an upper tolerance. Note, however, that they concluded that the majority of commonly used platforms for electronic music are worse than this benchmark. As to the amount of latency that can be experienced by musicians in gestural response, they state that even 4 msec may be noticeable, particularly as jitter. This is consistent with our experience.

Previous Work as Part of the Present ALPS Project

We demonstrated the viability of AL for positioning or tracking in artistic uses at the Klingt gut! Symposium on Sound in Hamburg in 2016: "Leluhelikvartetti" (Finnish for toy helicopter quartet), a homage to Karlheinz Stockhausen's Helicopter String Quartet, uses toy drones that spatialize the sound of the Free Improvisation String Quartet, using an algorithm based on the proof of concept (PoC) described by Schlienger (2016a). A short clip of the performance is available on video from https://doi.org/10.5281/zenodo.5608818. The performance was given an award for Excellence in Art, Design, and Production of Sound by the AES Student Section Hamburg. The AL tracking system used audible noise (around 15 kHz) for tracking, which was masked by the high-pitched whirring noise of the toy drones' four sets of propellers. Responsiveness was adequate for this performance, as the drones moved at moderate speeds. Although the principle remained roughly the same, the current implementation uses inaudible noise for tracking and is more responsive with considerably lower latency.

Earlier work explored how tracking technology could be applied to overcome limitations experienced with audio technology in practical application scenarios. Namely, the disconnection experienced by musicians when using electronic and electric instruments without built-in acoustic sound sources, and the gestural limitations of many electronic instruments are discussed in earlier publications (Schlienger 2016a,c), based on field notes from a free-improvisation workshop. (Other authors also describe the gestural limitations inherent to electronic instruments, e.g., Dean and Paine 2012; Mitchell and Heap 2011; Robinson et al. 2015; Salazar and Armitage 2018.)

The first author conceptualized the notion of the *kinaesthetic interface* in order to generalize the requirements for interfaces for spatially interactive sonic arts:

> It records kinetic events at the right resolution, over the necessary distances, at sufficient speeds, and with the necessary accuracy to make them relevant enumerations and encodings as parameters correlated to kinaesthetic experiences (Schlienger 2016c, p. 6).

Besides this qualitative notion, quantitative requirements summarized in Table 1 were identified with the help of an online survey, asking professionals in the field about their expectations of tracking systems. These quantitative requirements also provide the benchmark for the current implementation, as discussed below.

Schlienger (2016a) provided a PoC for an implementation for AL as a tracking device for automated panning (autopan) at the 2016 NIME conference, where it was successfully demonstrated. The demonstration also determined various application scenarios in which tracking technologies could help to create panning trajectories for moving sound sources in sonic arts. In the paper, four conceptual possibilities are identified. The following is an updated summary of the scenarios.

Two Spaces: Stage and Auditorium

This is the typical "concert hall" situation: Moving sound sources (i.e., musicians with instruments or singers) are in a different space from the audience. This other space can be a stage, in which case the automated panning reproduces (mirrors) the positions of the sound sources for the listeners in the auditorium. Even in the case of a classical music concerts, in which musicians remain seated throughout, it makes sense to replicate their spatial positions. In a broadcast situation, the listening space is even further removed from the musicians, but the logic remains the same. Here, and particularly in case of opera, musical theater, and other more spatially dynamic musical practices, automated panning can simplify the task of the sound engineer who otherwise would have to pan these sound sources manually. Conceivably, the scenario is the same if the musicians are in a studio. The difference in this case is that the position of a musician in the room does not need to be replicated, in fact it can be, and usually is, arbitrarily set without causing conflicting spatial impressions in the audience, who have no visual cue to the studio's setup. (So, in contrast to broadcasting and amplified live performance, automated panning would only have limited use in a typical recording studio situation.)

One Space: The Commons

A moving sound source is reproduced and amplified in the same space as the audience. For example, assuming a multiple-loudspeaker or surround-sound setup, and a very quiet, acoustic sound source with a closely placed microphone, which produces a signal that is simultaneously played back on the loudspeaker closest to it, or panned to its phantom position between multiple loudspeakers. Here "very quiet" can also apply to a laptop or another electronic or electric instrument. By locally amplifying it, the quiet instrument becomes the acoustic equal of a loud instrument, which probably does not need amplification and inhabits a distinct, localized space, with its sound radiating from it. Until now, the quiet instrument would have been amplified at a fixed pan position on one or several loudspeakers, possibly misrepresenting the instrument's actual position, particularly when the sound source is moving. Here, automated panning vastly increases the spatially interactive possibilities for a range of musical practices—even enabling new ones, in the fields of participant performances, performance art, festival commons, art games, and many more.

Virtual Sound Source

This is a typical example of a spatially interactive installation: A panning trajectory is created in real time (e.g., by visitors issued with a tracking device), so that sound follows their movements. The audio content being produced elsewhere or was prerecorded. A further possibility would be that a participant triggers certain events at certain positions. If the space is large enough for audio content to be heard selectively, depending on the participant's location, a panning trajectory can create narrative meaning: Participants will hear different sequences of events depending on their relative positions to each other. Last but not least, this is also a way to map offline content in real time to an online trajectory in space.

Trajectory and Sound Event Temporally Separated

Here the panning trajectory is produced in advance, and later used as a map for a musical event, which might also have been produced in advance. Although the trajectories could be generated by other means, tracking technology could provide an easy and organic way to achieve this.

The four application scenarios defined above are for spatial reproduction of moving sound sources in spaces from approximately 9 to 144 m^2. In earlier work it was shown how the same approach can also be scaled to smaller spaces, resulting in higher update rates, allowing gestural control of musical parameters that depend on quasi-real-time interaction (Schlienger 2016b). A Theremin-like pitch-control software device was demonstrated, along with a rudimentary percussive instrument—rudimentary being the operative word here. Yet theoretical possibilities are evident and the current research paves some inroads towards improving low-latency gestural control with AL.

System Overview

This section first introduces the ALPS algorithm and its implementations in the ALPS software Audio1 and al-Qt, written in C++ by the second author and available at https://doi.org/10.5281/zenodo.5602869. To achieve audio panning—the adjustment of relative amplitudes of audio signals on loudspeakers to achieve the impression of a phantom sound source lying somewhere between the loudspeakers—some information as to the position of that virtual sound source is required. Such positions are provided in ALPS by using AL techniques. Evidently, to use ALPS for audio panning makes sense as both processes, positioning and tracking with AL and audio panning, require an infrastructure composed of loudspeakers. But ALPS could be adapted to provide positioning data for a variety of uses other than panning, wherever loudspeakers are present as part of a setup (multimedia, virtual reality, auditoriums, surround sound, museums, etc.), for instance, to trigger items in an audioguide relating to museum exhibits, or similar.

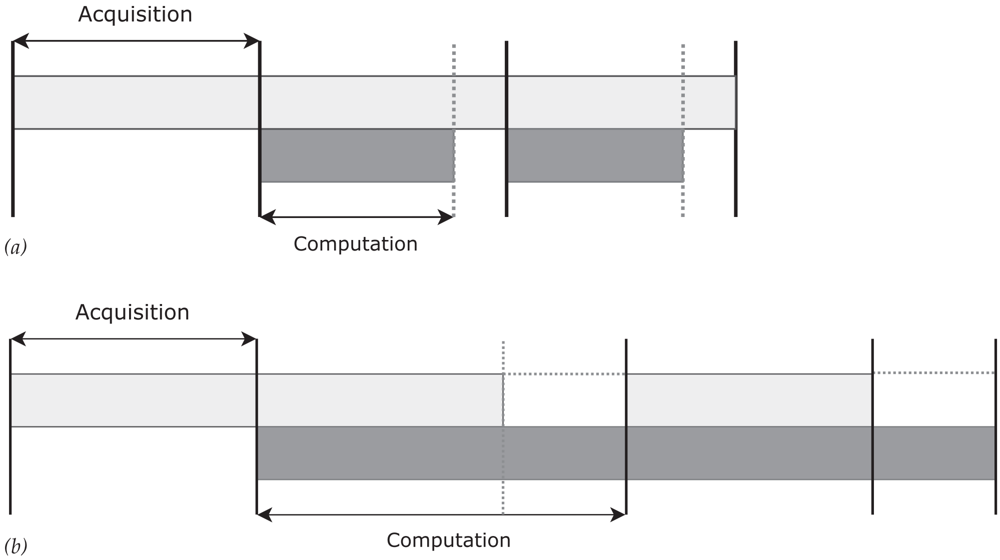
We then introduce the Autopan Max patch (https://doi.org/10.5281/zenodo.5607121), which utilizes ALPS to automatically pan content audio to the position of a person carrying a receiver (microphone) while moving in a room equipped with multiple loudspeakers. As such, it is a specific application of ALPS, which solves the seemingly technically orthogonal problems of *panning* and *positioning* by aligning them through the use of distance-based amplitude panning (DBAP, cf. Lossius, Baltazar, and de la Hogue 2009).

The ALPS Algorithm

To estimate positions of acoustic signal receivers (microphones) and senders (loudspeakers), ALPS uses AL. Acoustic localization techniques are applied in sonar, fish finders, and parking aids in cars, but also for medical purposes (e.g., sonography). Typically these technologies use dedicated hardware and ultrasonic signals (sound in the frequency range above human hearing). As it is known that sound moves typically at 343 m/sec, the distance between a microphone and a loudspeaker can be calculated by measuring the time delay of a known audio signal at the microphone in comparison to its original on the loudspeaker: $d = c \times t$ where d is the distance a sound wave travels, c is the speed of the sound wave, and t is time. Using several loudspeakers, or several microphones, the position of a sound source can be trilaterated, using the TDOA method. Trilateration calculates positions from distances, describing the points of intersections of spheres. To estimate the position of a moving object in 3-D space with trilateration, four synchronous distance measurements from known points are necessary. If the position to be tracked can be assumed to be on a plane—the earth's surface in GPS, for example—some simplifications are possible. Triangulation, in contrast, estimates a position from known angles between objects, without any knowledge of their relative distances and hence intrinsically inapplicable to TDOA methods.

The TDOA methods trilateration and multilateration are not discussed here in greater detail, as the application of ALPS discussed in this article, namely, autopanning, does not require them: The relation between distance measurements obtained through AL and the panning laws applied in DBAP means that trilateration is not necessary—this is perhaps surprising. The distance measurements, 1-D positioning so to speak, suffice. Two-dimensional estimates of relative positions were only necessary to validate the accuracy of the ALPS software in the experimental setups B1 and B2 discussed below: Positions were defined through Cartesian coordinates on a single plane to estimate the position of a device moving along a known trajectory. This was done with the help of the Pythagorean theorem implemented in the ALPS error calculator MATLAB script (https://doi.org/10.5281/zenodo.5607528).

Figure 1. Acquisition (light gray) and computation (dark gray). When computation is faster than acquisition, computation waits for the end of acquisition, thus the process happens in quasi–real time (a). On the other hand, when computation lasts longer than acquisition, acquisition waits for completion of computation, which results in additional latency but guarantees that computation always refers to the most recent acquisition (b).



Signal Processing in ALPS

Using the principle of AL, a measurement signal (e.g., band-limited white noise) is played on a loudspeaker and compared to an audio recording of it, a *capture block*, made on a microphone at a distance d. A sampling frequency and a window length in samples are chosen. The larger the window, the larger the area that can be covered with the system, but the longer it will take to obtain a measurement. Each sample has an index number, sequentially from first to last sample of a window. The next step is to take one window's length of the measurement signal and calculate its correlation with the signal of the same length that was recorded on the microphone at the same time. In the resulting correlation signal, the sample with the highest amplitude is found at the index number corresponding to the delay between the measurement signal and the signal recorded on the microphone. According to the relation $d = c \times t$, the distance can be estimated between the microphone and the loudspeaker.

The data acquisition records the signal for the length of time of the window, while it computes the previous window's correlation signal. If this computation lasts longer than the acquisition of the following length of audio, the system delays the acquisition of the next length until computation is completed (see Figure 1). If the computation takes less time than the acquisition time, the system runs in quasi–real time. As ALPS uses pulsed signals stepping through the loudspeakers one by one, the effective latency of the system is given by the length of the pulse multiplied by the number of loudspeakers, even when processing is achieved in quasi–real time. This is discussed in greater detail in the following section, The Measurement Signal.

Additional latency needs to be accounted for, introduced by concurrent processes on general-purpose computers, which tend to vary over time and make measurements unreliable (Lopes et al. 2006; Seob Lee and Yeo 2011). For ALPS, the following simple solution was applied (cf. Medvedev, Sorokin, and Khashchanskiy 1989): By physically connecting one output of the sound card to a reference input on the same card, round-trip latency (RTL) is measured for every window, covering all delays due to analog-to-digital conversion, operating system, and concurrent processes on the computer. In every window, ALPS deducts RTL from the length of the acquisition signal, so that the remaining length consistently corresponds to the time it took the sound to travel from the loudspeaker to the microphone.

For a structural schema of the ALPS code, see Figure 2.

Figure 2. Block diagram of the code used in the Acoustic Localization Positioning System (ALPS). The function main() sets hardware according to user defined configurations (a). It uses the function mainLoop(), which sends values via OSC and handles buffer blocking (b). This, in turn, uses the function compute(), which computes correlation and compensates for reference delay (c).

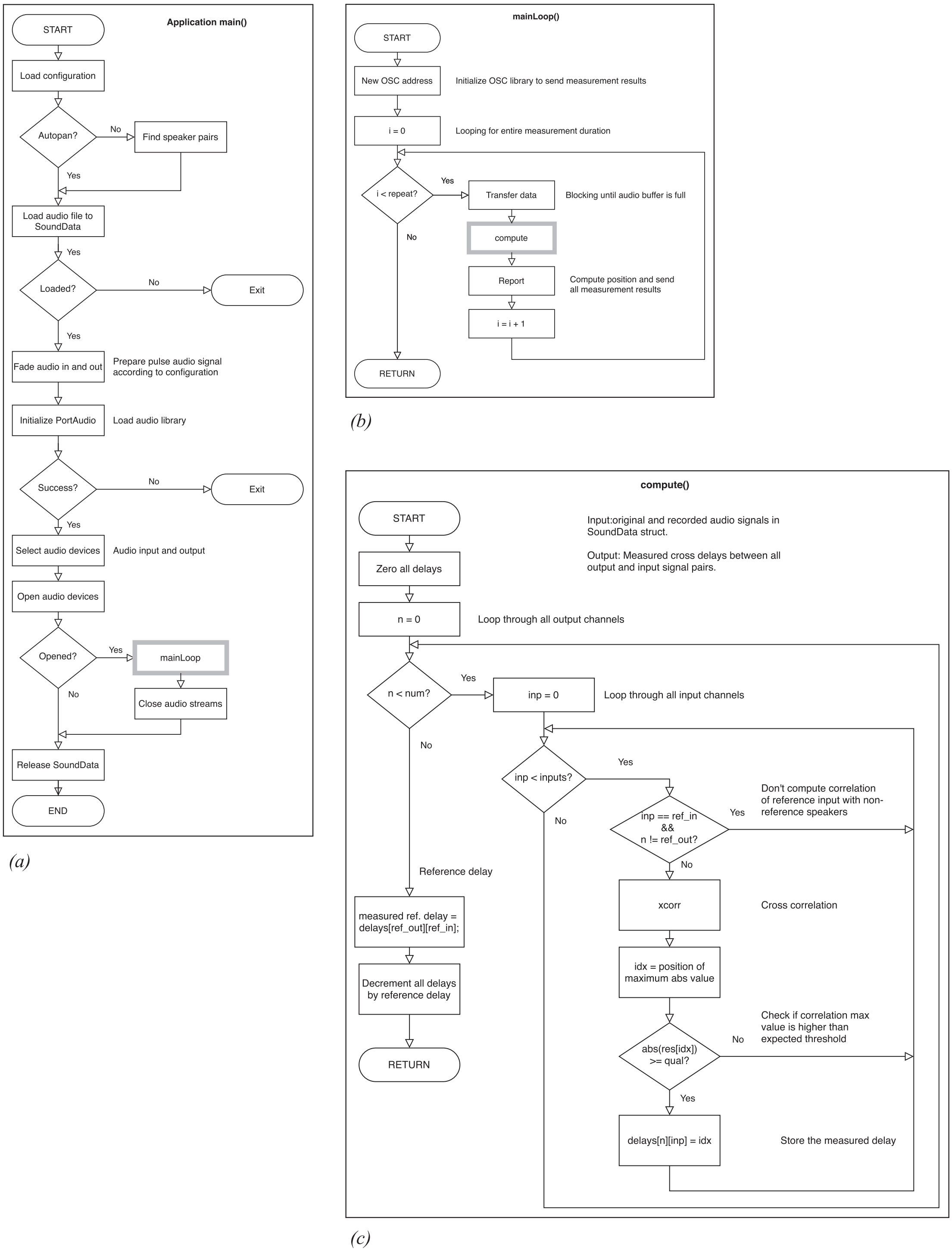
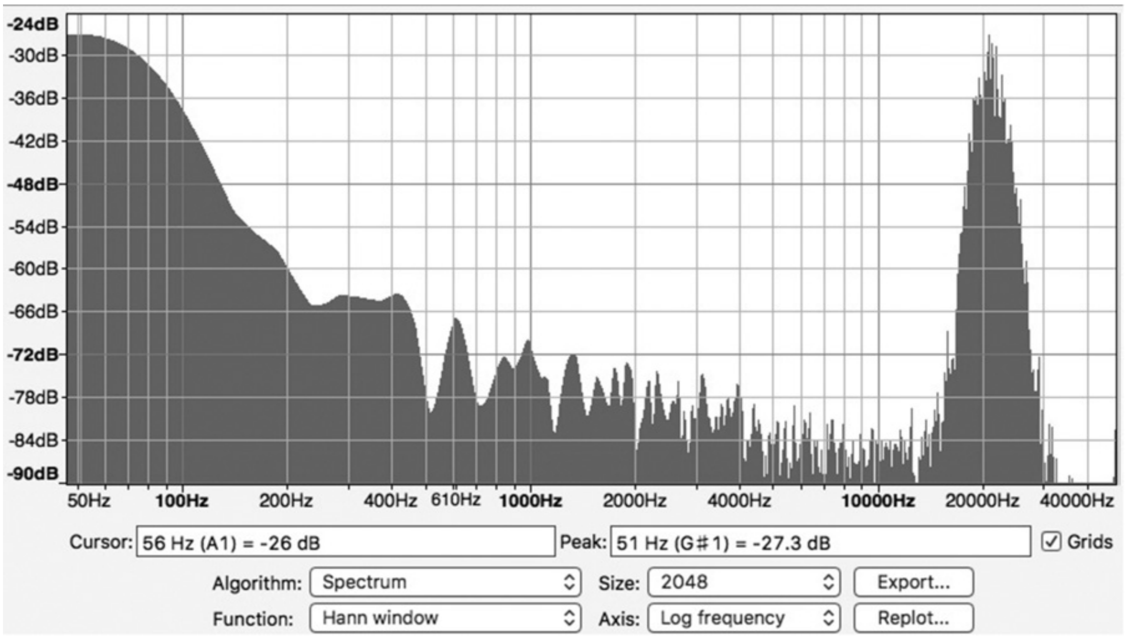


Figure 3. Measurement signal played on Genelec 1029A loudspeakers and recorded on DPA 4061 microphones.



The Measurement Signal

Arguably, the content audio itself could be used as a measuring signal. Yet, for the application in question, the position of the tracked device is also needed when there is no content sound present—for example, in musical pauses (silence). For that situation an inaudible signal is required. Thus, a signal is applied that is above the frequency range audible to the human ear, but still within the range of off-the-shelf loudspeakers. A pulsed measurement signal of random noise between 19 and 30 kHz is used in ALPS. If a loudspeaker's specifications state 20 kHz as an upper limit, this commonly refers to the frequency above which the response rolls off. Essentially, the frequencies in the roll off are still there, they are just quieter. Many off-the-shelf loudspeakers reach 30 kHz even before a significant roll off occurs. But we recommend creating a test signal for each hardware setup. In Figure 3, for example, one can see that there is indeed sufficient detectable signal for Genelec 1029A loudspeakers, recorded on a DPA 4061 microphone. If a loudspeaker's frequency range does not extend to 30 kHz, or the roll-off above the nominal upper limit is too steep, a narrower band could be chosen, fort example, 19 to 21 kHz, or a lower band, for example, 16 to 19 kHz. This might result in audible noise for some listeners, particularly younger ones. That the signal at higher frequencies has less energy does not constitute a problem, above the audible frequency range its power can be increased at will as long as the amplification does not cause distortion. For the experiments here, a white-noise signal was generated, with high- and low-pass filtering applied repeatedly and normalization in between. The resulting file was saved as a single multichannel audio file. The number of channels has to correspond to the number of loudspeakers in the system, and the file needs to be of sufficient length to cover at least a complete cycle across them. The pulse lengths and cycle duration are set in a configuration file.

In most cases, the same loudspeaker can be used both for localization and for content audio, as the measurement signal lies distinctly above the frequency band of content audio. To improve the signal-to-noise ratio, the content signal can also be "cleansed" in the band in question by using low-pass filters at a corresponding cutoff, for instance, 19 kHz. For the PoC shown by Schlienger (2016a), the measurement signals were sent to all loudspeakers at the same time, so that the correlation could be measured from a single capture block. This resulted in an abysmal signal-to-noise ratio: For every calculation, all other signals constituted noise. By pulsing the signal, that is, by taking turns for each loudspeaker, this is avoided here. The reversed approach is less problematic: When calculating the distance between multiple microphones and a single loudspeaker, only one measurement signal is necessary—in the same way that in everyday verbal conversations additional listeners do not make noise, only additional speakers do. For many applications this single-loudspeaker approach is therefore preferable. Unfortunately, the primary design here is for a multiple-loudspeaker situation.

Although Lopes et al. (2006) show that LOS issues are negligible, the effect of sound diffracting around objects is limited: The wavelengths at frequencies above 20 kHz will be shorter than 17 mm, and consequently reflect from larger objects. Still, assuming plane-wave propagation, the occluded wave is still available next to the obstacle: It will still be detected by an omnidirectional microphone, only delayed slightly with respect to the point source assumed for distance estimation. This additional error causes an inaccuracy in measurement, not a loss of signal. But this only applies to situations in which a plane-wave model is applicable. Zhang et al. (2017) presume the complete loss of LOS with every introduced obstacle. This might be sensible for their scenario, where the sound sources are small loudspeakers in mobile phones, because even relatively small obstacles create a near-field situation.

To play and record a random noise signal between 20 and 30 kHz, a sampling rate of at least twice the highest required frequency needs to be applied (Nyquist rate must be twice the highest required frequency to satisfy the Nyquist sampling criterion). So for a 30-kHz signal, the 41.1- or 48-kHz sampling rates commonly used in audio applications will not be sufficient. Owing to hardware restrictions in the initial phase of the project, all tests were run using a sampling rate of 96 kHz. In hindsight, 88.2 kHz could have improved processor performance and thus reduced computing time, at a minor loss in precision and update rate.

The ALPS Software: Audio1 and al-Qt

We implemented ALPS as a stand-alone application in C++ based on the PoC presented in an earlier publication (Schlienger 2016a). It was developed in cooperation with the participants of the workshop on Music, Space, and Interaction (MSI), and coded for this project in collaboration with the authors. There are two versions available. Audio1 forms the basis of most of the discussion here, unless indicated otherwise. It was tested under macOS 10.12 on a single-processor 2011 MacBook Air. For greater detail than what was shown in Figure 2, the reader is encouraged to examine the source code, available as a GitHub repository (https://github.com/spatmus/alps). The second version, al-Qt, which relies on a multiprocessor architecture, was tested for macOS 11.6 and provides the necessary update rates and low latency required for quasi-real-time applications. A precompiled binary of al-Qt tested for macOS 10.12 and macOS 11.6 is available from https://github.com/spatmus/alps/releases.

In ALPS, we use ubiquitous technology in the form of commercially available audio loudspeakers and audio microphones designed for frequencies within the human hearing range, that is, for speech and musical content. The processing power of a fairly recent laptop or desktop computer is adequate. This makes ALPS a straightforward choice for tracking or positioning in situations where loudspeakers and microphones are readily available, for example, surround-sound systems, virtual-reality applications, conferencing, live sound, and home theater.

Settings and Configurations

The settings for both instantiations of the ALPS software can be controlled via a configuration file (see config.ini in the GitHub repository for a generic example). In the al-Qt version, the configuration file can also be accessed via the dialog panel. The choice of adjustable parameters is intended to help in debugging and experimentation for application-specific performance. As a stand-alone application for 3-D tracking, ALPS uses trigonometrical Euclidean distance calculations for possible pairs of delayed signals. The positions of all loudspeakers need to be given as $x/y/z$ Cartesian coordinates in the positive quadrant. The three-dimensional position estimates are, however, not essential for use of ALPS in conjunction with the autopanning provided by the Autopan Max patch; the distance readings alone are sufficient.

To run the ALPS software with the Autopan Max patch, autopan needs to be enabled. This sends the distance readings directly via the User Datagram Protocol (UDP) and Open Sound Control (OSC) to a network port in Max, whose IP address and port number are also set here. In Max, the messages from ALPS are received via a `udpreceive` object. ALPS and Max can run on the same processor or over a network using OSC. Further settings are for the time offsets as well as for lengths and fade times of the pulses. The overall duration of the sequence can be set to compensate for the total latency of all processes involved and how many times the sequence will be repeated. In debug mode, the last repeat of the pulses, as well as the last instances of the signals on the selected inputs, are recorded as an audio file in the subdirectory set here. Generally, the challenge is to find the right settings for a balance between update rate, area covered, and latency for a particular situation: If the tracked device is ten meters away, around 0.03 seconds are needed before computation, but if it is only one meter, 0.003 seconds meet the requirements. The setting for maximum distance can limit erroneous outliers by setting it to a value smaller than the longest possible distance in the room. But it can also be set to a value arrived at by trial and error to mark the distance above which readings are no longer reliable, even if these values lie well within the room dimensions, for example, if the room is large. The quality setting allows one to filter readings that have a small amplitude and are hence more likely to be wrong. This effectively deals with early reflections, whose amplitudes are unlikely to be equal to or higher than the direct signal.

Autopan Max Patch

The Autopan Max patch utilizes the ALPS software to automatically produce the panning of a sound source in a multiple-loudspeaker setup by tracking a moving object that represents the position of that sound source. This has the advantage that virtual sound sources can be turned into spatially discrete sound objects that can change position, just as most acoustic instruments can. Examples of virtual sound sources are mobile devices, such as laptops and tablet computers, as well as electric and electronic instruments, but also many novel musical interfaces. Autopanning distinctly improves immersion in spatially interactive sonic arts, especially when virtual sound sources are played live along with acoustic instruments, as will be discussed in the Tests in Situated Use: Qualitative Evaluation section.

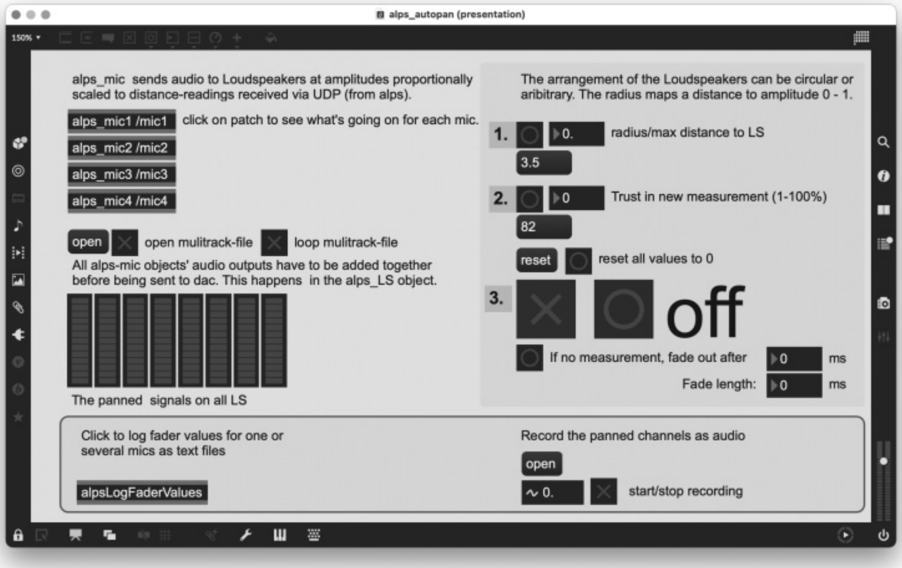
The Autopan Max patch and associated subpatches (tested for Max 7 and Max 8), as well as a standalone Max application that should work "out of the box" on all platforms, are available from https://doi.org/10.5281/zenodo.5607121.

The patch's main interface (see Figure 4) allows one to record the panned audio files and to save a log of the calculated trajectory in the form of error-corrected distances. A separate alps-player patch can recreate the distance readings. An additional alps-recorder patch is provided to write the raw data as received on the UDP port to a log file.

The smoother subpatch provides rudimentary filtering and error handling. When a distance reading fails to be updated from ALPS, the content audio starts to fade out on the loudspeaker in question (controlled via fade time and fade length). Once a new reading is available, the audio fades in again. This reduces responsiveness slightly but provides a high-pass filter for outliers during periods between signal updates. The filter is adjustable, in degrees of confidence in newly acquired values, by adding a user-defined percentage of the old measurement to the newest one.

The alps-mic subpatches (see Figure 5) show the smoothed levels of the distance readings in meters, measured for each loudspeaker. Further, the panned

Figure 4. The main window of the Autopan Max patch has three sections: The panel on the left provides access to the subpatchers of all tracked devices (microphones); selection of content audio in multitrack format and visual monitoring of levels on all loudspeakers. In the bottom panel, fader values can be logged as text files and the panned content audio recorded as a mutlitrack file. The panel on the right provides 1. Controls for loudspeaker layout, 2. Filter adjustment, and 3. A panic button (ON/OFF) and fade-out times for content audio in the absence of measurements.



content-audio source can be set here. The options are

1. on/off;
2. player, for a mono audio file which can be chosen via the open button;
3. audio in, to connect to the input of a connected sound card; and
4. mutrach, to choose one channel of a multitrack file selected from the autopan patch main window.

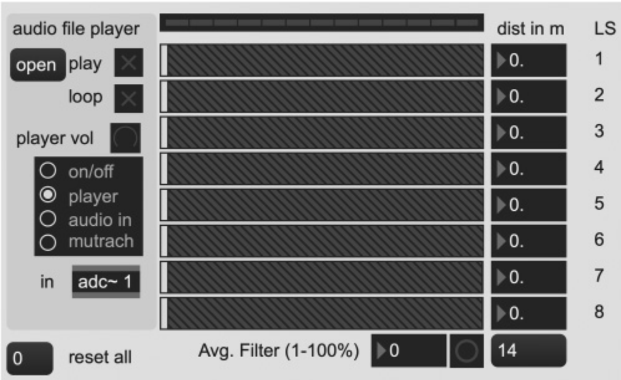
The reference microphone can also be monitored here (according to configuration setting for refIn in the ALPS software) and then, for example, routed to compensate for jitter.

Distance-Based Amplitude Panning

In an ideal world, every sound source to be reproduced could be represented by a real sound source in its place. In a multiple-loudspeaker setup this can be approximated by representing it by the nearest loudspeaker. If the ideal position lies between loudspeakers, it can be approximated on multiple loudspeakers according to amplitude panning principles. This is a simplified description of the DBAP algorithm by Lossius, Baltazar, and de la Hogue (2009). It lends itself to the use with ALPS, because the distance measurements are inversely proportional to the required amplitudes. In that sense, panning succeeds even if only one distance measurement is available, as usually happens when the distance between the tracked device and a loudspeaker is small. In that case, a single-loudspeaker representation is adequate. Presuming that rolloff $R = 6$ dB (in decibels per doubling of distance) equals the inverse-distance law for sound propagating in a free field, a direct mapping of distance measurements estimated by ALPS with the amplitude of the panned audio signal becomes possible.

The ALPS software can also calculate absolute positions in a Cartesian system. Although this is the sensible thing to do for most general positioning and tracking tasks, it is not necessary for audio panning with DBAP, as outlined in the The ALPS Algorithm section. This is in contrast to angle-based panning paradigms, such as vector-base amplitude panning (VBAP, cf. Pulkki 1997) or Ambisonics (Malham

Figure 5. The alps-mic subpatch provides visual monitoring of the measurement signals on each loudspeaker as "distance in meters." In the "audio file player" section the content audio to be panned can be selected: mono files, audio input, or the specific multitrack channel to be panned (of the multitrack file selected in the autopan main window); and individual adjustments of the averaging filter.



1998): To arrive at the required panning positions in Cartesian or polar coordinates, three known points are needed for trilateration in 2-D and four points in 3-D. With DBAP, the absence of a fourth, third, or even second measurement makes the situation less than ideal, but not undefined. Consequently, less hardware (i.e., fewer loudspeakers) is necessary when using DBAP. This is why the seemingly orthogonal problems of acoustic panning and acoustic localization align well here and inform each other in parallel.

Methods

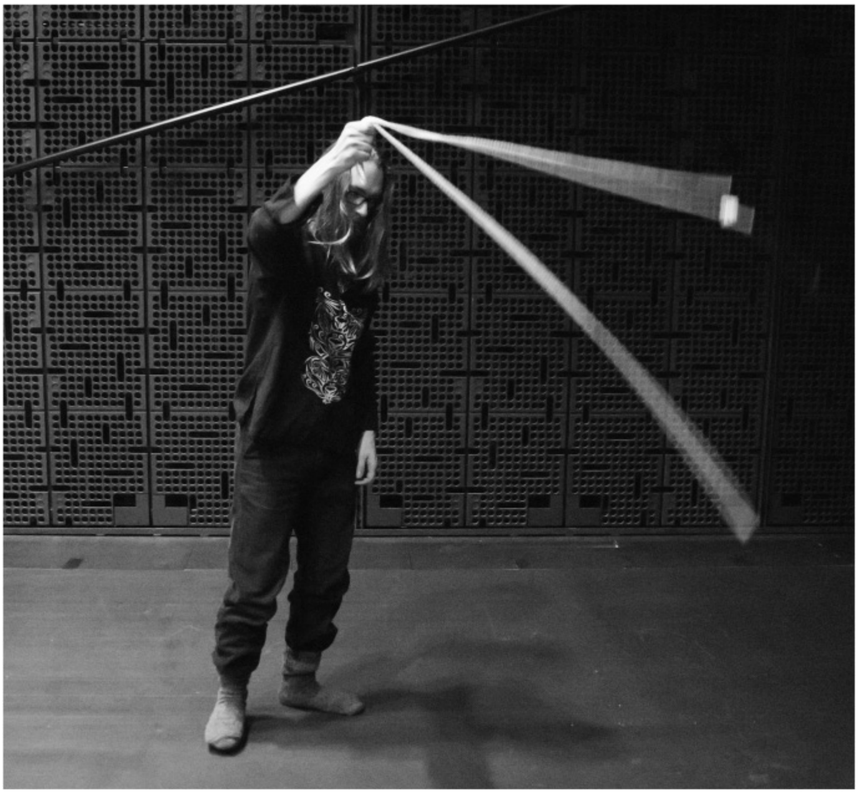
The development method of the ALPS project is an extension of participatory design principles though the addition of interdisciplinary improvisation (setup A). Besides this artistic-research approach, quantitative methods are also applied (setups B1 and B2), to gauge the extent to which the technology meets the requirements defined and evaluated in its artistic use.

Free Improvisation, Artistic and Qualitative Research

Interdisciplinary improvisation is an experimental practice bringing together practitioners of various disciplines to seek common ground, reduce significant differences, and identify challenges (Andean 2014). The Research Group on Interdisciplinary Improvisation was launched in 2012 at the University of the Arts Helsinki. In collaboration with some of its members, the MSI workshop was brought to life in 2013, with the idea of applying the practice of interdisciplinary improvisation to technology development, specifically in the field of spatially interactive sonic arts (Schlienger and Olarte 2016). It combines concepts of interdisciplinary improvisation with participatory design, whereby all participants in a development project are involved at all levels (Simonsen and Robertson 2013). Interdisciplinary improvisation, as applied in MSI, helps to find unexpected, simple, and sustainable solutions by prototyping a situation in a problem area, rather than finding the solution to a problem. To use a simple example, the situation "everything you find in the room is an instrument" is given as a score for the improvisation. This means participants move about to explore the room and its objects, inventing "instruments" through improvisation. This method could be applied to other fields. For example, as a participating town planner suggested, imagine a new housing development, where footpaths between the buildings need to be planned. One approach would be to do nothing initially but survey what paths the inhabitants choose if left to roam free. If this is not possible, the situation could be enacted as improvised theater, as in "let's pretend here is the bus stop, here the shopping center, and here the motorway . . ." The difference is that in art, which seems to be already situated on some metalevel with respect to reality, the element of role play is not necessary. There is no point of "let's pretend to play the violin" if one has a violin. Artistic situations intrinsically constitute this "let's pretend" level, making improvisation as a development method for technology in the arts even more felicitous. Figures 6 and 7 show still images of improvisations in MSI.

This approach represents a counter model, an antidote to conventional "consumer evaluations," in which a handful of test subjects test a product for 20 minutes, after which they are asked to "tick boxes" to select choices from answers to leading questions. Rather, the idea of MSI was to form new ideas for development over time,

Figure 6. Participant exploring the space–sound relations of a found instrument in an improvisation in the workshop on Music, Space, and Interaction (MSI). Photo by author.



developing practices and techniques along with the technology. Typically, participants were musicians and composers, scenographers, multimedia artists, sound designers, dancers, painters, town planners, choreographers, and so on. The workshop was open to students and professionals, the latter making up around 30 percent of the participants. The workshop underpinned the ALPS project from the beginning, as detailed in earlier publications (Schlienger 2016a,b,c).

Experimental Setups

Experiments were conducted in two settings: In setup A, a medium-size performance area was used for qualitative experiments oriented towards artistic research, whereas experiments in setup B were of a quantitative nature. Setup B was further subdivided into two variants, B1 and B2, to compare between the single-processor implementation Audio1 in setup B1 and the multiprocessor implementation al-Qt in setup B2.

Setup A: Stage-Size Performance Area

This is the setup for experiments with the ALPS software and the Autopan Max patch in the MSI workshop. Audio1 and the Max patches were run on two midrange MacBooks communicating over WLAN, one MacBook from 2011 and the other from 2012. For tracking we used omnidirectional AKG CK 55L microphones, which meet the requirements for AL in the frequency bandwidths above 20 kHz, despite the moderate price in comparison to the DPA 4061 microphones we used to record the audio content. Both types are small, unobtrusive lavalier microphones. The sound card, a MOTU

Figure 7. A typical scene in an Interdisciplinary Improvisation session in the MSI workshop. Photo by author.

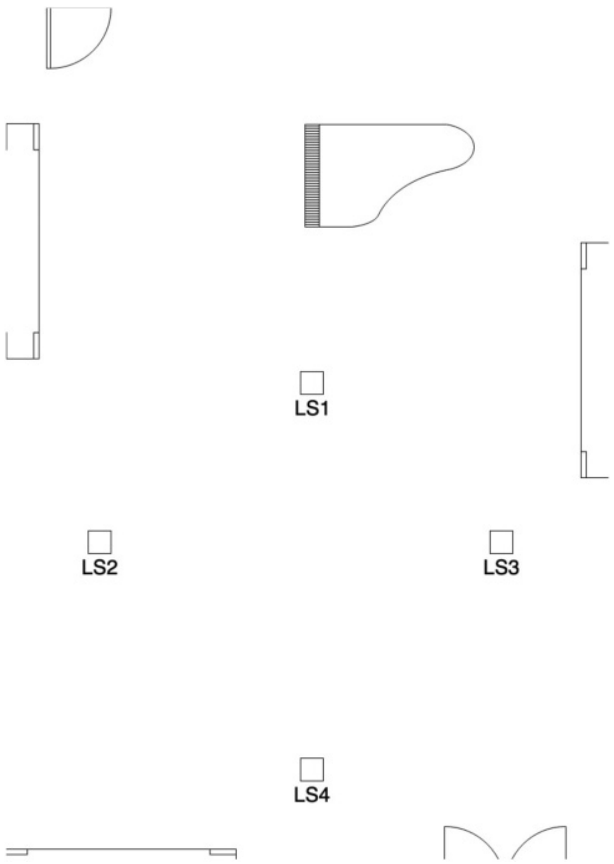
16A AVB, was chosen for its extremely low system latency (less than 2 msec) and flexible input/output options. The AKG microphones were connected to wireless senders and receivers that work in the analog radio-frequency band, with typical latencies in the microsecond range. For loudspeakers we used Genelec 1029As, which deliver sufficient volumes above 20 kHz, as was shown in Figure 3.

Setup A was located in a performance space approximately 10 × 16 × 3 m in size, treated with acoustic panels for classical music. First and early reflections were minimized, but the room was by no means dry. Four loudspeakers were spaced evenly around the center of the room at a radius of 3 m, on stands approximately 1.1 m high, making elevation negligible in relation to most acoustic instruments' positions in the room. Four to seven performers moved around freely with small acoustic sound sources, microphones, or electronic musical instruments lacking acoustic output. The electronic instruments' sounds were distributed via a second set of four loudspeakers positioned on top of the set of four dedicated to the tracking task. (For a schematic layout of the room, see Figure 8.)

The sessions in MSI usually took the form of exercises in free improvisation that were 10 to 40 minutes long, following a very loose score open to the participants' interpretation. For example, a group exercise might have the instructions "start in the middle of the room, spread out, but get quieter while you do so and listen to the others." After each exercise, the experience was discussed in the group. Based on these discussions, the techniques or technologies applied in the exercise were modified as needed, either immediately (adjusting playing techniques, "tweaking"), or between workshops (rebuilding instruments, recoding, developing).

For documentation, the audio content panned by the ALPS Autopan Max patch was recorded as an audio file. A selection of four-channel recordings can be downloaded from https://doi.org/10.5281/zenodo.5607027.

Figure 8. The room layout for setup A, a performance space of approximately 10 × 16 × 4 m, where we conducted the qualitative experiments. LS1–LS4 mark both the loudspeaker positions for measurements and those for content audio.



Setup B1: Performance Area Section

In setup B1, tests were run for situations having four microphones and four loudspeakers as well as having four microphones and eight loudspeakers, by running all equivalent computational processes but only actually distributing the measurement signals via three loudspeakers, and recording them on one microphone only. This way, a section of a larger performance area could be observed in detail, without having to completely set it up. The room is a living room, approximately 6 × 3 m in size, with laminate flooring, concrete walls, paneled ceiling, and some soft furnishings. The the test area covered 2 × 3 m of this room. The loudspeakers were placed on the floor, simplifying the setup by taking elevation out of the equation. A 0.5-m grid was marked on the floor and a video camera installed on the ceiling, which was approximately 2.5 m high. A setup having eight loudspeakers with the same spacing could cover an area of 6 × 3 m or 2 × 9 m. These were smaller distances than in setup A, which covered an area of 6 × 6 m with only four loudspeakers and accordingly increased the quality of the measurements. Further, the movement of the device to be tracked was automated by attaching it to a mechanical motorized vehicle. An electric toy locomotive (see Figure 9) was used for this, fitted with an AKG CK 55 L omnidirectional microphone. The track followed the line $y = 1$, with LS1 at $x = 0, y = 2$; LS2 at $x = 2, y = 3$; and LS3 at $x = 3, y = 0$. Elevation was negligible, as the loudspeakers were on the same level as the tracked device.

The distance measurements were recorded at a known rate in a log file for further analysis in MATLAB. Accuracy can be evaluated by comparing these measurements with predicted values based on the assumption that the electric locomotive runs at a constant speed on a straight track. An animation of the movement based on the measured data, in which the distance measurements are expressed as radii of circles with origin at the center of each loudspeaker, can be directly compared to a video recording of the experiment. Videos, data files, and the MATLAB scripts are available from https://doi.org/10.5281/zenodo.5607528. Figure 10 shows the merged stills of the animation and the video at 00h:00m:31s.

Setup B2: Performance Area Section

In a setup nearly identical to B1, B2 demonstrated the advantages of the multiprocessor architecture of the current implementation al-Qt over the single processor version Audio1 by testing it on a faster moving object (see Figure 11). A toy car and track of type Hot Wheels was fitted with an Audio Technica ATM350 cardioid microphone (see Figure 12). The car was accelerated by hand, so it decelerated during its journey. The experiments were recorded on video, available with the corresponding data from https://zenodo.org/record/5604446. The difference between setups B1 and B2 is that four physical

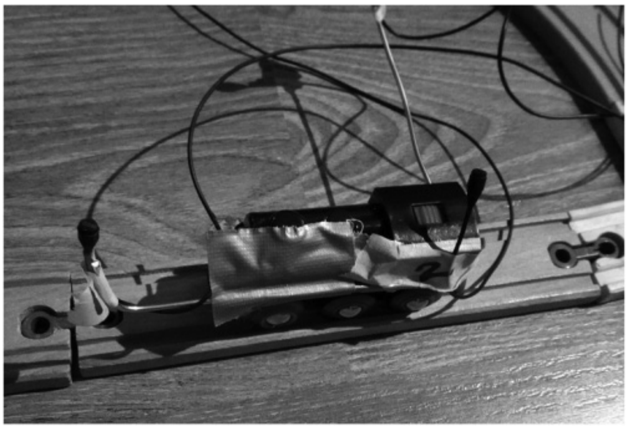


Figure 9. The "device of constant speed" in the picture, had three omnidirectional microphones attached, which in principle would also allow for 3-D tracking on one loudspeaker (multiple-microphone approach). For the experiments in setup B1, only recordings of one microphone were necessary (single-microphone, multiple-speaker approach). Photo by author.

loudspeakers were set up, rather than three as in B1. Furthermore, B2 is in a slightly less reverberant room with a slanted wood-panel ceiling and wooden floor boards, resulting in fewer early reflections.

In this setup, al-Qt ran on a MacBookAir (M1 2020), under macOS 11.5.2.

Results and Discussion

We now discuss the results of the experiments. First we analyze quantitative results and compare them to expected values from earlier works for each experimental setup. This is followed by a discussion of implications for low-latency applications. Finally, we look at qualitative results providing an evaluation of the implementation in the context of artistic practice.

Experiments in Controlled Environment: Numerical Evaluation

The data discussed in the following are available from https://zenodo.org/record/5604446. The experiments were recorded on video, available from the same URL. From the experiments conducted in setup B1 using the single-processor Audio1 version of ALPS, two sets of data are discussed, both sets generated by the Brio electric toy locomotive. One data set was recorded while processing audio for four loudspeakers, the other for eight loudspeakers. To ensure that the window length was sufficient for the whole range of loudspeakers, the signals were pulsed as if all loudspeakers were present, for example, LS1 received the first pulse, and LS8 the eighth. In B2, which looks at the improvements in the update rate and the latency of the al-Qt version of ALPS, two data sets are discussed. The first was recorded with four loudspeakers and a single microphone, the second set (marked "rerun" in the data set) with four loudspeakers and four microphones.

Experiments in B1

As the comparison between root mean square error (RMSE), mean average error (MAE), and median absolute deviation (MAD) in Table 2 shows, the system can be considerably improved by eradicating outliers, which are proportionally over-represented in RMSE, but have less weight in MAD. The difference between the eight- and four-loudspeaker sets, particularly the measurements for LS7 in the eight-loudspeaker set, is notable. Admittedly, it highlights a flaw in how the data were recorded: A measurement that was not updated at the end of a measured interval repeats the previous (and hence incorrect) reading until a new one is available. This adds error: In the eight-loudspeaker data set (https://doi.org/10.5281/zenodo.5607528), it can be seen that for nearly a third of the duration, no signal was recorded on LS7, repeatedly registering a distance of 0 m. This misrepresents the system's capability. A caveat here to future adopters of the current implementation: Better results would be achieved by filtering all distances that exceed a maximum distance setting of 2.5 m and by ignoring distances of 0 m distance, treating them as, for instance, the floating-point value NaN (not a number). In fact, looking at sequences of measurements in the data file, it is evident that where a signal does exist, there is little difference in error between the two sets. Consequently, performance could be further improved by extrapolating missing readings from direction and speed. The latency visible in Table 2 affects update rate only; it has no influence on the accuracy of the measurements (42.5 msec for four

Figure 10. Graph superimposed on video still of experimental setup B1, showing the tracked device at Y = 1 m, X = 2 m. Loudspeaker icons indicate Cartesian positions (x, y) on a plane with the tracked device (Z = 0) of the Genelec 1029A loudspeakers at (0,2) for LS1; (3,2) for LS 3 (LS7 in dataset 2); (3,0) for LS4 (LS8 in dataset 2). In the animation video, the distance reading for LS1 is shown in white; that for LS3 (LS7) is dashed white; and LS4 (LS8) is in black. Photo by author.

Figure 11. The layout of experimental setup B2 shows a toy car track made up of sections 0.3 m in length in a grid with markers every 0.5 m. Photo by author.

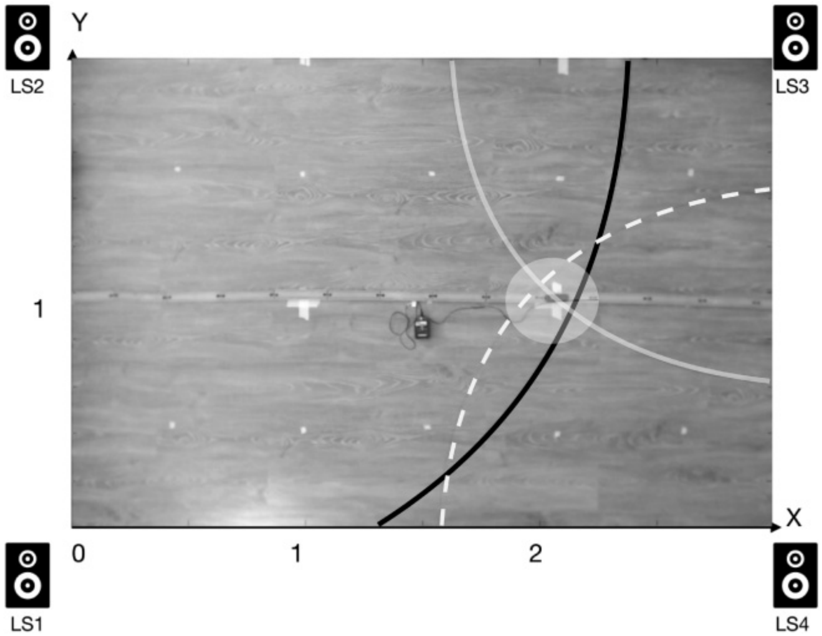


Figure 10.

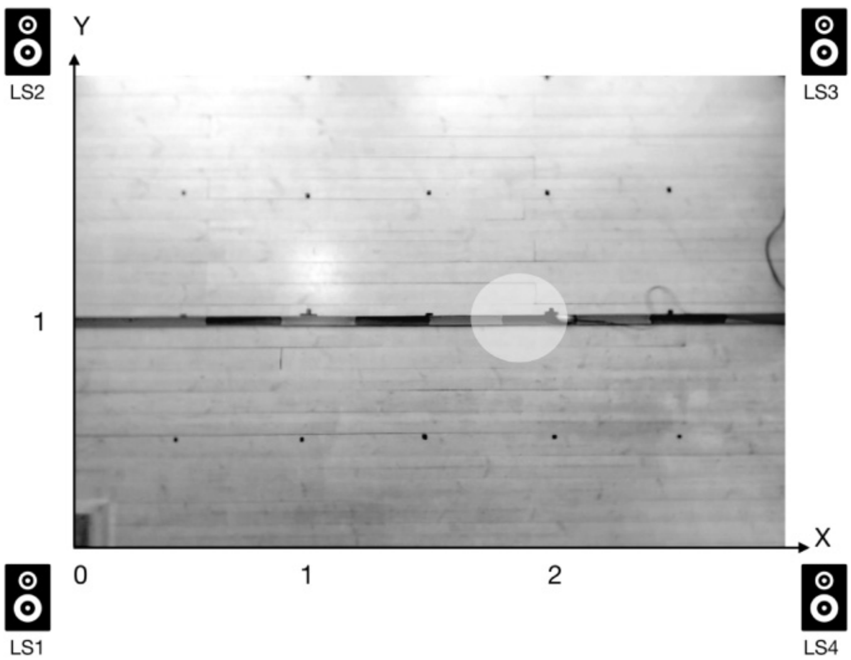


Figure 11.

Figure 12. The "faster moving device," a toy car with a microphone taped to it, was launched by hand and travelled at approximately 1.5 m/sec. Photo by author.

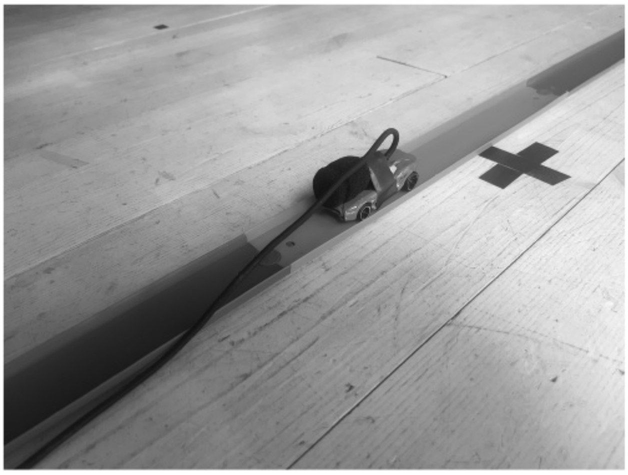


Table 2. Summary of Results from Setup B1

	MAD	*MAE*	*RMSE*	*Frequency (Hz)*	*Latency (msec)*
B1 LS1	2.8	9.7	10.7	4	250
B1 LS2	6.4	15.6	37.1	4	250
B1 LS4	2.5	18.5	56.9	4	250
Mean	3.9	14.6	34.9	4	250
B1 LS1	9.0	11.7	23.6	2	500
B1 LS7	48.7	110.3	163.0	2	500
B1 LS8	19.8	91.1	150.7	2	500
Mean	25.9	71.0	112.5	2	500

Results using ALPS Audio1. LS: loudspeaker number; MAD: mean absolute deviation (this and the next two columns in cm); MAE: mean absolute error; RMSE: root mean square error.

loudspeakers and 85 msec for eight loudspeakers, in relation to the actual computation).

Comparing the data from the experiments in B1 with the requirements set out in an earlier paper (Schlienger 2016c) indicates that "continuous update rate" can only be achieved at walking speed (approx. 1.5 m/sec). Faster speeds will result in a perceptible lag, as the measured update rate for four loudspeakers is 4 Hz, whereas for eight it is only 2 Hz. Running ALPS Audio1 in debug mode showed that the processor we used was not able to compute sufficiently rapidly to provide measurements in anything approaching real time. Yet the required accuracy of $\leq$0.3 m is met (see Table 1), as is the covered area, due to the scalability of ALPS. Round-trip latency (250 msec), on the other hand, is ten times higher than the requirement of keeping latency under 25 msec.

Experiments in B2

In setup B2, using the newer, al-Qt version of ALPS, no error analysis was undertaken, nor were error corrections applied to the data, as accuracy was less of a focus than latency and update rates. The results are summarized in Table 3. The higher update rates and lower RTL of al-Qt propose some interesting comparisons: Although still 6.6 times slower than world-class sprinters, who clock 100 meters in under ten seconds, the HotWheels-type car covered 1.5 m/sec and ALPS managed to catch between 13 and 17.5 measurements in that timeframe. This places it well within the requirements for update rates (cf. Schlienger 2016c). Latency between 57.1 and 77.5 msec is also four to five times less than for Audio1 in setup B1. For larger-scale applications, such as autopanning, these results improve on values that were already acceptable at a latency of 250 msec and update rate of 4 Hz. It might be worth noting that, from the listener's point of view, it is difficult to define a numerical limit as to what is an acceptable maximum latency. A minimum value could be derived from the speed of sound in air, which makes latency dependent on the observer's position relative to the sound source. Therefore, if the observer is 12 m away, a latency of 35 msec has to be expected. Yet, through the quality of the performers' actions—the "gesturality"—a link can arguably be established between a performative sound and its perception as such by the listener. And even then, perhaps thanks to a sense of syncresis (Chion, Gorbman, and Murch 1994), a causal link may be experienced beyond it being physically possible. Hence, we contend that the less-contestable measure of maximum latency should apply here: the one experienced by performers. If they manage to perform a causal relationship at a particular latency, this can be perceived as such by an observer.

Table 3. **Summary of Results from Setup B2**

Experiment	*Duration (sec)*	*Sum*	*NaNs*	*RTL (msec)*	*Systemic (msec)*	*Latency (msec)*	*Frequency (Hz)*
Experiments conducted with four loudspeakers and one microphone:							
B2.1	2.3	60	60	35.0	42.5	77.5	13
B2.2	2.0	46	58	35.0	42.5	77.5	13
B2.4	2.7	52	84	35.0	42.5	77.5	13
B2.5	3.5	69	111	35.0	42.5	77.5	13
Experiments conducted with four loudspeakers and four microphones:							
B2 rerun 1	2.7			14.6	42.5	57.1	17.5
B2 rerun 3	2.1			18.25	42.5	60.8	16.4
B2 rerun 5	2.2			21.0	42.5	63.5	15.7
B2 rerun 6	2.1			19.5	42.5	62.0	16.0

Results using ALPS al-Qt. Duration: duration of experiment in seconds; Sum: sum of measures taken on all loudspeakers during the experiment; NaNs: number of empty readings; RTL: round-trip latency; Systemic: latency due to configuration settings; Latency: RTL and systemic latency combined; Frequency: update rate.

The high quantity of invalid values in the data can partially be attributed not only to the cardioid directivity pattern of the microphone used but also to the decreased reliability of measurements at larger distances. By comparison, in the beginning the loudspeakers located behind the microphone are still within reach; but soon after the car is in motion, only the front-facing loudspeakers are measured.

Owing to an oversight, the experiments with the HotWheels car were initially run with only one microphone. This means that a direct comparison with setup B1, which was run with four loudspeakers and four microphones (plus reference microphone), is problematic. Therefore, it was decided to redo the experiments to clarify what influence the number of inputs has on latency. Puzzlingly, the latency was almost halved in the reruns (see Table 3 for values). It could not be conclusively determined whether this had to do with the way the macOS audio driver handles buffer sizes or with al-Qt itself.

The multiprocessor al-Qt version seemed to make good use of the highly optimized ARM64 system-on-a-chip architecture of the 2021 MacBook Air, so even when running Max on the same processor as al-Qt, none of the eight displayed cores showed significant use. What is more, there is a conspicuous incongruence in the fact that al-Qt performs worse for the simpler task of recording one microphone, but better for recording four. This might be anecdotal evidence that hardware limitations per se can be excluded as a cause.

Implication for Low-Latency Applications

The theoretical possibilities of AL indicate that gesture tracking for gestural control of musical instruments is possible. The idea at the beginning of this project was for a performance-stage-sized application, in which the position of participants and their trajectories through a room would be tracked with the accuracy and update rates given in Table 1. Be that as it may, the research was always also about gestural control of musical instruments, as one of its primary aims was to develop a generally applicable tracking tool for sonic arts. On examination of comparable examples in the literature, al-Qt provides fast and accurate measurements at higher update rates than most. Yet we seem to have reached an impasse: Aside from the fact that the approach of using a pulsed signal, stepping through pairs of inputs and outputs consecutively, limits the upwards scaleability of the principle (the measurement process takes longer with every additional loudspeaker or microphone), we observed that measurement cycles lower than 42.5 msec do not lead to lower latency or faster update rates. The data suggests that although RTL decreases, measurements are either repeated or left out (compare with data set SetUp_C available from https://doi.org/10.5281/zenodo.5607528). When decreasing duration below 42.5 msec, the de facto update rate does not decrease and remains the same.

Not taking into consideration the systemic latencies in Table 3, since they are significantly lower in applications for gestural control, RTLs between 14.6 and 35 msec at update rates of 15 to 18 Hz are an improvement over previous efforts in the PoC described elsewhere (Schlienger 2016b). These values are within the more generous recommendations quoted in Jack, Stockman, and McPherson (2016). But, as tantalizingly close as this is to the values set by the benchmark, other researchers are encouraged to explore this further: Yes, there are indications that gestural control with latency below 20 msec and update rates over 20 Hz should be possible using AL, but verifying this is unfortunately beyond the scope of this article. Monitoring the CPU activity on the MacBook Air shows that plenty of processing power remains unused, so the issue may well lie elsewhere.

Tests in Situated Use: Qualitative Evaluation

In participatory design, evaluation and development happens concurrently, in which case evaluation by "test subjects" who are not participants in the design project are neither appropriate nor necessary—nor, by this token, is a quantification of qualitative results, which questionnaires applying numerical scaling would arguably constitute! Based on extensive field notes and discussions that form an essential part of the workshop's practice, the following text summarizes the workshop participants' experiences with the Autopan Max patch.

On the principle that situations rather than solutions should be prototyped, countless sessions of the workshop were dedicated to the scenario of moving sound sources before we progressed to experiments using the Autopan Max patch. This allowed us to observe how the practice changed with the introduction of the technology. We noticed marked differences. It was striking how nonacoustic virtual sound sources suddenly blended in among the other acoustic moving sound sources. If, in improvisations without the system, nonacoustic instruments provided a nondiegetic background or background texture for what was happening with the acoustic sources, now they were perceived as equivalent to the acoustic sources, as musical spatial actants.

In free improvisation with amplified musical instruments, it is fairly common that musicians with unamplified, quieter acoustic instruments feel that they are not heard. For example, in several early sessions of MSI, electric guitars and electronic instruments like laptops were experienced as overpowering by violin players, flute players, or singers. In discussions, the guitar was described as "amplified," a term predominantly used with a negative connotation, for example, a participant felt that the "amplified sound drowned my sound completely" or others thought the "electronic sound is too loud." Yet, in practically identical setups after the introduction of the Autopan system, which localized the amplified instruments, this term was not used in this sense in any of the discussions. From these and similar reactions to experiments with the system, it became evident that the increased localization of the panned sound increased transparency in the overall sound and created more room for the other participants.

The use of the system also allowed for the creation of multiple localized areas of different reverberant acoustics within the same performance space. In earlier sessions, before introducing the system, whenever any amplified instrument applied some form of artificial reverberation, the whole performance space was immediately immersed in that uniform artificial space, which overrode the physical space's acoustic characteristics.

This was problematic for the specifically spatial interactive practice, as the artificial spatiality was forced onto all participants, leaving no choices. For example, participants with unamplified instruments could not create reverberant rooms of their own (if the artificial reverberation was set to "cathedral," everybody in the room had to be "in a cathedral"). In contrast, with the experiments with the Autopan system, when the reverberation was panned according to the instrument's position, the reverberation was experienced as localized. It was perceived as part of the distinct individualized quality of the sound associated with the spatial source for which it was intended.

Not only individualized, localized sound qualities have artistic validity. There are many spatio-acoustic phenomena used in musical contexts that do not rely on the listener being able to locate sound sources, but consist instead of diffuse sounds (Blesser and Salter 2007). But the workshop participants' experience was that localized qualities were hard to achieve in multiple-loudspeaker setups in improvisations before introduction of the ALPS system: It gave sound sources of both nonacoustic and acoustic origin equally spatial interactive roles.

A quadraphonic recording of an improvisation session is available for the sake of completeness (https://doi.org/10.5281/zenodo.5607027). The essence of a participatory event can only be experienced through participation; good free improvisations do not automatically make good recordings, and, as a piece of music, this would benefit from further editing. Nevertheless, it illustrates the panning trajectories.

Conclusions and Future Work

The Autopan Max patch in conjunction with the ALPS software overcomes shortcomings of standard audio technology for moving sound sources in a multiple-loudspeaker environment. It enhances immersion in spatially interactive musical performances. For the artistic idea from which this project started, ALPS and the Autopan Max system provided a relevant technological solution, at low cost, working on widely available equipment. Still, the necessary responsiveness for gestural control of musical instruments could not be completely achieved. In comparison with comparable examples from the literature, the ALPS system provides a competitive alternative and is indeed a step towards a general kinaesthetic interface.

Many optimizations are possible for the current setup—for example, tracking and filtering that are more advanced. But to achieve distinctly lower overall latency at higher update rates, even an embedded implementation would have to envisage alternatives to the data acquisition other than a pulsed signal. As one possible strategy, concurrent measuring of several output/input pairs should be envisaged.

For this project's initial artistic idea, in which accuracy, latency, and update rate must meet the requirements of spatial hearing, the autopanning system with ALPS now provides a viable and tested solution.

Acknowledgments

Many thanks to the participants in the workshop on Music, Space, and Interaction; all friends and colleagues at the Center for Music and Technology of the Sibelius Academy at the University of the Arts Helsinki, Helsinki; and Kone Foundation Finland, that has supported this project with a Researcher's Grant.

References

Aguilera, T., et al. 2017. "Broadband Acoustic Local Positioning System for Mobile Devices with Multiple Access Interference Cancellation." *Measurement* 116:483–494.

Andean, J. 2014. "Research Group in Interdisciplinary Improvisation: Goals, Perspectives, and Practice." In A. Arlander, ed. *This and That: Essays on Live Art and Performance Studies*. University of the Arts Helsinki, pp. 174–191.

Bazoge, N., et al. 2019. "Expressive Potentials of Motion Capture in the Vis Insita Musical Performance." In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pp. 266–271. 10.5281/zenodo.3672954

Blesser, B., and L.-R. Salter. 2007. *Spaces Speak, Are You Listening? Experiencing Aural Architecture*. Cambridge, Massachusetts: MIT Press.

Brena, R. F., et al. 2017. "Evolution of Indoor Positioning Technologies: A Survey." *Journal of Sensors*: Art. 2630413. 10.1155/2017/2630413

Chion, M., C. Gorbman, and W. Murch. 1994. *Audio-Vision: Sound on Screen*. New York: Columbia University Press.

Dean, R. T., and G. Paine. 2012. *Gesture and Morphology in Laptop Music Performance*. Oxford: Oxford University Press.

Dobrian, C., and F. Bevilacqua. 2003. "Gestural Control of Music Using the Vicon 8 Motion Capture System." In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pp. 161–163.

Filonenko, V., C. Cullen, and J. Carswell. 2010. "Investigating Ultrasonic Positioning on Mobile Phones." In *Proceedings of the International Conference on Indoor Positioning and Indoor Navigation*. 10.1109/IPIN.2010.5648235

Harter, A., et al. 2002. "The Anatomy of a Context-Aware Application." *Wireless Networks* 8(2–3): 187–197. 10.1023/A:1013767926256

Hightower, J., and G. Borriello. 2001. "Location Systems for Ubiquitous Computing." *Computer* 34(8):57–66. 10.1109/2.940014

Jack, R. H., T. Stockman, and A. McPherson. 2016. "Effect of Latency on Performer Interaction and Subjective Quality Assessment of a Digital Musical Instrument." In *Proceedings of the Audio Mostly Conference*, pp. 116–123. 10.1145/2986416.2986428

Janson, T., C. Schindelhauer, and J. Wendeberg. 2010. "Self-Localization Application for iPhone Using Only Ambient Sound Signals." In *Proceedings of the International Conference on Indoor Positioning and Indoor Navigation*. 10.1109/IPIN.2010.5648094

Lopes, C., et al. 2006. "Localization of Off-the-Shelf Mobile Devices Using Audible Sound: Architectures, Protocols and Performance Assessment." *Mobile Computing and Communications Review* 10(2):38–50. 10.1145/1137975.1137980

Lossius, T., P. Baltazar, and T. de la Hogue. 2009. "DBAP: Distance-Based Amplitude Panning." In *Proceedings of the International Computer Music Conference*, pp. 489–492.

Malham, D. 1998. "Spatial Hearing Mechanisms and Sound Reproduction." Available online at digitalbrainstorming.ch/db_data/eve/ambisonics/text02.pdf. Accessed February 2022.

Mandal, A., et al. 2005. "Beep: 3D Indoor Positioning Using Audible Sound." In *Proceedings of the Consumer Communications and Networking Conference*, pp. 348–353.

Medvedev, Y. P., A. V. Sorokin, and V. I. Khashchanskiy. 1989. Способ ввода информационных сигналов в цифровое устройство обработки [Method for Information Signal Acquisition Into a Digital Processing Device]. USSR Patent 1,508,275 A1, filed 9 October 1987 and issued 15 September 1989. Available online at yandex.ru/patents/doc/SU1508275A1_19890915. Accessed February 2022.

Mitchell, T., and I. Heap. 2011. "SoundGrasp: A Gestural Interface for the Performance of Live Music." In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pp. 465–468.

Nymoen, K., S. l. A. Skogstad, and A. R. Jensenius. 2011. "SoundSaber: A Motion Capture Instrument." In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pp. 312–315.

Peng, L., and D. Gerhard. 2009. "A Wii-Based Gestural Interface for Computer Conducting Systems." In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pp. 155–156.

Pulkki, V. 1997. "Virtual Sound Source Positioning Using Vector Base Amplitude Panning." *Journal of the Audio Engineering Society* 45(6):456–466.

Robinson, F., et al. 2015. "Gestural Control in Electronic Music Performance: Sound Design Based on the 'Striking' and 'Bowing' Movement Metaphors." In *Proceedings of the Audio Mostly Conference*, Art. 26. 10.1145/2814895.2814901

Salazar, S., and J. Armitage. 2018. "Re-Engaging the Body and Gesture in Musical Live Coding." In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pp. 386–389.

Schlienger, D. 2016a. "Acoustic Localisation for Spatial Reproduction of Moving Sound Source: Application Scenarios and Proof of Concept." In *Proceedings of the International Conference on New Interfaces for Musical Interaction*, pp. 407–412. 10.5281/zenodo.1176116

Schlienger, D. 2016b. "Gestural Control for Musical Interaction Using Acoustic Localisation Techniques." In *Proceedings of the International Conference on Live Interfaces*, pp. 161–167.

Schlienger, D. 2016c. "Requirements on Kinaesthetic Interfaces for Spatially Interactive Sonic Art." In *Proceedings of the Audio Mostly Conference*, pp. 162–169. 10.1145/2986416.2986441

Schlienger, D., and A. Olarte. 2016. "Carte Blanche, Right Now!" In *Proceedings of Art without Borders*, pp. 253–256.

Schlienger, D., and S. Tervo. 2014. "Acoustic Localisation as an Alternative to Positioning Principles in Applications Presented at NIME 2001–2013." In *Proceedings of*

the International Conference on New Interfaces for Musical Expression, pp. 439–442. 10.5281/zenodo.1178933f
Şentürk, S., et al. 2012. "Crossole: A Gestural Interface for Composition, Improvisation and Performance using Kinect." In *Proceedings of the International Conference on New Interfaces for Musical Expression*. 10.5281/zenodo.1178201
Seob Lee, J., and W. S. Yeo. 2011. "Sonicstrument: A Musical Interface with Stereotypical Acoustic Transducers." In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pp. 24–27. 10.5281/zenodo.1181432
Simonsen, J., and T. Robertson. 2013. *Routledge International Handbook of Participatory Design*. Abingdon, UK: Routledge.
Trail, S., et al. 2012. "Non-Invasive Sensing and Gesture Control for Pitched Percussion Hyper-Instruments Using the Kinect." In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pp. 1–4. 10.5281/zenodo.1178435
Zhang, L., et al. 2017. "Acoustic NLOS Identification Using Acoustic Channel Characteristics for Smartphone Indoor Localization." *Sensors* 17:Art. 727.

11

Prototyping Situations: Interdisciplinary Free Improvisation in Technology Development & Latour's Mode of the Technical (VII)

This article is a media rich exposition available from the Research Catalogue on: https://www.researchcatalogue.net/view/1690251/1690252

12

How Do We Experience Digital Arts? – An Exploration through Latour's Modes of Existence (VIII)

(Pending Publication in MIT Leonardo Journal)

How Do We Experience Digital Arts? An Exploration through Latour's Modes of Existence

Dominik Schlienger (Researcher), Centre for Music & Technology,
Sibelius Academy, Uniarts Helsinki, P.O. Box 30 FI-00097 UNIARTS.
Email: dominik.schlienger@uniarts.fi

Abstract

This article explores the divergence between practice and theory of technology observed through an interdisciplinary free improvisation workshop which critically engaged with (digital) technology. Bruno Latour's 'technical mode of existence' proposes an intriguing interpretation of this differentiation. What is it that we experience when we engage with 'digital art'? How does this bear on conceptualizations of technologies? The 'fictional' and 'reference mode of existence' further help to understand the notion of the digital as it pervades culture and media. Using examples from music, visual arts, and observations from the workshop, dystopian visions of technology are disentangled, re-configured. Embodied agency and kinaesthesia play a major role in this process.

Artistic Practice - Technophile Technophobes

When exploring technology in *free interdisciplinary improvisation* workshops that I ran on the subject of *music, space and interaction* [1,2], I was surprised at the number of participants describing themselves as technophobes. Many felt disenfranchised from digital technology, feeling that as artists and musicians they were forced into it by circumstance, in order not to be excluded from an increasingly digitalised, and technified practice. Ultimately, participants happily engaged in and with technology. However, many associated technology and *digital* technology in particular, with a sense of disembodiment, epitomized in *brains on sofas*, a dystopian trope they conjured up *verbatim*. Yet, the works they produced were almost exclusively what is generally considered to be *digital art*. The puzzling contradiction between technological *practice* and professed *conceptions* of technology seemed poignant, which motivated me to investigate it.

In the following I explore the trope of *brains on sofas*. Although the number of workshop participants was not sufficient to be representative for society at large, I believe that *brains on sofas* encapsulates a conception of technology which is quite common. But even if the workshop's experimental practice provides a valid case study for both the trope and my theoretical exploration of it, I am not aiming to "prove" neither the trope nor the theory right or wrong.

Instead I am trying to uncover the uncanny correlations and contradictions I see between the way we *conceptualize* technologies and the way we *enact* them.

Through the workshops, which took place in an arts-university setting between 2014 and 2018, I gathered information on both the conceptions of technology and technological practices: Interdisciplinary improvisation is an extension to participatory design principles [3], a *prototypical practice* from which new techniques and technologies emerge. The idea is that in a performative, situated, improvisational setting, immediate, simple solutions to technical problems can be found. Every session was followed by a discussion recorded in field notes [4] which captured not only technical ideas and reflection on our practice, but contextualised them in relation to a general discourse on technology and the (sonic) arts. Participants ranged from musicians, composers and dancers to scenographers, painters, poets, media artists, and others.
The discoveries from the workshop centred around the ideas that first, techniques and technology are not just related but, essentially, the same thing, and second, that new technology can be deconstructed into old technology. This is how I happened upon French philosopher, sociologist and anthropologist Bruno Latour's work which particularly reverberated with these experiences.

Latour's works provided the basis for an exploratory theoretical journey into technology. Of course, other theoreticians could have stood at the beginning. But I report on an *exploration* which started for me with Latour's concept of the *technical mode of existence* and the case study of the workshop's trope of *brain on sofas*. But during this exploration other tropes and other theories gained and lost importance, as I discuss below.

In *We Have Never Been Modern* [5], Latour attests to a discrepancy between theory and practice of the moderns: In modern theory, culture and nature are strictly separated, yet, in practice, they are inseparably mixed. According to him, modern theory was never put into practice — hence, we have never been modern. In *Inquiry Into the Modes of Existence, An Anthropology of the Moderns* [6], he revisits this problem, but rather than to state what we have not been, he approaches modern practices positively, and describes what we moderns *are*. Through a rigorous rethinking of modern theories, and an analysis of modern practices and values, he arrives at a multi-ontology of 15 nonhierarchical *modes of existence:* Beings exist in more than the binary modes of nature-culture or object-subject. The modes are empirical entities, so there might be more than the ones found so far. As different as they are, modes are defined by a plurality of truth-conditions set out in a "research protocol" [7]. In this article the *technical*, the *fictional* and *reference* modes are of importance.

Besides Latour, Alain Berthoz, Lucy Suchman and Carrie Noland, whose works I refer to primarily here, there were other important theoretical works, by Donna Haraway [8], Paul Dourish [9,10] and Anne Balsamo [11], for example. That none of them is prominent in media studies, is no coincidence: Media theorists like Manovich [12] and Grau [13], for example, assume that the digital has a quality, even a materiality that can be directly experienced. This was contrary to my findings and made me look for an alternative approach.

The Technical

In a memorable discussion, a workshop participant said: “All digital technology was developed for military purposes!” This is probably partially true, for example the internet was developed based on research commissioned by the United States Department of Defence in the 1960, and Alan Turing worked for the United Kingdom Government at Bletchley Park during WW II. Nonetheless, Turing did research before and after the war, and it was CERN’s development of the worldwide web which gave the internet the form we know. A list of technologies used for war and killing could be expanded *ad absurdum* and whatever technology’s origin — war is unthinkable without it.

I propose another path: Consider the spoon. Maybe you fed your child with one this morning, a model specifically shaped for the anatomy of the baby mouth, made of soft antibacterial plastic, or one carved of wood, treated with non-toxic, natural oils, maybe one made of stainless steel – most definitely a technological object through and through.

These materials too are linked to means of destruction, plastic is made of fossil fuel, a cause for war, metal a material for weapons and there is nigh nothing that cannot be made of wood, from spears to arrows, to catapults.
As Latour puts it, when somebody "However lazy he may be [...] is just shifting position in his hammock, it is through this hammock he must pass to keep himself up in the air. [...] everything, on this basis becomes technology. Not just the hammock but also the two solid tree trunks to which it is attached!” [14] — This is the *technical:* the materials do something for us, we delegate something to them: A spoon to feed a baby, a catapult to break a stone wall, or two trees to suspend a hammock — the material *wood* is delegated to perform the tasks of transporting food, hurl missiles or just keep you suspended in a hammock. The wood, however, is not the technical, not even the spoon, the hammock or the catapult. They are just “the objects the technical leaves in its wake” [15]. It is tempting to think of the technical as an abstract entity. Yet, I think it is, on the contrary, anything but abstract; not virtual, but actual — it has to be *performed* to be.

The technical is something we experience indirectly and through its failure. As long as it works, it does what it has been delegated to do: The bus brings us to work – if not slowed down by bad weather, in which case we’d send a text that we’ll be late – if the phone’s battery isn’t flat, in which case we’d run from the bus stop and make up for time lost – if we don’t sprain our ankle. If none of this happens, we just go to work. And that was that. — The *technical project* to get us to work requires the activation of a whole chain of objects which we only notice as *technical* if the activation fails. (A bus, on a good day is *red*. On a bad day it's *late*. — It's *on time* only if you are afraid it might be late!)

We take the technical for granted. Nevertheless, artists and engineers constantly think about possible failures. Of course, there is joy in the technical *not* failing, for example when watching a tightrope walker. Celebrations of the smooth workings of the technical refer dialectically to the same thing as failure: Memories of well-crafted art prolongs the novelty of a technical twist, but the technical itself remains absent.

What is it that we experience when we speak of *digital art*? – And how does this experience bear on conceptualizations of technologies? For now, I expect the reader to accept that *digital*, *technical* and *computer*-pertain to the same notion, like in everyday language. I shall elaborate when this is not applicable.

According to Wikipedia — referred to here, not as an academic source but to signify the state of common knowledge — "Digital art is an artistic work or practice that uses digital technology as an essential part of the creative or presentation process" [16]. On this basis, I define four types of digital art:

1. It has been created using digital technology
2. It needs digital computation to manifest itself
3. It is presented via digital to analog conversion
4. It is about the digital/about computers

In short, I believe that only type 4 is aesthetically digital art. Counter-intuitively, this is the type for which the use of digital technology is the least essential: James Bridle's *Autonomous Trap* (See Fig.1), I saw the first time in a Symposium on digital art. It makes my point, as the car could even be any *non-digital* car! The photo could be taken with a digital camera, digitally generated, the car photoshopped into the picture - its meaning remains the same. It is digital art *only* according to type 4, as I will discuss below.

Fig. 1 Autonomous Trap James Bridle 2017 (© James Bridle)

Similarly, digital recordings of classical music are aesthetically *classical music* not digital art. Yet *Man Machine* by Kraftwerk [17], produced on analogue synthesizers and published on vinyl in 1978, is undoubtedly *computer music,* i.e., digital art of type 4. Music made with Synthesizers sounds digital or electronic because it sounds different from "acoustic" music. We might have learned to group such sounds as *electronic* through habit: Early electronic music, has "taught us" to recognize electronic music as such, even if produced with analogue circuitry.

We don't have grounds to call *artificial* sounds any more technical than *natural* ones. One is as technical as the other; natural sounds are even more high-tech if compared to the 1500 years of development resulting in a violin or the 150 million years evolution resulting in the bird's voice.

Still, there is a *feel* to digital technology which is different to the feel of an acoustic or electro-acoustic musical instrument, or a brush, hammer, spade, or a spoon. And we know instinctively that Kraftwerk's work is *about* technology. So if it is not *technical*, what is it then? Is it *fictional*? – Latour would affirm this, emphasizing that fictional doesn't mean *merely* fictional.

The Fictional

"Fiction, designates not the field of art, culture, works of art, but the particular mode awkwardly designated by the adverb 'fictionally'." [18] To compare it to the technical, Latour describes a guardrail first as a technical entity: "the guardrail above a precipice that keeps you from jumping into the void keeps on protecting you with its steel uprights, whether you want it to or not." And then as a fictional one: "The hero threatened to throw himself into the precipice and was held back, at the last minute, by a guardrail of words." So for the fictional "The requirement of continuity is at once less strong than for the steel guardrail (you don't have to forge it) and stronger (you have to keep on holding it so that it will hold you!)." [19]

First, let's acknowledge that art is more than *just* technology. As was pointed out to me in a seminar, "A violinist can also perform as a traffic warden, demanding cars to stop on an up-bow, drive on a down-bow and turn left on a high note, right on a low note; this is technical, but it doesn't make it art." Interestingly, the professor used a fictional example, which shows that the fictional goes beyond the arts and assists in the finding of *objective knowledge* [20]. Latour shows that Einstein as well relied on fictional characters to prove his points [21]. Furthermore, the bowing traffic warden would have been a splendid artistic performance!

The white lines painted on the pavement in *Autonomous Trap* only carry meaning to a viewer who knows that autonomous cars follow such white lines demarcating the borders of driving lanes. So a whole complex of technical *figurations* [22] of cars, roads, sensor technology, robots, automation and AI is evoked: The *notion* of the technical and digital.

The narrative of the technical creation process can be essential to understanding an artwork. For example, the work in Fig. 2 is painted in menstrual blood. Not knowing about this situates the work in an entirely different way. The artist knows this and plays with it: On the one hand, she expects us to understand the figuration of *mother and child* as the motive of religious iconography. On the other hand, by letting us know about the technical creation process, she provokes an engagement on her own terms in subverting, for example, the doctrine of *immaculate conception*. Knowing about the technical process defines the narrative. Yet, the artist could have used *any* material. By *telling* us that she used menstrual blood, she gives us access to the artwork's meaning.

Fig. 2 Nainen ja Lapsi, Hanni Haapaniemi, 2013 (© Hanni Haapaniemi)

Similarly, Kraftwerk's album *Man Machine* (see Fig. 3), by its title, cover-art, alienated voices, and use of synthesizers *is about* technology, computers. Its aesthetic owes more to the futurism of the 1920s than to science fiction, despite the use of (non-digital) sequencers, which was pioneering for 1978.

As Reference

We are closer now to the experience of the digital. Still something is missing. Let's call it the e-card effect: The disappointment to have received on the promise "I'll send you a postcard!" an emailed picture rather than a slightly worn, stamped and mail carrier-delivered cardboard postcard. Sending it by post means a physical loss for the sender so that the postcard *materializes* in the receiver's post box. The digital picture is a nonmaterial *reference*, a *data-set* we can share, but not give away. The digital in this sense belongs to Latour's mode of *reference*, the referential chain that provides access to knowledge: The more elaborately referenced knowledge we have of something, the more universally replicable that knowledge is.

Even though the use of blockchain technology can seemingly make datasets unique, the uniqueness is not in the experienceability of what the dataset encodes, but solely the uniqueness of the code. So an identical copy of the encoded artefact would still be possible — the blockchain (like a signature) would need to be consulted to verify its uniqueness. In a simile, the non-digital *non fungible token* equivalent of the Mona Lisa would be an arbitrary number of identical paintings, where only one bears the original signature by Leonardo da Vinci.

Fig. 3 Man Machine, 1978 (© Karl Klefisch)

If we experience the digital in a recording of classical music, we either find the recording better than a non-digital recording, or we find it less accurate, a faulty reproduction. If the reference is of sufficient resolution, we cannot spot a difference between, say, a digital photo and a non-digital one: The resolution goes beyond what we can detect as pixelation with the bare eye. So we could imagine a world which is its digital copy: The world measured and digitalized, the world as its data – A reference to the world. Importantly, the world as its data, is not the same as a 1:1 copy of the world! If a map of the world is an exact replica of the world we would be as lost on the map as we are lost in the world [6, pp.69-96], as it *is* the world according to the law of identity. So the digital reference is not a copy, but a *discrete numerical* reference, which needs to be *converted back to materiality* to be experienced.

Conversely, this implies that whatever is digitized will only be experienced in analogue material form. What we actually hear, when listening to a digital recording, is a rapidly oscillating cardboard-cone connected to electromagnets in a loudspeaker.

Whilst readers may agree that the aesthetics of digital art do not stem from the technical aspects of the digital; that the *notion* of digital is *fictional*; that digital art of type 2 and 3 is a *reference* in the Latourian sense, — I won't hold it against them if they still feel something's missing. What we discussed so far cannot explain the sense of disembodiment experienced by the workshop participants. Nothing so far accounts for the specter of the *brains on sofas*.

It is intriguing: *Brains on sofas* is scary not because it stands for a loss of agency (arguably, as *brains on sofas*, we can activate a plethora of remote controls), but because the agency is disembodied! This contrasts to another popular specter, the tale of *AI-robots taking over the world*: These are juxtaposed extremes of technology-mediated agency. On the one hand, agency with no body (*brains on sofas*), on the other hand, body with no agency (*AI takeover* [23]). To explore if the notions of *brains on sofas* and *AI takeover* are valid concepts of possible futures requires a thorough analysis of agency.

Agency

According to Actor-Network Theory [24], the term *social* is problematic, and hence replaced by the notion that everything that is associated with something within a network is an actant: people, as well as things. I paraphrase this as *distributed agency*. If everything is an actant, so are we. The brain on a sofa's agency encompasses not just the brain but also the sofa, remote controls, machines and actants necessary to keep it there. If we assent to this distributed agency the specter of the *brains on sofas* as well as the *AI takeover* remain isolated from their networks and even as fictions, when fleshing them out, we would end up building intricate networks to make them believable. Will they ever become a reality?

Anthropologist Lucy Suchman, investigated this question for the trope of the "human-like machine" (which is somewhere halfway towards our monsters). As a senior scientist at the Xerox Palo Alto Research Center, she addressed the social and material practices which make up technical systems. In *Human-Machine Reconfigurations* [25] she shows how boundaries between persons and machines are discursively and materially enacted. Stating that people create meaningful action by improvising based on their social and environmental resources, she challenges common assumptions behind human-computer interaction design with the argument that human action is constantly constructed and reconstructed from dynamic interactions with the material and social worlds. She uses Donna Haraway's method of *figurations* [26] on the premise that "all language, including the most technical or mathematical, is figural."

Latour counts *figurations* as beings of fiction. This does not mean that they are not real, concepts, also scientific concepts can be accessed via the fictional mode. Latour explains this seeming contradiction: "Each mode grasps all the others according to its own type of existence [27]". So realities themselves are conceptually relevant fictions. This idea helps to disambiguate the many - possibly to this point contradictory definitions of digital I used so far. The digital as a *technical* being, as a being of *reference* and as a *fictional* being is wrapped up together in the figuration of the technical, because the way we experience the digital doesn't differ to other technical beings. What Suchman in this context is saying, is that *machines* are figurations which can not successfully be accessed via the technical mode, that it would constitute a *category mistake* to access them solely as technical beings. Of course there are many technical beings enacted in a machine, but principally as actants in a socio-material arrangement, which, *per se* is a figuration.

If "all language including the most technical or mathematical, is figural," human-machine configuration can thus be *reconfigured*: Distributed action provides a re-configuration of the human-machine situation, redefining machines as *situations*.

From the perspective of the machine as contingently stabilized interaction, where the person is essential part of a *socio-material* arrangement, the question wether machines one day might successfully mimic the "capacity of the autonomous human subject" seems slightly off, as it is not clear who or what is mimicking what or whom. Nor is it particularly interesting, as the real question is an ethical, rather than a technical one, because these arrangements' reiterations and/or reconfigurations are "the cultural and political project of design. [28]"

Evidently, this ethical dimension also raises the question of responsibility, which, according to Suchman "on this view is met neither through control nor abdication but in ongoing practical, critical and generative acts of engagement."

If we apply these insights to *AI takeover,* and consequently also to all forms of automated decision making, it becomes evident that the whole conception stands on the shaky legs of an isolated agency that could exist outside of an actor-network it essentially needs to persist. Even if the "autonomous human subject" is as much of a tall tale as our *specters*, (human autonomy is not self-evident if we agree on a distributed agency), it shows how unhelpful it is to think of ourselves positioned in a human-machine *opposition*. If, however, we think of machines as socio-material interactions, we integrate the machine into our practice without a need for hierarchical value systems.

As socio-material practices, machines are never autonomous either: As a configuration, even a robotic lawn mower relies on a lawn to be mowed, a charging station, spare parts in a store nearby, firmware updates, etc. If it avoids collision with a turtle, it does so in *response* to its negotiated design as a "socio-material practice", a "cultural and political project". Yes, in that sense the robot acts *responsibly*, but it is a responsibility distributed in a network of these "ongoing practical, critical and generative acts of engagement" [29]. Similarly, it is hard to see how a *brain on a sofa* could persist as an actor without a network.

All this doesn't still quite answer why we concocted that troubling vision of the *brains on sofas* during the improvisation workshops: Our worry was not about losing the power of decision-making, epitomized by *AI takeover,* but that our existence should become *disembodied*. Immediately, I hear protest rising from the audience that thinking with the brain is also an embodied act and yes, that is true too. Still, something is missing. I argue that it is the lack of kinaesthetic experience.

Tool or Machine?

According to Latour, the technical encompasses both techniques and technologies. Yet, how come that we experience a physical *tool*, like a spade or a guitar as fundamentally different to a *machine*, say a tape recorder, or a conveyor belt in a supermarket? Obviously, it is to do with embodiment, as the difference is that the machine takes a process out of our hands – literally when the conveyor belt transports our goods to the cashier.

In the field of *embodied action* there are many paths one could follow, for example, the works by Noë [30] or Gallagher [31]. I happened upon Carrie Noland's work *Agency and Embodiment* [32] first, and as it focuses on the embodied experience of agency manifested through gestural movement, it struck a chord with what we experienced in the workshop. Corporeal gesture is not a prominent concept in the works of Suchman and Latour. Noland makes the case that "kinaesthetic experience, acts of embodied gesturing, places pressure on the conditioning a body receives", as an alternative to constructivists' inability to produce a convincing account of agency." She asserts that "kinaesthesia – feeling the body move – encourages experiment, modification and, at times, rejection of the routine." Her most compelling argument for this embodied agency is her reading of neurophysiologist Alain Berthoz's *The Brains Sense of Movement* [33].

Berthoz is primarily interested in perception, but argues that perception neurologically constitutes an *action*. This is inherently a type of motor-neuron theory of perception, introduced by Berthoz with "William James's concept of an anticipatory neuronal pathway" [34] (see Fig. 4): If a sensory cell S is excited, it activates a motor neuron M which induces a muscle contraction. A kinaesthetic cell K detects the movement in the muscle and modifies the motor neuron. James proposed that there might be an additional path from the sensory cell *directly* to the kinaesthetic cell, enabling the kinaesthetic cell to modify the motor neuron even before it activates the muscle. The kinaesthetic cell can thus *anticipate*

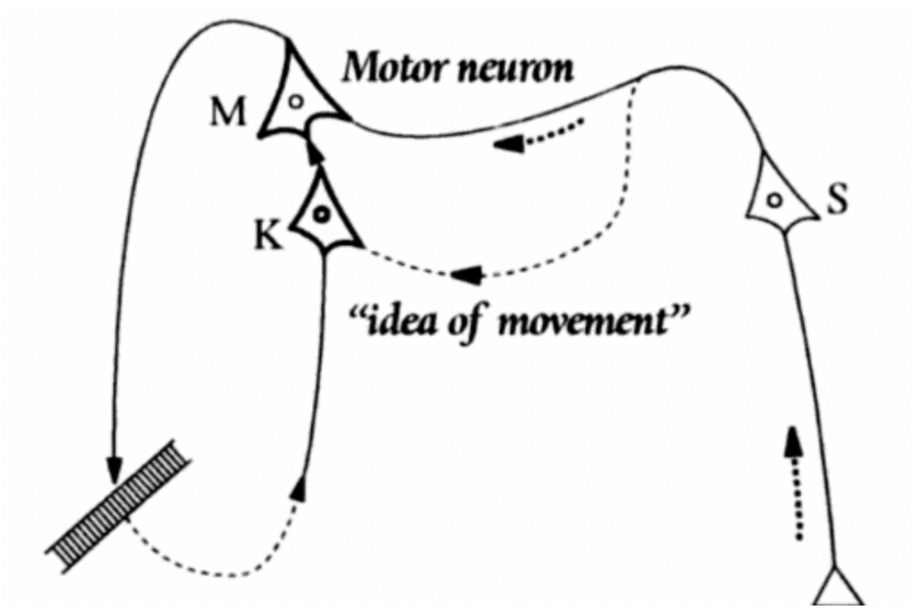


Fig. 4 William James's concept of an anticipatory neuronal pathway.
© Odile Jacob, 2005)

how to modify the motor neuron. Importantly, the kinaesthetic cell's actions are informed by the feedback from the muscle *as well as* from the sensory cell, in a continuous adaptive loop.

This '"enactive perception' the cornerstone of which is the agentic, decision influencing role of kinaesthetic sensation" [35] fills the conceptual gap Latour's modes of existence somehow left: Our disenchantment with a possible, if imagined reality of a world where the lack of kinaesthetic experience disconnects

us from active engagement is expressed in the *brains on sofas* trope because kinaesthetic proprioception *is* our sense of agency! If Perception is action, it intrinsically constitutes an agentic kinaesthetic experience: When we hear, see, or touch, we are active on the neuro-motoric level. If all action that we perceive as *agentic* is intrinsically *motoric,* i.e., a kinaesthetic experience, we have a sense of agency, when and because we experience something kinaesthetically. Maybe this heightened sense of agency can explain the experience of authorship, the sense of achievement we feel when we contributed to an art project or artefact [36].

Concluding Remarks

The notions *brains on sofas* and *AI takeover* express the fear of our own extinction. Poignantly, they were voiced in an improvisation workshop with the aim of developing techniques and technologies, and therefore endorsing machines! Instead of feeling relief that technology is doing our work, we are afraid of technology taking away our work. Work that as craftspeople, we enjoy doing. The workshop provided many examples for how to engage with technology is an empowering, enabling process:

Participants' unease with technology as it pervades cultural discourse was not borne out in their practice.

If we conceptualize our relation with technology through the popular discourses of technology of the type expressed through *AI takeover* and or *brains on sofas*, we forfeit our say, stake, and responsibility for how we design machines and technical systems. As a counter strategy, I have tried to show here step by step, how from a discrepancy between practice and cultural conceptions of technology, a re-thinking of the digital through Latour's modes of existence is possible. This is congruent with observations in the workshop: Participants did engage with (digital) technologies once they recognised them as not essentially different to their craft's *techniques*. This is how the rethinking of technology becomes possible — from within its practice.The fictional construct, the figuration of the *brains on sofas* cannot serve as an objective technical description of the digital. The digital as a *technical being* is a tool of reference, and escapes direct experience. Therefore digital technology cannot endow nor institute the dystopian concepts of *AI takeover* and the workshop participants' fear of *brains on sofas*. These specters are neither digital nor technical, but a fictional *tale* of the technical, a figuration. Interestingly, like the specters, machines too are figuration. Taking the tale apart, we see that agency in a machine is distributed across a network within which we find ourselves too, not helplessly entangled but as individual actants amongst others constituting the network, shaping and negotiating our relations with the technical. Workshop participants engaged in technical processes through actively con*figuring* machines and realising their stake in technical *actions*, rather than in the machine as an *object*.

By connecting to the neuron theory of perception, I hope to have reestablished that a sense of authorship arrives from kinaesthesia, which explains why we still feel pride in the instant of realisation of "I made this", even while being fully aware, and rightly so, that the agency necessary for the work's completion was distributed, and authorship only established after the *fact* of making the arte*fact*. In the workshop's focus on spatial interactivity, the gestural body as a *technical* entity was fundamental. As a collective, participatory practice, the distributed nature of the creative process was paramount. The sense of authorship nevertheless experienced by workshop participants provides evidence for this role of kinaesthesia in *technical action*s.

That we agree to socio-material arrangements wherein we are disenfranchising ourselves from our embodied practices — in the belief that *it's technology, so it's out of our hands* not only shows that we do not grasp the responsibility we have and should take in constituting these arrangements; it also exposes how we build machines and how we actively *mis*-conceptualize technologies. The call for an engaged, participative understanding of the *technical* was consensus amongst workshop participants, the experimental practice a step towards it.

References and Notes

[1] Schlienger, D., and A. Olarte. “Carte Blanche, Right Now!” In *Art Without Borders,* Proceedings, (Helsinki 2016) pp. 253–256.

[2] Andean, J. “Research Group in Interdisciplinary Improvisation, Goals, Perspectives, and Practice.” In A. Arlander, ed. *This And That*, Episodi 4. (University of the Arts Helsinki 2014), pp. 174–191.

[3] Robertson, T., and J. Simonsen. *Routledge international handbook of participatory design* edited by Jesper Simonsen and Toni Robertson. (Routledge New York 2013)

[4] Barz, G.F. and Cooley, T.J. *Shadows in the Field: New Perspectives for Fieldwork in Ethnomusicology*. (Oxford University Press 2008)

[5] Latour, B. and Porter, C *We Have Never Been Modern*. (Harvard University Press 1993)

[6] Latour, B. *An Inquiry Into Modes of Existence*. (Harvard University Press 2013)

[7] See [6] pp. 49-64.

[8] Haraway, Donna, *The Haraway Reader* (New York, Routledge 2004)

[9]Paul Dourish and Genevieve Bell, *Divining a Digital Future: Mess and Mythology in Ubiquitous Computing* (MIT Press, 2011)

[10] Dourish, Paul, *Where the Action Is: The Foundations of Embodied Interaction*. (MIT Press 2001)

[11] Balsamo, Anne. *Designing culture: the technological imagination at work.* (Durham, North Carolina, Duke University Press 2011)

[12] Manovich, L. *The language of new media*. (MIT Press 2002)

[13] Grau, O., *Virtual Art: from illusion to immersion*. (MIT Press 2003)

[14] See [6] p.214.

[15] See [6] p.219.

[16] Wikipedia contributors. "Digital art." Wikipedia, The Free Encyclopedia. [en.wikipedia.org/wiki/Digital_art].

[17] Kraftwerk *Man Machine* [Vinyl]. Kling Klang (EMI Electrola Düsseldorf 1978)

[18] Bruno Latour. [FIC], 2013. URL: http://www.modesofexistence.org/aime/voc/171.

[19] See [6] p.248.

[20] See [6] p.249.

[21] Latour, B. “A Relativistic Account of Einstein's Relativity” *Social Studies of Science*, Vol 18, Issue 1, (1988) pp. 3-44.

[22] See [8] pp. 7-45.

[23] Wikipedia contributors. "AI takeover." Wikipedia, The Free Encyclopedia [en.wikipedia.org/wiki/AI_takeover]

[24] Latour, B. *Reassembling the Social – An Introduction to Actor-Network-Theory* (Oxford University Press 2005)

[25] Suchman, L. *Human-Machine Reconfigurations: Plans and Situated Actions* 2nd ed. (Cambridge University Press 2006)

[26] See [8] pp. 7-45.

[27] See [6] p. 215.

[28] See [25] pp. 285-286.

[29] See [25] p. 286.

[30] Noë, A. *Action in perception.* (MIT Press, 2004)

[31] Gallagher, S. *Action and Interaction*. (Oxford University Press 2020)

[32] Noland, C. *Agency and Embodiment*. (Harvard University Press 2009)

[33] Berthoz, A. *The Brain's Sense of Movement*. (Harvard University Press 2000)

[34] See [33] p.10.

[35] See [32] p. 10.

[36] *Nota bene*, according to Latour, this sense of authorship comes *after the fact*, as “Competence, here again, here as everywhere, follows performance rather than preceding it.” in the technical [6, p.228].

Biographical Information

DOMINIK SCHLIENGER is a Composer-Researcher at UNIARTS Helsinki's Centre for Music & Technology. He received his MSc Audio Production from the University of the West of England (2012).

This book reports on a development of spatial interaction technology in sonic arts, the Acoustic Localisation Positioning System (ALPS) Moreover, it questions the societal role of technology by deconstructing the cultural figuration of technology. The research shows that through the omnipresence of code in form of computer languages, machines do not act autonomously. Rather, computers are to the brain what a hammer is to the hand — technical extensions of embodied human activity. This has consequences: it is no longer justified to bemoan the disenfranchisement from technology knowing that we essentially constitute it ourselves. It is our technical activity, not inaction, which defines us and pushes us towards the abyss. If we engage in participatory, improvisational practices instead —in technology development, but also in the way we organise our collectives — sustainable technology is possible. Whenever we act consciously-technically the question is if our actions are appropriate.

STUDIA MUSICA 92

PRINT
ISBN 978-952-329-286-4
ISSN 0788-3757

PDF
ISBN 978-952-329-287-1
ISSN 2489-8155

UNIARTS HELSINKI

SIBELIUS ACADEMY

APPLIED STUDY PROGRAMME
MuTri Doctoral School