DESIGNING FOR THE IMAGINATION OF SONIC NATURAL INTERFACES.

Tore Knudsen
DESIGNING FOR THE IMAGINATION OF SONIC NATURAL INTERFACES

Thesis Project II
by Tore Knudsen

Supervisor
Per Linde

Examinator
Henrik Svarrer Larsen

Time of Examination
Tuesday May 29th
9.00 - 10.00

Interaction Design Masters Programme
Malmö University
2018
Acknowledgment

Thanks to my supervisor Per Linde for his valuable feedback, enthusiasm and overall guidance throughout my process. Thanks to Andreas and Lasse from Støj studio for their technical guidance and inspirational views on creative machine learning. I would also like to thanks Bjørn Karmann and Nick Pagee for devoting their time to give me feedback during my design process. Thanks to my girlfriend Annika for her daily support and open mind during my work sessions at home where I’ve spend too much time talking to myself while training machine learning models with my voice.

Finally I would like to thanks my classmates for the inspiration and great times we have had the last two years. It has been a true pleasure.

Appendix and video links

Appendix folder:
https://www.dropbox.com/sh/zgnb9gj7xu60xz1l/AAAY7T8vgYVvIMpzV4OuGcaoa?dl=0

Video Folder:
https://www.dropbox.com/sh/17kmzqg3qpj4t3t/AAD-dskwYGDMItajLjnItN26a?dl=0
Abstract

In this thesis I present explorative work that shows how sounds beyond speak can be used on the input side in the design of interactive experiences and natural interfaces. By engaging in explorative approaches with a material view on sound and interactive machine learning, I’ve shown how these two counterparts may be combined with a goal to envision new possibilities and perspectives on sonic natural interfaces beyond speech. This exploration has been guided with a theoretical background of design materials, machine learning, sonic interaction design and with a research through design driven process, I’ve used iterative prototyping and workshops with participants to conduct knowledge and guide the explorative process. My design work has resulted in new prototyping tools for designers to work with sound and interactive machine learning as well as a prototypes concept for kids that aims to manifest the material findings around sound and interactive machine learning that I’ve done in this project.

By evaluating my design work in contextual settings with participants, I’ve conducted both analytical and productive investigations than can construct new perspectives on how sound based interfaces beyond speech can be designed to support new interactive experiences with artefacts. Here my focus has been to engage with sound as a design material from both contextual and individual perspectives, and how this can be explored by end-users empowered by interactive machine learning to foster new forms of creative engagement with our physical world.

**Keywords:** Sonic Interaction Design, Interactive Machine Learning, User-centred Machine learning, Natural Interface, Context Aware, Research Through Design, Ubiquitous Computing
Contents

06 Natural interfaces in 2018
1.1 Scope: Exploring the materiality of natural sonic interfaces

09 Technological imagination and material view
2.1 Material in Interaction design
2.2 Technological imagination
2.3 Machine learning
2.4 Envisioning new contexts (Related work)
2.5 Sound and interaction design

20 Method and Approaches
3.1 Research through design
3.2 Approaching sound as an expressive material
3.3 Sketching with material: Inspirational bits

24 Design Process
4.1 Designing a platform for prototyping with sound and ML
4.2 Exploring sound as inputs
4.3 Exploring the richness of sounds
4.4 Appropriating contexts
4.5 Process Reflection: Material guidance into new contexts
4.6 IML and sound as a storytelling mechanics
4.7 Final Prototypes and evaluation
42  Testing and Evaluation
   5.1 Testing Sounds on the farm
   5.2 Testing My foley book

47  Discussion and Reflection
   6.1 IML as a prototyping tool
   6.2 Material exploration as a way to facilitate creative thinking
   6.3 New opportunities in SID and natural interfaces
   6.4 Designing techno culture
   6.5 Limitations and challenges

51  Conclusion

53  References
   8.1 Appendix
1. **Natural interfaces in 2018**

The possibilities for how we interact with technology has through the last decades expanded a lot. We have seen the transition from text-based interfaces to mouse and keyboards, to ubiquitous mobile and touch screens, to gesture based and so called natural interfaces. Envisions of natural interfaces is often focused around the possibility to support common forms of human expressions (Jylhä, 2009), and as stated by Malizia & Bellucci; “letting people interact with technology by employing the same gestures they employ to interact with objects in everyday, as evolution and education taught us.” (2012, p. 36). Both advancements and accessibility in machine learning (ML) technologies for interactive systems plays a major role in the new possibilities of working with human gestures as input, and over the recent years it has also been adopted more actively by designers in the creation of digital artefacts and art that challenges and brings new visions of how we can use ML to interact with technology.

On of the most established and contemporary examples of natural interfaces that we see in our society is the increasing focus on voice recognition and speech interfaces. A new wave of digital artefacts is starting to occupy our private settings where home assistance (Amazon’s Alexa, Google Home, Apple’s Siri, Microsoft’s Cortana) and smart objects can be operated by human speech instead of button and screens and almost act as a new member of the family. This phenomena seems to continue and statements are also made that by 2018 voice interfaces will have changed the way devices and application are designed (Wired, 2018). This is further stated in the following quote:

“For years we have interacted with machines the way they have dictated, by touch – using a keyboard, screen or mouse. But this is not the natural way for humans to communicate. As humans, we prefer voice. In 2018, we’ll see more machines learn to communicate the way humans do” - Werner Vogels, CTO of Amazon (cited as in Wired, 2018).

This quote leaves as set of questions which summaries very well what will be the focus of this thesis project:
“*They have dictated*” relates back to the designer of the machines. By interacting with voice the designer behind the system is still the decision maker for how commands should be spoken and which cultural aspects of language to include. One example is how Alexa users receive emails every week with new commands they can try out with Alexa (See Appendix. 1). As Malizia & Bellucci (2012) states; HCI is full of examples of users adapting to designers choices and natural interfaces also seems to follow this pattern. However, some ML paradigms has shown the potential to facilitate end-user appropriations, which can let the user engage creatively with ML models to create interactive experiences and get a new level of agency over the interaction that before only was considered by the designer. This leads to the next question concerning the natural in natural interfaces

“The natural way for humans” to communicate is dictated by the designers of the systems and while language and speech is how we usual communicate as humans, we also live in a rich world of contextual sounds that is a big part of the embodied interaction that we experience the world through. Non-speech audio and everyday sounds as input to computational systems have received very little attention in interface design (Jylhäuser, 2009), and the natural of natural interfaces may be challenged or further developed when explored from an engagement with the richness of ambiguous sounds.

1.1 Scope: Exploring the materiality of natural sonic interfaces

These initial thoughts has outlined the setting for exploring a design space that aims to unpack how sound beyond the mainstream use of speech can be used and designed at the input side of natural interfaces, as well as how users can be empowered to use sound in expressive ways for interaction. I will engage in a material exploration of sounds positioned within Sonic Interaction Design (SID) and new design paradigms of human-centred machine learning, to present an approach that engages in a technological imagination of sonic natural interfaces.

With a design based research approach I focus less on technical performance and more on contextual, social and aesthetic meaning-making, I’ll in this thesis address the following research question:
How may natural interfaces within sonic interaction that goes beyond speech be imagined with a focus on end-user appropriation empowered by Interactive machine learning?

In addressing this main question I also engage in explorations that covers the following:

- How can machine learning based prototyping tools be developed to empower designers and “domain experts” in explorative design phases of sonic natural interfaces?

- How can the “natural” in natural interfaces be investigated and emphasised through creative engagement with expressive artefacts and interfaces within socio-cultural contexts?

To answer these questions I’ve explored ambiguous sounds and machine learning with a material view to unpack new potentials in this intersection. I’ll start chapter 2 by presenting the theoretical background that accounts for the material view that has been adopted in this thesis, and how it can be used to foster new technological imagination.
2. Technological imagination and material view

In this thesis I treat ambiguous sounds and ML as design materials in order to develop new perspectives on both SID and ML. A material view is central in the process of design and innovation, where design is defined in the intersection between what is thinkable and what is possible (Manzini and Cau, 1989). According to Manzini and Cau, the thinkable is close related to culture and models, while materials and development defines what is possible (ibid). A material view in the design process is also described by Schön (1992), who defines the main activity of design as the reflective engagement with the materials in a design situation, which creates a “talkback” from the material as the designer explores it.

2.1 Material in Interaction design

The same material practice is essential in interaction design, though with a focus shifting from: “..the visual presentation of spatial form to the act presentation of temporal behaviour.” (Hallnäs and Redström, 2006, p. 8). The temporal form and behaviour is one of the core material properties of digital artefacts which describes the state changing behaviour or form, and the formgiving process in interaction design is often oriented around this kind of “immaterial” forms (ibid). Close related Vallgårda and Redström (2007) also accounts for a material strategy to understand the computer as a material, which they define as computational composites. They see the computer as a material that needs to be part of a composite with other materials in order to come to expression on a human scale (ibid, p.3).

A clear definition of what design materials is in interaction design can however be hard to define, and Fernaeus & Sundström (2012) stresses that there is a underestimation of material knowledge when the computer becomes a design material due to the digital complexity; being the temporal form. While they present characteristics of digital material as: “..technology that can sustain an interaction over time with users (creating for a dynamic gestalt)” (ibid, p.487) they also state how this complexity calls for news ways of understanding these computational properties as materials (ibid). While a more fundamental view on the characteristics of a
material has been presented as “a physical substance that shows specific properties of its kind which can be proportioned in desired quantities and manipulated into a form.” (Vallgårda & Sokoler, 2010, p. 3), Fernaeus & Sundström (2012) states that there exist many different types of materials to be considered in the design of interactive systems. One example is the material nature of the human body as it moves in space when designing for movement-based interaction design (ibid, p.494). Structured material experimentations from an interaction design perspective is important to identify new opportunities and revolutionize the way we currently use technology (ibid). This view also related to Balsamo (2011) concept of technocultural innovation

2.2 Technological imagination

Technocultural innovation is described as a performance of two critical practices being the exercise of technological imagination and cultural reproduction (ibid, p. 6) Technological imagination is described as a mind-set that enables people to think with technology where the imagination is performative. This exercise engages with the materiality of the world that creates conditions for future world-making. Through the active engagement between humans and technology culture is also reworked through “…the development of new narratives, new myths, new rituals, new modes of expression, and new knowledge that make the innovations meaningful” (ibid, p.7).

Balsamo propose that designers act as cultural mediators by translating among languages, materials, and people” (ibid, p.11) and this refers to the articulatory and performative process of cultural reproduction. This practice suggest that for innovation to be comprehended it must draw on users and stakeholders’ already established understandings and circulations within the particular technoculture. At the same time it must: “perform novelty through the creation of new possibilities, expressed in the language, desire, dreams and phantasms of needs.” (ibid, p.10).

Inspired by this mind-set to think with technology as a way to foster technological imagination, I’ll in the next sections account for ML as a widely used technology and for how ML can be adopted by designers and users as a material and tool to create new interactive experiences.
2.3 Machine learning

Systems that can learn from their end-users is a widespread phenomenon that is used in more digital products and services than ever, and ML is a powerful tool for this learning process where data is transformed into computational models (Amershi et al, 2014).

While ML is not a new technology its public profile has raised with new advancement in algorithms and with the rise of Big Data ML algorithms is used with enormous datasets to make predictions in industries like healthcare, business and politics. In our personal devices and homes ML algorithms are also making decisions for us based on our behaviour. Some examples of where ML algorithms are predicting, curating or making recommendations for us is in our personal media feeds, streaming services and e-mail spam filters etc. (Dove, Halskov, Forlizzi & Zimmerman, 2017). In HCI research has also showed to work with ML to increase interaction possibilities such as gesture recognition and adaptive interfaces, which often is at the core of natural interfaces (Gilles, et al, 2016).

The user interaction with ML algorithms is however often indirect. The data that users provide as a by-product of their actions is used to train models that make future predictions and choices for the users. In this sense the user’s role is passive and non-engaging, and often this leaves ML as a black-box technology that users don’t know they are interacting with. Over the past few years a rising focus has been established around how ML can be designed to invite for more active engagement from the end-user; an approach that often is defined as user-centred machine learning (Gilles, et al, 2016).

User-centred machine learning (Interactive machine learning)

Gilles et al (2016) presents their view of human-centred machine learning as a suggestion to rethink how user-interfaces and ML algorithms can be designed to support human goals and be more useful for the end-users. Close related Dove, Halskov, Forlizzi & Zimmerman (2017) also states that designers should pay closer attention to the possibilities that ML offers as a new design material. One practice within HCI that has enabled explorations of such a user-centred ML approach is the ML paradigm of interactive machine learning (IML) defined by Fails & Olsen (2003). IML shows...
an approach which is based on rapid iterative cycles of the user training a model, testing its performance, and modifying the model with new examples to improve its performance if needed (Bernardo & Fiebrink, 2017). In this sense the user explicit shows the system new examples and decides how they should be labelled or created (Gilles, et al, 2016). This approach invites the user to become active in the decisions making of how a system should work and through a co-adapting relationship with the machine, iteratively reconfigure the system by providing new examples until a desired outcome is established.

Bernardo, Zbyszynski, Fiebrink & Grierson (2017), Gilles, et al (2016), and Amershi et al (2014) argues how IML can support new types of end-user customization and be a way to facilitate end-user innovation. Bernardo, Zbyszynski, Fiebrink & Grierson (2017) also claims how the right tools potential can democratize machine learning and make the benefits and use of ML algorithms realisable for a wide range of end-users without programming or ML expertise. A specific example of this is done by Katan, Grierson & Fiebrink (2015) who demonstrated how supervised machine learning tools can be used to empower disabled people to customize their own gesture based musical interfaces. This was done by letting the users demonstrate gestures to various gesture tracking cameras, which then was used as inputs to train an ML algorithm. The users could then perform these gestures to influence and create music. Katan, Grierson & Fiebrink (2015) demonstrates not only how IML allows end-users to make quick iterations and customization without applying programming but also how IML has potential as a design tool by: “accelerating the design process by allowing the quick translation of participant observations into prototypes” (ibid, p.4).

**IML Challenges**

While IML can empower end-user to gain more control over the behaviour of a system, it also presents new challenges. Interfaces that support sufficient feedback on how examples relates and how the model is trained, as well as a system’s limits, will be needed to enable the user to debug a model that doesn’t behave as expected (Gilles, et al, 2016). Another challenge is the balance of agency. As Amershi, Cakmak, Knox, & Kulesza (2014) points out, research has shown how users in general don’t want to be treated as oracles for machines and that a active labelling process tend to be annoying.
over a longer period of time (ibid).

2.4 Envisioning new contexts (Related work)

Related to Balsamo’s (2011) view of technological imagination, Dove, Halskov, Forlizzi & Zimmerman (2017) argues that often technology enters the market due to technical advancement with very little concerns on the design of it. Here designers work to re-understand the technology as it matures by inventing new forms that wasn’t imagined when the technology first was invented. “We see machine learning (ML) as a not so new technology that is ready for design innovation” (ibid, 2017, p.278). In the same paper they also suggest future investigations of ML to envision opportunities for apply ML in less obvious ways (ibid). The following section presents a set of projects that demonstrates new uses of ML in novel ways that empowers the end-user.

Fowl-language

One example that demonstrates the use of ML in a less obvious way is the Fowl-language project by Curtin, Daley & Anderson (2014). Their research is a proof of concept that shows how live-recorded sounds of broiler chickens can be processed with ML algorithms to predict the well-being of the chickens in a farm setting. Chickens have a “pattern of speech” and this reveals a lot about them and their environment. E.g. chickens makes more noise when under stress which can be due to rise in temperature, and a decrease in noise can be due to darkness either from night times or defective lightning. By finding and analysing these patterns the ML algorithm can inform the farmer about the chickens from remote.

Mogees Pro

The Mogees Pro (n.d) is a vibration microphone that through an app works as an midi controller. By placing the sticky microphone to any surface or object different types of strikes and gestures performed on that specific object can be mapped into MIDI controlled sounds (Figure 1.1).

The app provides an interface where the gesture mapping to instruments can be conducted, and an overview of how the given example is related in the model. When a new object is chosen the
user must train the app to distinguish between the different gestures by provide training examples. Mogees Pro is a good example of a consumer product that uses supervised IML to empower the user to appropriate different objects as music instruments through customization. See a video of the product here: https://www.youtube.com/watch?v=Eq7h8IwEJM

**Objectifier**

The Objectifier by Bjørn Karmann (Karmann, 2017), is a power socket with an embedded camera (Figure 1.2). Through an app the power socket can be trained to turn on and off in relation to what it sees through the camera.

The socket works with all 230V appliances, and opens up a new world of possible appropriations. Karmann presents several footages documenting users who trains home appliances for the first time, which ranges from turning on the music player with a specific dance move, to switching the light off when the user goes to bed. It also demonstrated a relevance in workshop settings where an industrial machine first will turn on when safety glasses is worn correctly, and therefore also turns off when the glasses are removed or misplaced. Karmann presents the notion of “spatial programming”, which is presented as a way to think of how the physical environment seen by the camera becomes the programming. A chair is suddenly a
variable in a system, which position can influence the output of the system.

Figure 1.2 - The Objectifier

Overall the Objectifier is a very strong example of how ML can be a way to provide users with tools that lets them "program" and customize products without writing code to gain agency by creatively train the ML model on how they want to interact with their objects.

**Machine learning: Conclusion**

The presented theories and projects shape an understanding of how IML can act as a design material. When a model is trained the temporal form is also shaped and acts as a material than can be shaped, changed, and influence the computed causality (Vallgårda & Sokoler, 2010) in the digital artefact or system. Where regular code often is used to shape the behaviour, IML allows for a rich form-giving process, which is embodied in the context of interaction when the training is done through live examples. While there exists multiple types of ML algorithms and paradigms I’ll be working specifically with supervised interactive machine learning, to work with classification where the outputs relate to a set of discrete categories (Hebron, 2016, p.19). This is the same type being used in
the presented examples.

However, by making parallels to Vallgårda & Sokoler’s computational composites, IML is only one part of the composite, while the chosen form of input acts as the other. From this point of view the chosen input type becomes a design material in the way it plays tightly together with the IML algorithm. This encourage for a material understanding of the input, being sound in my case. In the next section I’ll account for relevant roles that sound can have in Interaction design, and how its many properties can be seen with a material view in the design of interactive artefacts.

2.5 Sound and interaction design

Sonic Interaction Design

_Sonic interaction design_ (SID) (Rocchesso, et al., 2008) has emerged as a field often used in relation to ubiquitous computing and auditory displays. It is however not limited to this intersection, and the term is better used to describe all the possible aspects and practices around sound and the roles it can play in the interaction loop between users and artefacts. While sound has a strong history of being used as feedback mechanism in interface design (Caramiaux, et al, 2015), the field of SID encourage designers to perceive sound as a broader medium for interaction (ibid).

Several diverse descriptions of SID as a term has been presented through research papers and together these statements paints a picture of the potential sound has as a design material in the area of interaction design, which also call for new ways of using the sonic dimension. To highlight a few Caramiaux et al (2015) presents SID as design work that use sound as an active medium that can enable novel phenomenological and social experiences with and through interactive technology. Close related Ferranti and Spitz (2017) states that SID follows trends of third-wave-HCI that instead of focusing purely on functionality, has culture, emotions and experience at the scope of interaction between human and machines. These descriptions also implies how explorations of using sound as a material should go beyond its physics around pressure signals, frequencies and amplitudes, and into exploiting its expressive qualities (Delle & Rocchesso, 2014). In the following section I will present work that can help to understand sound as a expressive and
ambiguous design material.

**Sound as material**

Sound exists physically as movement of air pressure and vibrations that are invisible to our eye but can be experienced from our auditory ability and physical on our body as pressure at some frequencies. We can also find properties in sound such as tempo, pitch, timbre and rhythm, which often are attributes, used in music creation. Ferranti and Spitz (2017) compares these properties to the ones of colour (lightness, vividness and hue) and how both can map aspects of emotions. Similar Caramiaux et al (2015) also accounts for how sound can be an effective medium for evoking memories and emotions and that sound is an information medium that conveys clues about both materials, substances and physical environments as well as revealing a physical dimension of a space and its surface. In relation to this Caramiaux et al (2015) presents the concept of sonic affordances, which is a way to think about the corporeal response that a sound may invite on the listener.

This concept of the affordances of sounds is also explored by Gaver (1993) who presents a study on which sounds we hear and how we hear them. He puts the sounds of our surroundings as his focus and he states everyday listening as: “...the experience of hearing events in the world, rather than sounds per se” (p.1). Attributes in the everyday sounds concerns the sound-producing event where one example is how the sound of a car coming closer, can make us move away from the road and we can hear both its speed, size, location, and how it relates to the environment (Ibid).

Close related studies of sounds has been conducted by Schafer (1993) who purpose a framework which allows for studying the functionality and meaning of sounds, by classifying sounds according to different aspects. Schafer highlights the difficulty in this, due to aesthetic perceptions that has showed to be subjective and dependent on culture and context. One strong example of this by Schafer is presented in a experiment session where participants was describing the sound of an electric coffee grinder as “hideous”, “frightening” and “menacing”. However when the participants were presented to the actual object their attitudes mollified (ibid).

Within the field of SID, Delle & Rocchesso (2014) present the approach of procedural audio, which similar to Gaver’s concept
of sound-producing events, focuses on the underlying physical processes. Close related Jylhä (2011) also defines the concept of sonic gestures as sound-producing actions generated by humans. Still Caramiaux et al (2015) highlights how limited insights have been given to interaction designers on how to design and realize sound based interactions. Especially as an input.

**Sound as input**

While speech interfaces has been around for several years and has gained enormous attention with the trend of speech-based home assistants, the use of non-speech audio and everyday sounds have received very little attention in interface design as also argued by Jylhä & Erkut (2009). They present one example through a handclap interface for sonic interaction with computer (ibid). Another example is Dobson, Whitman, & Ellis’s (2005) Blendie; a blender that is controlled by mimicking the sound of a blender with growls. Some qualities of sound as inputs has further been covered by jylhä (2011) with the following four statements: 1) Sound does not require specialized hardware as most devices have built-in microphones. 2) Sonic gesture facilitates remote interaction. 3) Sonic gesture can work in situations where looking at the device is not possible. 4) Some sonic gestures can provide alternative means for accessing computers for people with motor impairments.

Another characteristic of sound in interactive contexts is how two levels of feedback exist when sonic gestures are performed. The sound and the sensation generated from the sonic gesture itself, and then the designed feedback from the computational system (ibid). When focusing on sound as input for interaction the temporality of sound is a vital property when considering the affordance of the sound. Related to this Jylhä (2011) presents a framework to characterize different sonic gestures according to properties of the temporality such as **impulsive, iterative** and **sustained** sounds. Different gestures provide different types of information, which are applicable for different purposes. (ibid), and this becomes relevant when designing interfaces around sonic gestures. One example is how sonic gestures such as table tapping is easier to perform rapidly than finger snaps, which demonstrated how different sonic gestures affordance different types of interactions.
**Sound: Conclusion**

Sound has physical properties in terms of waves with a certain frequency that travels through air with a certain speed. From this view it can be seen as a material with properties that we can shape with the use of other physical materials. One example of this is how foley artists apply instrumental techniques to create sounds which affordance specific perceptions of actions, gestures and materials. From this example sounds can also be seen as a material property of other physical materials, which relates to the concept of procedural sounds and how we relate to everyday sounds through the sound-producing event more than the sound itself. Sound can also easily be picked up by microphones and translated into a digital material that can be visualized, formed and manipulated through computational processes to a very high degree. And as also pointed by Ferranti and Spitz (2017), sounds can manifest emotions, contexts, cultures and memories that all ads to how sound can be used as a design material.
3. Method and Approaches

This chapter will account for the methodology approach I’ve conducted in order to do design research than has informed the overall research, reflections and analysis presented in this thesis. I’ll describe my overall stand within design based research methodology, followed by a presentation of approaches which I’ve applied parts of in my own research to explore sound and ML as design materials.

3.1 Research through design

My methodology has been guided by a research through design (RtD) approach that is characterized by employing design practices and processes as a method for inquiry to do research on the future (Zimmerman, Stolterman, & Forlizzi, 2010). Here designed artefacts and prototypes often becomes a type of implicit, theoretical knowledge contributions (E.g. conceptual frameworks, guiding philosophies) that often is framed as semi-abstract; meaning that it proposes a contribution for other designers to appropriate for other situations (Löwgren, 2007).

RtD is often seen as a way to challenge current perceptions on the role and form of technology and broaden the scope and focus of designers (Zimmerman, Stolterman, & Forlizzi, 2010, p. 311), and prior examples has showed: “…how conceiving of technology as a material allowed for a creative and inspirational rethinking of what interactive products might be” (ibid, p. 314). By sharing this view, I’ve mainly explored potentials in the intersection of ambiguous sounds and ML while extending a design space for possible new artefacts with these materials. This has been done through workshops, probes, iterative sketching (Buxton, 2007), building and expressing ideas through artefacts that has been evaluated both in-lab and ex-lab together with participants in domestic contexts.

Besides being a rich setting for sounds, the home context is chosen due to practical and ethical factors more than being a required setting for my research. Prototyping with sound as inputs requires a certain control of the soundscape for experimental control. Designing with sound should also consider the social acceptability
and the implications of contributing to noise pollution, which is a rising problem in public spaces (Jylhä, 2011). These aspects has made the home setting a good initial context for my design research to unfold and from their be guided by the generative characteristics of design based methods. These methods will be explained in the following sections.

3.2 Approaching sound as an expressive material

In order to design with sound and sonic gestures at the input of digital artefacts, I’ve applied explorative field research methods in order to unpack the potential of sounds and what sonic affordances means in design. The emergence of SID has resulted in a wide range of practices with the aim to understand sound within a design context here among the activities of Sonic Incidents, and synthesis-through-analysis.

Sonic Incidents

With a focus on sounds in the home I’ve adapted principles from the activity of sonic incidents (Caramiaux et al, 2015), which is an approach where sound, and users’ sonic experience is the starting point for where action-sound relationships are envisaged. The approach presents phases of ideation, which serves to generate ideas for action-sound relationships based on participants memories of sounds, and phases devoted to imagine possible gestural interactions with the sonic incidents. Here the focus should be on imagining what actions and reactions the sound may provoke. - Getting from the sound itself to its effect and to the actions that may cause (ibid).

Synthesis-through-analysis

Delle & Rocchesso (2014) presents synthesis-through-analysis as a exercise to open the ears by playful improvisation with objects to create sounds. The exercise is conducted as a group activity which aims to investigate the informative nature of sounds by putting the participants in a reflective state towards the affordances of sounds as well as the expressive potential. The setup requires an array of ordinary objects and recordings of sounds generated with the objects. The participants are then presented to the recordings and
asked to recognize the sonic process (what actions and objects makes
the sounds) behind the sound and recreate them with the available
items.

**Cultural probes**

My design methodology has also adapted principles from cultural
probes (Gaver, Dunne & Pacenty, 1999), which is a user-centred
approach to design research, where probes are designed and given
to participants for a period of time. The probes should be designed
to provoke inspirational responses from the participants in their
contexts as a part of conducting design research where stimulation
of imagination and inspirational data is preferred rather than
information for defining a set of problems. The main focus of
cultural probes should be on exploring functions, experiences and
cultural placements outside the norm as a ways to open up for new
design spaces (Gaver, Dunne & Pacenty, 1999).

As a part of my research I’ve used cultural probes to extend the
activity of sonic incidents, which is described above. The concept
of cultural probes has served as way to get very qualitative insights
from participants in order to use personal sound experiences as
inspiration for my design process.

### 3.3 Sketching with material: Inspirational bits

Inspiration Bits (Sundström et al, 2011) is an approach to become
familiar with the digital material as a design material. The purpose
is to open up the black box of a technology and experience its
properties first hand early in the design process as a way to become
familiar with the limits, possibilities and let these findings be a part
of the exploration that directs the design process. The concept of
inspirational bits has been demonstrated by Sundström et al, (2011)
through prototypes or bits that aimed to explore technologies such
as Bluetooth etc., and use these bits as a starting point for a design
exercise of brainstorming around the technologies.

Inspirational bits are presented as a rough way to feel and
experience a technology over time and space, and exposing it as a
material. In general these bits should be easy and quick to develop,
however a first foundational bit can take longer time, as soon as it’s
easy to convert it into different bits. The concept of Inspirational bits is about understanding the space of possibilities by exposing a technology's properties which then can inspire a space for creative thinking and guide the design process (Sundström et al, 2011).
4. Design Process

The following section will present the main design activities done in this thesis, which have been guided by the presented theoretical foundation and methodology. I’ll also account for how each activity has guided my design process further and contributed to my research.

4.1 Designing a platform for prototyping with sound and ML

Essential for my design process has been to work with a prototyping environment that enables quick explorations of sound as input to IML systems. I’ve been using Wekinator (Wekinator, nd), which is a IML tool for supervised learning that can receive inputs and send outputs over the OSC communication protocol, which makes it possible to use it in combination with a wide range of other programming environments. The Wekinator toolkit, allows for recording and training of data examples from real-time demonstrations and supports both regression and classification algorithms (Neural networks, AdaBoost, k-nearest neighbour, decision trees, support vector machines) (Katan, Grierson, Fiebrink, 2015).

Several applications for providing inputs to Wekinator already exist, including FFT sound analysers that use a computer’s internal microphone as input. From initial tests, I’ve found a lack of tools that includes the temporal aspect of sound and allows for adjustability of other parameters such as frequency spectrum and volume thresholds. To meet this lack, my first prototype has been a sound classification tool for the Processing environment (Processing.org, n.d), that makes live spectrograms of the microphone input, and sends the pixels as input data to Wekinator where the user can train a model with the data and make classifications of sounds. Creating images of sound allows to include information of both the volume, frequency, and temporal aspects as dimensions. In this way Wekinator makes image recognition on the incoming sound, which also is the basic of more advanced versions used in the industry of speech recognition (Smus, n.d).
This tool has served as a way for me to experiment with different features of sound and explore how different settings works for specific sound situations and desired outcomes. The tool comes with an interface with controls to change the analytical parameters, such as frequency spectrum, volume thresholds, and the time domain (Figure 2.1).

4.2 Exploring sound as inputs

This prototyping platform has enabled me to train sounds as inputs iteratively to control outputs of custom made experiments, which also has been made in Processing. As my initial bit, I have used this platform as the foundation for a set of experiments to explore different input-output relationships with sound at the input side. I’ll now present three main experiments I’ve conducted in order to explore sound as input.

Sound commands

One of the inspirational bits I’ve created in this environment has been a painting application sketch called Sound commands. A painting application was chosen as a context due to the large variety of different features often found in such programs that easily can be associated with physical and contextual gestures. I included several
features such as changing brush colour, increase and decrease brush size, and erase the canvas back to blank, which I’ve been able to control with sounds.

The IML part has allowed me to iterate different sound gestures to control the features and the prototype has in general acted as a way to extend the computers interface into the surroundings. Suddenly knocking on the table can change the brush to a random colour and scratching some paper can erase the canvas back to blank (Figure 2.2). Through this sketch it got very clear that the sense making when mapping sounds to function is highly dependent on the gesture producing the sound and how that specific gesture is related to existing conceptual models (Norman, 2010).

While the sketch propose new levels of multitasking by using voice and sound while operating the mouse and keyboard, I still found the sketch limiting in terms of exploiting qualities of sounds. Another property of sound is how it facilitates remote interaction (Jylhä, 2011) and that wasn’t exploited in this sketch. To elaborate on this aspect the remote interaction became guidelines for my next experiment.
DIY ubiquitous computing - Smart TV

This experiment was inspired by ubiquitous and ambient intelligence (Rogers, 2006) in smart TV’s and smart homes and with Processing I created a prototype that could mimic these systems by implementing a virtual keyboard robot into my Initial bit (Figure 2.3). Through a small interface I could train different sound gestures to execute self-chosen hotkeys on my computer while running other applications. E.g. this made it possible for me to pause, stop, increase volume, and shuffle between songs in Spotify and the same with Netflix. By connecting my computer to a Chromecast or a Bluetooth speaker, I was able to operate these applications from a distance by using self-chosen sounds. In this way I could simulate a smart TV, that responded to sounds, and I could quickly iterate the sound input to train different scenarios out.

![Figure 2.3 - DIY ubiquitous computing interface](image)

Through several iterations I experimented with how sounds could be used to stop and pause a movie playing on Chromecast with Netflix through three micro scenarios called: quick pause, never miss out, and Snack Rehab (Figure 2.4). Link to movie: [https://www.dropbox.com/s/lyw4l3xw7b6d4zc/three_experiments1.mp4?dl=0](https://www.dropbox.com/s/lyw4l3xw7b6d4zc/three_experiments1.mp4?dl=0)

Quick pause showed how the prototype was trained to turn on and off with the sound of finger snaps. Like a toggle switch, it would change
state every time the sound was detected.

*Never miss out* did use speech as input but not as specific commands. Instead the contextual indication of speech and conversation during the movie was used to pause and ensure that the users wouldn’t miss anything. When no speech was recognized the movie continued.

*Snack Rehab* took a similar approach to bring more contextual and ambiguous sounds into use, and here the sound of a chip bag was used to pause the movie. It worked as a frictional reminder that the user was eating too many snacks while watching a movie.

The overall point with the three scenarios was to experiment with sound as an information carrier to create a contextual aware system. The flexibility provided by IML to change behaviour of the system over time and context, showed to be vital to change the interpretation of the context. Pausing a movie while speaking would be good for one type of movie, but might not make sense in the context of a soccer match.

This also stresses how predefined rules would never be able to extract the contextual and often changing meaning of sounds. Here the IML empowered me to shape the functionality along with the change in context. Almost any action done in the room, which produced a sound, could affect the TV and it created a strong feeling of endless possibilities and space for exploration.
Another realisation from this prototype was how the rich expressiveness in human gestures often invited for a likewise more expressive and sustained output. Something that most home devices does not support. The simple command of shifting from “on” to “off” was strong in relation to finger snaps but to really engage with the expressive and temporal qualities of sounds my next prototype was constrained to include this and explore a more tightly connected and dynamic relation between input and output. This was also motivated by Jylhä’s (2009) suggestion to make future work on exploring continuous parameters in sound as input.

3.st experiment: Dynamic output

![Swinging the salad swing to light-up bulb](image)

This bit aimed to explore the temporal quality of sound and how it can control a dynamic output with a expressive character similar to the input. To create a bit that would express possibilities rather than utitillian usage I combined a hackable desk lamp with my initial prototyping platform and made it possible to train a sound to trigger the light bulb (Figure 2.5). To get a more dynamic and temporal interaction relation the volume of the specific sound was mapped to control the light density. One experiment used a salad swing and the sound from its rotational movement to be the input for the lamp. The faster the swing was rotated the louder the procedural sound would get, and the brighter the bulb would light up.

This experiment showed how volume of sound often follows the
force/energy source that creates the sound. In this experiment it created a feeling of a mechanical relation between the salad swing and the light source. Due to the ML it only reacted to the specific sound of the salad swing, which gave a feeling that it was the object more than a sound than was causing the light. Link to movie: https://www.dropbox.com/s/99qyq8diz3caaxn/DSC_0967.MOV?dl=0

The tight relation between physical movement and light output created an interesting interaction mainly due to the fact the two objects (salad swing and lamp) is well known objects but far from each other when it comes to original functionality and use context. By ignoring this “gap” and instead creating a strong relation between the non-familiar objects, gives an experience of creative freedom and sense-making in the situation. The rotating movement of a salad swing suddenly gives the feeling of creating kinetic energy when the lamp starts to light up.

**Reflection on experiments**

By exploring sound and IML as material through the previous sketches I got first-hand experience on how frequency spectrums, temporality and thresholds direct influences how the IML models is shaped. This is one perspective of the material of sound and IML as a composite. By exploring sounds and sound-gestures as inputs it also got clear how a material nature appears when contextual sounds becomes variables in systems that carries gestural and dynamic information about embodied interaction, as e.g. seen with the salad swing.

**4.3 Exploring the richness of sounds**

To compensate for the technical experiments explained in the previous chapter I’ve conducted two more human-centred experiments to gather qualitative and cultural understandings of how we as humans can relate to sounds in the home. I’ve conducted two workshops, the first called *Guess a sound*, which is inspired by Delle & Rocchesso’s (2014) Synthesis-through-analysis and the second called *Which sounds do you hear*, which was inspired by Caramiaux et al’s (2015) experiment of Sonic incidents, and extended with a cultural probe and a follow-up workshop.
Guess a sound

This experiment has been executed as a group workshop together with two families where the focus has been on the individuals’ experience of creating sounds, guessing sounds and reflecting on which actions and reaction a sound may provoke. One important constraint has been to remove the sonic gesture from the visual modality in order to allow the imagination to fill in the gaps and explore the affordances of sounds.

Procedure

The experiment took place at the home of the families and the general activity was to generate sounds with everyday objects and try to identify how we related to the sounds, when the sound itself was isolated from its physical form and the sonic-gesture. First the participants were told to gather 3-4 items they could interact with and bring with them (Figure 3.1). The objects was then placed on a table for everybody to see and on turn one person generated a sound by using the objects, and with closed eyes the rest had to describe the sound and talk freely about the thoughts, memories, emotions and imagination that the sound provoked. The participants were also asked to reflect on which kind of response or effect they natural associated with the sound.

We continued generating new sounds with the items (Figure 3.2) on turn until all participants including me had tried at least two times. During the sound creation and guessing I asked general questions to get a conversation going on what the sounds made us think of.

Outcome

I processed each session by categorizing the sounds based on the conversations we had at the workshop around the sounds. Inspired by sound classification schemes by Schafer (1993) and Jylhä...
(2011) I described the sounds in a list form that explained the objects, gestures, acoustics, temporal form, semantics, and aesthetics around the sounds (Appendix. 2). The main findings from this experiments and the classification was:

1) It was highly subjective how the participants experienced the sounds. E.g. some found a sound cute while another found the same sound annoying.

2) Repetitive sounds with a rhythm, was often associated with machines and craftsmanship especially the older kind of machines such as old sewing machines that normally would require a lot of manual work.

3) Sounds that were more sustained, temporal and dynamic were often associated with kids and play and one participant also stated: “The playful is often in the dynamic sounds that are temporal and varies. There follows a narrative to those sounds”.

**Which sounds do you hear? - A cultural probe**

This experiment was conducted as a probe that was given to six participants as a way to get qualitative insights around sounds in the home and the context they exists in. The probe was an A3 poster (Figure 3.3, Appendix. 3), which was placed at a visible spot in the home of the participants. The poster encouraged the participants to be sensitive to sounds and sound-producing events that occurred in the home, which could be anything from a slamming door to branches hitting the window.

The poster had three sections; the first was four guiding steps that explained the participants how to use the poster, then a set of inspirational adjectives was presented and the third section was made of four empty sound-tracking templates. Here the participants could write down the sound, the sound producing gesture and then describe each sound from a set of attributes self chosen or taken from the list of inspirational adjectives. With only 4 spots, the participants was encouraged to track the 4 sounds that they found to be the most noticeable or interesting. After delivering the poster the participants had minimum 2 days to fill out the poster before a follow-up interview sessions and small workshop was conducted using the sounds as a starting point at the participants homes.
The intended outcome from the probe and interviews was to get a more qualitative understanding of the sounds in the home, as well as understanding why the participants had chosen a specific sound over other sounds. Getting from the sound itself to its effect and the context, meaning, and personal relation that a participant had to a specific sound, has helped me throughout the process to keep a repertoire of sound events and examples of which kind of feelings, meaning and rituals sounds can be a part of in a home (Appendix 3). The following insights was also gained:

- Meaning and rhythms around sounds grow with time.

- While we can relate to many sounds and identify them, they can often have very personal meanings in each home. It often becomes a part of a ritual. E.g. how the sound of an toothbrush can be a distressing and be a ritual for either ending or starting the day.

- A lot of information through sound forms our social relations, our routines, and can make us act. E.g. how the sound of coffee brewing in the morning can indicate that its time to sit down and eat breakfast.
- The characteristics that were given to the sounds, we often more related to the actual context around the sound, than the sound itself.

- A lot of sounds in the home become so natural that we don’t notice them until we are explicit told to listen for them.

As a part of the follow-up workshop I also tested my 2nd prototype: *DIY ubiquitous computing* into the home and context of the participants where they were asked to take one or several of the tracked sounds from their poster and explore how they could be used to create interaction. Some of these sessions will be explained in the next chapter

### 4.4 Appropriating contexts

One strong quality of my DIY ubiquitous computing prototype, was the openness of the concept. Training sounds to execute hotkeys could appropriate all kinds of applications running on a computer, and it was fast for the participants to suddenly control games or other programs with sounds. As earlier experienced, one of the strong qualities in using sounds for interacting with existing application was in the remote and ambient properties. Therefore most of the experiments with participants were focused around controlling Spotify, Netflix and games in combination with Chromecast and Bluetooth speakers.

**The magic toothbrush**

One example of how the prototype was used together with participants, was in the combination of an electric toothbrush that one participant had noted on the poster. With a strong characteristic sound when turned on, we could use that sound as input for the prototype. Here the participant choose to create a prototype where the TV would stop while the toothbrush was turned on and start again when turned off (Figure 4.1). Suddenly the toothbrush was acting like a remote control and could function from anywhere in the small apartment. The participant described the feeling to be a bit “magical” and that she could see very good opportunities to prank friends with this kind of “function mapping”.

Design Process
Randomness and punishment

Together with another participant we explored Spotify on a Bluetooth radio together with a set of dices. Dices often become a symbol for randomness and has a very significant sound feedback when being rolled. Here the participant mapped the sound of a rolling dice to shuffle the Spotify playlist to a random song. While enjoying a good cup of coffee the computer was hidden away and our attention was on our casual conversation, until we wanted to hear a new song, which we could accomplish by rolling a pair of dices (Figure 4.2).
This quickly involved into a classic tumble tower game where we placed the dices on top of each other to build a tower (Figure 4.3). The person making the tower collapse would be the loser; something that the tension and sound of the falling dices really supports. The sound is often a core mechanic in the “punishment” in these kinds of games, however in our game the digital output also became a consequence of collapsing the tower.

This relates back to the quality of sound as input which jylhää (2011) highlights as the two levels of feedback that exists when sonic gestures are performed; the sensation from the sound itself, which in this game became the panic and release of tension when the tower collapsed, and the digital output that in our case changed the song. This opened a potential design space of sound and ML as input and mechanics in physical game design.

4.5 Process Reflection: Material guidance into new contexts

My design process has in general followed principles of Inspirational bits, which has allowed the properties of sound and IML to play a strong role in how the outcome and design direction has been shaped (Sundström et al, 2011, p.2). Most of my prototypes has started with an utilian purpose as seen in sound commands and the DIY ubiquitous computing experiments. However when explored with participants the chosen input often took a playful direction as seen with the magical toothbrush and the dice game. Besides finger snaps for convenient control of on/off states, it was very hard for participants to find good utilian uses of sounds as inputs that justified to the use of the technical setup. The dynamic relation that was explored in the Dynamic output sketch showed a rich interaction that had a expressive sonic input as well as a output that followed the same expressive richness. However, the experiment was conducted with no contextual setting in mind and served only to get a feeling of the technological possibilities and the relation between input and output.

These experiments has functioned as creative bits (ibid) that has triggered spin-off ideas and guided the design in a direction that aims to embraces the limitations of the materials and turns them into features that can articulate new meanings and uses of sound.
and ML as design materials. To manifest the materials findings that I’ve accumulated through theory and experimentations, my design process and material exploration has continued by appropriating the context of entertainment media for kids with a concept around interactive narratives and books.

This choice is grounded in the attempt to first of all find a context that can manifest and embrace the expressive and dynamic interaction that sound enables. The playful attitude of participants during most of my experiments has also encouraged me to approach a context beyond rational settings for letting the technological imagination grow. The context of interactive narratives is a result of this where I attempt to present the general design qualities that sound and IML fosters as design materials in a more holistic manifestation.

In the next chapter I’ll follow up on the framing around interactive narratives for kids and how previous bits has let to this as well as how the design process has continued.

### 4.6 IML and sound as a storytelling mechanics

This contextual framing is a part of following the technological imagination that the creative engagement with IML and sound as materials has fostered. My explorations around IML has really shown how the practice of end-user customization can be a highly creative and explorative one, which also have facilitated a lot of playful attitude when testing with participants. The performative act of training the IML model has itself often been a playful act that has included a lot of laughter and experimentation in the process and within the context of kids narratives I’ve seen potentials to incorporate this meta-state of training as an active part of the narrative experience.

From the Guess a sound workshops it got clear to me how rich sounds can spark imagination. E.g. was the sound of a shaking matchbox associated with rain, which also stresses the creative practice of creating sonic expressions with materials and gestures - just as seen with the foley effect. The probe of Which sounds do
you hear? also demonstrated how the home setting is a rich place for sounds that can become very personal in meaning. To sum, a lot of interesting aspects and qualities has been revealed through my design process and with a desire inspired by Balsamo (2011) I aim to present new possibilities and technological imagination, expressed in a language and through a form that adds to the reproduction of meanings of new technologies. This is another aspect of my decisions to manifest my exploration in interactive narrative artefacts for kids. These thought has let me into paper sketching sessions (Figure 5.1) to explore different concepts around narratives for kids than would use these properties of sound an IML as key mechanics.

This let to two prototype ideas that both manifest potential for more refined concepts with a communicative quality of meaning-making with the technology. I’ve called the two prototypes for Sounds of the farm, and My foley book. Both concepts have been tested with kids through three sessions with 4 kids (2 pair of siblings: one 3-years-old, one 6-years-old to times, and two 7-years-old twins). One parent to both groups of kids has signed a consent, which has allowed me to record videos and images of the kids during the tests (Appendix. 4).

The following sections will present the two prototypes, and afterwards follows an account for the testing sessions while presenting reflections and an analysis around the results and the insight gained from them.
4.7 Final Prototypes and evaluation

Sounds of the farm

The Sounds of the farm prototype is inspired by classic animal recognition books for kids (1-3 years old) where pictures of animals are showed and able to play the sound of the specific animal when touched (Figure 5.2). By switching the concept around this prototype takes form as a simple screen based interface that allows kids to give sounds of their choice to four animals by clicking on images while making the sound. Here the kids are actual training a ML model that learns to associate the produced sound to the targeted animal (Figure 5.3)

After sounds have been given to the animals a play mode is enabled where the animals are hidden until the associated sound for one of the animals is performed and first then will they appear back on the screen. This prototype is imagined as a digital environment for the kids to express themself sonical to explore the system and get a feeling of ownership and agency over the experience. The prototype was made in Processing using my first initial bit as back-end to classify the sounds.

This prototype had two main goals. 1) To introduce
the design materials of sound and IML into a context of kids entertainment to get a first hand experience how kids would react to this kind of interactivity, and 2) To explore if the concept of training the model or “giving sounds to animals” was a concept that kids understood and could get a meaningful experience out of.

**My Foley Book**

*My foley book* is the second concept and final prototype in my thesis project. The concept aimed to incorporate most of my material findings from previous experiences regarding sound and IML.

The concept is build around an interactive playbook that takes principles from the foley effect to let kids use IML as a creative tool to associate and give sounds from their surroundings to objects in the book. The given sounds is then used to start and control animations of the objects and in that way bring small stories to life that the kids can create themself by moving objects around the canvas.

![Figure 5.4 - My foley book interface](image)
This prototype was also created in Processing but is imagined to work as an Ipad app for easy mobility and touch interface. The interface consists of an item bar on the left, where the objects can be trained with sounds by clicking on them. The implemented objects are a cloud, a pair of shoes and a door. After receiving around 100 training examples (5-6 seconds) the target object will appear on the canvas and can be dragged around with the mouse. When the associated sound is performed it will control the animation, so e.g. the cloud would start to rain as seen on figure 5.4.

The volume is also mapped to allow for a dynamic relation as explored in the Dynamic output sketch, which gives a way to control the amount of rain from the cloud according to the volume level of the sound gesture. The choice of objects is a strategy to both include objects where a fitting sound is easy to find but also more abstract objects that requires imagination and creativity to create sounds to. A door is found in most homes, while a cloud on the other hand calls for imagination and exploration of sounds in the home.

The objects are also chosen to support animations that can express the three types of temporalities that Jylha (2011) accounts for being sustained, iterative, and impulsive. The cloud (raining animation) allowed for sustained temporality, the shoes (walking animation) for iterative, and while not limited to, the door (open/closing animation) encouraged more impulsive temporality.

The next chapter will present the testing sessions that have been conducted around the two prototypes and how they each has given empirical and analytical insights into a technological imagination of how IML and ambiguous sound can be used to create meaningful interaction within a context of interactive narratives for kids.
5. Testing and Evaluation

Video link:
https://www.dropbox.com/s/r1h32ezt3tvrer4/mashup%20video.mp4?dl=0

5.1 Testing Sounds on the farm

All four kids tested the sounds on the farm prototype and in general all the kids showed a curiosity and an understanding towards the system. The 3-year-old kid was able to come up with sounds for the animals as long as I facilitated the training. After a first session with the 3 and 6-years-old kids (who are brothers) it got clear that some kind of narrative was necessary to frame the training phase as something engaging.

![Figure 6.1 - Testing with the 6-years old.](Image)

To meet this in further sessions I introduced the prototype to the kids as a small narrative that was build up around the animals fleeing from the farm, and that we had to learn the animals’ languages to call them back. This clearly showed an increase in engagement from the kids and created a more holistic experience.
During the process of giving sounds to animals it got clear how many different ways a animal sound can be mimicked and that there is a certain skill set. This was especially clear with the duck where the 3-year-old said “rap rap rap” while the 6-year old had learned himself to say like Donald Duck being a much more advanced technique. This showed a strong scenario of how IML could be used to customize the experience and let the kids choose their own way to interact with the system. In a collaborative setting between the kids the IML model could also be trained to react on both sounds, which shows how IML can be used to support both inclusion and personal expression at the same time.

Figure 6.2 - Testing with 3 and 6-year-old

Recording sounds and giving sounds to animals (in reality training a model) also showed to be easy for the kids to understand. In fact they already had toys and other Ipad apps where they could record sounds and videos, and they showed awareness when a possible noisy sound was among the training examples. Another quality of IML that gets vital here is how it allows for rapid iterations if the intended behaviour is not met through the first training examples.

The openness of choosing sounds was also appropriated by the 6-years old who started to say other sounds and nick names to the animals and trained them to respond on that. A more creative idea by him came through when he asked if there was a dog in the game, because then he wanted to sing Who let the dogs out (by Baha Men)
and with that chorus make dogs appear. This really showed how the openness provided by IML and the expressive quality of sounds-gestures made the 6-years old imagine new ways of engage with these technologies in creative ways to create enjoyable experiences.

5.2 Testing My foley book

My foley story was tested in two sessions, an early version with the twins (7-years old), which only had the cloud, and a more refined version with the 6-years old with all three objects. Again I presented the prototype through a improvised narrative about the cloud which had lost its sound, and that we needed to find a new sound to it and again it showed how the activity of training the IML model could be integrated in the experience an be a playful mechanic rather than a meta state of being an oracle.

![Image](image_url)

Figure 6.3 - Twins making rain sounds with cardboard box and LEGO bricks

After presenting the task of finding a new rain sound for the cloud the two twins had a very creative and enthusiastic approach where they went to there rooms and started exploring their toys. They were discussing different objects in order to find a rainy sound and ended up taking an empty cardboard box, which they filled with
Lego bricks. Shaking it was their association to rain and similar they also broad a jar of beads which created a similar sound (Figure 6.3). The 6-years also had the shoes and the door in his prototype and besides the obvious choice of findings his dad’s big shoes to create footstep sounds he also explored a wide range of objects like an old dial phone, and two walkie talkies grinded together. It showed a very explorative and reflective approach towards objects and materials in the home to discover sounds that potentially could be used in the narrative creation (Figure 6.4).

When “giving” (training) sounds to objects the prototype worked the way that the mouse had to be pressed on the object in order to start the recording. This was a mechanism to keep the kids attention while recording, however it got clear that the kids needed both their hands to create some sound. This gave obvious insights that the interface design should support the use of two hands and maybe even the whole body in the creation of certain sounds. In the situation we solved it by team working where one kid pressed the cloud while the other shaked the object. This showed how the experience also could be a collaborative activity and that some sounds may even only be able to create with the hands and coordination of multiple people.
The same collaborative performance was also experienced with the twins when they discovered that the cloud was movable with the cursor. While one made the sound gesture the other could move the cloud around, and this simple functionality also showed a freedom that facilitated the kids to play out their own micro narratives hence their only was a cloud and small interface on the screen. The kids also understood the dynamic relation between the volume density and the amount of rain, which made the kids experiment with the shaking patterns and rhythms in order to explore the system.
6. Discussion and Reflection

The presented design process has resulted in reflections and suggestions that will be the focus of this last chapter. I’ll present reflections on my approaches and the insights I’ve gained throughout my design based research, supported by the presented theory.

6.1 IML as a prototyping tool

I’ll start this reflective section to highlight some insightful anecdotes from my process, which elaborates on Katan, Grierson & Fiebrink (2015) statement of how IML can accelerate the design process. Quick iterations and the possibility to include feedback on the spot with participants without programming has been a strong focus for I’ve aimed to adopt. The strongest example of this is the DIY ubiquitous computing tool that makes it possible for designers, and even end-users to play around and create their own scenarios of how e.g. a “smart” TV or speakers should be controlled. Using hot keys showed to be a very effective way to keep an openness supported by an interface where no programming is required; not in the ML part, but also not in the choice of outputs.

A very strong example of how this easy appropriation can foster the creative thinking is how the toothbrush was used to control the TV when tested with participants. This example opened up a design space and a new interpretation of what shape changing interfaces could be like when the toothbrush suddenly was enhanced with the functionality of a remote control. Through sound this functionality could quick be bequeath to another object or gesture that can produce a sound. This example stresses how non-speech sound can gain new potentials when paired with ML and act as an information carrier or variable to hand out functionalities and complement solutions in Internet of things and ubiquitous computing systems. As one contribution I’ve made the design tool for classifying sounds in Processing available for other designers (Appendix. 5)
6.2 Material exploration as a way to facilitate creative thinking

In this thesis I’ve conducted design based research with an initial focus on the design materials rather than a specific context or target group. With this view I’ve allowed possibilities and use situations to grow from my material exploration. This is visible in how relevant context and potential target groups has emerged and been manifested in my design experiments. To list the most prominent ones I suggest: Prototyping tools for designers, ubiquitous DIY environments for tech enthusiasts, game mechanics, and several kids entertainment media concepts. While all these areas calls for more focused and user-oriented research, which has been beyond the scope of this project, I do present them as possible design spaces to appropriate for future work.

Using the material exploration to guide the design process also relates to Sundström et al’s view on how exposing a technology’s properties as well as its technical constraints, can inspire a space for creative thinking (2011). In my exploration of IML I’ve experienced the property of active training to be both on of the biggest qualities as well as constraints. As Amershi, Cakmak, Knox, & Kulesza (2014) also points out, it can create a bad user experience if the end-user is position as the oracle for the machine. This meta state of training is often detached from the intended interaction in IML and during my “smart TV” experiments I iterated a lot and had to take account for similar sounds to the one I used. While this problem probably could be solved with more advanced ML techniques the constraints did foster creative thinking of how this limitation could be used.

This got manifested in my design of the interactive child book concept where training of the algorithm was framed as an active and vital part of the narrative. The training is suddenly used to facilitate the playful engagement of creating and experimenting with sound-gestures in the surroundings. Likewise could the at times unreliable results from a bad trained model be framed as the object in the narrative being “confused” or “forgotten how it actual sounds” instead of being a technical error.
6.3 New opportunities in SID and natural interfaces

SID suggest that sound offers rich potentials as a broader medium for interaction also beyond utilitarian usages (Caramiaux et al., 2015). This calls for both new understandings of sounds properties as well as new interaction methods and mapping strategies for interaction designers to use (Rocchesso et al., 2008). This is especially valid when sound is used as input and Jylhä (2011) also states how a body of work that demonstrates how sonic gestures may be useful as input is missing and that the field of HCI could benefit from design patterns and guidelines within this area.

To meet these lacks I’ve through my experiments presented new interaction methods and material explorations of sound beyond speech that aims to exploit how contextual, aesthetic and informative qualities of everyday sounds can gain new potentials through ML and be used to design new interactive relations between humans, the physical world, and digital artefacts. Many sounds in our everyday life has become so natural that they have disappeared into the background of our consciousness and through workshops and experiments I’ve taken a designerly stand towards sounds that often are taken for granted. By bringing these sounds to the foreground of our consciousness and explore how we relate to them, I’ve engaged in a design space than suggest new interpretations of natural interfaces and how this consciousness can foster technological imagination.

Through this design space I’ve questioned the natural in sonic natural interfaces by unpacking the complexity when both the sonic affordance and aesthetic interpretation of sounds is subjective and dynamic in relation to context, culture and other variables. This is especially highlighted in the never miss out “smart” TV scenario, that shows how the natural and context awareness is tightly connected. Referring back to critiques of natural interfaces by Malizia & Bellucci (2012), they highlight the same problematic design pattern of interpreting gestures’ meanings and predefining them. IML offers ways to rethink this, which was strongly highlighted during the Animals on the farm experiments, where the kids each had their own natural way to express the sound of a duck. To sum, I suggest new ways of designing with ML and sound that supports subjective, contextual and aesthetic experiences of sounds.
which is highly supported when the end-user is encouraged to explore sounds with tools that enables a creative practice of bridging the physical world with the digital.

6.4 Designing techno culture

A final reflection, which also encourage future work is some of the thoughts that has been manifested and provoked through the exploration of sound and IML in interactive narratives for kids. Today's kids will be the first generation to grow up with artificial intelligence (AI) as a normalized technology that is used in countless domains. AI technologies are also starting to occupy kids entertainment medias as seen with e.g. Amazon Alexa programmes for kids.

Parts of my material exploration has aimed to contribute to envisions of how design can engage in a technocultural formation that through artefacts can unpack the black box of emerging AI technologies and empower the new generation with tools and toys that position them as active creators instead of passive consumers of these technologies. The concept communicated through My Foley Book aims to make kids familiar with the concept of training computers to recognize sounds, by actively give kids the tools to apply such technology for their own purpose in creative storytelling and play. With this contribution I hope to engage in the meaning-making practice that Balsamo (2011) describes in Designing Culture where “...meanings of new technologies are reproduced, structured, manipulated, hijacked and sometimes contested.” (p.49)

6.5 Limitations and challenges

All my technical experimentations have been limited to supervised classifications of sounds. This has mainly been due to the technical accessibility and easy grasping concept that makes it suitable to explore together with end-users. However many other forms of machine learning algorithms exits that are capable of doing more advanced computation. ML is still an area that requires a lot of expertise to work with the state of the art algorithms, and my own knowledge in this area has been a limitation for how advanced I’ve explored ML as a material.
I’m also aware that the material view that I’ve adopted to guide my design process has let to designs for a wide range of context and different users. While I present conceptual prototypes in my design process, I’m full aware that designing ethically and with the user’s needs in focus, will require a more theoretical foundation on best practice design patterns and stronger focus on the users and the context that the technology.

7. Conclusion

This thesis has presented a design space with the aim to explore how sonic natural interfaces beyond speech can be imagined and how IML can support this. Both as a way to accelerate the design process, and to challenge current design patterns of natural interfaces. Through design based research, guided with a material view on IML and sounds, I’ve aimed to expand the limited body of work in this intersection and develop new perspectives for designers engaging in the intersection of SID and ML. Through a theoretical foundation I’ve accounted for how both IML and sound can be understood as design materials for interaction designers, and especially the expressive and ambiguous qualities of sounds, and how they can be used as input for interaction.

Part of my design work has been a prototyping platform, which makes it possible to classify sounds and create interactive prototypes without any prior machine learning expertise. Such a tool can be expanded in many directions where one example is my DIY ubiquitous computing prototype, which shows how IML can be used to create open and adaptable prototyping tools that can be used by users without prior programming expertise. I have also been demonstrated this through my workshops with participants, who were able to prototype new modes of expressions with interactive artifacts as seen with the magical toothbrush. I suggest future work for such tools to be conducted, especially within research of co-design methods as a way to facilitate more inclusive design processes that allows for technological experimentation without technical expertise as a requirement.

With a focus on SID and IML as design materials, my research does also suggest how the informative quality of everyday sounds can gain new potentials in combination with ML by being exploited as
a input parameter for contextual awareness systems and be away to combine the physical and digital by transforming sound-gestures into digital variables. Working with sounds as input has also showed several challenges mainly due to the ambiguous characteristic of sound. The meaning of sounds are defined from multiple parameters such as context, previous experience and culture and are not static, which makes it nearly impossible to predefine rules around what a sound affordance. I see this as tapping directly into the natural in natural interfaces and the complexity that the design ideal is facing. This is also stressed through the workshops and probes that have aimed to unpack the meaning of sounds in the everyday home.

Instead this challenge has guided my design-based research to explore sounds beyond rational usage. Here I suggested how the expressive and cultural qualities of sounds together with IML can offer potentials as interactive mechanics in a broad scope of fields and genres, here including interactive entertainment media for kids, games and potential shape changing interface. Here I have shown how qualities of IML such as active learning, can be appropriated by end-users to provide their own meaning-making to sound-gestures instead of having pre-defined solution as often seen in most contemporary natural interfaces. This has especially been visible in the sounds on the farm prototype. Testing with kids have also shown how IML and sound can facilitate creative engagement with the technology as seen with My Foley Book, which becomes a tool for creation and a way to explore new modes of expression and interactions. In this relation I’ve also suggested how the oracle state, which often is seen as a downside of active learning, can be exploited as a more integrated part of the user experience, as seen when the kids were searching for sounds in the home to drive the narratives.

Instead of working with a specific target-group or context as a starting point, I’ve used my material exploration as guidance to steer the project. My research presents several potential contexts, which encourage future work to conduct further and more focused appropriation with these design materials with a strong focus on the context and target group. With this thesis I hope to have contributed to a design space that can inspire to a technological imagination in the intersection of SID and IML that differs from the contemporary mainstream of sonic natural interfaces.
8. References


Buxton, B. (2007). Sketching user experiences: getting the design right and the right design, Amsterdam: Morgan Kaufmann.


Fernaeus, Y., & Sundström, P. (2012, June). The material move how materials matter in interaction design research. In proceedings of the designing interactive systems conference (pp. 486-495). ACM.


Löwgren, J. (2007) Interaction design research practices and design research on the digital material. Published at webzone.k3.mah.se/k3jolo on March 16, 2007


Smus, B. (n.d.) Audio features for web-based ML. Retrieved April 27,


8.1 Appendix

Appendix 1 - What’s new with Alexa email
Appendix 2 - Guess a sound - sound classification
Appendix 3 - Which sounds do you hear posters
Appendix 4 - Parent consents.
Appendix 5 - Software sketches
  Appendix 5.1 - Initial bit
  Appendix 5.2 - Sound commands
  Appendix 5.3 - DIY Ubiquitous computing
  Appendix 5.4 - Animals on the Farm
  Appendix 5.5 - My Foley Book.