HÖGSKOLAN
DALARNA

Thesis

Master's Degree

Jamming with the Plot Twister

Designing virtual worlds for improvised performances

Author: Einar Christoffersson
Supervisor: Johnny Wingstedt
Examiner: Per-Anders Forstorp
Course examiner: Åsa Pettersson
Subject: Audiovisual Studies
Course: AU3001
Points: 22,5 cr
Date of examination: 18-05-2018

Dalarna University
791 88
Sweden
Tel: +46 23-77 80 00

Abstract

Using a research through design approach, this thesis presents a first prototype of *Plot Twister*, a novel interactive, audiovisual interface specifically designed for solo musicians improvising alongside computers. This highly simplified and intuitive interface allows the performer to simultaneously manipulate visual and musical content using one single controller, thus enabling the other hand to play an instrument. The design is based on the concept of *adaptive music* used in video games, where progression of musical form and expression over time is produced through movement in virtual 3D-space. This thesis describes the design process of two interface iterations and audiovisual compositions which are first evaluated using interviews, and then through auto-ethnographic multimodal interaction analysis by stimulated recall to identify its functions and potentials in the context of improvised live performances. The results implicate different ways how the physical interface can both influence as well as enable the user to effortlessly adapt musical form and content to fit her improvised musical exploration. Both audience and performer share access to the visual interface that provides real-time visual feedback of the musical content in a “gamified” style. The technical framework strongly favours looped, continuous, gradual and non-linear progressions of computer-based media where the concept of *Plot Twister* potentially can be implemented in other interactive contexts in the future.

Keywords

interface, audiovisual, improvisation, performance, electronic, music

Acknowledgments

This thesis is dedicated to my wonderful family and the loves of my life, Elena and Joselito.

Special thanks to Johnny Wingstedt for your continuous support, feedback and inspiring conversations. It has been a true honour having had so much of your valuable time dedicated to discussing my theories and questions.

Extra much love to Nora, Oscar, Ture and Folke, who always welcomed me with open arms into their lovely home, so far away from my own.

As always, Alex saved my life and I’m forever thankful.

Table of contents

Introduction

> “When a sonic interactive system is created, it is not ‘‘born’’ until it comes into use.”
> (Stowell et al., 2009:962)

What do video game music, computer-music and jazz improvisation have in common? First of all, they are all human activities involving sound reproduction that requires highly interactive tools and systems to function, be it digital or not. Second, this thesis attempts to unite these seemingly contrasting areas in order to create a completely new way of interacting with sound and image in the context of improvised musical performances.

The result is *Plot Twister*, a single-parameter controller to manipulate musical and visual content at once, allowing musicians to perform audiovisual live shows using only one hand, thus freeing the other to play an instrument. *Plot Twister* is especially useful for improvising one-man-bands where the single musician must control sound and image simultaneously while still finding the time to play an instrument. To make the audiovisual interaction more intuitive the visual interface has been “gamified”, as it borrows conceptual ideas from video games where the players movements inside a virtual world controls the musical progression (for example: from forest to city sounds, or from battle to exploration music). Bearing similarities to visual projections in today’s popular electronic music scene, *Plot Twister*’s visual interface becomes a *camouflaged interface* to provide status of the musical processes as well as to help the performer in structuring musical form.

For today’s solo musicians in computer-based (or *electronic*) music performances, there is a lack of tools that allow them to easily control large amounts of digital content while simultaneously finding the time and peace of mind to engage in improvised interaction with their instruments and audiences. With sometimes up to hundreds of different virtual instruments and musical parameters to manipulate in real-time in order to perform a song, it seems like an impossible job. No wonder then why electronic musicians chose to perform alongside VJ:s to create audiovisual performances. Not only must the songs be written and recorded, but customised tools and systems have to be created to control all necessary sounds and effects. For the improvising musician, now turned engineer/programmer, trying to retain focus on enjoying and interacting with the music in a live performance is challenging. The long preparation time and level of technical knowledge necessary is no doubt one of the main reasons why traditional musicians are discouraged to perform with the "virtual ensembles" of the computer.

So, what can be done to help the improvising musician to go on the stage to perform digital audiovisual compositions? This thesis argues that the solution first lies in one essential step; dramatically constraining the options of controlling the audiovisual content into an over-simplified device, *so* intuitive that the musician's brain power can be dedicated to the instrumental performance, like improvising on the keyboard. But what would such a device look and behave like to easily integrate in a live performance context, while at the same time avoid losing artistic and musical value? And how could all the necessary visual feedback be offered to the performer without disrupting the interaction with the instrument and audience?

This research project attempts to contribute to the constant development of adapting technology to the human processes, rather than vice versa, and to facilitate human artistic expression with the help of computers. *Plot Twister* aims to expand the scarce selection of intuitive tools available to the audiovisual performers of the future, making the interaction with digital media more seamless and less demanding cognitively, which in turn can help shifting the focus to improvisation and other activities.

Background

The author's background as a jazz pianist, now turned producer and computer-music performer, has greatly inspired this research project. Years ago, the earliest ideas for *Plot Twister* were sprung from the frustration of transitioning from improvised jazz performances into the digital music domain. As a solo artist, many attempts were made since then to solve the issue of dealing with massive amounts of musical content in a live performance context, while still being able to play an instrument. The discovery of how music in video games were adapted to the player's interaction went on to inspire the design of a new interface, and subsequently, a new way of interaction with audiovisual media.

Improvisation in the jazz ensemble

It might be hard to imagine for a non-musician, but because improvising with a jazz ensemble is a such complex interaction, stepping inside the world of computer-based music meant having to reconfigure the author's own expression and style of composing and performing music. Georges Lewis defines musical improvisation as "an interaction within a multi-dimensional environment, where structure and meaning arise from the analysis, generation, manipulation and transformation of sonic symbols" (Lewis, 1999:101). The typical example of an improvising group in music is the jazz ensemble (MacDonald & Wilson, 2006), where the musicians can assume roles, alternating between the lead soloist and the accompanying rhythm-section in a highly interactive collaborative improvisation so that "a creative product emerges that could not even in theory be created by an individual" (Sawyer, 2007).

After the theme has been played the solo instruments take turns playing solos and the performer is "free to play over as many choruses as he or she wants" (Braasch, 2011:987), thus making the structuring of musical form dependent on cues to communicate musical intentions, eg. end of solos or beginning of the song's main theme. To facilitate expressive improvisation in the traditional jazz ensemble, a set of rules determines the framework of a complex interaction between the rhythm-section (drums, bass, piano/guitar) and soloist; eg. taking turns playing solo, within a certain key, harmony or tempo etc. These conditions set the constraints for the space of possibilities that improvisors use to interact and express themselves within. The score is "a recipe for possible music-making" (Bailey, 1992:81) where the rules set by it can be re-negotiated by the musicians at any time. George Lewis says that "the whole point in improvising together is to make the other guy sound good" (Perkis, 2008:3). Thus, the rhythm-section's main task is to encourage and support the soloist's exploration by actively adapting their style and expression. Both soloist and rhythm section are constantly adjusting musical parameters in their performances (such as tempo, key, loudness) as a response to what she sees and hears from the ensemble, in a kind of feedback loop (see *figure 1*).

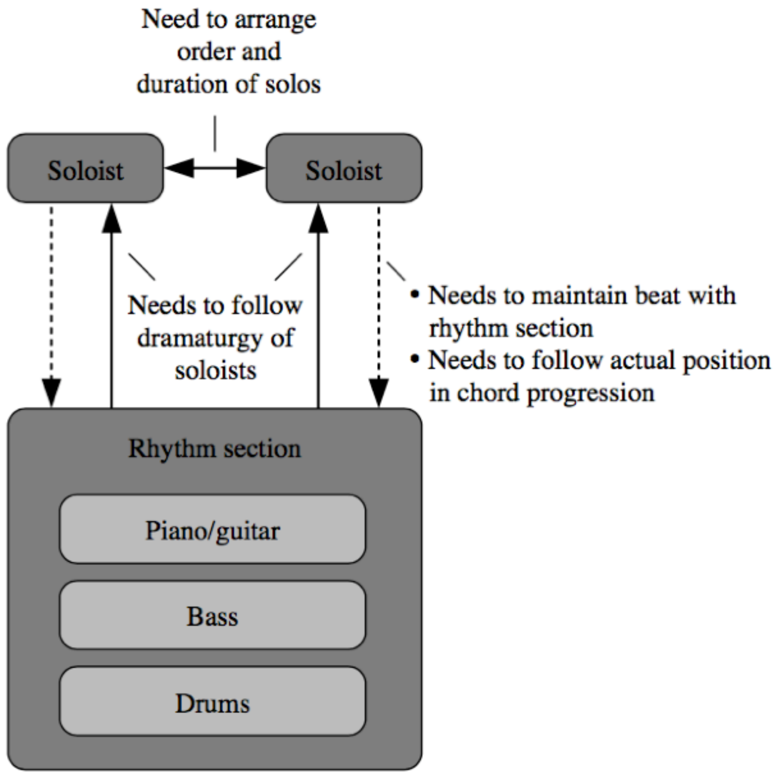

Figure 1. Schematic communication scheme for a traditional jazz performance (Braasch, 2011:988).

For experienced musicians, the adaptation is a seamless and continuous process where musicians pick up on each other's audiovisual cues. The interactions range on a scale from slow development to sudden discrete changes. The soloist can be given prompts by the rhythm section in order to initiate action, for example when engaging in a call and response type of activity. This is true for most contexts of group improvisation, where "an actor reacts to another actor's offer of a newly created symbol or utterance by imagining an interpretation and thereby creating a new offer" (Binder et al., 2011:127). As discussed above, musical improvisation in the jazz ensemble can be a demanding task, requiring full cognitive attention to one's own instrumental performance, while also being highly aware of the musical interaction as a whole, both as soloist and member of the rhythm-section.

Today, electronic music is widespread in global culture and computers are used in varying degrees in practically all musical genres, for production, recording and performance. Although producing and performing music using a computer offers seamlessly endless possibilities in sound and style, interaction with digital music is usually a one-way communication; the human controlling the machine through the interface. The backside of great technological advancements is often that the human interactions develop to accommodate it, rather than the other way around. Unless programmed to contain complex systems of artificial intelligence[1], an improvised musical interaction approaching the one of the human jazz ensemble is difficult to achieve. Is the human destined to remain in the rhythm-section, adapting her own expression to fit the computer's?

The control offered by the interface should attempt to mimic the non-linear, seamless, intuitive, gradual and continuous adaptation between the rhythm-section and the soloist, while at the same time leaving one hand free for playing the keyboard. After extensive search, no such interface seems to exist to the author's knowledge. Thus, an interface must be designed to fit the author's particular requirements. With no previously existing interface designs fitting the criteria, one solution is to look to other computer-based, interactive media for inspiration. How can musical content be adapted to a user's continuous interaction that is both non-linear and intuitive?

To imitate the real-world behaviour of sound while moving in a 3D environment, video games employs a technique called *adaptive music.* Adaptive sound mechanics are implemented to automatically adjust the selection of sounds (music or sound effects) to the player's

---

1 Some examples include the robots *Haile* (Weinberg, Driscoll & Mitchell, 2005) and *Shimon* (Hoffman & Weinberg, 2010).

interaction within the virtual game world; actions (eg. moving, talking), environment (eg. indoor, outdoor), game state (eg. exploration, battle) and so on. Similar to modern (electronic) music, most video games use loops; never-ending musical fragments that repeat seamlessly. These come in handy in interactive video games, where every player can spend as much time as she needs to complete a certain level without the soundtrack ending. Sweet (2015:62-63) describes how video games traditionally structures *dynamic music*, or “interactive scores”, according to player interactions.

Techniques similar to those used for controlling musical material in interactive video games will form the basis for the design of the interface. This in turn adds a new layer to the design work; creating a virtual space, a *visual* interface, that can be seemingly navigated through in a “gamified” type of interaction. By incorporating a virtual space into the interface, structuring musical form can be transformed into a metaphor of visual movement inside of this space. In combination with using musical loops, the musical adaptation can be done seamlessly, in a non-linear, continuous fashion, so that all the performer can be concerned with is the action at hand: instrumental improvisation. Because musical production itself becomes the goal of the interaction, and not merely a narrative element to enhance the visual experience, the roles of the visual and musical modes, found in conventional video games, are reversed. Instead, here the visuals should primarily play a supporting role by representing musical content in a *virtual* world, with the purpose of facilitating improvisation in the *real* world.

Objectives

While insisting that "technology is both the problem, and the solution", the aim of this study is to expand on the knowledge of audiovisual interface design to solve problems arising in the context of improvised interaction with computer-based music. This thesis can hopefully offer a better understanding for the type of features that such a system requires in order to make the experience more fun and seamless, ever so slightly approaching a musical interaction similar to the one of a jazz ensemble.

Research questions

1. *How can an audiovisual interface be designed to facilitate interaction with digital music in improvised performances?*

2. *What functions and potentials are afforded through the interaction with the interface?*

Method

To answer the first research question an auto-ethnographic approach to research through design method was used. The evaluations used was first done using semi-structured interviews for iteration 1, and then an auto-ethnographic interaction analysis using stimulated recall with iteration 2. This chapter describes the following methods in more detail.

Research through design

As a starting point this thesis will use the guidelines on research through design (RtD) by Löwgren (2015) when designing and evaluating the two iterations of the interface. RtD is "an approach to scientific inquiry that takes advantage of the unique insights gained through design practice to provide a better understanding of complex and future-oriented issues in the design field" (Godin & Zahedi, 2014). The design process has become an accepted means of creating types of knowledge that cannot be accessed by studying user responses alone (Gaver, 2011, Bowers, 2012). Löwgren (2015) describes RtD as requiring both "creative moments" as well as "practical design work", and emphasises that it is not the same to simply know how something could be represented, as it is to actually *create* an artefact as a representation of that "possible future"; "a statement of how something could be" (Löwgren, 2015:9).

One major limitation of RtD is seen as explicating tacit knowledge, since "designers are not used to accounting for what they know or do" (Pedgley, 2007:466). Therefore, a key point in presenting the findings, according to Löwgren, is to carefully describe, or *articulate*, the full design process in so much detail as to enable the reader to recreate the process. This is simply because in RtD, validity cannot be evaluated by the reproducibility of the results since "[t]here can be no expectations that two designers given the same problem, or even given the same problem framing, will produce identical or even similar artefacts" (Zimmerman et al., 2007:499). Instead RtD has its own validity criterion to make-up for replicability: *recoverability*. This means that the designer/researcher must make sure that "the process is recoverable by anyone interested in subjecting the research to critical scrutiny" (McNiff, 2013:18).

According to Godin & Zahedi (2014), Findeli offers the "simplest yet most elegant validity criterion of RtD", suggesting that if the resulting knowledge is good, the process is too (Findeli, 2003:170, cited by Cournoyer, 2011:55). In other words, the designer/researcher is using the project as her or his field for data collection and the validity of the choice of this field comes with the success of the design project. Simply put: in accord with Biggs and Bütchler suggests that the rigor of the process validates the outcome (2007), if the project works and the artefact produced is acceptable, then knowledge produced through the process is also valid.

RtD is not concerned with the "true", but with the "real". Jonas (2006) describes the three main phases of the RtD process as *analysis* (how things are currently, *the truth*), *projection* (how things could be, *the ideal*) and *synthesis* (how things will be, *the real*). Löwgren (2015) also divides RtD into three design phases, that can be progressed iteratively, and describes them as following:

- *Pre-study;* limiting the scope of research by identifying the target user as well as the effect desired and how it could be evaluated. Also included is to review shortcomings of similar solutions to gather inspiration and ideas. Empirical pre-studies include field studies, interviews and smaller experiments with users. Analytical and theoretical pre-studies can identify "useful terms, techniques and methods in the academic literature".
- *Design work;* based on both the empirical and theoretical/analytical results of the pre-study, more defined ideas and directions can be made. This material will influence the divergent phase of the design process, where an overview of many different solutions are analysed and evaluated, so that a few of those can be refined and developed further in the following convergent phase. Löwgren (2015:10) suggests that the relation between pre-study and design process is iterative and flowing, to allow new more relevant questions to appear and be tested again.
- *Evaluation;* here the target user gets a chance to try out the proposed solution created in the design process to provide qualitative data used in a subsequent theoretical (or analytical) evaluation. The design can at this point be anything from a primitive prototype to a nearly finished product, as long as it generates a real sense of its functions.

Interviews

The evaluation of iteration 1 was done using semi-structured interviews (Dalen, 2015) with participants as they were interacting with the composition 'Skyfall' through the interface. The main focus of the evaluation was to receive feedback on the interface, composition and interaction, where the results could potentially inform the continued design process of the second iteration (the prototype) and provide an empirical base for its auto-ethnographical evaluation. The secondary objective was to observe *how* the participants interacted with the physical and visual interfaces, although this was never explicitly explained, to minimise negative researcher effects. The participants were instructed to "think aloud" (Dix et al., 2004:343), as they were using the interface, a technique normally applied in human-computer interaction usability evaluation. The interface was available to interact with freely during the whole interview, except in those cases where technical difficulties arose.

Semi-structured interviews are commonly used in human-computer design evaluations where the participants are encouraged to discuss their experiences of using the interface in free and guided exploration (Preece et al., 2004). The interviews were made in pairs, four in total, for several reasons:

- so that the participants could take turns assuming the roles of active interacting and the spectating, so that interaction between themselves also could be studied.
- to stimulate spontaneous discussions in-between the participants as they were actively and passively experiencing the interaction, to achieve similar benefits as in a focus group. Group discussions can produce "more and different discussion around a topic and can demonstrate the group negotation of categories, labels, comparisons and so on" (Stowell et al., 2009:963)
- the control device and 15" computer screen showing the visuals were best adapted for a smaller number of participants.

All participants were students of Dalarna University in Falun, Sweden, that previously knew each other, and they were randomly selected as they would enter "Mediehuset", the main building on the campus for media production students. All four studies were done in a large, open, common area inside "Mediehuset", in the afternoon when it was almost empty and quiet. The method of selecting the participants and location of the study were highly pragmatic and made it easier to find participants within reasonable time, many of them with special interest in media production. However, the data generated generally tends to be more precise the more experienced participants are in the field of the study at hand (Berg, 2012:203).

As a semi-structured interview, the interviewer followed a small set of chosen questions (Alvehus, 2013) giving the participant greater possibility to control the conversation. The following questions were made into an interview guide (Dalen, 2015:34), used to stimulate conversation between the participants, in the hope that they themselves could reach conclusions.

- What do you think?
- What are you doing? / What is it you *think* you are doing?
- Is there a difference?
- What can this be used for?

The questions would become more specific towards the end of the interviews if certain topics had not yet emerged naturally in the discussion, regarding live performances and digital instruments. All interviews started without questions, but simply advising that the participants now were free to "try the interface". When the participants interaction or discussion came to a natural stop, open questions were asked to stimulate further interaction or dive deeper into emerging topics, as well as requesting feedback on the qualities and potential future developments of the interface. The participant's interactions that were studied focused on the types of physical movement of the controller made with their hands (which resulted in navigating the composition), listening to the sounds and gazing at the visual interface.

Ethical considerations

During this study, the ethical guidelines established by Vetenskapsrådet (2013) have been followed. All participants were informed of the purpose of the study as well as their confidentiality.

Auto-ethnography

To carry out the design work and to evaluate the second iteration, an auto-ethnographic approach will be used. Auto-ethnography is a research, writing, and storied method of self-analysis that connects the personal, cultural, social, and political (Ellis, 2004). It is a self-reflexive form of research (Marechal, 2010) that should allow the reader to step inside the mind of the researcher (Crouch & Pearce, 2012:91) for a first hand account of her socially situated experiences. Not only has research through design been established as a scientifically viable method, so has auto-ethnographical design (Ellis et al., 2010) and auto-ethnographic practices (Neustaedter & Sengers, 2012) also been introduced as valid approaches, "despite their apparent violation of the norms of scientific research" (Taylor et al., 2017:4825).

There are several ways in which so called "self-situated performance research" can lead to extending the traditional practices in HCI (Human-Computer-Interaction), as it can potentially lead to "providing new means of making sense of interactions" as well as "permitting researchers to refine and shape their work even as it is being conducted" (Taylor et al., 2017:4825). As both the practitioner, designer and researcher, one can therefore place oneself fully within the design frame, experiencing the benefits and risks to one's own practice of each element of an iterative design process.

However, as a method of studying oneself the demands for reflexivity (Schön, 1983) becomes crucial to the reliability of the results. To exercise a reflexive mindset, the researcher must take a step back, and critically examine her practice from an analytical point of view. In this thesis, critically analysing one's own practice is seemingly the only way of finding the relevant design criteria, since the chance of succeeding in creating a useful tool requires awareness of one's own unique behaviour in the context of music interaction. Because this study is focused on analysing an interactive process and physical interaction is necessary, using an auto-ethnographic becomes necessary for the researcher to get first hand data. Physical interaction with interface will be necessary to collect data about its functions and reach immediate insights that cannot be achieved merely by only studying other people's use.

Using theoretical frameworks such as social semiotics and methods such as stimulated recall becomes a way for distancing oneself from the interaction. To observe one's own interaction in a video recording and transcribe it applies a professional and analytical lens. There are obvious limitations in first designing the context for an interaction and then investigate it, as if having a fresh perspective. By implementing improvisational and generative processes, processes triggered by chance, the conditions changes from each interaction to the next, thus potentially making the pre-designed situation less predictable and open to surprising discoveries.

Additionally, by further combining roles like the composer and performer, the design process can potentially be made more effective when creating the system that will be used to perform it. In discussing the development and evaluation of complex, interactive music systems, Emmerson feels the need to address this point directly:

> Composers are often idealistic at best, at worst ambiguous about where the real control in a performance lies. This is a potential source of much frustration between performers and composers (if the two are not the same person): it is not that all performers want total responsibility for the work, but that adequate possibilities of response and control over the unfolding of the work are available to them. (Emmerson, 2009:171)

As the tools aimed to be designed in this thesis are based on the specific creative context of the author, there is an apparent requirement to possess the skills in order to properly evaluate such an artefact, in other words, one must be an expert in it to design expressive tools for it. Hummels, Overbeeke & Klooster argues for the case in movement-based interaction for HCI and meaning making (semiotics):

> ‘Interaction creates meaning’ does not only hold for users during interaction but also for designers when generating ideas and developing concepts. Therefore, we postulate that if one truly likes to design for movement-based interaction, one has to be or become an expert in movement, not just theoretically, by imagination or on paper, but by doing and experiencing while designing. In order to do so, we believe that designers need design tools, techniques, knowledge, awareness and skills that support their search for expressive, rich behaviour. (Hummels, Overbeeke & Klooster, 2007:677)

As a hybrid artist, researcher and practitioner, with skillsets from a range of multiple disciplines, the author finds himself in a unique position to provide a perspective on RtD through auto-ethnography that will not only assess the interaction through/with the created artefact in question, but to potentially expand the knowledge of the field of multimodal interaction to the benefit of all electroacoustic live musicians struggling to design meaningful improvisatory tools.

Moreover, after considering the particularity of the design goals described in this thesis, it would seem unsatisfactory to merely observe how other participants might find use for it. Perhaps it would be the focus of future research once the prototype has been developed further. Instead, the author stresses the crucial advantage of being able to design and test, perceive and analyse, as both researcher and practitioner, to provide highly specialised and critical insights on the complex, multi-disciplinary project at hand.

Finally, Simpson & Archer (2017) argues for the benefits of combining an auto-ethnographic methodology with multimodal social semiotics. By drawing insights from a study in civil engineering, they present well grounded arguments how three main aspects of research inquiry that can be enhanced, including: “a greater understanding of social context and its impact on meaning-making: an awareness of the multimodal nature of meaning-making: and an understanding of the ways in which the “interest” of the researcher (as participant) manifests in texts” (Simpson & Archer, 2017:664). One of the most important points concerns how the limitations of auto-ethnographical method (researcher effect and objectivity) can be avoided by approaching it with multimodal social semiotics; incorporating multimodal social semiotics turns the focus towards identifying underlying patterns of meaning-making. In so doing, the combination of the two approaches overcomes the issue that sign-systems quickly come to be taken as "natural", thus hiding from view the fact that they have been learned and internalised (Goodman, 1996). The challenge of becoming "blind" to how meaning-making takes place is thus minimised by making the very "blinding" process the object of inquiry" (Simpson & Archer, 2017:664).

Multi-modal (inter)action analysis

In *Analysing Multimodal Interaction: A Methodological Framework* (2004), Sigrid Norris lists types of *actions* as "communicative modes" used in multimodal interaction; *lower-level action* (smallest interactional meaning unit), *higher-level action* (multiplicity of lower-level actions) and *frozen action* (higher-level actions that are entailed in material objects; eg. recorded music). This model will suit to classify and analyse the types of interactions available to the performer, interpreted as the following:

- *lower-level actions;* performing or manipulating properties of any of the input modes available in the system, eg. speech, musical instrument.
- *higher-level actions*; overarching manipulation of lower-level and frozen actions, as well as progression through musical structure.
- *frozen action*; pre-recorded music or commands.

Multimodal interaction transcription

There are various ways of transcribing multimodal interactions, depending on the study's choice in highlighting certain *modes*, a socially and culturally shaped set of resources for making meaning, such as speech, gesture or image (Bezemer & Mavers, 2011:196). Transcription is 'semiotic work' (Kress, 2010), where the researcher represents how different modes operate alongside each-other. Transcribing actions to text results in *transduction*, the process of transforming one mode into another, "through which analytical insights can be gained and certain details are lost" (Bezemer & Mavers, 2011:196). Transcripts are partial, like any representation, making certain elements *salient* that are relevant to the academic context, "viewing the 'original' activity through professional lens which is, inevitably, different from the lens through which the participants in the 'original' activity constructed it" (Bezemer & Mavers, 2011:194).

Stimulated recall

Stimulated recall is a method of documenting the activity of interview participants through video or audio recording, and to then show them the material to allow them to comment on it (Haglund, 2003). The recorded material functions as a way of stimulating and reminding the participant of her thought processes during a certain period (Alexandersson, 1994). Stimulated recall will be used as a method for the author to observe his own interaction with the interface in the evaluation of iteration 2.

In the context of pedagogy, Keith (1988) emphasises the importance of the environment of recording, technical equipment, whether or not the participant had access to the material before being interviewed as well as the methods of interview itself. According to Calderhead (1996), the researcher must direct the focus of the study either on representing the participant's knowledge and beliefs, or her interactive thinking.

The method has been used to analyse human interaction, focusing on how events are created and in what order verbal and non-verbal communication unfolds between people (Jordan & Henderson, 1995; Goodwin, 1994). Stimulated recall has also been used as a complimentary method in combination with others to "fill the gaps" when uncertainties appear in verbal interaction analysis (De Grave et al., 1996).

Keith claims that it is hard to argue that a recording can give the same stimulus as when the interaction is happening in real-time, asking if "the viewer is reliving the original event" or if "a new event created at the time of the viewing and interview" (Keith, 1988:11). Yinger (1986) means that the participant viewing the video recording recreates a similar situation and that the subsequent interview primarily deals with the stimuli that the recorded material affords, and secondary, the actual thoughts at the time of the recording. Calderhead (1981) and Nespor (1985) adds to the list of researchers questioning the method's validity of studying interactive thinking, thus implicating clear limitations with the method.

Theory

Social semiotics

In order to evaluate the resulting interactions made possible by the designed artefact, a social semiotic perspective will be applied, with the focus on multimodality (Kress, 2010; Kress & Bezemer, 2016), especially directing attention to semiotic modes such as sound (van Leeuwen, 1999), images and gaze (Kress & van Leeuwen, 1996). Applying this theoretical foundation of social semiotics will provide effective tools in analysing the different layers of "meaning/sign-making" made possible by interacting with the artefact.

Social semiotics is concerned with "meaning in all its appearances, in all social occasions and in all cultural sites" (Kress 2010:2). Similarly, a multimodal approach provides the "means to describe a practice or representation in all its semiotic complexity and richness" (Iedema 2003:39). Multimodal social semiotics offers a meta-language for analysing symbolic forms as modes constituted by semiotic resources. At the core of this theoretical framework is the notion of the signmaker: the motivated combination of signifier and signified (Kress & van Leeuwen, 2001, 2006).

In "Multimodality, learning and communication" Bezemer & Kress (2016) discuss constraints in the mediation of technological resources, by making some modes available and excluding others. Further arguing that "one important task in social semiotic analysis is to account for the way in which sites — whether online platforms or physical spaces — shape communication and learning, by providing specific resources for making meaning and by re-presenting, ordering and arranging signs in specific ways" (Bezemer & Kress, 2016:35). Special focus will be aimed at the following key terms, including:

- *time-scale*; human interactions taking place within and across time and space, at many different timescales (Lemke, 2000), are often associated with multimodal analysis of digital environments such as music and gaming. Lemke (2005) distinguishes between interaction with media over time as either a *trajectory*, within the same homogenous media, or *traversal*, as across different media spaces.

- *affordance*; potentialities and constraints of different modes – what it is possible to express and represent or communicate easily with the resources of a mode, and what is less straightforward or even impossible.
- gaze; the direction of orientation that people display through the positioning of their head, notably their eyes, in relation to their environment.
- *agency*; the capacity, condition, or state of acting or of exerting power.
- *prompt*; a request to initiate a reaction.
- *transduction*; transformation of content from one mode to another (Kress, 2001:51)

Figure-ground-field model

In *Speech, Music and Sound* (1999), Theo van Leeuwen presents an expanded three-stage model for organising, sound that are heard simultaneously, into three hierarchized layers; *figure*, *ground* and *field*. *Figure* is always something that "the listener must attend to and/or react to and/or act upon", while 'ground' is "heard but not listened to" and only plays a supportive and minor part compared to that of the *figure* (Leeuwen, 1999:16). Field is simply the "place where the observation takes place, the soundscape", only noticeable by the ambient noise of the environment, which is not in the listener's social but physical world (Leeuwen, 1999:23). Any sound may be *figure*, *ground* or *field*; it all depends on the position of listener, the perspective. The opposite of perspective is immersion, wrap-around sound. Certain sounds like the bass are harder to tie to a particular spot and seem to come from everywhere at once. van Leeuwen's high-level concepts of sound offer tools for analysing and creating meaning, transferrable across different modes and applicable to musical parameters in real-time manipulation. It is the author's ambition to enable this thesis' emerging audiovisual performance system to act as a type of multimodal meaning maker, providing the musician with set of modes and gestures for direct interaction within the composition, such as instrumental improvisation as well as full control of transitioning in-between perspectives of *figure*, *ground* and *field*.

Gestural agency

This analysis of the final iteration will regard the interaction space between the performer and the interface as a *gestural space,* using Mendoza & Thompson's theories on gestural agency in human-computer music interaction. Their ideas can serve "as a framework to analyze musical interaction that integrate humans, traditional musical instruments and newer electronic musical instruments." (Mendoza & Thompson, 2017:418). Human performers and computer-based instruments (or 'virtual performers') are agents in a musical ecosystem that are connected by multi-modal signals (Françoise, Schnell, & Bevilacqua, 2013), signals that function as interfaces between agents (DiScipio, 2003) and "carrying information that we call *gesture*, regardless of the nature of the signal" (Mendoza & Thompson, 2017:416).

Gesture can be "visual, kinetic, auditory, or of any other kind, as long as it is information that could be perceived by an agent as meaningful and thus having musical influence" (Mendoza & Thompson, 2017:416). Here, the site of interaction for the musical ecosystem is called a gestural space, and they go on by claiming that "there is always gesture, and therefore agency, in the various communication channels between those who are making music." (Mendoza & Thompson, 2017:416).

In the interaction between the agents the production of musical sound can be seen as navigating through this space (Chadabe, 2002; Choi, Bargar, & Goudeseune, 1995; Schwarz, 2012), where "the amount of dimensions in gestural space could be large, providing the musician with a rich set of possibilities." (Mendoza & Thompson, 2017:417). Here the human can be viewed as more of a participant than a user (Kaipainen et al., 2011), and thus where the "production of signals is linked to a function, a role of each agent in relation to the other agents" (Matyja & Schiavio, 2013).

Mendoza & Thompson go on to describe gestural agency as "the influence that an agent exerts over other agents within a musical ecosystem". The extent of this influence is a means of power that an agent has on shaping its musical environment, including the behaviour of other agents" (Mendoza & Thompson, 2017:416). Mendoza & Thompson also mentions how, because of gestural agency, agents can “demand[s] action to be taken” that could lead to challenges in the interaction to “satisfy demands from the musical environment or to accomplish individual musical goals” (Mendoza & Thompson, 2017:416). Here a musical machine also can demand “a certain gesture from the human, to which the human has to respond, thereby changing internal connections to adapt to this requirement” (Mendoza & Thompson, 2017:416). According to the authors:

> …both human and machine have the potential to become agents exerting equal or unequal influence upon each other to produce music behavior is affected by the conditions imposed, demanded, or proposed by other agents (Bown, Eldridge, & McCormack, 2009; Gurevich & Trevino, 2007). The process starts with exploration and discovery of the intentions of the other agents, gradually turning into an objective-based task as a musical aesthetic emerges (Caramiaux, Françoise, Schnell, & Bevilacqua, 2014). In this way, the machine resembles an entity more a musician than a musical instrument (Van Nort, 2011). (Mendoza & Thompson, 2017:416)

Previous research

This thesis will focus on designing audiovisual tools that musicians can use in combination with their instruments in the context of an improvised live performance, a multidisciplinary endeavour bridging research fields such as human-computer interaction, music interaction and audiovisual production, among others. To the author’s knowledge, tools, or controllers, for manipulating audiovisual content in this context have not been designed, nor widely researched. Typically, the International Conference on New Interfaces for Musical Expression, or NIME[2], is regarded as being on the forefront of research in the field of interactive music systems since 2001 (and several articles from them are used throughout this thesis). However, music has so far been the focus making audiovisual interfaces much less prominent.

A good example of a research project being closely related to this study is *Cinejack*, where the authors designed a system to “allow musicians to control narrative visuals while also engaged in the act of performing”, that would be easily integrated into their existing setup, allowing

[2] NIME 2018: http://nime2018.icat.vt.edu/

"for differences in style and content" and would also "add significant value to their performances" (Schofield et al., 2014:209). Paradoxically, although being the most similar study, it uses the exact opposite approach to this thesis, the aim being to use "live music to control narrative visuals". In *Cinejack*, the designed framework consisted of software modules that could "translate musical meaning and expression from users' instruments into cinematographic actions such as edits and stimulated camera movements", where "music controls, rather than supports film". *Cinejack* was intended to "link narrative visual content and live musical performance in real time without the need for a dedicated VJ" (Schofield et al., 2014:209). The performers taking part in the interaction were not the authors themselves, thus explaining one of the most important contributions of the study being how it "stimulated a far deeper understanding of other musicians' performance practices in context, in a way which we believe will ensure the appropriateness and usefulness of the final set of tools" (Schofield et al., 2014:217). The overall feedback from the musicians were very positive, but the authors regarded the development as "challenging" since it forced them to leave "the relative comfort and convenience of the laboratory" and structure the "development process entirely around our participants' schedules" (Schofield et al., 2014:217).

In his dissertation *Improvisation, Computers, and Interaction: Rethinking Human-Computer Interaction Through Music*, Henrik Frisk discusses topics closely related to this thesis:

> What are the essential qualities of sensibility and sound in a context that includes the computer? In what ways may they change? Is something like sensibility at all compatible with the computer? Or even with the digital (as opposed to the continuous)? What is it to prove sensible towards a machine that is insensible by its very nature: Should I alter my own sensibility? Or attempt at altering the computer's possibility for mimicing or responding to sound and sensibility? (Frisk, 2008:12)

Because "the digital is static and unchangeable, trapped in its own formality", the only reasonable solution seems to be for the human, the only agent "capable of altering herself", to "approach the digital through the interface" (Frisk, 2008:93). The interface enables the interaction to take place, as it binds the digital and the user together. Metaphorically speaking "the interface *creates* interaction" (Frisk, 2008:93). Thus, the interface could potentially improve the improvised interaction by giving the user the power over the computer to mimic the continuous adaptation to her performance, similar to the rhythm-section in the jazz ensemble. Frisk calls this process *blending*; "a constant negotiation in which tone color, intonation, energy, volume, articulation, etc. has to be perpetually altered" (Frisk, 2008:10). He emphasises that *blending* is not only a mutual exchange of information between agents in the interaction, but that it happens "in between", and that it "remains hidden if the musician-computer interaction is not truly continuous" (Frisk, 2008:10).

Frisk's suggests moving the focus from Human-Computer Interaction as a methodology of control, or *interaction-as-control*, to a more dynamic and reciprocal mode of interaction, which he names *interaction-as-difference* (Frisk, 2008:88), interaction as an activity concerned with inducing differences "that make a difference". While regarding the computer as a 'player' in the musical interaction with himself he writes:

> A 'player' system approaches a circular cause-and-effect system, a container for feedback, a cybernetic system. If what comes out of it is not satisfying according to some standard—a standard imposed and set in runtime rather than, as is the case with an 'instrument', pre-runtime —it is up to either part to alter its behaviour and attempt at creating the desired change to approximate the criterion, to initiate a big enough difference to modify the balance and cause a system change. *Interaction-as-difference*. (Frisk, 2008:89)

While most of Frisk's results fits the scope of an artistic research project it still bears great resemblance to the ideas on improvisation and musical Human-Computer Interaction that has motivated the design goals of this thesis.

Design & Results

This chapter will describe the complete design process leading to the final interface prototype of *Plot Twister*, as well as the two compositions 'Skyfall' and 'Spaced'.

Design process outline

In order to evaluate the design iterations of the interface, compositions of audiovisual content are needed to interact with. Because the interface is a completely new design, all compositions have to be customised and adapted to it. Thus, the design work consists of the parallel creation of audiovisual compositions and the technical software and hardware framework that enables interaction with them. Below follows an outline of the sub-chapters in *Design & Results*, including a brief description of purpose and methods used.

- *Technology;* following the conclusions of the Music interaction chapter this next phase identifies the necessary components and resources to use for the design of the first iteration.
- *Design - iteration 1*; a full description of how the technical framework and the composition 'Skyfall' was designed.
- *Evaluation & Results - iteration 1;* using semi-structured interviews, the first iteration is evaluated. The results are discussed and summarised to form a blue print for the next design iteration.
- *Design - iteration 2;* a full description of how the technical framework and the composition 'Spaced' was designed.
- *Evaluation & Results - iteration 2;* using a combination of auto-ethnography, multimodal interaction analysis and stimulated recall, the second iteration is evaluated with a focus on its functions and potentials in an improvised performance. The results of the interaction are first analysed and discussed, and then finally summarised and illustrated using the theory of *gestural agency* (Mendoza & Thompson, 2017).

Technology

As discussed earlier in the Background section, the concept of *adaptive music* was identified as a fruitful starting point, which consequently turned the designed artefact into an audiovisual interface. To fulfil the criteria of the audiovisual design, this chapter briefly describes the necessary components for the interface design as being Ableton Live, Resolume Arena, 'ABL-Resolume Plug' and 'Multi Parameter Curves 2.0', as well as controller, mapping and transparency.

- Ableton Live 9; digital audio workstation (or DAW) Ableton Live (Ableton, 2018) offers simple tools for processing, editing, triggering and organising audio loops. It was chosen for the design due to prior knowledge and its compatibility with third-party Max-for-Live (Max-for-Live, 2018) plugins, which offer extra features necessary to create the technical framework for the design.
- Resolume Arena 5; software for visual performances used in VJ:ing[3] (Resolume, 2018), that organises, triggers and manipulates visual material. It was also chosen because of its compatibility with the same third-party plugins for Max-for-Live (Max-for-Live, 2018).
- 'ABL_Res Plugin 1.5'; a third-party plugin suit created by OOEEVV (2018), available as a free download. This plugin allows Ableton Live and Resolume Arena to communicate, enabling simultaneous reproduction of sound and visuals necessary to design the audiovisual interface.
- 'Multi Parameter Curves 2.0'; a third-party Max-for-Live plugin created by Tom Cosm (2018), available as a free download. This plugin creates a so-called divergent mapping, allowing one single parameter to control up to five others. Curves can be drawn to customise non-linear changes in the parameters over time. This plugin will allow the one physical controller to manipulate all musical and visual parameters simultaneously.

[3] VJ stands for "video jockey", derived from DJ, "disc jockey".

- *Controller*; a physical MIDI-controller (hereafter referred to as the controller) used to manipulate the audiovisual parameters in Ableton Live and Resolume Arena. There are many kinds of controllers affording different actions to the user; buttons, sliders, encoders etc.
- *Mapping*; most controllers are highly programmable and their relationship between the change in *sensor value* and the resulting parameter change is called *mapping* (see *figure 2*).

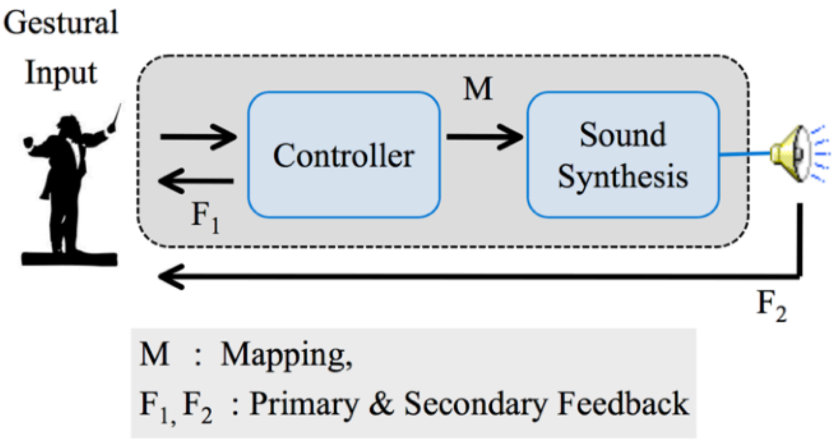

Figure 2. Scheme of the relation between the performer and sound synthesis of a musical interface (Miranda & Wanderley, 2006).

- *Transparency*; the quality of a mapping depends to a large degree on *transparency;* the player's and audience's "understanding of the mapping between the player's actions and the sounds produced" (Fels, Gadd & Mulder, 2002). The mapping, and the ease of understanding it, are therefore critical to determining the success of an instrument. The more transparent the mapping is, the more expressive the device can potentially be. The opposite of transparent is opaque, where the production of sound remains hidden to the audience and/or the performer. The below model (*figure 3*) can be used to classify instruments on a transparency-opaque scale (Fels, Gadd & Mulder, 2002), ranging from opaque for both audience and player (OO) to transparent for and/or opaque for one of them (OT/TO) or optimally, transparent for both parties (TT). Metaphor is one technique to facilitate moving from an opaque mapping to a transparent mapping, creating an intuitive and logical cognitive link between player's control effort and the audible result (or audiovisual, as in the case of this thesis). By grounding a mapping in the *literature* ('common knowledge'), it is made transparent to both audience and player. Metaphor restricts and defines the mapping of a new device and through it, transparency increases, making the device more expressive (Fels, Gadd & Mulder, 2002).

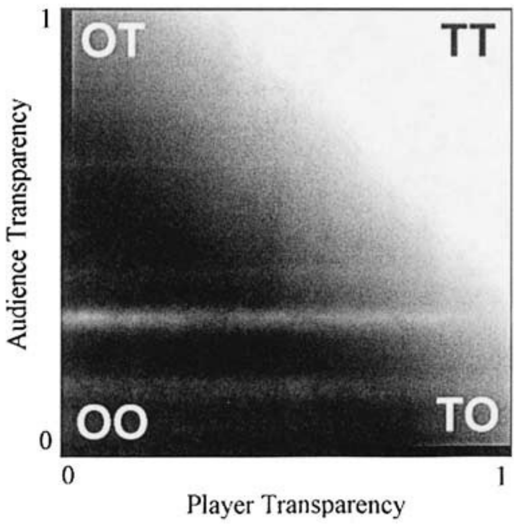

Figure 3. illustrating the transparency/opaque space (Fels, Gadd & Mulder, 2002).

Design - iteration 1

This chapter describes the design process of the first interface iteration, as well as the composition, ‘Skyfall’, that was created as audiovisual content for the interaction. The purpose of the design of ‘Skyfall’ was to:

- attempt to apply some of the ideas of a continuous and non-linear interaction necessary for improvised performances.
- establish and test the technical framework and communication between Ableton Live and Resolume Arena, where the user would be afforded basic, simultaneous navigation in a still image and looped music using a single controller device.
- evaluate the iteration in the next chapter by conducting interviews with participants testing the interface.

The ‘Skyfall’ composition contains (1) a still image, (2) music composition and (3) a controller device for navigating through it. The design work consisted mainly of timing the unfolding of musical parameters to specific segments in the image, as well as setting up the technical framework necessary for combining the manipulation of musical parameters in Ableton Live with the visual parameters in Resolume Arena. Using the ‘ABL_Res Plugin’ (OOEEVV, 2018) to establish communication between Ableton Live and Resolume Arena, one controller could be used to control parameters in both softwares at the same time.

Visuals

After the communication framework was created using online tutorials by the creator (Sanz, 2016), the focus was moved to the visuals, before the music. The reason for choosing this order of design was that, as an experienced musician, it would be much easier to create sound that fit the image, rather than vice versa. A still image (see *figure 4* in *Appendix*), taken from the Japanese anime-serie *Your name* (2016) was used, as visual material for the interface. The main reason being its dimensions; wide enough to be displayed in high resolution on a 15” screen, and most importantly, so long that moving through the image would provide different segments for parts of the music to be associated with. Starting from the top, the image consisted of three main segments; space, clouds and city. The long format of the image would allow for a longer, continuous travel, or *trajectory* (Lemke, 2005) through it, in both directions. Using a still image, instead of video material, the design time was reduced considerably, allowing more time for fitting the musical material, as well as planning the subsequent participant study. After the image was imported to Resolume Arena, the controller was mapped to its Y-axis parameter, so that the perceived movement in the image could be created by moving the image vertically, both up and down, with the speed and distance determined by the controller input.

Music

To further save time, loops from an existing composition by the author, were first assembled into three main parts — intro, break and chorus — and then mapped to the controller so that each musical part corresponded to a segment of the image. In this process, the visual properties of the image's three segments would dictate what part of the song would be heard, as far as the musical material would allow. Practically, this meant that the visual segments were combined with the musical parts; space with the intro, clouds with the verse and city, with the chorus. This was done by connecting the input gesture from the controller, that already controlled the Y-axis location of the image in Resolume, to a range of musical loops playing in Ableton Live.

Since each part of the song contained up to five different loops playing on separate tracks, a divergent mapping had to be created using the 'Multi Parameter Curves' plugin (Cosm, 2018), that would allow the manipulation of multiple parameters over time, at different speeds and timed at specific "locations" in the image. Four instances of these plugins were used to manipulate fifteen parameters of the composition. Three arcs, 'ARC1', 'ARC2' and 'ARC3', represented the three parts of the song; intro, verse and chorus, each controlling five different parameters such as sound level or effects. The 'MASTER ARC', controlled in what order the sub-arcs would be faded in (see *figure 5* in *Appendix*).

Evaluation & Results - iteration 1

This chapter first describes how the study was carried out, then sums up the most relevant results by the emerging topics, rather than separate interviews, and finally discusses strengths and limitations of the study as well as conclusions for the design of iteration 2. A short demonstration of the composition 'Skyfall' can be seen in the video[4] in the *Appendix*.

Evaluation

All interviews were done in Swedish, except #3, that was done in English, and therefore didn't need to be translated. Unless specifically noted, participants would spontaneously bring up topics and ideas without being questioned beforehand. In these cases any preceding questions are noted using R, as the alias for the researcher. Non-verbal interaction is transcribed in square brackets, eg. […].

The interviews were recorded using a mobile phone audio recorder and had average durations of 40-50 minutes, therefore only selected parts were transcribed and analysed. Notes were also taken during the study to document any non-audible observations and interactions. Because of the tight integration between sound and image in composition, it is possible to "hear" in the recording exactly how the participants were using the physical interface and what part of the image they were seeing during the discussion.

The participants were shortly briefed on the author's role, background and what context the study had as part of the master program in Audiovisual studies. They were then assured that the study followed all ethical aspects of research and that the interview would be recorded but that only their age and gender would be used when presenting the results. Once the students accepted the conditions, the physical and visual interfaces were introduced. The Midi Fighter Twister (DjTechTools, 2018) was used in the first two interview sessions, whereas the final two used the modulation wheel of a generic MIDI-keyboard. The two types of physical interfaces afford two different ways of movement in the controller; turning encoder left and right vs. scrolling large wheel up and down.

[4] See the first part of the "Plot Twister Interaction Video", link in *Appendix*.

Two interviews consisted of a particular test where the participants were asked to interact with the interface without the screen, using only the physical controller. This way the combination of sound and image could be investigated closer. Due to technical difficulties this was not possible to do in interview #2 and #3.

The following list details the order of interviews, the aliases set for the participants, as well as their age and gender. Because of anonymity, the students' specific educational programs are not detailed, but they included ongoing bachelor degree studies in audiovisual production, film studies, music production, and sociology.

Participant information in 'Skyfall' study

Interview	Age	Gender	Alias
#1	22	Male	M1
	22	Male	M2
#2	24	Male	M3
	21	Female	F1
#3	24	Male	M4
	22	Male	M5
#4	21	Female	F2
	32	Male	M6

Total: 6 Male, 2 Female

Age span: 21-32 years

Average age: 23,5 years

Results

The following section presents the results from the interview transcripts and observations, compiled related to the topics of interaction, interface, future development and live performance.

Interaction

The initial interaction phase — the time the participants first interact with the interface before stopping — varied between the four interviews, where the four groups spent a total time of 3:20, 2:10, 4:55 and 4:45, respectively. On average, the first participant of each group interacting with the interface spent more time using it than the other. Worth noting is that the second participant of interview #3 (M5), spent the shortest time using the interface (15 s.), after the first participant of the same group spent a total of 4:40 min. In each interview, the first person interacting, of the two participants, would always move all the way to the end of the composition (and controller), before attempting to go in the opposite direction. However, the style of movement varied in terms of speed, distance, rhythm, as well as the process of exploring ways of progressing the interaction.

After arriving smoothly at the bottom, without talking, using longer, continuous gestures, M1 returned all the way up at similar pace. He then started moving back and forth in smaller, irregular intervals, while trying out different ideas of progressing. He then attempted to make quicker "jumps", using faster and shorter twists of the fingers to seemingly teleport between locations in the composition, while commenting:

> M1: "If you could navigate through sections… one controller, that take you directly to one section and another that fades, for example jumping from break, to intro, to verse… maybe similar to drops."
>
> M2: "Like this you could create the entire production, jumping there and then going a bit further down".

M2 then continued using the interface, more smoothly, seemingly looking for parts that M1 not yet had explored enough. In interview #2, F1 used a much more careful pace of progression, while short time after starting asking if it was possible to “turn around”, referring to changing the viewing perspective in any other direction than vertically. When arriving at the bottom M3 excitedly asked if it was possible to go up again, and when it proved to be possible, F1 exclaimed “Wow!”, in a similarly enthusiastic tone. She then proceeded to alternate between different sections at different speeds, explaining that she was looking for “where the different parts fade into each other and are heard at the same time”, before finally identifying her “favourite part” located around the clouds, where birds could be heard. When it was M3’s turn, he progressed in a similar way, casually browsing through the composition but without describing his thoughts.

In interview #3, M4 moved down quicker than the other participants, before going up and down a few times, while “thinking aloud” and carefully describing what was represented, how & why image and sounds were combined, and speculating on a story behind it.

> M4: “…the concept that one has to transport oneself through the piece of art, gives one a different relation to the art, since you have to invest more into it. I’m pretty fascinated by things that loop and come again, so I could imagine myself dedicating a lot of time scrolling up and down. I would’ve been more fascinated if there was more modulation, in the way that, you could identify the different areas where the sections merge. As it is now the transitions feel abrupt. It would feel more like moving along a spectrum, rather than jumping between sections.”

After 4:40 of testing and describing, M4 turned to M5:

> M4: “Would you like to try it as well?”
>
> M5: “No, I think Ive seen it…”
>
> M4: “You don’t want to feel the trip to the underworld?”

M5: (laughing) “I felt it when you did it. I was the passenger.”

In the last interview, both F2 and M6, started by making faster movements, as if they were scrolling through a website, stopping only long enough to read the major headlines. At the same time, the participants would review the content of the composition and how well audio and visuals were combined.

M6: “This was the best one.”

F2: “This one was also good.”

M6: “Does the music fit with the image? - No.”

F2: “But the transitions are nice.”

R: *“What are we doing here?”*

M6: [using the interface] “From an image production perspective, I’m matching image with music, thats what I’m doing. Im moving…”

F2: “Changing perspective.”

M6: “Changing perspective. Yes, thanks. Very smart. From what I feel, [changing tone of voice] ‘Well, I like the sound in this image, then I will stop here’.” [stops moving controller]

On turning the screen off

Because of technical difficulties, only two couples of participants was subjected to trying the physical interface without the visuals on screen. The most notable result was that both participants, both the one interacting and the spectator, directed their gaze to the physical controller, while trying to reevaluate the feeling of the controller by repeating movements done previously.

R: “Is there a difference?”

M1: [using interface] “Working without image makes it less coherent, doesn’t feel as if it is one and the same thing, although the fades are working fine. If I would sit like this, I would get bored pretty quickly, or maybe not… the more I’m using it the more I get used to it. I’m trying to find sections to jump in-between quickly.”

F2: [as spectator] “Transitions are not as smooth, you don’t see or know when it will change, and suddenly it’s new. But with the image you are prepared for something changing to a new sound.”

M6: [still using interface] “It’s true… really rough transitions.” [surprised]

Removing the visual interface not only seemed to change the perception of the sound transitions, but also the physical interface itself:

M6: “For me, its the opposite now, this is now the bottom.” [pointing to controller] “I know it is the top of the image, but only because I’m turning the controller all the way back, this becomes the bottom now!” (laughing)

F2: “Yes.”

M6: “Because if you pull the controller upwards, you are also moving up, in a way. Because I’m tilting the image up, my direction is up. Now that I am all the way down, [referring to controller] I am also all the way down in the image.”

R: *“You mean you didn’t think of the inverted controller before?”*

M6: “No, because at that time I saw the visual part, like: ‘now I’m moving up, now I’m moving down’. I didn’t think of the inverted controller.”

Without the image, M2 claimed he “imagined a different story” from a movie he had recently seen, and that “it felt like the interface visuals blocked this new mental image” he had. When asked what images the participants were imagining in interview #4, F2 didn’t “see a real world” but rather “a game or fantasy movie”, whereas M6, jokingly, could “imagine sitting down in the Falun copper mine now, staring into a black wall”.

> R: *"Is it the sound complementing the image, or the other way around?"*
>
> M6: "It could be! Because when you turned off the screen, it had that effect. I heard the music in a different way, and when the image was back I got the feeling of 'understanding the sound'".
>
> F2: "It's a better effect with the image, the sound complements the image."
>
> M6: "When you see this with the image, you understand the context."

On potential applications

When asked what it could be used for, suggestions ranged from contexts such as art exhibitions and interactive installations (M1, M5, M6) to social media (M5), audiovisual performances (M3), VR/360/3D videos or images (M1, M4, M6, F2) and video games (F1). The idea of turning the style of interaction into a "Where's Waldo?" - kind of game was mentioned in two separate interviews. Suggesting that children could benefit from using the interface was also mentioned by two different groups:

> M1: "My little sister would've loved this, navigating and looking and hearing music. She's very young, she loves picture books, I think she would like this. If you could move freely, hear differences and feel like as if you're doing something… interactively."

> M6: "I have worked 17 years in kindergarten, and this could be used in education, or like a relaxation thing for kids, and adults, in my opinion."
>
> F2: "I agree, it could be used in education and get people interested in the audiovisual."
>
> M6: "I see that this would work in the pedagogic environment that exists in pre-school and elementary school ages, where you have the audiovisual and can use a physical

controller to get an experience. The learning might not be so evident, but the experience could teach something 'in the back of the head.'"

The concept of web layout and commercials were discussed considerably in interview #3:

> M4: "First thing I thought of was a page layout for a website. For example for a company that has text blogs of short stories, in a "radio theatre" kind of way. It could also work for commercial adventures, like a travel agency, showing different destinations, like: 'This is how it feels being in Rhodos in the daytime, and here is the exciting night life when the sun goes down'."
>
> M5: [agrees]
>
> M4: "A lot of web design goes to a direction of a more seamless interaction."
>
> R: *"Please elaborate."*
>
> M4: "Before when the web was first the designed you clicked a lot to move between tabs, and the transitions were very obvious. Now it feels like if design almost wants to hide that the website is segmented, you don't want the code to be visible. Today there are many websites that is using the function of 'infinite scrolling', that you never feel like you come to the end of a page and have to click yourself further, but so that it feels like a waterfall. I think that this aesthetic would fit."

M4 and M5 then both made comparisons to the, so called, 'carousel' post format on the social media platform Instagram; "a single post where the user can scroll past images and videos that belong together", the benefit being that "the spectator himself could discover it". M4 emphasised that the interaction was "undemanding" but that the users "have to interact with it for a change to happen", and that it could be done "at your own pace". Often the ideas discussed were placed in the context of advertising:

> M4: “Since it would require engagement from the user, the ad would have to be for something more exclusive, where you assume that the user has the kind of time to invest in interacting. Most ads you are just ‘thrown out in the duck pond’, and anybody can just experience it without even paying attention actively. It could work using this format, but it would have to be pretty specialised.”

In a similar fashion, one participant in the first interview, explains the idea for a music video, very similar to a video game, but where the only goal is to explore the sounds through moving in a virtual space:

> M1: “This would be a really ‘very cool’ as a music video, releasing a song as a music video, so that one could choose how the song progress. I’ve seen something similar with a character where you could walk around in a small, simple room and the music was playing differently depending on where you were located. But this is cooler [looking at screen], because depending on where you want to go the music changes, whereas what I saw only distance or reverb changed. I only used it because it was fun to play in that little world they had made, listening to the song several times, even though I didn’t like the song very much.”
>
> R: *“Why do you think you liked the experience so much?”*
>
> M1: …
>
> M2: “I think one likes having some control even if it is just a little bit, you become more interested in controlling it”
>
> M1: “I also believe that people who are not music producers can feel like as if they are music producers.”

On developing the interface

The participants were asked to suggest improvements, or the addition of new features, to the interface. When asked about the physical interface, M1 confessed that gestural interaction with an encoder "feels familiar", most likely because of his experience in digital music production, but suggested that a "joystick would be really cool, or two encoders". In the second interview, when F1 mentions a "touch controller", M3 says "VR headset", but quickly adds that "it could be messed up (laughter). Going from there and then instantly turn that way and then 'pofff!'" [gesturing quick movements with head].

M4 complains how the mod-wheel "feels a bit flimsy", and M5 agrees that "it feels loose, if you compare with a normal computer mouse scroll". M4 further adds that "when you arrive at the bottom of the controller, it stops, and it feels like the frame of the experience. But if you had a mouse scroll you could potentially do it forever". F2 thinks it "would be nice to use hands, like a Wii-remote[5]. So you can move around objects using movements of your hand" . When asked specifically about the mod-wheel, she adds that "it'd be nicer to turn a knob, instead of going up and down. We are so used to doing it on the iPhone, so getting a knob would feel new."

Suggestions for development often regarded different types of movements and dimensions in the visual interface, such as moving sideways or turning around (M1, F1), zooming and depth of the image (M1, F1, F2) as well as movement and animation of the visuals itself, independent of the spectator (M4, M6).

> F1: "If you could zoom in the city, and then go around, to look at the buildings, and maybe see through the windows. Explore more, and when you explore you hear different kinds of music. I connect it with video games where they tell a story, and you go deeper and deeper into the story. Like a puzzle game, where you press a button and you go into another reality."

---

5 The Wii-remote is a wireless controller for the Nintendo Wii video game console.

> M3: “You could do it like chapters, like first start in one kind of environment, then go into another chapter to dig deeper whats going on there, to get this audiovisual experience throughout different environments.”

Similarly, M1 mentions navigating “through different sections of one large image”, where “maybe you could have an in- and out-zooming capability. So that you can enter a part of the image that you like, and the music would change with it.” On the other hand, M4 “likes the distance”:

> M4: “I believe it would have been reduced a bit if you could click yourself in on details. It would’ve ruined the seamless feeling.”
>
> M5: [agrees]
>
> M4: “But it would maybe have been interesting to move sideways.”
>
> M5: [agrees]
>
> M4: “But that depends on if you want to explore a large image or go deeper into it.”
>
> M5: “If you go sideways, you can play with the stereo effect.”
>
> M4: “Yes, then it would feel more as if you explore something.”

Similarly, M6 seems content with the depth of the visual experience:

> M6: “I see it as a painting, and then you have to use your own fantasy to imagine what is there. I have a vision in my head that I can enjoy. I am where I want to be, both in my mind and what I see.” [looking at F2] “The feeling *you* require is to go inside the image, I already have it inside here.” [pointing to head]

When asked if any visual objects, to represent instruments, were missing in the image, like musical instruments, M2 suggests that “some type of animation would be nice”, but that “sometimes it’s good to leave some for the imagination”. The participants in interview #2 follows a similar line of thought as M6:

> R: “*Are u missing any action in the image? Do you connect music with city or do you need movement?”*
>
> F1: “I connected immediately. It gives an idea of going into the city, I pictured Rio (de Janeiro).”
>
> M3: “I got my own pictures in my head, I didn’t need the display to show me, I got to picture it myself, how it looked like in the city, people walking around, it was very lively. I could use my own imagination to picture myself in the scenario, in the town: I could see myself in there with the music.”

For the participants in interview #4, matching seemingly human expression with human representation is important:

> M6: “To see something that flows naturally, that you can relate to. That doesn’t get too stale. For example, the audiovisual performances here in the school. Really nice stuff, but many times it has been too electric in relation to the music. There hasn’t been any relation to the human aspects of the music that I can relate to.”
>
> R: *“So if you hear something human, you expect seeing it represented visually by a human as well?”*
>
> M6: “Yes, if I hear somebody running, I want to see somebody run. I hear and see it [referring to first example], but what I feel inside doesn't work.”

On movement

In interview #3, the participants discussed movement in the image beyond their control, as something that could motivate interaction:

M4: “I’m imagining a fixed background, but with smaller parts that move and develop over time, for example lights turning on, trains moving, an airplane in the sky…”

M5: “…or clouds moving in a loop…”

M4: “…to engage someone to stay, to imply that things are…” [searching for words]

M5: “…happening.”

M4: “But making it pretty slow, so that it requires you to stay and observe. But implying that there are obvious differences in watching the image now and in five minutes.”

M4: “For example, you scroll down and see something little changing, and then when you scroll back up something else has changed in the other parts. I think it gives a more organic feel, to arouse interest how things are ‘flowing’ slowly and change form.“

M5: “I agree with that. Even a small thing would be enough.”

M4: “If characters are moving or sound is turned on. Implying that something is happening, in what you see, but beyond yourself, to motivate shifting between the different levels and not only one’s own interest.”

When asked what F2 would like to do in the virtual environment, she mentions being able to “look close, go into the city” or “look closer at the comet”. M6 confesses that “computer animation would be nice, but it’s not a must” and that “it’d be like going to the museum, sitting and watching art. Getting into the image, what is the image saying?” He continues:

M6: “To make it extremely interactive, would be nice to have divine power over the image, so that you can grab the comets and throw them around.”

F2: “Yes exactly, move the clouds around.”

M6: “To create a new image.”

F2: “To feel like you can change things.”

R: *“Why is movement important?”*

F2: “Keeping interest.”

On live performance

Towards the end of the interviews, unless the discussions hadn’t already touched on the interface as being a tool for music performance, more direct questions were asked about the hypothetical situation of seeing or using the interface on stage during a live concert:

M1: “It would’ve felt a bit fake if I saw you [looking at researcher] using this controller live, unless you would’ve shown us in advance how it works. It’s hard to know the technology behind it, so I’d assume that it was pre-recorded.”

M2: “Maybe if you use a bigger knob it would be easier to understand the connection as an audience.” [laughing and gesturing in the air, as if using a large steering wheel]

The participants in interview #4 also responded skeptically:

M6: “It feels a bit cheap.”

F2: “You don’t have to give so much, just turning the knob.”

M6: “Now we are thrashing all DJ’s around the world.” (laughter)

F2: “It wouldn’t be the same experience as a live band. It would be fun to get peoples attention, but after a while I’d feel I don’t have much more to give.”

M6: “Then maybe you can use two knobs?” (laughter)

R: “*So it wouldn’t feel live?”*

M6: “It is live, but the live feeling is very thin, because it’s so electronic, it’s *too* electronic, compared to somebody who’s playing bass, guitar, drums or singing on stage, where you can see the instrument physically being played vs. something heard from a synth.”

R: *“Why is it more fun to see somebody play an instrument?”*

> F2: "Because I like seeing what I hear. If somebody hits a drum, you like seeing it too."

Both F1 and M3 says that if they paid to see a concert, it wouldn't be important to know the truth about what the artist is actually doing on stage.

> R: *"Do you think, if an artist was using this interface, that you would see the connection between the controller movement and image/sound in a live performance?"*
>
> F1: "I went to a concert, and because the artist was *so* into the performance I understood he was in fact controlling it. I felt he was really doing something, not just pressing buttons."
>
> M3: "When I'm at a concert I don't know shit what the panel (interface) does, so when I see someone turn and the music and image changes I think, 'he is only changing he music, and the image has been animated beforehand'. Maybe if you could do it in a way to make it obvious to the audience that you are controlling both at the same time".

When asking if adding a live instrument to the interface would change the experience, F2 replied that she "could imagine herself playing an instrument together with this, especially if things were moving (in the interface). It would help if you'd really want to communicate a specific feeling or a space". While it reminded M6 of some performances where, because of missing dedicated light operators, instead, the artists, would use foot pedals on stage to trigger moving lights and colours. F1 and M3 are positive to adding a real instrument to the interface but would still be wondering how it all is working.

> F1: "If there was an instrument I would think, 'OK, he's playing an instrument and changing images with the other hand', it would be more obvious in some way. But

with this, it's just like 'What is he doing, it's amazing, how is he doing it?' It's more curiosity."

M3: "Yeah maybe, but I'm not good with this kind of technology. I would still be wondering what is going on with that panel (interface), what is this person actually doing on stage. I would still be questioning."

M4 prefers the traditional roles in the live collaboration between the musician and the VJ, but finds the concept of control afforded by the interface interesting:

M4: "I would want sound and image to be controlled separately, not that somebody else decided the combination beforehand. For me to see it live on stage, I'd like less ready-made content. Image and sound could be ready-made separately, but what I see in real-time is how they are combined. The event would then focus on somebody trying to combine the two elements."

R: "*Could you imagine yourself going on stage and using this to perform music?*

M4: "I think that if using this for DJ:ing, one would definitely want much more nuances between the levels and the possibility to move sideways, to modulate it a bit more. It could be an innovative way of controlling how music is moving from beginning to end, and then it would definitely be interesting to move sideways. Vertically, it's not different states of song, but rather different songs. In that way I don't feel like I am moving from beginning to end, but rather levels, so that if you had more control of the structure on the separate levels as well, like a more 3D coordinate system. Musical parameters changing sideways could be changing tempo or key, or pre-recorded music material. It's like a colour circle; with different nuances going into each other, where you're not creating the separate segments but rather how you place them, managing the transitions of the song elements. It'd be fascinating; instead of being a song creator, being a song *navigator*."

Discussion of results

At the beginning of the interview sessions, when the participants were asked to "try it out and think aloud", without being asked questions, many reviewed the composition in terms of a combination of music and image, attempting to identify their favourite section. One of the reasons for this could be that the participants expected this to be the purpose of the study. Furthermore, as media production students in the university, it is common practice to critically analyse audiovisual productions as a method of learning. The participants' initial focus on exploring and giving feedback on the audiovisual content made the physical interaction with the interface easier to observe by the researcher, and confirmed the importance of the decision to not initially inform the participants that the purpose of the session was to observe *how* the interface was interacted with.

Interaction

During the interviews, the interface was normally first used to explore the composition, and then to quickly navigate between certain parts of special interest or preference, to support discussions and debates in-between the participants. The types of gestural interaction, or *navigation*, varied in style between all participants. Movements were alternating between slow progression, smooth or rough, random or precise jumps and quick browsing or calm exploration.

The intention behind the composition was to create a continuous *trajectory* (Lemke, 2005), a seamless navigation in-between parts of the same song, gradually moving from intro, to verse, break and chorus. However, this progression and organisation of musical structure was perceived very differently by the participants. Some perceived the transitions between the parts as rough and therefore viewed the composition as consisting of several levels of isolated songs, an interaction more similar to a *traversal* (Lemke, 2005). Often extended control was requested to afford more options for interacting with a particular segment, rather than to achieve a more 'smooth spectrum' of progression. To other participants, the experience appeared more seamless and sometimes too smooth. Here, alternatives to the gradual progression was instead suggested by dividing the composition into chapters or to enable quicker jumps from one location to another.

Sound and image

In all interviews, the participants understood the interface as a means for presenting visual content, and thus regarded the interaction to be with the image itself, rather than the music. More revealing than hearing *what* the participants were thinking, was *how* they were explaining it. They would speak as if they themselves were moving inside the virtual world of the image, and not simply manipulating the position of image. Furthermore, the physical controller was described primarily as a tool for changing the image (eg. vertical movement, navigation in space) and sections of the audiovisual composition was almost exclusively referred to in terms of its visual content (eg. around the clouds, in the city). When the sound was mentioned, it was many times in debating whether it had fulfilled a secondary function of fitting the image, rather than the other way around. Seemingly, the image afforded a ‘visual map’ of the musical parts, from where visual associations and language could be derived and used in discussing the composition as a whole. Not surprisingly, the image seemed to contain more familiar and salient features (such as city, clouds, stars), compared to the music (with the exception of birds), that in turn provided helpful points of reference for the discussion. Thus, the musical process of transitioning from verse to chorus, instead became a progression of metaphorical movement from above the clouds, down into the city.

Screen on/off

In the two interviews where the screen was disabled half-way through the interaction, both the active and passive participants would start looking at the physical controller instead. Participant F2, as the spectator, then perceived the transitions as more sudden, speculating that the image must have prepared her in anticipating the changes in the sound. In the same interview, M6 surprisedly explained how the perception of direction in the composition had changed completely after realising that the controller mod-wheel was in fact inverted, which undoubtedly also seemed to make the image-less interaction less intuitive.

Once the visuals were turned back on, participants would raise their gaze back towards the visual interface, and would not look back at the physical controller unless discussing material aspects of it. For M6, the previously so blatant discrepancy between the gestures of the inverted controller and the visual interface, then ceased to be a noticeable issue. Previous to the screen on/off test, none of the participants would recognise the practical aspects the image afforded in visualising musical structure. However, when the image was removed, its functions were made clearer as a way of helping the participant's in anticipating future musical events or giving context to the sound. Here it's fitting do draw parallels to today's electronic music live performers, as they too use their controllers and machines as interfaces for similar reasons as just mentioned. It would seem that, following the implications of this study, once the more preferred interface was missing, the physical controller itself became the second best source of visual feedback available, both to the active performer and the spectator.

New features

The different preferences of interaction soon prompted requests on how to develop the interface further. All participants, except M6, expressed a wish for additional ways of interacting with the composition, visually and/or musically. Often, the additional features, such as manipulating other musical parameters, were described as increased freedom of movement within the virtual world to explore more; by moving sideways or go deeper, by zooming in and out. Others, seemed content with the interface not revealing too much to instead let their own imagination do the work.

Movement did not only involve intentional gestures by the spectator, but moving of visual objects in the screen itself, as it could create "interest", according to one participant who also said she would be more inclined to use the interface if "things were moving". Some requests were made for visual movement beyond the user's own control, especially where the scene's appearance would change over longer time, to motivate a more detailed move-stop-move type of exploration. Interestingly none of the participants mentioned the most salient generative component, the kick drum, that would continuously recreate rhythmical patterns and pauses according to predetermined amount of chance of execution.

Musical performance

Based on knowledge, interest and personal experience, suggestions were made for a whole range of potential contexts of applications similar to that of a concert. Overall, the interactive aspect remained an important feature; being able to navigate yourself and being in control, where the user (child or adult) is required to actively take part in creating the experience through free exploration. As mentioned earlier, the interface was more often referred to as a visually based interaction than a tool for live music performance, unless when the participants were being explicitly asked about it.

As a hypothetical scenario, the participants were asked to imagine and rate the controller in the context of a live performance, but it was rejected as being “cheap” or “fake”. It was speculated that the simultaneous manipulation of image and sound would have to be explicitly apparent to the audience, or the experience would remain a mystery. In the same hypothetical live context, the interface was criticised as the performer would not have to make an effort. Although, the interface’s evidently lacking transparency was expected to be regarded as a weak point by the author, its effortlessness was not. In most cases, however, it was seen as a strength, enabling amateur users to feel like music producers or DJs, where instead of creating the musical content, they could assemble its pieces by assuming the role of “song navigator”.

Strengths and limitations of the interview sessions

Without reaching any solid conclusions, the study resulted in a wide range of feedback and interaction styles, providing the researcher with diverse insights, implications for improvements and new ideas for the continued design process. One of the most apparent insights regarded the instability of the technical framework, as it would crash in two of the four interviews, after having been used actively or left idle for around 30-40 min. Since previous lab-tests normally would not exceed 15 min of continuous use, this input was valuable for when pursuing longer live performances or installations that could span over several hours.

The selection process limited the participants to media production students at Dalarna University, with access to a technical institution. A majority of the participants were aspiring to become, or were already, working in the media industry, who would form part of the target group of a final prototype. By studying students with a potentially better knowledge of media production than that of the average person, the feedback received during the interviews were often detailed and applied in their own fields and interests.

When choosing interviews over surveys, a chance was missed to allow participants to criticise the design anonymously without fearing to hurt the feelings of the researcher/designer. Many times, overwhelmingly positive feedback suggested a situation where the participants intended to please the researcher. However, more often, the participants showed seemingly genuine interest for the ideas they had themselves generated when giving feedback on how to develop the design further. This researcher effect may thus have had lesser impact when observing how the interaction was done with the interface.

Interviewing the participants in couples benefited self-generating discussions in-between them without much researcher interference, as well as being able to observe a small scale performer-spectator interaction. However, as evident in the results, in three of the four interviews, one participant would speak more frequently and interact with the interface considerably longer than the other. Perhaps this could have something to do with the test setting, where the most dominant participants would normally be seated next to the researcher and was also the first one to take initiative to start the interaction. Perhaps, to reduce group dissymmetry, a better placement of researcher could have been being seated in-between participants or behind the screen. Since the initial interaction was the most interesting to describe, as the second person then would have pre-conceived knowledge of the interaction at hand, the first participant would also dominate the presentation of the results.

Possibly, because the first participant had already thought aloud when testing it, the second would feel less prone to share her own view if it would contradict the first. If the study had focused only on the interaction, doing the video recorded interviews one at a time, it would have allowed for a more detailed investigation of each individual's personal interaction, without influence from a co-participant. However, this technique would potentially have its downsides too, such as a need for longer interview sessions and more active involvement by the researcher in gathering verbal feedback.

Conclusions evaluation iteration 1

To create a basic design blueprint for the second and final iteration, this chapter will briefly summarise six of the main ideas discussed in the previous chapter's evaluation of 'Skyfall', and suggest ways of implementing them. These points were mainly selected to fit this thesis' focus on musical improvisation, but also on the amount of time necessary for the author to acquire the necessary skills and resources for implementation.

- *Challenge and/or additional ways of interacting with the composition, beyond simply navigating through it.*

As mentioned earlier, oftentimes music was perceived to fill a secondary function, even though the author's intention for designing the visual interface is to facilitate live improvisation in computer-based musical performances. In the 'Skyfall' study, when M4 explains his preference for the classical combination of VJ and musician, it is motivated by the artistic values of the challenge of human performers attempting to fit one with the other, in real-time. But since this thesis primarily focuses on the visual interface's role as serving musical performance, the next logical step would be adding a musical instrument to extend the interaction. Now, the challenge is instead focused on fitting one's own musical performance in a process similar to *blending* (Frisk, 2008), using the new instrument, with the sound of the "virtual ensemble".

Here the originally intended use of the interface comes further into focus; representing digital, musical content audiovisually, perceived to occur simultaneously, for the purpose of facilitating improvised musical performance, in such a coherent and precise way that no VJ or musician could achieve together. This could potentially allow the improviser to assume the additional role of, as M4 put it, a song *navigator*, rather than composing all virtual instruments in real-time, as the song *creator*.

- *The composition could contain a larger degree of 'mystery' or actions beyond the performer's intention, that change over larger time periods, which in turn could create interest and motivate longer interactions*.

In order to potentially stimulate and influence musical improvisation, both music and visuals could be given generative features; triggering of events based on probability. Depending on the complexity of these rules, the audiovisual results of these processes would remain unknown to the designer, and could thus offer an interesting way of creating surprise and interest, motivating longer interaction and exploration of certain static locations in the composition.

- *Using moving pictures or animation, instead of a still image, to better represent the movements in the music and thus possibly create more interest.*

To further increase interest in the improvisational interaction and approach the real-life human ensemble setting of gestures generating sound, moving images could be used. Resolume Arena, which normally deals with videos, contains a rich assortment of tools and effects for manipulating the material. When transitioning from a static image to moving pictures, not only will it be challenging finding visual material that can represent its musical counterpart, but moreover; the video needs to allow for being infinitely and seamlessly looped.

- *Increased freedom of movement in the virtual world as spectator, directed inwards or forwards, thus requiring a 3D environment.*

To create the sensation of depth in the visual interface, as in our own three-dimensional physical world, a single 2D video would not suffice. Instead a collage of multiple objects would be necessary to create relations in space over time, to reflect the musical elements' progression in the soundscape.

- *Gradual musical progression of a single song, that let the performer anticipate future musical events, seemed to fit the interactive format better, as sudden jumps were harder to achieve by the participants.*

The virtual interface needed for the second iteration, so far (in theory) consisting of looped audiovisual material moving in a three-dimensional world, had to be set in an environment that could represent continuously evolving musical progressions, where one could see the status of the song as well as anticipate future events by simply looking at the interface. The song composed for this format would have to allow for infinite and seamless looping, similar to 'Skyfall', and this time, to use Lemke's terms, emphasise the musical navigation as a *trajectory* (2005); movement within the same homogenous media.

- *Need for a more stable software framework to reduce risks of system crashes.*

When reviewing the list of improvements above, it becomes clear that the computational demands will increase with the higher complexity of real-time graphical rendering. Thus, more than before, the underlying reason/s for the software crashes appearing after 30-40 min, of active and idle use, had to be investigated and identified to allow for an interaction long enough for the final evaluation.

Design - iteration 2

This chapter describes how iteration 2, ‘Spaced’, was designed. A preliminary design blue print for iteration 2 is based on the conclusions made and discussed in the previous chapter “Conclusions iteration 1”. The design process was in many ways similar to ‘Skyfall’, but instead added 3D models, visual movement while reducing the musical content.

Software

In the hope of reducing crashes due to unstable third-party software, ‘Multi Parameter Curves 2.0’ (Cosm, 2018), was replaced by the more recent plugin ‘MultiMap Pro 1.0’ (RutterStudios, 2018), which offered practically the same divergent mapping functions, but increased the total amount of simultaneous parameters from five to six. Unfortunately, this change did not affect the overall performance since the heavy graphic processing would lead to constant crashes. A work-around was to take turns designing the audio in Ableton Live and visuals in Resolume Arena, and to then check the synchronisation by running both at the same time. Because of the heavy processing needed, this was only possible to do in shorter intervals, while keeping an eye on the status of computer to avoid overheating. Only being able to design in short intervals, with the constant risk of crashes and loss of unsaved progress, proved to be time consuming and frustrating at times.

Controller

Instead of the keyboard’s mod-wheel, the Midi Fighter Twister (DJ TechTools, 2018) was chosen as a physical controller, because of several reasons:

- more precise in selecting values of much higher resolution.
- the encoder provided some friction when turning it.
- easy to adapt its location to optimise the performance setup.

To make the gestural movement of the physical controller translate smoothly into changing values in the software, the third-party plugin 'Smoothie' was also used here (see *figure 6*) for all arcs. Here a 'ramptime' could be set between 1-9999, deciding at what speed the change in values would transition[6]. The value was set low, generally between 100-400.

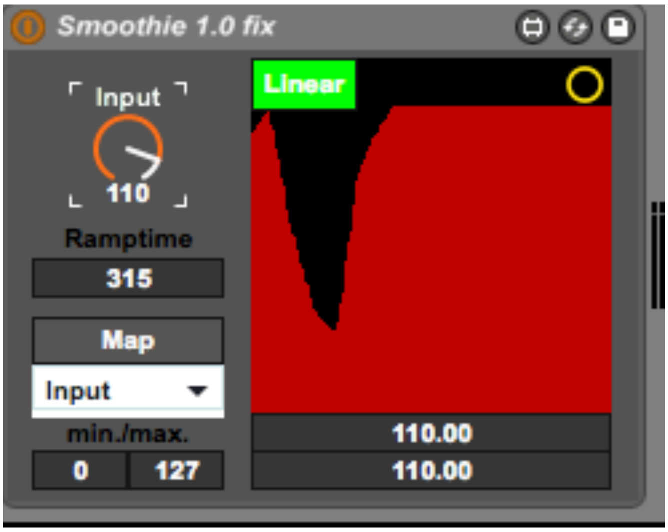

Figure 6. Smoothie plugin.

To make the experience more like a *trajectory* (Lemke, 2005), the aim was making 'Spaced' a more homogenous environment, set in outer space, with much less diverse audiovisual content overall, and with less musical material than 'Skyfall', but instead more visual objects. One of the reasons for reducing the music material, as well as real-time effects, was to save CPU processing power, which had been linked to crashes.

The goal of the audiovisual design was to translate a musical composition into a visual, virtual world, in which each object clearly corresponded to a sound in the song, where the performer could structure the musical story form through navigating from start to end in the visual scene. Moreover, the musical content would have to be designed to offer space for improvisation, thus preparing segments with minimum melodic content, where the soloist's performance could be the main character.

[6] Experiments were done with 'Smoothie' to create something like an auto-pilot, where the performer could indicate input a gestural movement from the beginning until the end of the composition that would take several minutes, thus allowing the performer to be transported without further interaction.

The most important function of the visual interface was not to create a stunning experience, but rather to function as a practical and easily over-viewable map of the musical composition, that would facilitate "reading" the status of it at the glance of an eye. Basically, to fulfil the need of an interface in a traditional sense, similar to other existing physical digital music controllers; organising and making digital music content accessible in a performance context, without the need for looking directly in Ableton Live, using mouse and keyboard.

Audiovisual content

The audiovisual design process started by finding a visual story, and then designing sounds to fit it. The reason for this was simply because the author, who had much more experience in sound production, had very limited knowledge of creating visual material, and was thus limited to existing resources, available for free online.

First, a large collection of visual material (videos, gifs, 3D-models, still images) was collected and categorised according to certain story locations that they would fit into, for example; forest, ocean, car and so on. Then, the material was tested and manipulated inside of Resolume Arena where basic ideas for an audiovisual story would start to form. Finally, one scene was selected and developed further. The decision to select 'Spaced' was very much motivated by the availability of 3D-objects relating to the space theme. Because of its high compatibility with Resolume Arena, 3D models created in Quartz Composer (Apple Computer, 2018), could not only be imported, but also manipulated in complex ways. This useful feature make QC-patches highly customisable compared to standard, pre-rendered video material, which do not allow for specific parameters to be adjusted independently. One could say that these models allowed Resolume Arena to function as a basic 3D-designing software where objects could be placed and, some of its features from Quartz Composer, changed over time.

To link the physical controller to changing both musical and visual parameters over time, the 'MultiMap Pro' plugin was used. Similar to the design process in 'Skyfall', one 'MASTER

ARC' (see *figure 7*) module was first created to control other sub-arcs, as well as three parameters of the performer's keyboard input ('track volume', 'reverb level', 'distortion'). But instead of organising each arc into parts of the song (intro-verse-chorus), as when designing 'Spaced', one Arc per audiovisual element was used; 'Star Drone', 'Beat Ball' and 'Sun Choir'. This was a more apt way of organising the parameter curves, since all audiovisual content were continuously controlled from the beginning to the end, and not only in some parts of the composition. In some cases, parameters from another object arc would have to be moved to another arc because of lack of space.

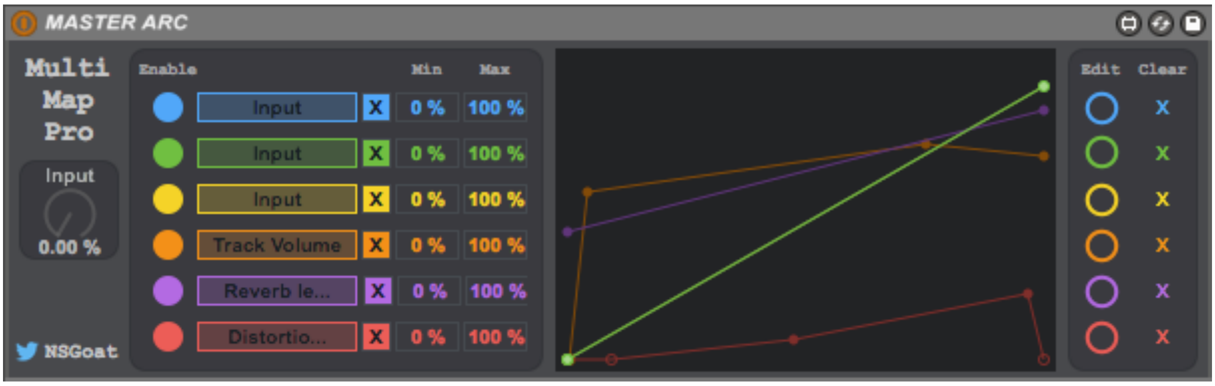

Figure 7. The 'MASTER ARC', where the first three curves follow the same linear progression.

An important first step was approximating when and how long the parameters would change over time, as an example; how long distance does the 'Beat Ball' need to be faded in? When will the 'Sun Choir' first be seen? When should it climax before the end? These decisions came down to an essential bottleneck of the composition; the physical controller's resolution of 128 steps (0-127). Because the minimum gestural input was +/- 1 step, and maximum 128, the whole audiovisual story had to fit within this digital range of values, but more importantly, the physical range of the encoder. As a reference, the eleven LED lights were used on the Midi Fighter Twister controller to get a visual idea of when events should be triggered, for example; after turning the encoder a distance corresponding to one LED light, the 'Star Drone' would fade in quickly. After that the 'Beat Ball' would use 3-4 LED lights to fade in slower etc. This was an important part of the design phase, since the distance of the encoder corresponding to the physical movement of the fingers would ultimately afford the interaction more or less precision in fading in and out sounds. This was clearly shown in the 'Skyfall' study, where some participants commented on the rough transitions between the song's parts.

In ‘Spaced’, the progression would need to be more gradual and discrete to create a more seamless experience. Before describing the audiovisual content in detail, here is a brief summary of the story of ‘Spaced’ in chronological order:

1. Start: No image, no sound.
2. ‘Star Drone’ is introduced.
3. ‘Beat Ball’ fades in slowly from the distance to then remain in the bottom right corner towards the end.
4. At the same time, ‘Sun Choir’, is slowly approaching.
5. After arriving much closer, ‘Sun Choir’ overtakes the sound and image completely.
6. End: No image, no sound.

Star Drone

The Quartz Composer patch ‘starfield’, combined with an ambient drone-like sound created the first element, ‘Star Drone’, and thus set the location to outer space. The seemingly never-ending ocean of randomly located, approaching stars creates a sensation of constant forward-moving, rather than a stationary, spectator, even when the physical controller is not used. Thanks to Resolume Arena’s high compatibility with patches created in Quartz Composer, some 3D parameters, such as the speed of the stars and the rotation, could be manipulated and were set to low values. The sound was created using a one minute, looped, synth pad sample called ‘numbed.wav’[7], also used in ‘Skyfall’, but now pitched down four semitones, with the intention to simulate the continuous, low frequent noise of distant stars and galaxies passing by.

Using the previously described bridge between Ableton Live and Resolume Arena, volume of the sound track and the ‘opacity’ of the ‘starfield’ layer could be mapped to the same physical controller via the ‘Layer-Dashboard’ module, so that both sound and image were faded in and out simultaneously, using the ‘MultiMap Pro’ plugin (see *figure 8*).

---

[7] The origin of this sample is unknown and has been in the author’s possession for many years.

Figure 8. 'ARC STAR DRONE' also contained the filter frequency for the Beat Ball, as it did not fit its own arc. Note that the sound (blue, 'track level') and image (green, 'Dash1') are not completely aligned, as it was required to create the sense of audiovisual connection.

Beat Ball

The second element that can be heard is the 'Beat Ball', a sphere-like, white object rhythmically bouncing to the sound of the drums. Using the 'Audio-Layer-Position' module, its movements were triggered by the sounds of drum-kit and the distance of the movement is directly relative to the sound level, so that a louder sound triggered a longer offset jump on both the X and Y-axis. This feature is especially relevant since the 'Beat Ball' is the only object representing a rhythmical performance.

The 'Layer-Dashboard', parameters such as 'opacity', 'scale' and 'anchor z', were used make the object appear, come closer and grow in size. These changes were directly linked to sound parameters such as 'track volume' and low-pass filter, simulating the natural lack of high frequencies in sounds further away. The instance of 'MultiMap Pro' (see *figure 9*) also controlled the 'dry/wet' parameter for the reverb on the whole composition, that towards the end makes all the sounds mix together, including the keyboard input, as the Sun Choir grows to dominate the image and create chaos.

The 3D object 'brainsplitter'[8], also created by 'eighteight' in Quartz Composer, allows Resolume Arena to manipulate several parameters such as size, offset, length, diameter and rotation on the Z-axis. Using Resolume Arena's automatised Timeline feature, these

---

8 Available for free download here: https://github.com/eighteight/theatre-assets/tree/master/Goodbye%20Quartz%20Compser

parameters can be set to slowly vary over time, thus creating a constantly morphing and turning movements, as if the object was flying through space.

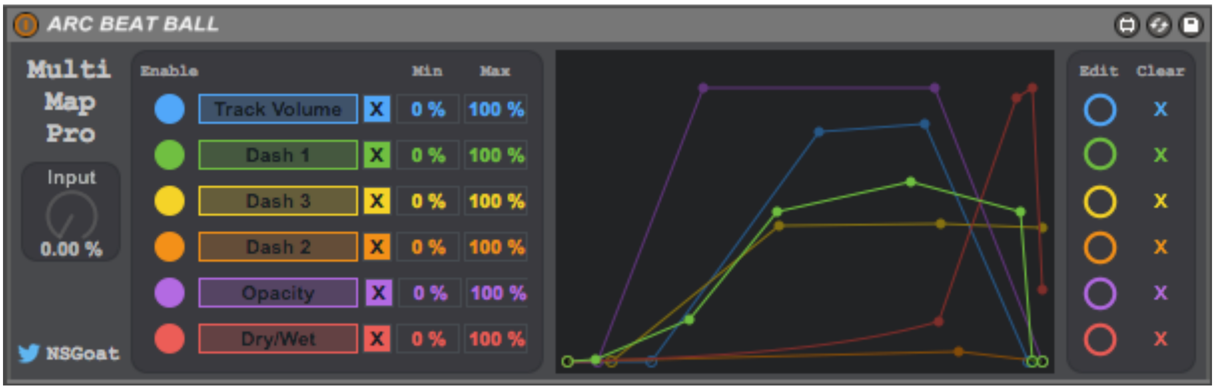

Figure 9. 'ARC BEAT BALL'.

The drum samples were taken from the author's personal library of sounds and were then triggered by *Stochas* (Clissold, 2018a), a stochastic MIDI-sequencer plugin for generative music, that lets you program probabilities for certain musical events to be executed, which in turn can create variation and unexpected results. As seen in *figure 10*, *Stochas* consists of sequencer, where the probability for a MIDI-event can be adjusted for each event, allowing for complex rhythms to be created.

MIDI-event can be adjusted for each position in time, in this case 16 steps per measure of a 4/4 bar. For 'Spaced', the drum samples (kick-drum, rim-shot, clap1, clap2 and hi-hat) listed from the bottom up, have all been given unique probabilities to be triggered, with higher values on the 1st, 3rd, 5th, 7th row (and so on). The red boxes represent the events with the highest probability of being triggered (from 85%-98%), kick-drum and hi-hat, to keep a simple rhythm going most of the time. In-between, on the 2nd, 4th pulse (and so on), yellow (and orange) events adds extra rhythmical variation and syncopations. Note that some events have a chance of being triggered as low as 1% (one in one hundred), such as the last yellow kick-drum event before the next red event. Furthermore, no event has a probability of 100%, thus not guaranteeing any drum hit to be played every time. As a reference to jazz music, the drums were given a basic, jazzy touch, using a salient cymbal rhythm and snare drum syncopes, typical of swing music.

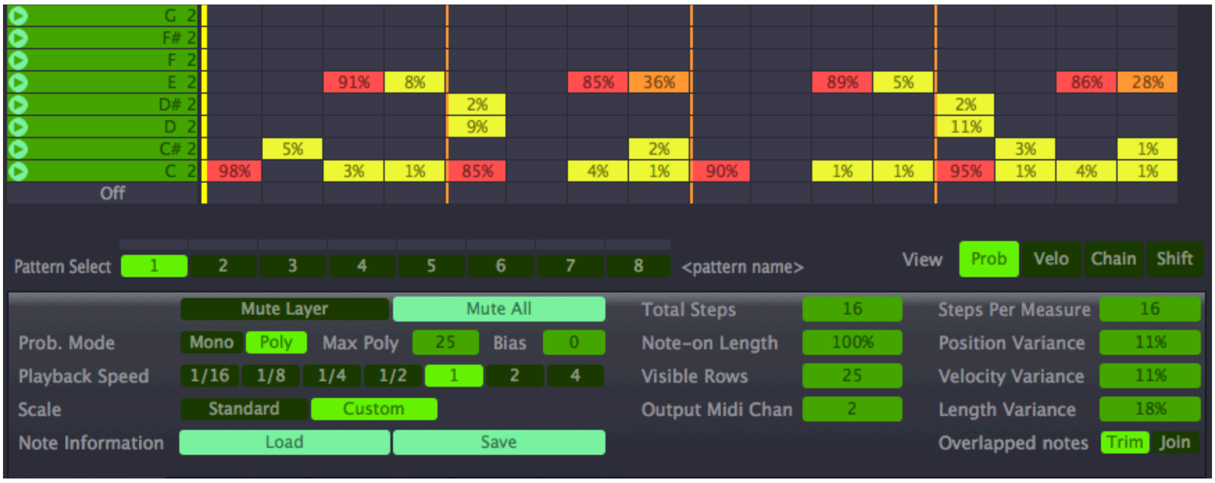

Figure 10. *Stochas* interface.

Also velocity and individual shifts in milliseconds can be set for each MIDI-event. In order to create even further variation in timing, intensity and duration of the individual notes, *Stochas* can also apply small position, velocity and length variance, here seen set to global values of 11%, 11% and 18% respectively, in the bottom right corner. The *Stochas* user manual explains that "the variance options add a humanised feel for the layer by changing the position start, velocity and the length of the note, by any value between a + and - of the number chosen" (Clissold, 2018b). Which in the case of the given position variance above, the start position will vary randomly between + / -11% of a step length.

Going deeper still, *Stochas* allows for setting the probability of more complex chains of events, for example: if event x triggers, only then proceed to event y. Negative chain lines can be drawn too, which instead declare the opposite: if event x triggers, then do *not* allow for event y to trigger. In 'Spaced' this function has been applied to all yellow kick drum hits, to keep the amount of extra kick drums to a minimum and avoid potential situations where all trigger in the same measure. The endless options for generating variation in rhythm and 'musical expression' in *Stochas* could potentially reduce the robotic repetitiveness of digital drums and stimulate a slightly more humanised feel, resembling a jazz ensemble interaction.

Sun Choir

Although being the second object to appear visually, but instead last to be heard, next up is the "Sun Choir", consisting of a burning three-dimensional sphere, accompanied by the constant sound of multiple female voices singing overlapping notes and melodies. This was the only object created from a seamless video loop (originally converted from the file 'sun.gif'[9]) that, using the Resolume Arena plugin 'stingy sphere', could be transformed into a spherical 3D object. The sound of the sun was created by combining eight separate tracks of song performances by four different female singers, previously used for other personal music productions. The individual singing performances were played back using a generative process similar to the one popularised by Brian Eno in the 1970's; in Ableton Live, all recordings started playing at the same time, but would, because of their varying individual durations, loop back at different points, thus creating a seemingly infinite amount of unexpected combinations over time. Additionally, different ranges of rotation in all three axises (X, Y, Z) were slowly changed over time to create barely noticeable movements in the object, in an attempt to somehow match the random, generative processes happening in the music. The visual parameters 'opacity', 'scale' and 'sphere size' were mapped to the physical controller together with the musical parameters 'track volume' and 'frequency' of a low-pass filter (see *figure 11*), to simulate an approaching object, similar to that of the Beat Ball. To create the explosive and burning characteristic of the climaxing sun, the parameter 'extrusion', in the 'stingy sphere' effect, was used to its fullest, coinciding with the maximum volume level and full frequency width of the voices towards the end of the story.

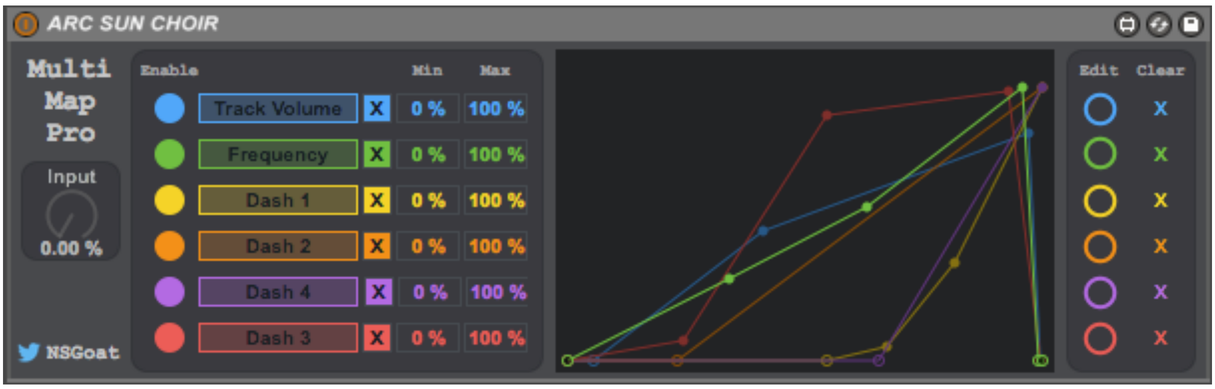

Figure 11. *'ARC SUN CHOIR'.*

9 Available for free download here: https://gph.is/1kKoOua

Evaluation & Results - iteration 2

Evaluation

To evaluate the second iteration, the author performed and filmed an improvised interaction with the composition ‘Spaced’ (see link in *Appendix*). A projector was used to display the visual interface on a wall, located in front of the performer, keyboard and physical controller. The camera was placed so to represent a potential audience’s perspective (see *figure 12*). At times, the low lighting made the physical interaction with the controller hard to see, but can best be understood by identifying gradual changes in audiovisual parameters, size of objects or sound level.

Figure 12. Image from the video recording of the improvised interaction (link to full video in *Appendix*).

Using stimulated recall, the multimodal interaction was transcribed and analysed. The goal of studying the interaction process was to present some functions and potential interactions that the tool can afford in an improvisational context using a keyboard. Observations and excerpts from the multimodal interaction transcript will be presented and discussed as examples of features of the interface. The modes focused on, was the performer's physical interactions and gaze, as well as the audiovisual results of the interaction.

Results

This chapter will first present and discuss the interactions observed from the video observation (see link in *Appendix*) by exemplifying some functions of the interface which affords the user to influence the interface, and vice versa. Finally, the chapter is summarised and discussed in relation to the theories of *gestural agency* by Mendoza & Thompson (2017).

Space, time and form

By using a single, physical controller, characteristics of the pre-designed audio and visual modes, or *frozen interactions*, can be manipulated by the performer. As seen in the video, the first action afforded by turning the controller encoder is initiating the composition; both sound and image appear simultaneously. The visual interface generated through Resolume Arena functions as a visualization, or virtual "embodiment" of, the digital music processes reproduced by Ableton Live. Here, the different elements of the music are represented by animated 3D-objects in a virtual space. The user is given certain power to influence the composition. In 'Spaced', story time is malleable, and it allows the user to continuously travel through it, forwards or backwards, stopping at any point, or *moment*, frozen in story time, where one can spend any amount of real-world time. Changing story time is done by turning the encoder, resulting in effects on the audiovisual content contained in the visual interface, that can be viewed both by the audience and the performer.

A story "is a series of events recorded in their chronological order", where as "a plot is a series of events deliberately arranged so as to reveal their dramatic, thematic, and emotional significance" (Burroway, 200:39). Before showing the filmed interaction with 'Spaced', the video (see link in *Appendix*) shows an one minute demonstration of the composition from beginning to end, without playing the keyboard. If this, "original form", is regarded as the *story*, then the performer's own path, deviating from it in any way, can be called a *plot*. In reality, the story is a set of audiovisual parameters programmed to change in different ways, depending on incoming values from a controller. The plot is then *how* the user chooses to manipulate these parameters over time when navigating between two points in story time. Physical interaction with the interface is required for a plot to be constructed out of the story.

In the study, the interface was used as a tool for structuring form in an improvised musical performance. By progressing forwards slowly in short intervals of story time, the performer spends around one minute before reaching the "virtual drummer", 'Beat Ball'. After consistently having progressed the story time forwards in shorter intervals, at around 2:37, the linear story line is deviated from, by going back in story time at a faster rate than before, thus returning to an earlier moment in the story. Here, the move in opposite direction functions as a way of framing a musical part, for advancing the plot into a new segment of the musical form, where a "post-theme" melodic phrase is performed.

Scale of control

The moments in story time can be imagined as being spread out over, and limited to, a continuous spectrum of 128 values (0-127), the maximum range afforded by the interface encoder. In the example of 'Spaced', the single controller enables the performer to seamlessly browse through the unfolding of 18 parameter settings used to construct the audiovisual story. The scale of control is best demonstrated at the end of the interaction, starting at 3:46, when the performer changes all parameters of the composition at the same time using only one encoder.

Instead of pressing buttons to activate specific segments of the composition, all material is constantly being played on a loop 'behind the scenes', and the controller enables a way of 'scrolling' through the content, fading them in and out over time. Thus, this type of interaction benefits gradual musical progressions of looped material, rather than unique harmonic progressions that cannot be repeated.

Although the controller mapping affords the simultaneous manipulation of multiple parameters, during some parts of 'Spaced', only a few of those are perceived to change. For example; when revealing the 'Star Drone' (0:02), slowly fading in the filtered rhythm of the 'Beat Ball' (0:25-0:54), or decreasing the intensity of the 'Sun Choir' slightly, but while still keeping the drums sounding (2:35-2:45). In these situations, because of how the mapping was designed, the performer is temporarily given more specific control of a reduced number of parameters, resembling the precision that one would have using dedicated controllers for single parameters.

Keyboard interaction

Because the controller interface required only one hand in use, it enabled the performer to improvise on a keyboard with the other. Therefore, as reflected in the multimodal transcripts, the context of the interaction with the interface is best understood in relation to the keyboard performance. All keyboard performance were improvised in the moment, including the recurring variations of the "theme". During the improvised performance, interactions with the controller can be divided into two types: using while simultaneously playing the keyboard with the right hand (ie. 0:36, 2:18, 3:46) and alternating between the actions, in a pattern of playing-moving-playing (ie. 0:26, 1:07, 2:07), which occurred more often when two hands were used to play the keyboard. The two actions were used interchangeably throughout the whole interaction to continuously adapt to the constantly re-evaluated direction of the improvised plot.

The types of gestures used to interact with the controller during the performance were most commonly directed forward in story time, slowly and short distances (ie. 1:07, 1:58), and less often; long (ie. 2:18, 3:46) or fast and long (ie. 2:35). In-between these actions, the performer remained in a 'static' moment of story time, normally from a few seconds of listening and playing, up to longer intervals (0:58-1:07). The performer's appearance and physical actions also form part of the visual performance, as playing the keyboard requires physical gestures that are visually observable. This interaction is not visually represented in the interface since the physical interaction with the keyboard is naturally embodied.

Interface as conveying information

The screen displays visual representations of the virtual instruments that otherwise would remain hidden, affording the representations with visual characteristics such as colour, shape, size and movement, according to the designer's own choice. During the interaction process described above, oftentimes the performer would look at the screen while adjusting the controller. In this context, a function of the visual interface was to convey information or status of the musical processes in the composition. An example from 'Spaced', is the 'Beat Ball', whose movements, triggered by sound intensity, is perceived as the bouncing to the rhythm of the beat. By gazing at the screen, these visual cues can contain musical information such as pulse and tempo, and potentially contribute to a more holistic impression of the 'virtual performer'. The results of the generative processes, triggering musical events in Ableton Live, is here given a visual representation, thus making the visuals generative too, although features afforded by *Stochas* are not available in Resolume Arena by default.

Because the *figure-ground-field* model (van Leeuwen, 1999) is transferrable across semiotic modes, the audiovisual objects (3D-model + sound) can be designed to correspond to similar, perceived locations in their respective soundscape and virtual space (as an example; if the 'Sun Choir' sound is *field*, then the visual representation should be too). During the interaction, the performer could potentially anticipate events and gather information from the screen about the composition as a whole or separate parts of it, without having to rely on the LED-feedback on the physical controller. In the case of the Sun Choir in 'Spaced', its visual size is designed to correspond to its sound intensity, thus allowing the performer to use a visual mode for gathering a better understanding of the audible mode, and potentially, vice versa[10]. *Transduction* (Kress, 2001:51), the process of transferring meaning from one mode to another, is also how conventional music interfaces work; ie. a loud noise triggers a red light on the mixer board, signalling clipping. But in 'Spaced', the visual feedback afforded by the interface is presented in a format that is more accessible than the technologies controlling it.

[10] In the perfectly designed composition, one deaf and one blind performer could perform equally well.

Shared interface

The hypothetical audience's gaze is, in the above described video recording, represented by the camera's view, and therefore share a view similar to the performer's. The controller is the most discrete object, and cannot easily be spotted in the video. Moreover, connecting the interaction with the controller with audiovisual changes on screen are not as evident as the keyboard performance. However, other than potentially giving the audience a better understanding for how things are working (a transparent mapping, as opposed to opaque), the controller's primary function is to translate the physical movement into digital. As mentioned, the audiovisual interface is the performer's most important source of audiovisual feedback from the composition. Thus, although the performer and audience (but to a lesser extent) can see the controller, it is not used as an interface for visual feedback. This separation is made possible due to the disconnection between controller input and source of resulting audiovisual feedback, an essential characteristic of digital musical instruments and controllers that is not shared with conventional, acoustic instruments. In 'Spaced', combining the — to the performer — informative screen, and the — to the audience — aesthetic, visual experience into a type of *camouflaged interface*, that is equally visible to all parties, can potentially unify the performer and audience in raising their gaze, away from the opaque technology on stage.

Perceived interactions with the virtual ensemble

A true musical interaction, similar to that in a (human) jazz ensemble, is impossible to achieve without applying a complex form of artificial intelligence. In 'Spaced', exerting influence through musical or visual modes is a one-way communication, directed from the audiovisual interface to the human performer. A key concept in improvised musical performance is the soloist's (or soloists') status as *figure*, the main character, and rhythm-section as *ground*, supporting the soloist's musical exploration. In performances with larger orchestras, the conductor's responsibility is to guide and shape the collective musical expression, such as tempo, dynamics, emotion etc.

In 'Spaced', the roles assumed by the performers can be fluid or static. The composition has, by design, certain pre-determined conditions for how the improvised interaction can unfold, but enables the performer to assume the role of the conductor, thus allowing for these roles to change over time, using the controller. This way, the performer is given the ability to gradually adapt the conditions for the improvised interaction, by simulating a limited range of gestural responses that an intelligent (human) rhythm-section normally would need no assistance in doing.

The one-directional interplay of roles and communication can be demonstrated in the performer's interaction with 'Spaced', by applying the *figure-ground-field* model (van Leeuwen, 1999) to the active agents; both virtual and human. Any single sound being heard alone automatically becomes *figure*, as in the case of 'Star Drone' at the beginning of the composition. As the first notes are played on the keyboard (0:08), the agent now assumes the position of *figure*, as its sound is more salient and therefore demands the attention of the listener. At 0:54, the drums of the 'Beat Ball' has approached from a *field* position, close enough to be heard loud and clear, ready to assume the role of *ground* or *figure*, depending on the performer's keyboard input. Here, without progressing the story time, two opposite scenarios unfold over 15 seconds, where the generative musical processes contribute to a switch between the (perceived) roles of performer as soloist (*figure*) and 'Beat Ball' rhythm section (*ground*):

> (00:56) Looks down on keys, as if planning next melody line. Looks up again at the Beat Ball. Performer remains passive.
>
> (00:58) Suddenly, a loud snare drum hits on a syncope, for the first time in the interaction.
>
> Performer thinks for moment, then looks at keys, moves hands into new position and hits two notes once (1:00), attempting to imitate the intensity and timing of the snare drum.
>
> In that same moment, 'Beat Ball' hits the same snare drum syncope twice in a row (1:00).
>
> Performer looks for new position on keyboard, then plays a single chord twice, once again intending to imitate the timing and intensity of Beat Ball's. (1:02)
>
> Performer smiles and looks up at 'Beat Ball' again.
>
> (excerpt from multimodal interaction transcript)

Because of the performer taking on a passive role and leaving space (00:56) after having improvised already as *figure*, a sudden salient snare-drum hit produced by 'Beat Ball' is perceived as an intentional *prompt*, a call to initiate a reaction[11]. The performer responds (1:00) by imitating the snare-drum hit, and thus engages in a call-response interaction. Immediately after, to the performer's surprise[12], the 'Beat Ball' seems to respond by playing the same snare-drum twice, thus initiating a new call, whereto the performer once again responds through imitation. Even if just for a short time, because of the perceived call and response interaction, 'Beat Ball' assumes a lead role, as *figure*, dictating the conditions for the musical conversation, seemingly 'demanding' the performer to adhere to its new rules of engagement. For the performer to not imitate, or engage in the call and response interaction, would be to ignore the 'Beat Ball's perceived attempt to interact.

Just a few seconds later (1:13), the "same snare hit is played once again" but is this time ignored by the performer "in favour of completing a melodic phrase already started". Here, the roles have changed compared to the previous interaction at 0:58, and, with it, the conditions for interaction; the performer has been actively playing since a few seconds, long enough to assume the role of soloist as *figure*, made possible as 'Beat Ball' has kept playing a less salient rhythm since the double hits at 1:00. However, when that same snare drum sound returns at 1:13, it is not produced by the (perceived) lead role in a call and response interaction anymore, as 'Beat Ball' now has assumed a supporting role to the performer, in a more conventional and static soloist/rhythm section relation.

These are two examples demonstrating how generative music could potentially contribute to extended interactions with the interface, while staying in the same moment of story time. These type interactions strongly depend on chance and context, which is constantly changing, and therefore could've occurred differently for any other musician, at a different time. Although moments of perceived intentional agency were brief in the studied interaction, simply knowing there is a chance of it occurring, encouraged the performer to pay closer attention to what was being played by the other virtual agents.

---

[11] The drum hit is triggered by the generative drum sequencer *Stochas* and has a 5% chance of being triggered.

[12] The relatively low chance of both events to trigger in the same round were 1% and 5% respectively, which further reinforced the surprise and excitement in the performer.

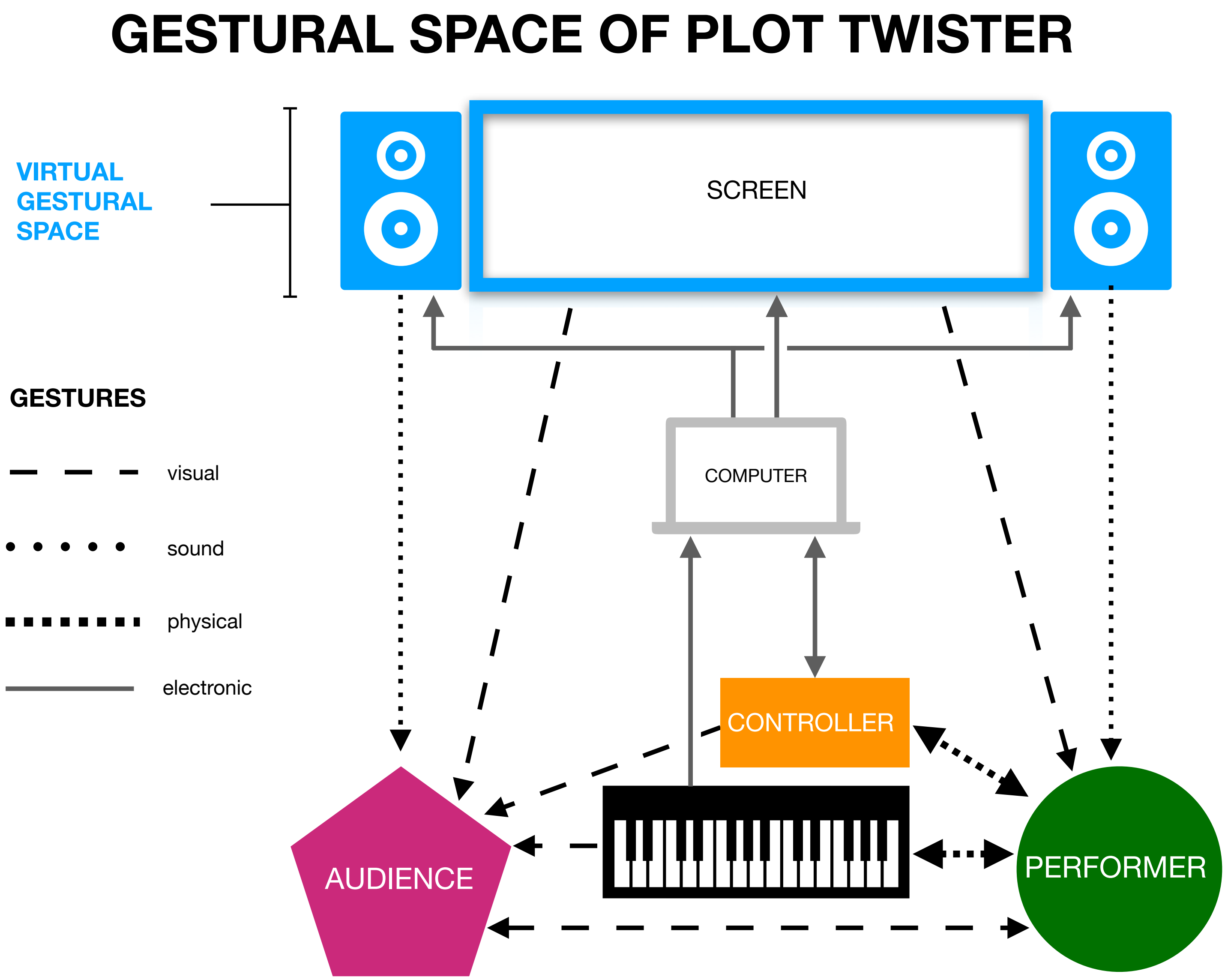

Figure 13. Scheme of Plot Twister as a *gestural space.*

Summary of interaction as gestural space

To summarise the findings from the previous section, an illustration of the interaction site, or *gestural space*, is created and discussed in relation to the theory of *gestural agency* in human-machine musical interaction (Mendoza & Thompson, 2017:412). In *figure 13* above, the interaction with the interface has been organised in two gestural spaces; the physical space of the interaction itself and the virtual space contained within it, represented through the screen and loudspeakers. The audiovisual interface (screen and loudspeakers) contains the gestures of the computer-based, virtual instruments, and can therefore be called a *virtual gestural space*. To produce audiovisual content, the interface enables the performer to use physical gesture to navigate the virtual gestural space, constructing a plot out of the story material. The perceived navigation is created by changing *perspective* (van Leeuwen, 1999:12) of the audiovisual gestures in the interface, in relation to the spectator.

The influence that are exerted by the agents in the musical ecosystem are unequal, since only the human agent can perceive audiovisual gesture as meaningful, as opposed to the interface, which is limited to digital input. In the interaction with 'Spaced', the interface could both *have* gestural agency — in the ways it influenced the performer's behaviour — as well as *afford* gestural agency, by enabling the performer to shape his own musical environment through the interface. The human, as a participant rather than a user, can assume certain roles in relation to the other virtual agents. In the interaction with 'Spaced', the performer perceived the virtual agent to demand a certain gesture, and consequently changed his musical goals to adapt to this requirement. In the improvised context, the interface could temporarily enable the human and virtual agent to assume the (perceived) roles of either soloist or rhythm section, moving in-between *figure*, *ground* or *field* (van Leeuwen, 1999). The interaction consists of a *gestural feedback* loop, for example; the performer uses the controller and perceives the change in the interface, thus prompting a new gesture to be produced, be it physical or musical. In the improvised context with 'Spaced', this loop was utilised by the performer to adapt the interface's gestures to support the development of a musical plot.

The dimensions of the gestural space provides the performer with a pre-designed set of *semiotic resources,* or audiovisual gestures, and ways of manipulating them over time. Although completely pre-designed, the story in 'Spaced' leaves some gestural production to chance through generative processes, potentially making the *frozen interactions* seem more fluent and less pre-determined. In the gestural space, low-level actions, consist of single gestures produced by the agents during the interaction, such as keyboard notes, sound events or visual movement. A high-level action, such as structuring musical form, can be regarded as a *hyper-gesture* — a gesture of gestures — a complicated structure of nested gestures (Mazzola et al., 2016:168). Progressing the story time using the controller could result in manipulation of between one to eighteen different audiovisual parameters at once. Thus, using the physical interface could result in both high *and* low-level action, depending on the timing of the interaction itself.

Conclusions

This chapter summarises the results of this thesis to answer the research questions.

1. *How can an audiovisual interface be designed to facilitate interaction with digital music in improvised performances?*

The design process detailed in this thesis resulted in an interface prototype called *Plot Twister*, that transforms the process of structuring digital musical composition into a “gamified”, visual experience of navigation through virtual space. The design combines simple musical and visual components with affordable or free consumer software and hardware to develop a novel way of interacting with audiovisual media.

2. *What functions and potentials are afforded through the interaction with the interface?*

The physical interface enables the user to effortlessly adapt musical form and content to fit her improvised musical exploration. Both audience and performer share access to the visual interface that provides real-time visual feedback of the musical content. The technical framework strongly favours looped, continuous, gradual and non-linear progressions of computer-based music compositions. Because of the extended mapping of musical parameters, the musician is afforded great amount of control using only one single encoder device and can therefore dedicate her full attention to other activities, such as instrumental performance. Using generative audiovisual processes to add unpredictability to the interaction shows potential for motivating longer interactions with the "virtual ensemble".

Discussion

The design work in this thesis has resulted in a prototype of an interactive system; the combination of a physical (controller) and visual interface (screen), including sound reproduction (loudspeakers), that should be regarded holistically as a unified tool. But more than a material and digital outcome, the main contribution of this work has been a novel *way of interacting* with audiovisual media. The interaction as a whole can then be summed up as the user applying the physical gesture of turning, or *twisting*, the controller encoder to shape audiovisual content into a *plot*. Acknowledging that this plot can, literally, take an unexpected turn at any moment, there seems to be no better name for the interface than the *Plot Twister*.

Parallel to the technical design work, the compositions 'Skyfall' and 'Spaced' were designed to be interacted with and so that the second research question could be answered. The author's need for performing instrumental improvisation in computer-based music led to the designing of a simple tool to simulate the expressive, seamless manipulation of musical parameters of an improvising rhythm-section. *Adaptive music* techniques from video games inspired to design work and consequently turned the musical performance into a "gamified", audiovisual experience. *Plot Twister* enables the improviser to assume the additional role of, as one interview participant put it, the song *navigator*, rather than having to perform all virtual instruments, as the song *creator*. The interaction with digital audiovisual content has been reduced to a single control device, which continuous movement makes the interaction feel more gradual and fine-tuned, like adjusting the frequency on a FM radio.

To minimise complexity of the interaction, the resulting system is constrained to afford only one control of one audiovisual story at a time. The creative process as a whole has been an active and conscious effort in setting musical rules by designing constraints to available functions and parameters, while still allowing a large-enough space of musical possibilities. By scrolling through the composition in less than a minute, the performer can review the content and available actions afforded by it, so that the space of potential expressions can be identified, necessary to adjust her own performance to the conditions set by the system.

Within the frames of the interface, the performer was afforded freedom to adapt the musical form in a way that suited the musical goals for that specific moment, and thus the constraints could be used as a source for creative exploration.

A major flaw in the design turned out to be the stability of the software framework itself. In the design phase, the first iteration would randomly crash and did so during two of the interviews after 30-40 minutes. For the second iteration, after replacing a plugin that was thought to be the cause, other issues arose. This time related to processor overload, which made the crashes more predictable, but would only allow short intervals of continuous design work and use. Clearly, these conditions make this design method unsustainable for longer live performances, where demands on stability are even higher. Thus, for future iterations, a better solution must be found where processor-heavy audiovisual material can be triggered in steps, rather than being constantly running in the background. Incidentally, this is exactly how professional video game music software works. However, these are not compatible with Resolume Arena and instead requires advanced knowledge in professional 3D modelling software, which in turn are even heavier for the computer to handle.

The auto-ethnographic approach in combination with stimulated recall resulted in preliminary implications of the interface's functions and interactive potentials, that could benefit the development of future design iterations. But to potentially draw any general conclusions requires further studies with other participants, in real-world performance environments. The same level of unpredictable input received through the interviews, when evaluating the first iteration, would no doubt have given a fresh perspective on the interaction with the second iteration, most likely far beyond the author's own expectations. Since the goal eventually will be to spread the design to other people in different interactive contexts, this study would have benefited greatly from second-hand feedback from participants active in other disciplines and arts.

However, thanks to the first hand experience with the interface, afforded by the auto-ethnographic approach, some of the more complex dimensions of the interactions could be described in great detail and would often have a direct influence on the iterative design work, potentially making the process more effective in terms of time and flexibility. Although, facing limitations as practitioner/researcher, the improvised nature of the interaction combined with the generative musical processes in the second iteration contributed to reduced level of predictability that would lead to at least one genuinely surprising situation during the filmed performance (see video at 0:58). When investigating the possibilities of using participants for the final evaluation, it became clear that the author's unique position of having one foot in improvised jazz tradition and the other in digital music production, was a rare combination and it could have required extensive preparation to identify potential candidates that fit the profile for the study.

Looking back now, considering the unstable computer performance in both iterations, the first iteration still turned out to be a better option for the longer, semi-structured interviews. Perhaps, by adding a musical instrument to the interaction with the interface, not all participants would have kept the same open mind for potential applications, thus risking a reduction to the wide range of helpful and interesting feedback that the discussions contributed to. However, in the efforts to develop the prototype further, continuing to subject it to other contexts and users will be an important part in gaining new insights and ideas for improvement. In this thesis, the interface has been applied in the context of keyboard improvisation, but as implicated by the participants during many of the interviews, there seem to exist a diverse spectrum of ideas for potential applications in other interactive contexts that are well worth looking into.

References

Literature

Adams, T., Holman Jones, S., & Ellis, C. (2015). *Autoethnography*. Oxford University press.

Alexandersson, M. (1994). *Metod och medvetande*. Göteborg: Acta Universitatis Gothoburgensis.

Alvehus, J. (2013). Skriva uppsats med kvalitativ metod (1st ed.). Stockholm: Liber.

Bailey, D. (1993). *Improvisation*. New York: Da Capo Press.

Berg, J. (2012). Elva bidrag till ett nytt kunskapsområde. In J. Wingstedt & G. Ternag, *På tal om musikproduktion* (pp. 197-210). Göteborg: Bo Ejeby förlag.

Berliner, P. (1994). *Thinking in jazz* (1st ed.). Chicago: University of Chicago Press.

Bezemer, J., & Mavers, D. (2011). Multimodal transcription as academic practice: a social semiotic perspective. *International Journal Of Social Research Methodology*, *14*(3), 191-206.

Biggs, M., & Büchler, D. (2007). Rigor and Practice-based Research. *Design Issues*, *23*(3), 62-69.

Bowers, J. (2012). The logic of annotated portfolios. Proceedings Of The Designing Interactive Systems Conference On - DIS '12, 68-77.

Bown, O., Eldridge, A., & McCormack, J. (2009). Understanding Interaction in Contemporary Digital Music: from instruments to behavioural objects. *Organised Sound*, *14*(02), 188.

Braasch, J. (2011). A cybernetic model approach for free jazz improvisations. *Kybernetes*, *40*(7/8), 984-994.

Burroway, J. (2000). *Writing fiction* (5th ed.). New York: Longman.

Burzik, A. (2002). Practising in Flow—The Secret of the Masters. *Stringendo*, (24(2), 18-21.

Binder, T., Michelis, G. de D., Ehn, P., Jacucci, G., Linde, P., & Wagner, I. (2011). Design things. MIT Press.

Calderhead, J. (1996). Teachers: Beliefs and knowledge. In D. C. Berliner & R. C. Calfee (Eds.) Handbook of educational psychology (pp. 709-725). New York: Simon & Schuster Macmillan.

Calderhead, J. (1981). Stimulated recall: a method for research on teaching. *British Journal Of Educational Psychology*, *51*(2), pp. 211-217.

Caramiaux, B., & Tanaka, A. (2013). Machine Learning of Musical Gestures: Principles and Review. *Proceedings Of The International Conference On New Interfaces For Musical Expression (NIME)*, pp. 513-518.

Chadabe, J. (2002). The limitations of mapping as a structural descriptive in electronic instruments. *Proceedings Of The International Conference On New Interfaces For Musical Expression*, pp. 38-42.

Choi, I., Bargar, R., & Goudeseune, C. (1995). A manifold interface for a high dimensional control space. *Proceedings Of The 1995 International Computer Music Conference*, pp. 385-392.

Cournoyer, M. (2011). Le rôle de la rhétorique dans le projet en design : une exploration du concept d’argumentation de Perelman et Olbrechts-Tyteca (M.Sc.A). Université de Montréal.

Dalen, M. (2015). *Intervju som metod* (2nd ed.). Malmö: Gleerups utbildning.

Dahlstedt, P. (2008). Dynamic Mapping Strategies for Expressive Synthesis Performance and Improvisation. In Sølvi Y., Kronland-Martinet R., Jensen K. (eds.), Computer Music Modeling and Retrieval. Genesis of Meaning in Sound and Music (pp. 227-242). Copenhagen, Denmark: Springer.

De Grave, W., Boshuizen, H., & Schmidt, H. (1996). Problem based learning: Cognitive and metacognitive processes during problem analysis. *Instructional Science*, *24*(5), pp. 321-341.

Di Scipio, A. (2003). ‘Sound is the interface’: from interactive to ecosystemic signal processing. *Organised Sound*, *8*(03), pp. 269-277.

Dix, A., Finlay, J., Abowd, G. D., Beale, R. (2004). *Human-computer interaction* (pp. 327-357). Harlow [u.a.]: Pearson Prentice-Hall.

Ellis, C. (2004). A methodological novel about autoethnography. Walnut Creek, Calif: AltaMira Press.

Ellis, C., Adams, T., & Bochner, A. (2010). Autoethnography: An Overview, Qualitative Social Research. Retrieved from http://www.qualitative-research.net/index.php/fqs/article/view/1589

Emmerson, S. (2011). Combining the Acoustic and the Digital: Music for Instruments and Computers or Prerecorded Sound. *Oxford Handbooks Online*.

Fels, S., & Hinton, G. (1998). Glove-TalkII-a neural-network interface which maps gestures to parallel formant speech synthesizer controls. *IEEE Transactions On Neural Networks*, *9*(1), pp. 205-212.

Fels, S., Gadd, A., & Mulder, A. (2002). Mapping transparency through metaphor: towards more expressive musical instruments. *Organised Sound*, *7*(02), pp. 109-126.

Françoise, J., Schnell, N., & Bevilacqua, F. (2013). A multimodal probabilistic model for gesture--based control of sound synthesis. *Proceedings Of The 21St ACM International Conference On Multimedia - MM '13*, pp. 705-708.

Frisk, H. (2008). Improvisation, Computers, and Interaction: Rethinking Human-Computer Interaction Through Music. Diss. Musikhögskolan i Malmö.

Gaver, W. (2011). Making spaces. Proceedings Of The 2011 Annual Conference On Human Factors In Computing Systems - CHI '11.

Goodwin, C. (1994). Professional Vision. *American Anthropologist*, *96*(3), pp. 606-633.

Graddol & Goodman, Sharon, 1964- & Graddol, David & Lillis, Theresa M., 1956- & Open University (2007). *Redesigning English* (2nd ed.,Revised). Routledge ; Milton Keynes : The Open University, London, pp. 38-72.

Gurevich, M., & Treviño, J. (2007). Expression and its discontents. Proceedings Of The 7Th International Conference On New Interfaces For Musical Expression - NIME '07, pp. 106-111.

Haglund, B. (2003) *Pedagogisk Forskning i Sverige.* 8(3). Stockholm: Stockholms Universitets Förlag, pp. 145-157

Hallström, L. (2017). Att designa en tjänst för föräldrar till barn med funktionsvariationer: En designprocess med avsikter för användning som startpunkt (Masteruppsats). Linköpings universitet.

Hayano, D. (1979). Auto-Ethnography: Paradigms, Problems, and Prospects. *Human Organization,* 38(1), pp. 99-104.

Heathcott, J. (2007). Noise Orders: Jazz, Improvisation, and Architecture - Edited by David P. Brown. *Journal Of Architectural Education*, *60*(4), pp. 65-66.

Hoffman, G., & Weinberg, G. (2010). Gesture-based human-robot Jazz improvisation. *2010 IEEE International Conference On Robotics And Automation*, pp. 582–587.

Hummels, C., Overbeeke, K., & Klooster, S. (2007). Move to get moved: a search for methods, tools and knowledge to design for expressive and rich movement-based interaction. *Personal And Ubiquitous Computing*, *11*(8), pp. 677-690.

Iedema, R. (2003). Multimodality, resemiotization: extending the analysis of discourse as multi-semiotic practice. *Visual Communication*, *2*(1), pp. 29-57.

Jonas, W. (2007). Research through DESIGN through research. *Kybernetes*, *36*(9/10), pp. 1362-1380.

Jordan, B., & Henderson, A. (1995). Interaction Analysis: Foundations and Practice. *Journal Of The Learning Sciences*, *4*(1), pp. 39-103.

Kaipainen, M. & Ravaja, N. & Tikka, P. & Vuori, R. & Pugliese, R. & Rapino, M. & Takala, T. (2011). Enactive Systems and Enactive Media: Embodied Human-Machine Coupling beyond Interfaces. *Leonardo* 44(5), pp. 433-438.

Keith, M. (1988). Stimulated Recall and Teachers' Thought Processes: A Critical Review of the Methodology and an Alternative Perspective. *Annual Meeting Of The Mid-South Educational Research Association*.

Kress, G. (2010). Multimodality: A Social Semiotic Approach to Contemporary Communication. London: Routledge.

Kress, G. & van Leeuwen, T. (2001). Multimodality: A social semiotic approach to contemporary communication. London: Routledge.

Kress, G., & Van Leeuwen, T. (2001). Multimodal Discourse: The modes and media of contemporary communication. London, Edward Arnold.

Leman, M., Lesaffre, M., & Maes, P. (2017). Crafting the playing body as an infrastructure of “immediate” and “mediate” embodied music cognition in an academic jazz program. The Routledge companion to embodied music interaction.

Lemke, J. (2000). Across the Scales of Time: Artifacts, Activities, and Meanings in Ecosocial Systems. *Mind, Culture, And Activity*, *7*(4), pp. 273-290.

Lemke, J. (2005). Multimedia Genres and Traversals. *Folia Linguistica*, *39*(1-2), pp. 45-56.

Lewis, G. (1999). Interacting with latter-day musical automata. *Contemporary Music Review*, *18*(3), pp. 99-112.

Macdonald, R., & Wilson, G. (2006). Constructions of jazz: How Jazz musicians present their collaborative musical practice. *Musicae Scientiae*, *10*(1), pp. 59-83.

Makela, M. (2008). The Practice of Live Cinema. *Media Space Journal*.

Maréchal, G. (2010). Autoethnography. In A. J. MillsG. Durepos & E. Wiebe (Eds.), *Encyclopedia of case study research* (pp. 44-45). Thousand Oaks, CA: SAGE Publications, Inc.

Matyja, J. R., & Schiavio, A . (2013). Enactive music cognition: Background and research themes. *Constructivist Foundations, 8(3)*, pp. 351-357.

Mazzola, G., Mannone, M., Pang Clark, Y., O'Brien, M., Torunsky, N. (2016) All about the music. The Complete Ontology: Realities, Semiotics, Communication, and Embodiment. Berlin: Springer

McNiff, J. (2013). *Action research*. Abingdon, Oxon: Routledge.

Mendoza Garay, J. I., & Thompson, M. (2017). “Gestural Agency in Human-Machine Musical Interaction” (pp. 412-419). In M. Lesaffre, P.-J. Maes, & M. Leman (Eds.), *The Routledge Companion to Embodied Music Interaction*. Abingdon, UK: Routledge, Taylor & Francis Group.

Miranda, E., & Wanderley, M. (2006). New digital musical instruments: control and interaction beyond the keyboard. Middleton: A-R Editions.

Moore, F. (1988). The Dysfunctions of MIDI. *Computer Music Journal*, *12*(1), pp. 19–28.

Nespor, J. (1985). The Role of Beliefs in the Practice of Teaching: Final Report of the Teacher Beliefs Study. Washington, DC: National Inst. of Education (ED).

Neustaedter, C., & Sengers, P. (2012). Autobiographical design in HCI research. *Proceedings Of The Designing Interactive Systems Conference On - DIS '12*, pp. 514-523.

Norris, S. (2004). Analyzing Multimodal Interaction: A Methodological Framework. New York: Routledge.

Pensoneau-Conway, S., Adams, T., & Bolen, D. (2017). *Doing autoethnography*. Rotterdam: SensePublishers.

Perkis, T. (2008). *Some Notes on my Electronic Improvisation Practice* [Ebook]. Retrieved from http://www.perkis.com/wpc/perkis_dean_book_rev2.pdf

Preece, J., Rogers, Y., & Sharp, H. (2002). *Interaction design*. New York, NY: J. Wiley & Sons.

Pruyn, M., & Holman Jones, S. (2018). *Creative Selves / Creative Cultures*. Cham: Palgrave Macmillan.

Simpson, Z., & Archer, A. (2016). Combining autoethnography and multimodal social semiotics: potentials for theory and method. *Social Semiotics*, *27*(5), pp. 656-670.

Sawyer, R. (2007). Group genius: The Creative Power of Collaboration (1st ed.). New York: Basic Books.

Schofield, G., Green, D., Smith, T., Wright, P., & Olivier, P. (2014). Cinejack: Using live music to control narrative visuals. *Proceedings Of The 2014 Conference On Designing Interactive Systems - DIS '14*.

Schwarz, D. (2012). The Sound Space as Musical Instrument: Playing Corpus-Based Concatenative Synthesis. *New Interfaces For Musical Expression (NIME)*, pp. 250-253.

Stowell, D., Robertson, A., Bryan-Kinns, N., & Plumbley, M. (2009). Evaluation of live human–computer music-making: Quantitative and qualitative approaches. *International Journal Of Human-Computer Studies*, *67*(11), 960-975.

Sweet, M. (2015). *Writing interactive music for video games* (1st ed.). Upper Saddle River, NJ: Addison-Wesley.

Swift, B. (2013). Chasing a feeling: Experience in computer supported jamming. In S. Holland, K. Wilkie, P. Mulholland, & A. Seago (Eds.), *Music and human computer interaction* (pp. 85–100). London: Springer.

Taylor, R., Spence, J., Walker, B., Nissen, B., & Wright, P. (2017). Performing Research. *Proceedings Of The 2017 CHI Conference On Human Factors In Computing Systems - CHI '17*, pp. 4825-4837.

Van Nort, D. (2011). Human:machine:human: Gesture, Sound and Embodiment. *Kybernetes*, *40*(7/8), pp. 1179-1188.

Vetenskapsrådet (2014). Forskningsetiska principer inom humanistisk-samhällsvetenskaplig forskning. Retrieved from www.vetenskapsrådet.se

Weinberg, G., & Driscoll, S. (2007). The interactive robotic percussionist. Proceeding Of The ACM/IEEE International Conference On Human-Robot Interaction - HRI '07, pp. 97–104.

Yang, H., Malaka, R., Hoshino, J., & Han, J. (2010). *Entertainment Computing - ICEC 2010* (pp. 32-42). Berlin, Heidelberg: IFIP International Federation for Information Processing.

Yinger, R. (1986). Examining thought in action: A theoretical and methodological critique of research on interactive teaching. *Teaching And Teacher Education*, *2*(3), pp. 263-282.

Zimmerman, J., Forlizzi, J., & Evenson, S. (2007). Research through design as a method for interaction design research in HCI. *Proceedings Of The SIGCHI Conference On Human Factors In Computing Systems - CHI '07.*

Websites

Ableton Live. (2018). Retrieved from https://www.ableton.com/en/live/

Clissold, D. (2018a). Retrieved from https://www.audiovitamins.com/stochas/

Clissold, D. (2018b). “Stochas Documentation – Audio Vitamins”. Retrieved from https://www.audiovitamins.com/docs/stochas

Cosm, T. (2018) “Multi Parameter Curves version 2.0”. Retrieved from http://www.maxforlive.com/library/device/4172/multi-parameter-curves

DJ TechTools. (2018). Midi Fighter Twister. Retrieved from https://store.djtechtools.com/products/midi-fighter-twister

Max for Live. (2018). Retrieved from http://www.maxforlive.com/

OOEEVV - Orquestra Electrònica del Vallès. (2018). “Abl Resolume Plug V1.5”. Retrieved on 01-03-2018 from http://ooeevv.com/webwp/controlling-resolume6-from-ableton-live-via-osc-v-1-6-1/ (plugin has since updated to 1.6.1)

Resolume VJ Software. (2018). Retrieved from https://resolume.com/

RutterStudios. (2018) “Multimap Pro 1.0”. Retrieved from http://maxforlive.com/library/device/4502/multimap-pro

Sanz, D. (2016) “Abl Resolume Plug V1.5 Manual”. Retrieved from https://youtu.be/vJshd65TXVQ

Quartz Composer User Guide. (2018). Retrieved from https://developer.apple.com/library/content/documentation/GraphicsImaging/Conceptual/QuartzComposerUserGuide/qc_intro/qc_intro.html

Appendix

“Plot Twister Interaction Video” (6:56)

Video includes short demonstrations of “Iteration 1 ‘Skyfall’” and “Iteration 2 ‘Spaced’”, as well as the full “Improvised Interaction with ‘Spaced’” that is analysed in the chapter “Evaluation & Results - iteration 2”.

Link: https://bit.ly/2jObGWP

Figure 4. ‘Skyfall’ picture.

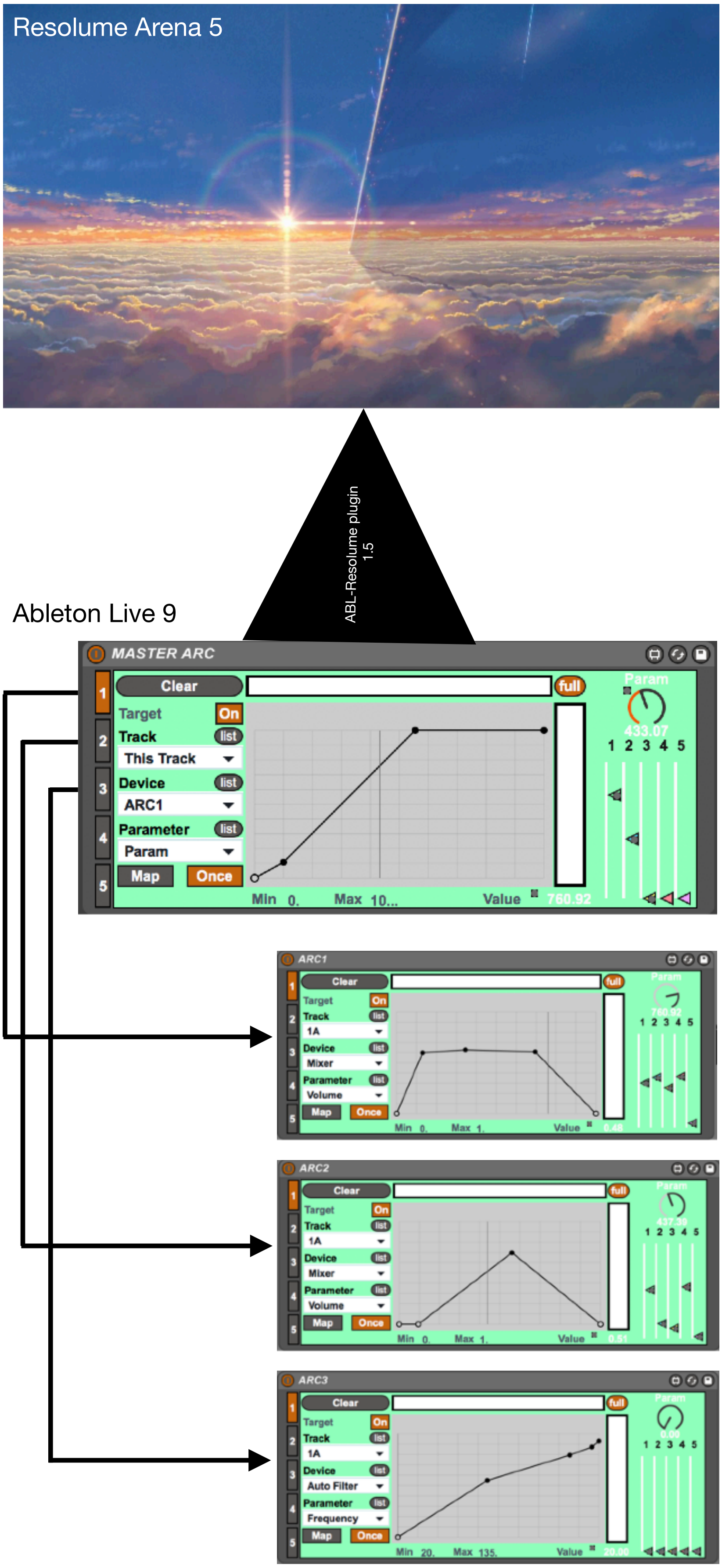

Figure 5. MASTER ARC connecting ARC1, ARC2 and ARC3 to Resolume Arena.