Designing a Dynamic Music System for Narrative Audio Storytellers

An HCI design approach to Creative Audio Software

Frederik Bjørn Blauenfeldt

fbbl@itu.dk

Software Design Thesis

KISPECI1SE

Supervisor - Lousie Barkhuus

A Collaboratory Study with Hindenburg Systems

0.1 Abstract

Narrative audio storytellers rely on music and soundscapes to shape mood, emotions or to build up to narrative emphasis points, yet the tools available to them are either built for musicians or around static music assets, which are misaligned with their technical capabilities and creative practices. Existing research has explored ways to support the manipulation of static music as well as how to create interactive music systems, though the domain of game audio. The field of dynamic music systems within the domain of narrative audio stories remains unexplored. Consequently, as narrative audio storytellers are forced to use these manual and cognitive extraneous workflows, it hinders their exploratory interaction by diverting attention away from the main goal, namely their narrative work of setting scenes, creating moods and evoking emotions.

To investigate this problem, the thesis adopts a design-oriented HCI approach, involving a within-subject qualitative study, comparing current static methods with our proposed dynamic music system. The results revealed that a dynamic music system is very capable in invoking creative exploration and serendipitous discovery while simultaneously lowering extraneous cognitive load, keeping the narrative storytellers in high-level goal-oriented operations. From an HCI perspective, this work suggests that moving creative audio tools away from static workflows, towards more dynamic and exploratory workflows, which align with the mental models and processes of the narrative audio storytellers, can better support narrative-focused creative practices.

keywords: narrative audio, dynamic music systems, soundscapes, creative software, cognitive load, human–computer interaction.

0.2 Table of Contents

1. Introduction

All media files, including audio, video, documents and images, can be found on our website, frederikbjorn.com.

1.1 Narrative Audio Storytellers

Narrative audio storytellers, working as podcasters and radio producers, are seeking to create compelling narratives that actively engage listeners through sound. Within this process, music and ambient sounds play a crucial role in shaping emotions like establishing mood, reinforcing key moments, and helping to guide the listener's interpretation of narrative meaning. These sonic elements do not function as foregrounded musical works, but they act in support of the speech and storytelling, shaping the atmosphere and mood without drawing attention to themselves.
Despite this, the tools used to create and manipulate music and soundscapes are largely designed around the needs, terminology, and mental models of musicians and audio engineers. As a result, narrative audio storytellers often rely on the finished productions of static music tracks and workflows that are not aligned with their competencies. Rather than supporting an exploration and storytelling flow, these tools introduce friction by requiring technical skills that are outside the storyteller's skillset. This mismatch leaves narrative storytellers with limited ways to shape narrative atmosphere, and not in ways that reflect how they think, work, and create.

1.2 Practice-Based Motivation

After working in the narrative radio and podcast industry for two plus years, and working in the radio and podcast software industry for about two years as well, we have repeatedly observed the need for soundscape creating tools, specifically tailored to narrative audio storytellers. As this domain is a rather peculiar one, mostly occupied by journalists and professional storytellers, who have no formal training in either music or audio production, the tools available are often only adjacent domains, mostly adhering to the domain of music creation, where overlaps exist, but which is also a much more technically demanding field with a different focus.

After observing narrative audio storytellers struggle in their interaction, such as finding suitable compositions, looping, extending, ending and creating variations in their musical passages, we wanted to research and explore an alternative. Drawing on a limited knowledge of music systems for the field of Game Audio, we wanted to emulate these aspects of player interaction with an algorithmically built dynamic music system to create a background music generation tool, where the focus would be on a simple interaction, rather than the music industry standard of complexity is freedom, and the result would be to only generate music which fit the setting of narrative audio stories.

1.3 Problem Statement

To address these challenges, we set out to develop a dynamic music tool that shifts the focus of existing music-centric workflows away from static, low-level editing and toward a more dynamic, exploratory system, one that better serves the needs of narrative storytellers. We will then explore how this tool actually fits into the narrative storytellers' workflows in practice, by comparing the current workflow and our solution in a within-subjects qualitative study. In addition to testing the viability of such a music system, we will also examine a few core aspects of what such a tool encompasses. Firstly, the differences in working with a dynamic music system, as opposed to the currently static music-oriented solutions. Secondly, by using cognitive load as a measurement, how can we determine which of these systems is the most suited? And thirdly, how this proposed tool would function as a bridge between musical composers and narrative audio storytellers.

1.4 Positioning in HCI & Creative Software

This thesis will follow the approaches of Human-Computer Interaction (HCI), especially in the evaluation of creative software through design and user-centred analysis. It takes a design-oriented empirical exploration perspective, exploring the possibilities of creating a viable commercial product. It will explicitly explore the interaction between user and creative software and the design of such, not visual or aesthetic design in general. The scope is deliberately limited, as it will only focus on the specific needs of narrative storytellers and the design choices that would be suited for them, not an overall exploration of features that are common in creative software in general.

1.5 Research Gap

To analyse the field, we will address the limited amount of research in the field and look at the few proposed tools for working with music for narrative audio stories, as well as address the current solutions from the industry. To supply and frame the limited amount of research, we will supply research done on creative software applications for visual media. Research exclusively on the viability of a dynamic music system, as we are proposing, is yet to be seen.

This leads to the research questions as follows:

1.6 Main Research Question

How can we design an interactive, dynamic music system for narrative audio storytellers, enabling them in the creation of suitable soundscapes and background music for their narrative audio productions?

1.6.1 Sub-Research Questions

Sub-research question 1: Working explicitly with dynamic music
How can we replace and improve the current state-of-the-art workflow of static music with a workflow of dynamic music?

Sub-research question 2: Cognitive load & serendipity
How can we create a workflow which increases the time spent on the high-level overarching goals, enhancing serendipitous exploration, while simultaneously lowering the extraneous load of the user?

Sub-research question 3: A bridge between musical composers and narrative audio storytellers
How can we create a tool which can be used as a bridge between musical composers and narrative audio storytellers?

1.7 Approach & Contribution

This thesis adopts a design-oriented HCI approach to explore how a dynamic music system can better support narrative audio storytellers. It proposes dynamic music, not as a technical feature, but as a separate paradigm, as compared to the current static music-based workflows, one which aligns with the cognitive and creative practices of narrative audio production. To investigate this, the thesis presents the design and implementation of the MusicBox, our prototype of a dynamic music system. The system is guided by cognitive load theory, semiotic perspectives on musicology and user-centred design principles, and it's intended to support exploratory, creative and serendipitous interactions rather than technical manual editing. The viability of this approach is examined through an empirical, within-subject qualitative study, which compares the current static music workflows with the MusicBox. Through conceptual framing, design exploration and empirical evaluation, this thesis contributes insights into how dynamic music systems can reduce extraneous cognitive load, support serendipitous exploration, and bridge the gap between musical composers and narrative audio storytellers.

1.8 The Role of Hindenburg Systems

This study was conducted in collaboration with the company, Hindenburg Systems, which has developed the narrative audio storyteller-specific digital audio workstation, Hindenburg PRO, one of the few commercial tools available to narrative audio storytellers. As Hindenburg PRO does not offer functionality for the generation and manipulation of background music and soundscapes, this tool will provide a reference point for a future integration of such a system. Hindenburg Systems contributed to the project in an advisory and supportive role throughout the entire process. The involvement has consisted of ongoing dialogue and feedback, and they have stated no specific functionality requirements or asserted any influence over the design. Decisions regarding research questions, methodology, and system design were made independently by the authors.

1.9 Thesis Structure

This thesis is structured to establish both the theoretical and practical foundations to address the research questions. It begins by stating the state-of-the-art from both a practical and theoretical standpoint, as well as our theory of research, adhering to the subjects of semiotics within musicology, cognitive load theory and UI in creative software. Building on this foundation, a problem analysis is conducted, analysing stakeholders' pain points, current limitations, and a conceptual framing of the problem space, providing context for our research questions. Subsequently, a methodology section, stating our intended study, followed by a user guide to showcase what our proposed application, the MusicBox, can do, further expanded by a description of its technical aspects and functionalities. The results chapter then analyses the findings from our study, with a focus on the different aspects of the MusicBox's functionality and how they were perceived. Finally, the thesis concludes with a discussion, reflection and conclusion on these results, their implications and limitations and outlining directions for future work.

2. Background

To frame this thesis, we have chosen to supply a background section which includes the current state-of-the-art approaches to soundscapes in narrative audio stories, both solutions from the industry and proposed solutions from academia. To aid our understanding of the subject, we have looked at literature from three different fields: semiotics within musicology, cognitive load theory and UI in creative software.

2.1 State-of-the-Art

We have chosen to separate our state-of-the-art section into two different categories: The first, practical solutions, comprises four contemporary approaches. Each option presents distinct, often overlapping issues, primarily concerning their provision of static music and prevailing industry practices.
The second, theoretical solution, are papers which try to answer parts of our research questions. Here we have found two possible candidates, LoopMaker, as proposed by Shi and Mysore (2018) and UnderScore, proposed by Rubin et al. (2012, 2013). An important point here is that both of these solutions utilise static music assets for their workflows.

2.1.1 Practical Solutions

Based on our own observations of current practice in the domain of narrative audio storytelling for podcast and radio productions, we identify four primary approaches that narrative audio storytellers rely on when sourcing, manipulating or producing soundscapes and background music. These are royalty-free music, generative AI music, self-composing music, and hiring a composer.

2.1.1.1 Royalty-Free Music

Royalty-free music is ubiquitous across the industry. Most often, producers will have access to a royalty-free music library (e.g., Epidemic Sound, Soundstripe and Splice), which they will browse until they find something suitable. This is what we consider to be the most common workflow for sourcing background music and soundscapes. The appeal is clear: a low-cost subscription, no royalty complications, and fully online access. The main issues, in our view, relate to quality and suitability. Much of the music on these royalty-free platforms is produced to suit a broad range of media rather than radio and podcast production specifically, which can make it too overt and blunt for these formats. See our description of background music in the terminology section for reference (2.2.2).

2.1.1.2 Generative AI Music

Generative AI (GenAI) music is still relatively new in narrative audio productions and, as a result, less prevalent. In our view, it inherits many of the same drawbacks as the royalty-free music libraries: content is often created for broad, cross-media use and typically yields a single, linear and static piece of music. The workflow can also be frustrating for producers, i.e. time spent crafting prompts and waiting for processing, whereas many prefer rapidly auditioning large libraries, deciding on suitability within a few seconds of listening. That said, Generative AI has yet to show its full potential.

Emerging systems are capable of outputting stems or loopable sections, introducing more dynamic possibilities (e.g., Suno). Improvements in speed, reliability, and finer creative control might make GenerativeAI a viable option in the future.

2.1.1.3 Self-Composing

Producers composing music themselves have several obvious advantages - it makes the music much more fitting to the specific setting, it has no limitations in terms of flexibility, and it can all be done by the producer.
But as obvious as the advantages are, the least as obvious are the invariably of this solution for most producers. Learning how to compose music like this takes immense time and skill. It requires experience and knowledge in a whole host of fields, from music theory and composition to producing music, which most producers do not possess.

2.1.1.4 Hiring a Composer

For a production to hire a composer, the advantages are quite similar to self-composing, except maybe for a smaller amount of flexibility. The composer will be able to create custom background music exactly tailored to a given production and aesthetic view of the producers.
However, given the industry's size, as compared to film or games, many of these composers are more used to creating music for other media, which can lead to compositing not fitting the subtle tone of background music for narrative radio. Also in relation to the industry's size and their budgets, this option will always only be available for the biggest of productions, of which there are quite few.

2.1.2 Theoretical Solution

In contrast, research within HCI has explored systems designed to streamline and better integrate musical workflows into narrative production.

The two solutions we examined, the UnderScore (Rubin et al., 2012; 2013) and LoopMaker (Shi & Mysore, 2018) systems, were introduced as research prototypes in publications at the HCI conferences of UIST[1] 2012 and 2013 for UnderScore, and CHI[2] 2018 for LoopMaker.

2.1.2.1 Shi & Mysore's LoopMaker

The LoopMaker tackles a problem we've seen very often when working with static music: finding and creating seamless, also called perfect, loops from a single, pre-recorded track. Its focus is to create a tool which can minimise manual editing and

[1] ACM Symposium on User Interface Software and Technology.
[2] ACM Conference on Human Factors in Computing Systems.

find high-quality loops in background music for bedding. It helps to save time by spotting a clean loop within a chosen region and repeating it to whatever length you need. The aim is to outperform quick manual looping; as the authors put it, the system "automatically and semi-automatically find[s] segments of music that sound natural and seamless when played continuously" (Shi & Mysore, 2018).

2.1.2.2 Rubin et al.'s UnderScore

The project UnderScore (Rubin et al., 2012, 2013) is a tool designed for building a musical underlay from pre-recorded music. It does this by first letting the user input an emphasis point, i.e a key moment in the story. The UnderScore system then applies a pre-fade rising before the emphasis point, where the music hits a noticeable change point, and the speech pauses, and then it applies a post-fade as the speech comes back in. The system automates the manual editing parts: you mark an emphasis in the speech and pick a musical composition, and then UnderScore does the rest. It sets the volume ramps, according to predefined gentle curves, then creates a quick rise into the chosen segment, suggesting a shift, then does a short predefined segment of music, before the speech resumes. The results in a polished, reflective pause, without the low-level task of creating the fades manually.

While these two approaches move closer to supporting the needs of narrative audio storytellers, they remain limited as they are mainly focused on a single issue, offer minimal opportunities for serendipitous discovery, assume a degree of prior musical or technical knowledge, and both work exclusively with static music.

2.1.3 Related Creative Software Tools and Applications

As for the research done before our survey, we tried to assemble relevant papers with a similar goal to ours, i.e. creating a product for creative exploration in creative software, which is to be a part of a bigger ecosystem of tools.
In addition to the LoopMaker & Underscore papers, we looked at four papers and a PhD, all by C. A. Fraser (2016a, 2016b, 2019, 2020a, 2020b). These projects are not from the audio domain, but rather from the graphical and visual design domain. We analysed these proposed solutions as a way to find inspiration for our project and survey. We analysed the two projects, DiscoverySpace (Fraser et al. 2016) and ReMap/RePlay (Fraser et al., 2019, 2020a).

DiscoverySpace is an add-on tool for Photoshop, which provides high-level suggestions for a series of effects to add to a photo. The goal of DiscoverySpace is ultimately to keep the user in the high-level editing mode, utilising community-created macros as opposed to designing a longer series of low-level operations from scratch.

The project of ReMap/RePlay is an example of building tools for better assessing information regarding a specific problem domain. These observations were especially useful for us in analysing creative software through the lens of cognitive load theory.

2.2 Audio, Musicology and Semiotic Terminology

As this thesis explores the crossroads between musicology and software development, we have chosen to include field-specific terminology to help readers gain a deeper understanding of the concepts. These terms relate to both the audio narrative part of audiology, musicology and sound engineering, as well as terms related to the industry and profession. The terminology used in our application, the MusicBox, is explained in the technical description (6.3).

2.2.1 Placement and Terminology of Non-Speech Sounds.

The practice of enriching narrative audio with sounds beyond speech, like background music, ambience, soundscapes and sound effects, has a long and varied history. Given the proliferation of overlapping labels, this section first considers where such sounds sit in the listener's attentional field before settling on terminology. As background for the musicology and semiotics theory of the thesis, we have mainly used three books: *Speech, music, sound* by Theo van Leeuwen (1999), *Music's meanings: A modern musicology for non-musos* by Philip Tagg (2013) and *Audio-vision: Sound on screen* by Michel Chion(1994)

In audio storytelling, non-speech sound is used to deepen the listener's experience. Although terminology varies as A.E. Beeby distinguishes "Immediate/Support/Background" sounds and Walter Murch's "foreground/mid-ground/background" fields, the underlying principle is consistent: concurrently heard sounds are hierarchised cognitively, so that some are demanding attention while others are 'heard but not listened to'. Following van Leeuwen, we adopt R. Murray Schafer's tripartite schema of figure, ground, and field, throughout this thesis (van Leeuwen, T, 1999).

Figure is the current focus (typically speech or an audio cue); ground supports and shapes perception without competing for attention; field is the wider soundscape or setting conceptualised as the background. For example, a city ambience may open prominently to establish place (figure or ground), then be faded to operate as field beneath a voiceover, which then occupies the place of figure. A car then quickly passes, occupying ground for a brief period before fading out to field.

2.2.2 Audio and Musicology terms

2.2.2.1 Soundscape

Soundscape is a very broad term, mostly adhering to the placement of field. As stated by ISO, a soundscape is "sound generated by nature or human activity" (ISO, 2014). For this thesis, we have adopted a more specific usage: "Soundscape composition is a form of electroacoustic music characterised by the presence of recognisable environmental sounds and contexts" (Truax, B, 2012). It remains important to include the non-diegetic[3], musically oriented dimension; as Schafer notes, a soundscape is the sound or combination of sounds that forms or takes shape and surrounds the listener within a given environment, encompassing both the natural acoustic settings and composed or designed sound (Schafer, R.M., 1977). In what follows, we sometimes use the term soundscape as a synonym for background music, acknowledging the practical ambiguity between these terms.

2.2.2.2 Background Music

Background music, also known as musical underlay/overlay, most often occupies the space of ground or field. It is quite a broad term used for a wide variety of music, as stated by Merriam-Webster dictionary, background music is: "music to accompany the dialogue or action of a motion picture or radio or television drama" (Merriam-Webster, n.d.). It should be noted that background music for narrative radio and podcasts exhibits distinct qualities compared with film and games. It is generally subtler, less variable, often drone-like in character, and closer to a musical non-diegetic soundscape, as opposed to diegetic soundscapes, which are sound prevalent in a scene, i.e. real-world sounds and music from an active source (Chion, M. 1994), more on this in the compositional guide (6.4).

2.2.2.3 Ambience

Ambience is a more narrow definition; real-world sounds, all adhering to the placement of field and sometimes short periods of ground, consisting of ambient recordings of the environment or longer passages of foley. Sometimes people use the term musical ambience to describe synthetically or musically derived sounds which serve the same purpose as real-world ambience, but are more artistically and aesthetically oriented. We will not favour this interpretation, as we will instead rely on the term soundscape for that.

[3] Non-diegetic sound, as opposed to diegetic sound, refers to audio elements that do not originate from within the story world and are not perceivable by the characters, such as background music, narration, or sound effects added for the audience's emotional or narrative guidance.

2.2.2.4 Bedding

Bedding (or simply bed) is a more technically oriented term and refers to the figurative bed, which speech or dialogue can be placed atop. The term bedding exclusive occupies the placement of Field.

2.2.2.5 Jingles

Jingles are short, branded musical cues, often a part of a bigger music composition. It occupies the placement of the figure.

2.2.2.6 Sound Effects

Sound effects (or SFX) are real-world sounds or emulations thereof, most often occupying the placement of a figure.

2.2.2.7 Royalty-free Music

Royalty-free music is a term commonly used to describe background music that can be used across different forms of media without incurring royalties. In practice, it means that once a user has acquired the track, it can be incorporated into any production and distributed on any platform without additional fees. Because music rights are often restrictive and complex, this guarantee of safe, unrestricted use is a key part of the appeal.

2.2.2.8 Narrative Audio Stories

Narrative audio stories or narrative audio storyteller is the common nomenclature of producers of narrative content for radio or podcasts. It differs from the simpler term, podcaster or podcast producer, in that these terms are a more overarching and simpler definition. As an example, a producer of a Talking-Heads podcast, i.e. just two or more people having a conversation, is not a narrative storyteller, but simply a podcaster.

2.2.3 Static vs. Dynamic Music

Our definition of static music is quite simple: an audio file containing a single continuous stream of a single musical composition.

Our definition of dynamic music is derived from the field of game audio and specifically from the writings of Michael Sweet (2014) and Winifred Philips (2014), who use the term dynamic music in tandem with interactive music, as in computer games, you often interact dynamically with music. These principles go all the way back to the first examples of algorithmically composed music, the Musikalisches Würfelspiel, i.e. a composition based on the chance of a die, as composed by Mozart in 1793 (Alpern A., 1995).

In this thesis, for simplicity, we will use the definition of dynamic music, as music incorporating any type of chance or algorithmic behaviour working in tandem with any interactive input from UI.

2.2.4 Variations

Variations play a crucial role in the development and composition of dynamic music. Following Åke Parmerud (2009), a variation is “the most basic way in which we can change the identity of a given musical object, or a larger portion of the music, without losing the original aspects that created this identity”. In other words, variation extends and/or develops material while preserving its character.

2.2.5 Audio Editing Tools

Audio editing has a long and varied history, from the codification of music in western notation, through physical splicing and cutting of magnetic tape, to today’s digital workflows. As tools have evolved, they have diversified by use case: notation editors (e.g., Sibelius, Dorico); music production and multitrack editors (e.g., Ableton Live, Pro Tools, Logic Pro); audio and DSP programming environments (e.g., Pure Data, Max/MSP). Collectively, these applications are known as digital audio workstations or simply as DAWs.

This study focuses on a single commercial product, Hindenburg PRO, a DAW purpose-built for radio, podcasting, and narrative audio, prioritising speech-oriented editing and publication workflows.

Most modern DAWs support third-party tools called plug-ins (e.g., Audio Units, VST), enabling users to extend functionality to suit specific needs and practices. The application developed for this project, MusicBox, is delivered both as a VST plug-in and as a standalone application.

2.3 Creative Cognition

In creative cognition, we have devised three different categories: cognitive coad theory in general, and as a subset of this, the cognitive processes of divergent and convergent thinking and the two different macro-micro granularities of level of operation.

2.3.1 Cognitive Load Theory

To serve as a measurement of the success of a given piece of creative software, we choose to measure and analyse it through the lens of cognitive load theory. We have drawn on a few different concepts, as part of the cognitive load analysis, the concepts of extraneous and intrinsic load and the analysis of means-ends, as well as the more independent measurabilities of divergent and convergent thinking and level of operation.

Extraneous load, as proposed by Sweller (1988), is the concept of how a user of a given system tries to achieve their goals.
The extraneous load is the portion of the total cognitive load that is attributed to the interface design, including instruction material (Cooper 1998), as opposed to the intrinsic cognitive load, which is caused by the intrinsic nature, i.e. how difficult it is to learn a specific task.

Means-ends analysis is a cognitive process that focuses on reducing the difference between the current problem state and the goal state. It is a problem-solving strategy, primarily used by novices when they are faced with conventional problems that have clearly defined goals. However, the mental effort required to manage the goal stack, the current task, and the differences between these, consumes working memory capacity. This consumption leaves fewer resources available for schema acquisition, i.e. learning. However, if when analysing means-ends, the extraneous load is found to be too high, a strategy of nonspecific goal (goal-free effect) can be deployed. This technique removes the specific goal, forcing the learner to focus on the information provided and work forward. This forward-working path is much simpler, imposes very low levels of cognitive load, facilitates learning and engages serendipitous discovery.

2.3.2 Divergent and Convergent Thinking

The terms divergent and convergent thinking describe two different ways of thinking that determine the user's interaction with creative software (Matyas, K., n.d.). A study of visual designers comparing digital tools with pen-and-paper sketching during ideation reported that 86% of designers used pen and paper for this phase (Frich et al., 2021). In the auditory domain, our participants do not have the same option to switch to pen and paper for rapid ideation; therefore, it is imperative that a tool can support this exploratory flow. In this study, we use divergent thinking to denote exploration and ideation for establishing overall concept and mood, and convergent thinking to denote exploitation and refinement towards finalisation (Frich et al., 2021; Zhou et al., 2021)

2.3.3 Level of Operation

The Level of Operation refers to the granularity of the task at hand and can be understood as shifting vantage points. The high-level task is the overarching goal; here, the participant is to assess whether a given sound or composition matches the storyline, mood or setting of the production. By contrast, low-level editing consists of individual, detail-oriented actions that are often highly application-specific (Adar et al., 2014). In the creative domain, we most often require a set of low-level operations to complete one overarching goal (Fraset et al., 2016b). Within audio work, low-level editing is typically the most time-consuming and skill-dependent part of the process, which helps explain why many producers seek tools that automate or assist with these tasks (Rubin et al., 2012).

2.3.4 Serendipitous Discovery

The term serendipitous discovery is used throughout this thesis as the exploration of happy accidents through a medium. This is a common practice for creative software, where the goal is to give the user the means to create exactly what they deem necessary, through discovery of the medium. The term is described by C. Busch (2022) as 'Active Luck'. While serendipity does not explicitly reduce extraneous cognitive load, it exerts the discovery of new possibilities and hereby engages the user in creating new ideas and concepts (Fraser CA., 2020b).

2.4 Creative Software and UI

In this thesis, we will favour the term creative software, as opposed to alternative labels such as creative software applications, creative tools, or creative computing software, drawing inspiration from the work of C.A. Fraser (2016a, 2016b, 2019, 2020a). This definition differs from Fraser's notion of Complex Software (2016a) and from purpose-built systems that require specialised training or advanced technical expertise.

2.4.1 UI Guidelines and Background

To ensure our design aligned with established usability principles, we drew on Nielsen's Ten Usability Heuristics (1994a) and Norman's book *The Design of Everyday Things* (2013), a foundational piece of work on user-centred design. These readings served as background literature on UI.

3. Problem Analysis

3.1 Context and Stakeholders

Narrative audio production is a challenging domain for practitioners. Informed by our own practice, informal discussions with producers, and the literature reviewed earlier in this thesis, we identify three recurring factors that shape the conditions for our stakeholders: their knowledge and skills, the tight time constraints under which they work, and their characteristic background as narrative storytellers rather than technical specialists.

The first factor is simply knowledge or skill. As users, professional or not, are most often inexperienced in music production, but understand the complexity of audio production in general, they often come to the conclusion that music-making is a black magic they will never understand.
The second factor is due to the tight time constraints, especially for the professional producer. This often leaves them with very short time to make their productions, and since finding soundscapes is usually done at the end, this results in a very stressful and tight process.
The third factor is the nature of the producer, as they often come from a more journalistic or narrative background, as opposed to a more technical background; they orient their focus to the high-level task at hand, therefore, making a dive into the low-level editing setting adds extra strain on their cognitive load.

Given this context, the core tension addressed in this thesis is that audio storytellers must make nuanced musical decisions regarding soundscapes without having the necessary skills, time, or conceptual frameworks to do so. Any proposed solution must therefore be technically simple, have an intuitive design, be cognitively lightweight, and use the language of the audio storytellers, not technical or musicology-oriented terms.

3.2 Current Workflows and Pain Points

Across the four dominant practical solutions (royalty-free music, generative AI, self-composing, and hiring a composer), a common structural issue emerges: all rely on static, linear music. Whether professionally composed or algorithmically generated, these assets take the form of a fixed audio file that can only be changed with manual editing (not taking into account new AI-generating systems, as these are yet to show their full potential).
As a result, audio storytellers face the same set of recurring problems: that is, locating loopable regions, cutting at musically plausible boundaries, using crossfades to mask

imperfect seams, adjusting timing to match the story's emphasis points, manually reducing or shaping musical density using fades and rejecting otherwise suitable tracks because of a single element which can ruin the whole thing.
These points obviously lead to extraneous cognitive load as they demand technical effort that is unrelated to the storyteller's actual goal of shaping moods and narratives.

3.2.1 Limitations of Research Prototypes

As mentioned in our background, we also analysed the research-based systems of the UnderScore (Rubin et al., 2012) and LoopMaker (Shi, Z., & Mysore, G. J., 2018). As stated, these systems alleviate specific technical tasks like loop detection or automating emphasis curves, but inherit the same dependence on static music. They require users to reason in terms of beats, sections, transitions, and harmonic alignment, all of which assume musical schemas and knowledge.
This reveals a deeper systemic issue; existing solutions, from industry or research, are built around the mental models of musicians, not the mental models of narrative audio storytellers, leading to discrepancies.
In addition, we also see the same static music editing cognitive load pattern, as shown in Figure 1.

Figure 1. The static music editing cognitive load pattern.

3.3 Conceptual Framing

3.3.1 Cognitive Load Theory

So, to analyse the two points, discrepancies in the mental models and the static music editing cognitive load pattern, cognitive load theory provides a useful lens for understanding why they impose disproportionate cognitive demands on narrative audio storytellers. Working memory has a strictly limited capacity (Cooper, 1998), and performance deteriorates when the combined demands of a task exceed this capacity (Albers & Tracy, 2006).

As static music forces narrative audio storytellers into a series of low-level editing operations which do not contribute meaningfully to the creative or narrative goals, it adds an extraneous load, which is then mental effort spent not on storytelling but on compensating for tool and skill limitations. Because working memory must be shared across all concurrent tasks, every moment spent solving these extraneous load tasks it reduces the cognitive resources available for high-level narrative judgement.

Sweller's (1988) analysis of means-ends strategies underscores this point. Tasks that require stepwise, detail-oriented problem solving impose a higher cognitive burden than tasks oriented toward broader, nonspecific goals. In our context of narrative audio, low-level editing resembles a means-ends strategy: the producer must continually evaluate small technical steps to reach an intended outcome of a single task. High-level narrative reasoning resembles a nonspecific goal strategy and is therefore cognitively lighter and more compatible with creative ideation.

3.3.2 Analysis of Divergent vs. Convergent Phases

To further analyse the problem domain, we also focused on divergent vs. convergent thinking. As audio storytellers seek to create early stages of soundscapes for their productions, this requires divergent exploration and ideation. Exploring textures, moods, feelings, atmospheres and more, without committing to a detailed structure. Static music constrains this because every possible exploration risks becoming a technical editing task.
We choose to think of it as divergent and convergent phases of operations, as one half of the double diamond model (Design Council, 2007), as shown in Figure 2.

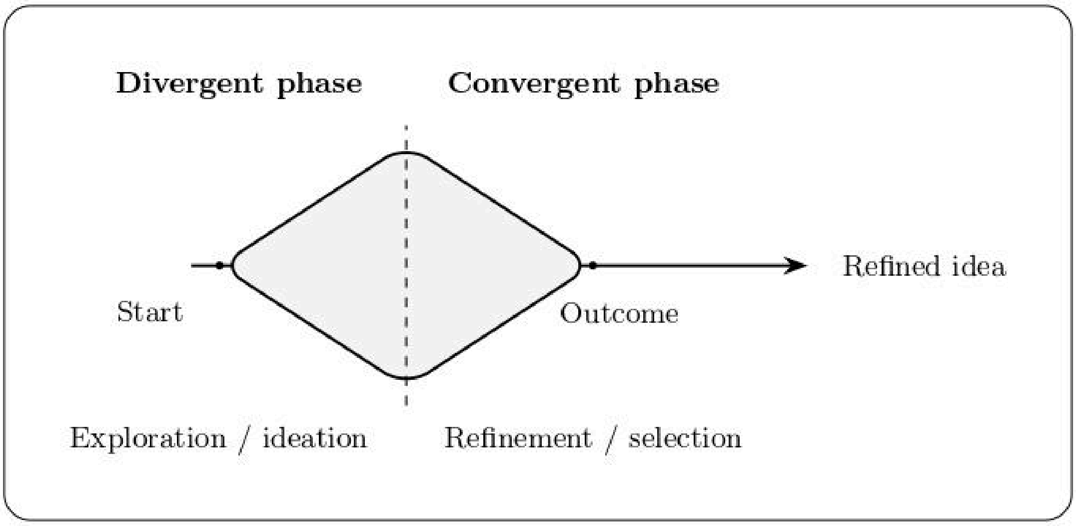


Figure 2. Divergent and Convergent Phases Schema

To conclude, static music requires a lot of this continual micro-editing. Audio storytellers are repeatedly pulled out of the divergent, exploratory thinking and into narrowly focused, extraneous technical problem-solving. This shift not only increases extraneous load but disrupts the flow and continuity of narrative reasoning, undermining the very creative decisions the music is meant to support.

3.4 Research Questions and Context

To further expand on our main research question and sub-research questions as stated in the introduction (1.6). We have here supplied purpose, objectives and scope.

3.4.1 Purpose

The purpose of this thesis is to investigate how a dynamic music tool can better support the creative, cognitive, and practical needs of narrative audio storytellers by proposing the application, the MusicBox, aimed at reducing technical barriers and enabling exploratory interaction for the creation of soundscapes.

3.4.2 Objectives

Development & Design
Develop the MusicBox, a dynamic music system which can compile to a stand-alone application as well as a VST for interactions with DAWs like Hindenburg PRO.

Conceptual Contribution
Articulate a theoretical framing linking creative cognition concepts like cognitive load theory, convergent and divergent phases, as well as the levels of operation in narrative audio workflows.

Establish Terminology
Create a terminology set to use both in the conceptualising and the development of the MusicBox.

Create a guide for soundscape creation.
Create a schema for musical compositions to create soundscapes, both fitting the technical constraints of the MusicBox as well as the domain-specific conventions of narrative audio productions.

Empirical Evaluation
Compare the MusicBox with a standard royalty-free music library-based workflow, using within-subject user testing.

Design Implications
Derive design recommendations for future creative audio tools based on the concepts, terminology and empirical data analysis derived in this thesis.

3.4.3 Scope

This thesis is constrained to the design and evaluation of a prototype system, the MusicBox v0.1. It focuses on narrative storytellers, who are experienced with Hindenburg PRO but lack musical knowledge.

The scope is limited in several ways. Firstly, the project addresses only the creation and manipulation of soundscapes; it does not encompass full soundtrack composition, melodic scoring, or the broader domain of music production. Secondly, the evaluation is a study, which is restricted to short sessions as opposed to a longer and more thorough domain-specific analysis. Therefore, these conditions enable comparison between workflows but do not replicate long-term, real-world production cycles, collaborative editorial processes, or the use of personalised sound libraries.

Furthermore, the thesis does not attempt to assess cross-platform plugin deployment, long-term user adoption, integration into any commercial sphere, or any comprehensive post-production tasks such as editing, mixing, and mastering, which will remain the responsibility of Hindenburg PRO.

Finally, the system is assessed solely by using musical material composed by us, the research team; therefore, the testing and evaluation of third-party compositions fall outside the scope of this thesis and are considered only in terms of their technical structure and accompanying guidelines, see the compositional guide for reference (6.4).

4. Methodology

As described in the background section, after reading numerous studies (Adar et al., 2014; Fraser et al., 2016a, 2016b, 2019, 2020a; Rubin et al., 2012, 2013; Shi & Mysore, 2018) which tries to measure the usefulness, effectiveness and sometimes cognitive load of creative software, we designed our own study to collect empirical findings on our proposed tool, the MusicBox.

4.1 Study Type

We employed a mixed-methods within-subjects, non-counterbalanced experimental design to compare the usability and mental workload of MusicBox with a prevailing workflow based on editing royalty-free music. Each participant completed two tasks: (i) a baseline task using the standard workflow and (ii) a condition treatment with access to the MusicBox. We chose a within-subject design because podcast producers vary a lot in experience, aesthetic taste, specific fields of interest and technical know-how, so

having each subject do both tasks in the same session allowed us to compare them to themselves instead of to other subjects.
The task order was fixed across participants (A-B) as the baseline case, as well as the condition case needed a bit of explaining, so as not to overwhelm the subject with too much information at once, we would brief the subjects before the baseline treatment about the survey and experiment and brief about the MusicBox before the condition phase. We did, however, employ a break in between to minimise fatigue. In both tasks, the participants performed a brief audio-editing and sound-bedding exercise representative of routine practice with which they were already familiar.

4.2 Participants

The study was deployed on six participants who all adhered to the following characteristics;

- Working daily with narrative podcast and radio production.
- Is not highly musically inclined, i.e. no former professional musicians.
- Is familiar with the digital audio workstation Hindenburg PRO (version 1 or 2) and is capable of using the program at a professional level.

The participants were selected mostly out of our own network, and are either working for Danish media companies, like Jyllands-Posten, Munck Studios or Danmarks Radio (DR), or studying podcasting at the Danish podcasting school Æteren, or are working as podcast teachers at Krogerup Højskole or Æteren.
Before each of the sessions, we had a small chat surrounding their normal workflow to ensure that each subject had a high level of knowledge and experience in adding ambience and bedding, as well as creating soundscapes for podcasts and radio.

On a note of ethics, we ensured that all participants were given informed consent to be interviewed and were told they could stop at any time with no penalty. They were also informed that we would save and use their work from the tasks for this study.

4.3 Material & Setup

The study was conducted using two briefs that both included the following source pack:

- Access to the newest version of Hindenburg PRO (version 2.05).
- A 12-minute-long storyline, consisting of either narration via speech and two sets of dialogues between a set of actors (brief 1) or descriptive narrations framing an interview (brief 2).
- A premade SFX pack of both synthesised and real-world samples relevant to the specific story, accompanied by some more default and cinematic sounds, consisting of a total of 30 sounds.

- Access to the sound library Soundly [4](a built-in feature of Hindenburg PRO).
- As for the baseline only, a royalty-free music pack consisting of 28 descriptively titled musical tracks in various lengths (20 seconds to 3 minutes), all consisting of different musical passages within each musical track.
- As for the condition treatment, access to the MusicBox v0.1, with all its features.

The subjects were all given the same hardware, a Windows laptop with all the data and applications as stated above, as well as a pair of headphones or monitors (depending on location), to ensure a high quality of audio.

The audio material, i.e. narration, SFXs and ambience, for the two briefs, was all from our own sound libraries. The royalty-free music pack for the baseline task where explicitly found for this study, through a generous post on a podcast forum, on the website Reddit[5].

4.4 Task Flow and Procedure

The main part of the study, the baseline and condition treatment tasks, was where the subject created a small storyline, anywhere from 1 to 5 minutes, with the explicit focus on quality over quantity, that they themselves were pleased with, using the assigned story brief.

The task flow of the survey was the following;

1. *Briefing the subject on the study (10 min)*
2. *Recap - (0 to 5 minutes)* - If the participant is a Mac user, a short recap of Hindenburg Operations, see below.
3. *Baseline Task (30 min)* - First task, create a narrative story from either Brief 1 or 2.
4. *Questions (10 minutes)* - A few questions from the observer to clear up for the log.
5. *Fill out the Creative Support Index (CSI) questionnaire for baseline (5 min)*
6. *Break (15 min)*
7. *Briefing and introduction to the MusicBox (5 to 10 minutes)*

[4] A complete sound effects platform that provides a searchable cloud-bsed library.
[5] https://www.reddit.com/r/podcasting/

8. *Condition Task (30 min)* - Second task, Create a narrative story from the other brief, now using the MusicBox.
9. *Questions (10 minutes)* - A few questions from the observer to clear up for the log.
10. *Fill out the Creative Support Index (CSI) questionnaire for condition 1 (5 min)*

A total survey time of approximately 2 and a half hours

When briefing the participants on the task at hand, a strong emphasis was put on the sound laying and bedding aspects of the task. The participants were allowed to do whatever they wanted with all the material, but the importance of their work was adding SFX's sound and especially music, in contrast to editing the spoken word or focusing on the narrative aspects. They were also encouraged not to try to add sounds to the entire project, but rather to create a passage that they were satisfied with.

While the baseline and condition treatment tasks were being done by the subject, the observer would log their process, further expanded in the data collection segment (4.5), as well as assist users if certain technical problems arose.

The recap was only for the subjects requesting it, and was a way to minimise the friction of working with different operating systems, as the shortcuts are different and as these are often muscle memory, this introduces more frustration. The observer would also be of assistance to the subject during the tasks, if any unnecessary friction occurred.

4.5 Data Collection

For each participant, we assembled a dataset to enable robust, within-subject comparisons between the baseline and the condition treatments.

Firstly, we maintained an observational log while participants worked with both tasks. The purpose of this log was to capture how their work unfolded, their sequences of actions, points of hesitation or breakdown and serendipitous behaviour.

After each of the tasks, the questions would help us to clarify specific points of interest and give the subject the chance to explain parts of the process, both in relation to their thoughts during the task as well as the final result. These reflections suppled us with the participants' reasons for their choices, perceived affordances, limitations of the tools, and any contextual factors that might not be clear to the observer, e.g., prior habits or unfinished ideas - due to the time constraints of the study.

For every task, we administered the creativity support index, see the creative support index questionnaire segment (4.6).

Finally, we archived each completed production as a Hindenburg PRO session, preserving audio stems and any available session markers. This format enabled us to analyse participants' working practices, examining session structure and tool use within Hindenburg PRO, in addition to evaluating the rendered audio.

4.5.1 The Creativity Support Index Questionnaire

For recording ratings of mental effort, enjoyment, exploration, expressiveness, immersion, results worth effort, and collaboration we deployed a slightly modified version of Cherry, E., & Latulipe, C (2014)'s Creativity Support Index (CSI) questionnaire (see Attachment 1), where we've added the entry of an optional description of a serendipitous discovery. The CSI measures on a seven-point Likert scale and is a proposed standardised tool for measuring creative software as based on an older NASA TLX load index (Hart, S. G., & Staveland, L. E., 1988). As Cherry, E., & Latulipe, C (2014) state; *"Researchers are interested in comparing the creativity support of two similar tools, so they design a study in which participants complete the same creative task in both tools"*.

Although our tool, the MusicBox, does not explicitly support collaboration, we retain all CSI items to preserve the tool's standardised structure. Cherry and Latulipe (section 4.6, 2014) emphasise that even when collaboration is not relevant, the full CSI should be administered so its weighting system and overall index score remain comparable across studies, for maintaining methodological consistency supporting future comparison.

Our rationale for using the CSI is, hence, twofold: it offers a validated, standardised instrument for creative-support evaluation, and the 1–7 scale provides a way to detect the within-participant differences between conditions. These quantitative measures complement the qualitative materials and assist us in remembering each participant's differences, in addition to the observer logs.

4.5.1.1 The Serendipity Self-Report Addition

In addition to the CSI, we gathered a self-report of serendipity for each task. Noting whether any serendipitous event led to a change of plan. We used a four-point scale (from not at all to slightly, moderately and significantly) to encourage decisive responses. As the subject stated the event, a few follow-up questions would inevitably arise, and the observer would log further information for clarification.

The serendipity self-report is to serve two purposes: to prompt participants to reflect explicitly on the concept so that it can be meaningfully discussed in the post-task questions, as well as to log their serendipitous discovery with a note as to whether it made them change their plans of production.

4.6 Observer Log and Focus

As the participant undertook the tasks, an observer was present to log notable activities, including moments of frustration, workflow patterns and serendipitous discoveries. In particular, the observer attended to two continually shifting focuses in the participant's behaviour. The first concerns the mode of thinking: whether the participant was engaged in divergent thinking or convergent thinking (Frich et al., 2021; Zhou et al., 2021; Matyas, K, n.d.). The second focus concerned the shifting granularities in the level of operations: whether, at a given moment, the participant was addressing a higher-level task or a lower-level editing setting (Adar et al., 2014; Rubin et al., 2012).

Importantly, these dimensions are not mutually exclusive; a participant may, at any instant, combine modes (e.g., be both divergent and high-level, or divergent and low-level), and may fluidly transition between them over time as well as change these modes of focus at an instant.

4.6 Limitations of the Study

This study prioritised the measurement of participants' workflows. However, several features of the experimental set-up may limit ecological validity. I've put these into two different categories: structural limitations and technical limitations.

4.6.1 Structural Limitations

As outlined in the study design, we employed a within-subject, non-counterbalanced design. A within-subject approach was chosen because the research question hinges on comparing each participant's established workflow with our proposed system, the MusicBox. Given the varied nature of participants' prior experience, treating each participant as their own control offered the most reliable basis for evaluating differences in process and outcome between the baseline treatment and the condition treatment.

As for the decision not to counterbalance, this was primarily structural. All participants completed the baseline task first, working with royalty-free music, followed by the condition, where they worked with the MusicBox. Because the tasks were relatively complex, a substantial briefing was required before the baseline, and an equally extensive introduction to the functionality and workings of the MusicBox was necessary before the condition. Reversing the order would have required delivering both the task briefing and the MusicBox training up front, risking information overload and impeding participant performance. Preliminary pilot trials with MusicBox alone suggested that such front-loaded instruction was overly demanding, which informed the choice to maintain a fixed order.

In narrative podcast production, practitioners typically curate personal libraries of ambience, SFX, and musical materials and develop deep familiarity with these assets. In the present study, standardisation required all participants to work with the same briefs and source materials, with which they had no prior familiarity. Although filenames followed a descriptive convention (e.g., “ambience_bus_driving_by_in_HøjeTaastrup”, “Inspirational Piano”), participants necessarily expended additional time and mental effort to interpret the material before it could be used. This constraint may have increased extraneous load and reduced efficiency relative to participants’ usual practice.

Participants were provided with a pre-defined narrative in which the timeline and spoken-word elements (i.e. interviews or speech) were already fixed. Working *in* medias res in this way is atypical, as producers are ordinarily involved in discussions about content and structure. The imposed brief, therefore, constrained editorial latitude and may have altered the kinds of decisions participants would normally make.

Taken together, these factors likely placed participants outside their comfort zones and routine workflows. To contextualise results, we therefore administered brief follow-up questions after each condition to capture participants’ preferred and customary ways of working. Findings should be interpreted with these constraints in mind, and future studies might mitigate them by allowing limited personal-library use, introducing a familiarisation phase with the shared assets, or permitting greater involvement in shaping the narrative brief.

4.6.2 Technical Limitations

The briefs were constructed by the research team using material from prior productions to construct simple, coherent narratives. Two genres were selected: true crime and personal storytelling, on the basis that they are prevalent in contemporary podcasting and, in our judgement, likely to be familiar to most producers. While this choice promoted comparability across participants, the briefs may not fully capture the diversity and complexity encountered in professional contexts.

The MusicBox application evaluated here was an early-stage build (v0.1) and was therefore tested exclusively on a single workstation. To reduce development complexity, no cross-platform validation was undertaken; consequently, all sessions were conducted on the same Windows machine. Because keyboard shortcuts in Hindenburg PRO differ between Windows and macOS, macOS-native participants faced potential interactional friction arising from unfamiliar key bindings and operating-system conventions. As stated, we provided a brief familiarisation period before each session and offered assistance if these OS-related issues arose.

4.6.3 Limitations of Results

As for the limitations of our results, an immediate and structural limitation was due to the nature of narrative audio storytelling, and the subjects' relation to their sound effect and music libraries.

There is no single "normal" workflow for producing narrative audio; practices vary with genre, cultural background, professional experience, and training. If one were to generalise, a typical sequence might involve ideation, recording (speech, ambience and SFX), development of the narrative structure, the editorial phase using personal libraries and/or royalty-free music, and final detailing before release. Our study intentionally measured only a narrow segment of this pipeline: the editing phase. The apparent limitation of our study is that editorial choices are often conditioned by upstream activities; they might have recorded a special SFX during the recording phase, or they might have sounds in their personalised sound libraries which they have been wanting to utilise since the ideation phase.

These contextual factors were not reproduced: participants worked with a fixed brief and standardised library, without prior influence and knowledge of the recorded material or the story in question. Consequently, decisions observed in the study may not fully reflect participants' customary practices, limiting external validity.

4.6.4 Different Possible Study Design Approaches

One seemingly straightforward and quite naïve solution would be an end-to-end field study in which participants create a complete narrative production from scratch over several months. While this could be done in collaboration with an educational institution, the design would be highly resource-intensive, likely unattractive to busy professional producers due to the length and workload of the study, and would introduce numerous uncontrolled variables, complicating qualitative comparison.

A more practicable direction is a within-subject, counterbalanced, longitudinal study, focused exclusively on professional podcasters and radio producers currently working in the field. Each participant would complete two production periods (e.g., ~6 months each): one using their normal workflow (baseline) and one having access to the MusicBox (condition), with order counterbalanced across participants. Mixed methods (self-logging plus interviews at the start and end of each period) would capture behavioural and experiential change while preserving the realism of the process. This design would mitigate many of the present study's constraints, albeit at the cost of longer timelines and the need to manage carryover effects.

5. User Guide

Our User Guide can be found on frederikbjorn.com/userguide

To demonstrate the functionality and usecase of the MusicBox, we have made a video explaining the different functionality, the UI, as well as showcasing how it works. It features a standalone example as well as an example of its intended use as an integration into Hindenburg PRO.

Timestamps:
0:00 - Explanation of the UI.
1:35 - Starting to play and manipulate music in the MusicBox.
6:47 - How to use the MusicBox in a Hindenburg PRO Session, example 1.
10:50 - How to use the MusicBox in a Hindenburg PRO Session, example 2.

Both examples of Hindenburg PRO sessions are the briefs from our study, the once used by our participants.

6. Description of the MusicBox

To explain the use of the MusicBox, we have made a description of the application, split into three sections. Firstly, the terminology of the musicBox(6.1), which is built atop the terminology described in the background section. Secondly, the interaction with the MusicBox(6.2), which outlines and explains the functionality, interactions and UI of the application. Thirdly, the technical description(6.3), which explaines the requirements, data and audio flows and the logic behind the different systems at play. In addition to these three sections, we have included a compositional guide (6.4), which is an overview for musical composers who are to contribute to the system.

6.1 Terminology of the MusicBox

In the development of the MusicBox, to structure the different systems and elements of the application, we created an application-specific terminology which draws on a mixture of musicology (Cage, 1969; Kane, 2014; Landy, 2008; Schafer, 1994; Truax, 2012), music semiotics (Chion, 1994; Parmerud, 2009; Tagg, 2013; van Leeuwen, 1999), as well as game-audio literature (Dean & McLean, 2018; Sweet, 2014).

6.1.1.1 Mood

The term mood, also described as “emotion words” by Philip Tagg (2013) or “expression of affect” by van Leeuwen (1999), is a term describing the name of a given composition. The MusicBox v0.1 contain five different Moods, all of which are emotionally defined pseudowords; see the compositional guide (6.4) for a full explanation.

6.1.1.2 Modes

The modes of the MusicBox are varying terms for a specific instrument. They are rhythm, chords and pad. These three modes are all to be subjectively interpreted by the composer, but signify a few differences. They are chosen on the basis of apstractivity, so as not to conjure up recognisable cultural meaning (van Leuween, 1999), as exemplified in Schaefer's acusmatic music tradition, which deliberately hides the instrument, foregrounding the sound object instead (Kane B., 2014).
Rhythm signifies movement and forward drifting, which in a domain-specific context can be used to build up to an emphasis point, either narratively or musically. Chords signify background, tonal character and timbral base and are a common term from musicology. Pad signifies the essence of a soundscape, as “pads holes and fills gaps in the music’s overall texture” (Tagg, P., 2024, page 494). Some synonyms of our interpretation of pad include: harmonic bed, background layer or ambient layer.

These three modes also leave room for the subjective interpretation of the composer.

6.1.1.3 Theme

The theme can terminologically be thought of as an extra mode, albeit with a difference in functionality. The term theme can, in many instances, be synonymous with terms like jingle, sting or signature, but the vague character of the term theme was ultimately deemed better suited.

6.1.1.4 Musical Segments

The term musical segment comes from musicology and is traditionally used to describe parts of a musical arrangement, such as the intro, verse, and chorus. In the context of the MusicBox, however, decisions about localisation and context awareness of the arrangement are made not by the composer but by the producer. Therefore, we chose to define the two musical segments of the MusicBox simply as A and B.

6.1.1.5 Perfect Loops

The term perfect loop, is most often used in the context of game audio. It describes a method for creating seamless loops of music and ambience, allowing a segment to repeat indefinitely (Sweet, 2014). The method used to achieve this varies depending on the context; see the perfect looping system segment (6.3.6) for details of our implementation.

6.1.1.6 Randomisation Algorithm and Loop Iterations

To introduce further variation and to work in tandem with the perfect loops system, we implemented a randomisation algorithm that generates differences between successive loop iterations. We designed this in line with Parmerud's definition of variation as "changing the identity of a musical object or even a whole section of music without losing the original aspects that created this identity" (Parmerud, 2009, page 7).

A loop iteration is a single loop of a mode, including fade-in, main loop, and fade-out phases. For a more detailed explanation of the randomisation algorithm and loop iterations, see the mood system, musical segment selector and randomisation algorithm segment (6.3.4). All loop iterations have a length of 18 bars, 1 pre-fade bar, 16 main bars and 1 post-fade bar, as specified in the perfect looping system (6.3.6).

6.2 Interaction with the MusicBox

To better understand the use of the MusicBox, here is a detailed description of each state, function and feature in the application and how it affects the different streams of audio.

6.2.1 The Different States of the MusicBox

As the MusicBox is first launched, either as a VST or Standalone plugin, the user will be greeted by the default state of the application, as shown in Figure 3.

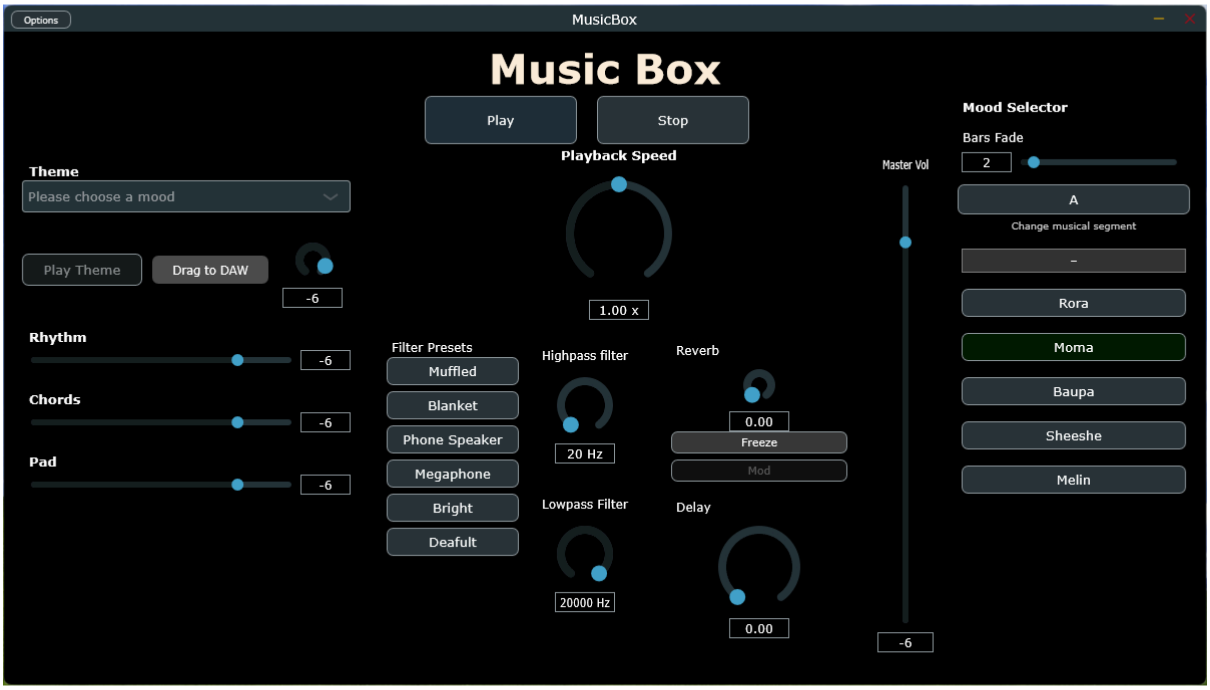


Figure 3. The MusicBox in its default state.

Firstly, the default state does not play any audio until the 'Play'-button is pressed, in which the default state of Moma (as depicted by a slightly darker tint) will begin. Alternatively, we can choose a mood, which then unlocks the theme dropdown selector and then press the 'Play Theme'-button, to play just a single instance of a theme. As part of the default set-up, all faders related to volume, i.e. the mode faders of Rhythm, Chords and Pad and the Master Vol fader are all set to a default value of -6 dB. We can also observe that all dry-wet post-processing, i.e. reverb and delay, is set to 0, that the filers are all open, i.e. highpass filter at 20Hz and lowpass filter at 20kHz and that the playback speed is set to the default value of 1.00x. We can change all these values dynamically, see Figure 4 for an example of the playing state of the MusicBox.

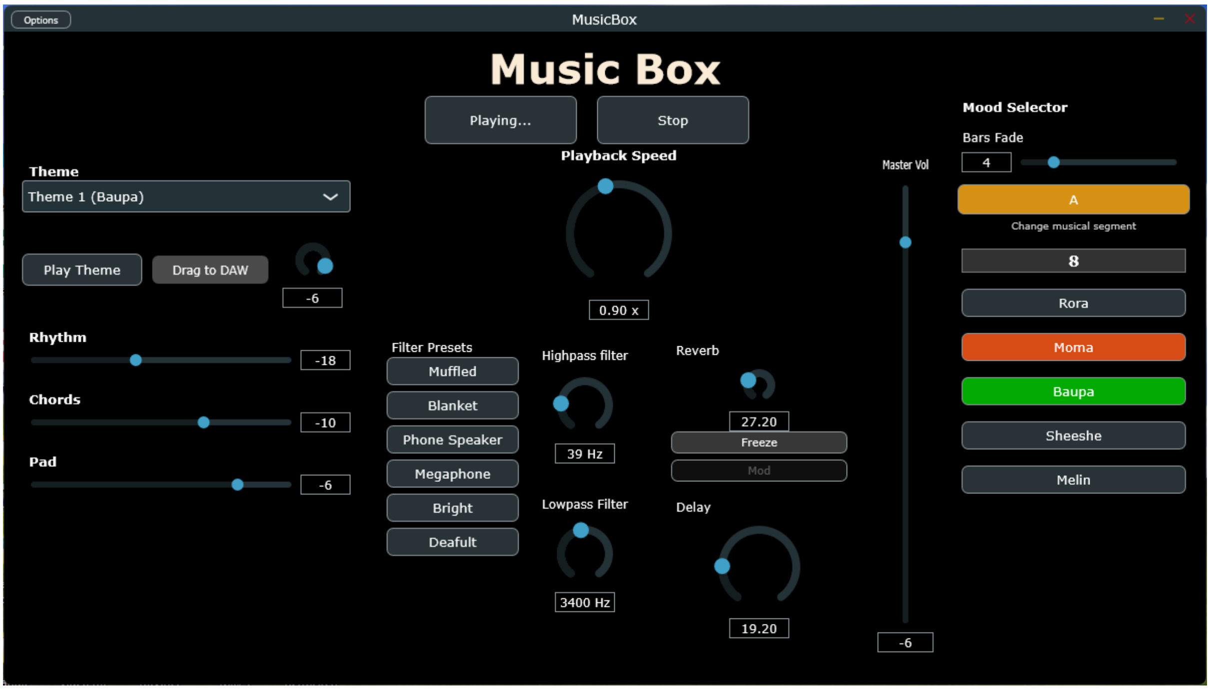


Figure 4. The playing state of the MusicBox.

For interaction with the MusicBox, we can segregate functionality and use into three categories: the background stream of audio, the theme stream of audio and the post-processing of audio.
As depicted in Figure 5, the audio stream for which the background stream plays and the themes play are different, but they are both funnelled into the post-processing elements.

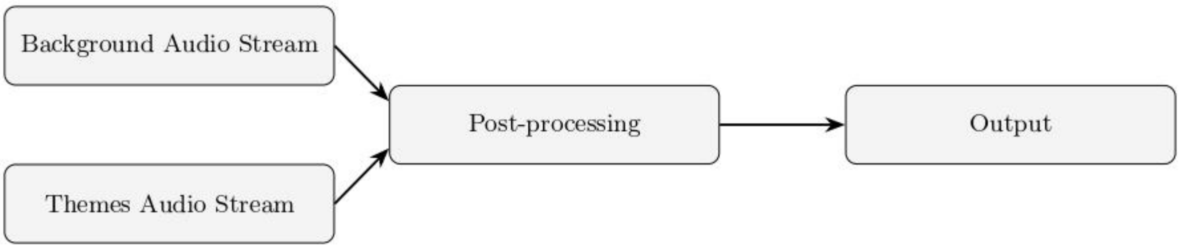


Figure 5. Relationship between the Background and Themes audio streams and the shared Post-processing stream in the MusicBox.

6.2.2 Background Stream of Audio

The background stream of audio is the main feature of the MusicBox. It is the stream which interacts with the elements of mood selection, randomisation logic, the perfect looping system, musical segment switching, post-processing, playback speed and the master volume. To initiate a background stream of audio, the user is to press the 'Play

'-button. Now the user will see the text of the 'Play'-button change to '*Playing...*' indicating the prevalence of sound. If no mode was selected prior, the mood of *Moma* will play.
The three mode faders of Rhythm, Chords and Pad are streams of audio that play within the background audio stream, and their volume can be changed independently. Each of these modes always follows the same structure and behaviour of the given mood selection, musical selection, post-processing and playback speed.

6.2.2.1 Interaction with Mood Selector

The fade time between each mode is determined by the Fade Bars value. When a new mood is selected, the existing fading-out mood will be represented by a red flashing light, and the entering mood will be represented by a green flashing light, see Figure 4. When the red and green lights stop flashing, the switch is over, and the currently selected mood is then represented by a slightly dark tint, see Figure 3.

6.2.2.2 Interaction with Musical Segment Selector

When a playing state has been initialised, a counter will display the current number of musical bars until a change in musical segment can occur, see the counter displaying the value of '8' in Figure 4. Since all loop iterations have a main duration of 16 bars, the counter will always start at 16 and then count down to 1.
When a selection is made (by pressing either the A or B button), the button begins to flash, indicating that a change in musical segment will take place at the next possible transition point.

6.2.3 Themes Stream of Audio

The themes are an independent stream of audio, which, in contrast to the background stream, is not dependent on most of the elements of the MusicBox. The themes only interact with the mood selector, post-processing, playback speed and master volume, in addition to having their own drag-to-DAW feature.
The themes are selected via a drop-down, as shown in Figure 6. The themes available in the selector are dependent on the selected mood. As shown, the mood *Moma* has been selected, and therefore, the pool of available themes is all *Moma*-specific.

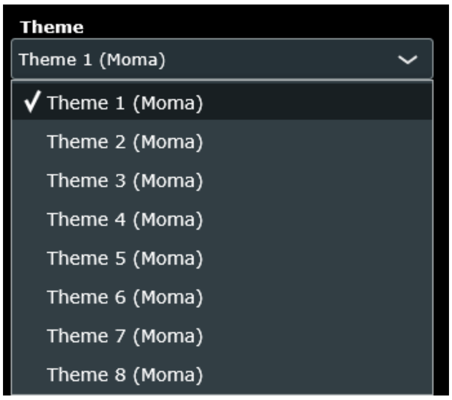


Figure 6. Theme selector initialised

After a theme has been selected, the drop-down will close, we can now interact with the two buttons, 'Play Theme' and 'Drag to DAW', as well as the volume knob for the themes. As shown in Figure 7, the 'Play Theme'-button will simply play the theme right away with no rhythmic syncopation, and in accordance with the playback speed, through the post-processing chain.

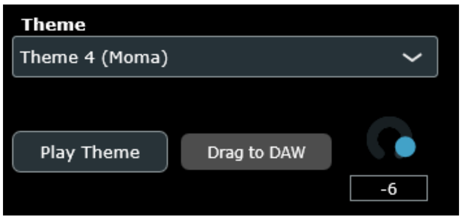


Figure 7. Theme UI when a Theme has been chosen.

6.2.3.1 Drag-to-DAW Feature

The drag-to-DAW feature is the only theme-specific feature. It renders out the currently selected theme via drag-and-drop. It can do so directly to a track in Hindenburg Pro as well as directly to a folder or desktop.
The theme will be rendered only in accordance with the specified playback speed, and will not apply any of the post-processing. The rendered file will receive the name with information of the theme number, from which mood it came and the playback speed, i.e. *"MusicBox_Theme_4_(Moma)_r1.63"*.

6.2.4 Playback speed

The playback speed is a single button with a value indicator underneath. It goes from a lowest possible value of 0.25x to 2.00x times the initial playback speed. The default value is 1.00x. The playback speed does not change any aspects of the randomisation of mood logic, as these are all calculated in musical bars, which do not change depending on the playback speed.

6.2.5 Post-processing

The last part to be determined is the post-processing part. As shown in Figure 5, both streams of audio are being funnelled into the post-processing stream before being sent to the master volume and finally to the output.

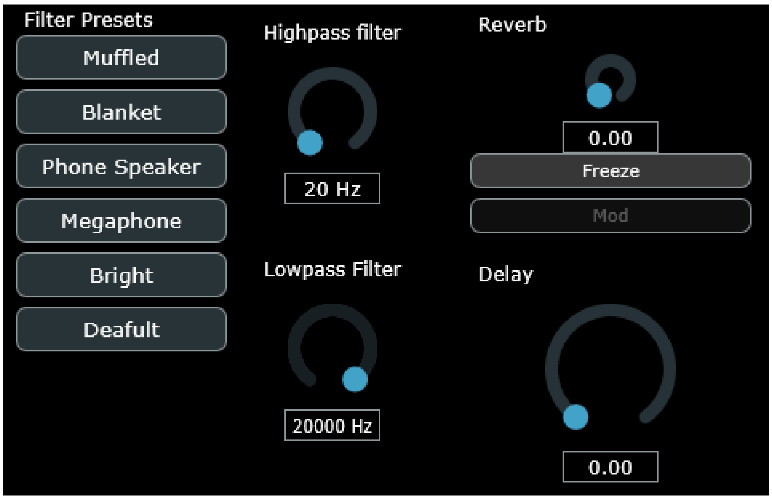


Figure 8. The Post-Processing elements.

In Figure 8, we see the three different types of post-processing: the filters, consisting of highpass and lowpass, the reverb and the delay.

6.2.5.1 Delay

The delay is quite straightforward. It is a dry/wet knob, where the user can choose between 0 - 100% of a simple pre-defined delay type.
The delay type being added is a tempo-synced stereo delay set to a constant 1/4th, it has a cross-feedback, also called ping-pong, type of delay, where the echo energy alternates between the left and right channels each delay period. It is single tap, meaning just a single stream of repeated audio which has equal power in the dry/wet mix - the knob controlled på UI.

6.2.5.2 Reverb

The reverb works in much the same way, with a dry/wet knob going from 0 to 100% of a pre-defined reverb type.
The reverb type is an algorithmic stereo reverb; it runs as a parallel path, where the amount is determined by the UI knob and the parameters are fixed to a Room size of 0.85, damping of 0.60, and stereo width of 1.00. The reverb has two additional functionalities, Freeze and Mod, as in modulation of the freeze.

When the Freeze function is enabled, it sets the reverb to infinite sustain, capturing the ambience at a specific moment in time, to create a pad-like hold of the audio. When

Freeze is active, a chorus modulation can be enabled via the 'Mod'-button on the wet path only. This adds gentle motion by modulating the volume and pitch of the freeze capture and avoids a static pad.
The volume of the sound generated by the Freeze and Mod is determined by the reverb dry/wet knob, as these functions only affect the wet signal.

6.2.5.3 Filtering

The filtering can be controlled in two different ways, either via the presets or via the highpass and lowpass knobs. The highpass is set to a default of 20kHz and goes down to 20Hz (limit of audible frequencies), whereas the lowpass knob does the opposite and starts with a default of 20Hz and can go up to 20kHz.
The filter preset selection simply changes the values of the knobs, giving the user a visual indication of the change that has taken place. The presets and their values are shown in Table 1.

Name	**Highpass value**	**Lowpass value**	**Note**
Muffled	20Hz	400Hz	
Blanket	30Hz	800Hz	
Phone Speaker	300Hz	3400Hz	
Megaphone	500Hz	5500Hz	
Bright	3500Hz	20kHz	
Default	20Hz	20kHz	The initial state

Table 1. - The UI parameter values of Filter presets.

As the names suggest, they all correspond to an emulation of a naturally occurring filtering, i.e. being under a blanket or listening through phone speakers.

The chain of post-processing work in accordance with the schema in Figure 9.

6.3 Technical Description

This part of the description will focus on the technical aspects of the MusicBox.

6.3.1 Requirements, Goals and Scope.

At the onset of development, we devised a short list of functional requirements as well as non-functional requirements which the MusicBox should include, and put them on a

MoSCoW[6] list. The MoSCoW list was helpful for the beginning of our process, not only to determine the prioritisation of our different ideas, but also to sift out potential functionality which was out of our scope. For the full list, please see Attachment 2.
When devising these requirements, we largely drew on experience from a similar project done a few years ago, which assisted us in entering this process with a number of predefined observations. An example of this would be the design of the moods: the intentionally chosen broad, non-instrumental terminology rather than fixed instrument categories (e.g., Synthesiser, Piano, Guitar, etc.) was an observation made before the start of this project.

6.3.2 System Architecture

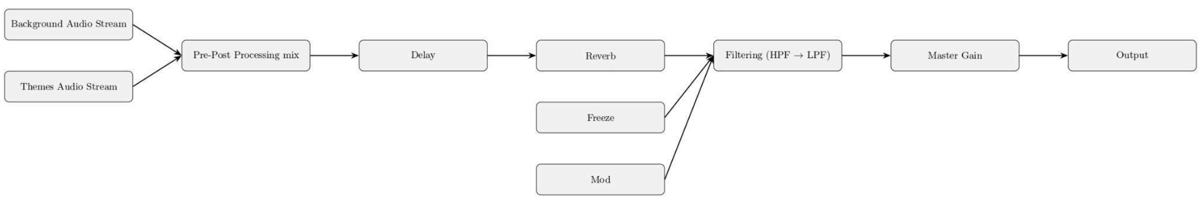
To provide an overall perspective on how the MusicBox is organised, this section presents a high-level overview of the system architecture. The following diagrams summarise how components are structured and how data and audio signals are separated and coordinated within the system.

6.3.2.1 Architectural Overview

The MusicBox is structured around a modular architecture that separates control from audio rendering. User input and interaction influence the system state and scheduling decisions, while audio generation and signal processing operate independently. As stated, audio content is produced in parallel streams, the backgrounds and themes streams, and then combined, ensuring a single output, while allowing the most flexibility in generation and processing.

6.3.2.2 Audio and Data Flow

Figure 9 illustrates the audio flow of the MusicBox, showing how background and theme audio are generated independently, combined, and routed through a shared post-processing chain before reaching the output. Figure 10 shows the data flow, highlighting how user input, timing information, and system state propagate through the system, coordinating, scheduling and routing, without directly manipulating audio signals.
The detailed behaviour of each subsystem is described in the subsequent sections; the diagrams are intended as a reference frame for understanding the system as a whole.



[6] A prioritazation tools; *M - Must have, S - Should have, C - Could have, W - Won't have.*

Figure 9. Post-processing chain of operations a larger version can be found on frederikbjorn.com/figures.

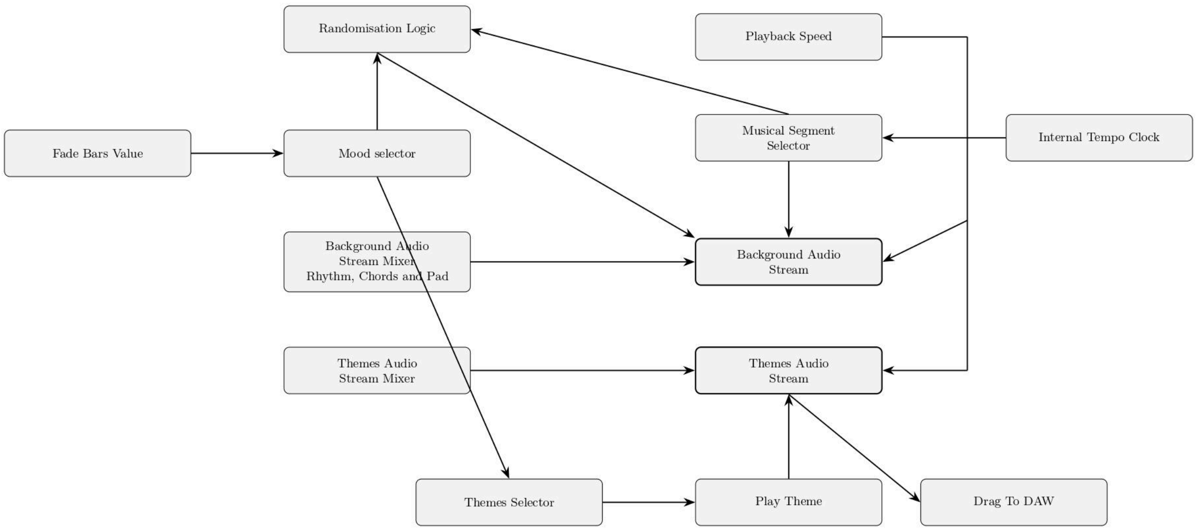


Figure 10. Data flow schema of the MusicBox,
a larger version can be found on frederikbjorn.com/figures.

6.3.3 Mood system, Musical Segment Selector and Randomisation Algorithm

To illustrate the operation of the mood system, the musical segment selector, and the randomisation algorithms, we developed a schema (see Figure 11) that visualises how the different loop iterations are executed and how the algorithm transitions when a new musical segment is selected. As previously described, the mood system plays loop iterations from the three modes, rhythm, chords, and pad, according to the selected mood. When the MusicBox is initialised, the randomisation algorithm begins by selecting one loop iteration from segment *A* for each mode. For example, if the mood *Sheeshe* is selected, the algorithm might choose Sheeshe_Rhythm_A1, Sheeshe_Chords_A3, and Sheeshe_Pad_A2.

Transitions between moods may occur at any time (see the mood transitions segment (6.3.5) for further detail).

The randomisation algorithm follows two simple rules: due to its interaction with the perfect loops system, the algorithm must select a different loop iteration once the current loop reaches its endpoint and it does so with an equally weighted change, i.e, if A1 is currently playing, the algorithm will choose between A2 and A3 with an equal probability.

As noted earlier, each mood contains two musical segments, and switching between these is user-controlled via the interface. As shown in Figure 11, segments A and B are logically and structurally identical: they contain the same number of loop iterations of identical duration, differing only in musical content. However, because the musical segment selector operates in conjunction with the perfect poops system, segment changes can only occur when the active loop reaches its endpoint. Consequently, the user interface displays the number of remaining bars until the main portion of the currently playing loop set is completed, thereby indicating when a segment transition may take place.

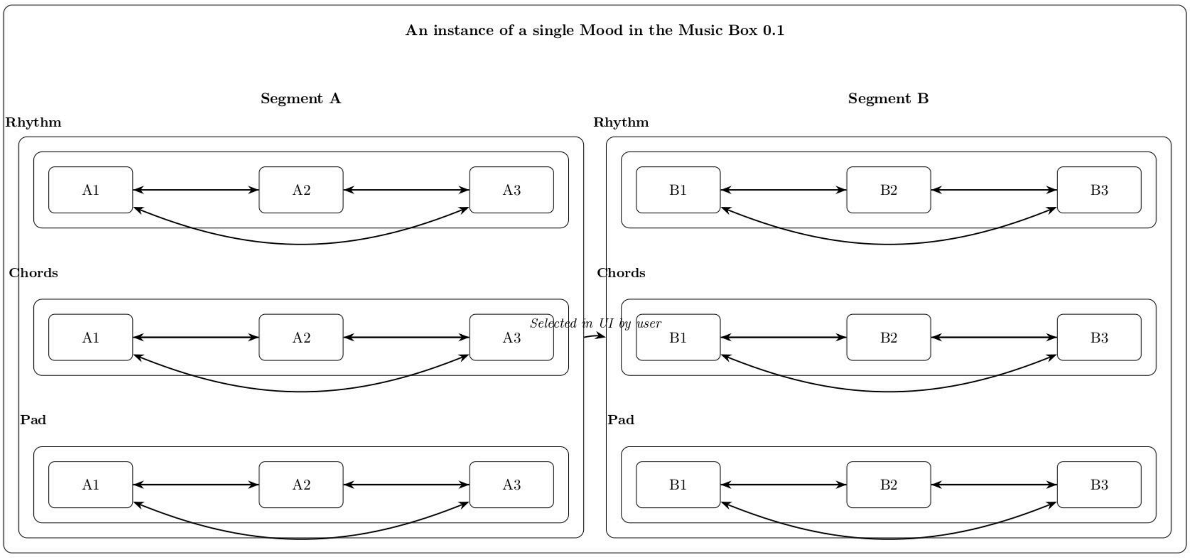


Figure 11. Showing a single instance of the mood, containing segments A and B. All arrows are determined by the randomisation algorithm, except for the switching between Segment A and B, which is upon user input, a larger version can be found on frederikbjorn.com/figures

6.3.4 Mood transitions

For transitions between moods, the MusicBox keeps all mood states running continuously and controls audibility via their mixer gains. When a new mood is selected, the crossfade engine quantises the change to the next bar boundary and then applies an equal-power crossfade between the current and target mood. The fade duration is specified in musical time, the fade bars, so transitions can be short or stretched over several bars while remaining seamless, rhythmically aligned, and free of audible level drops (see Figure 12).

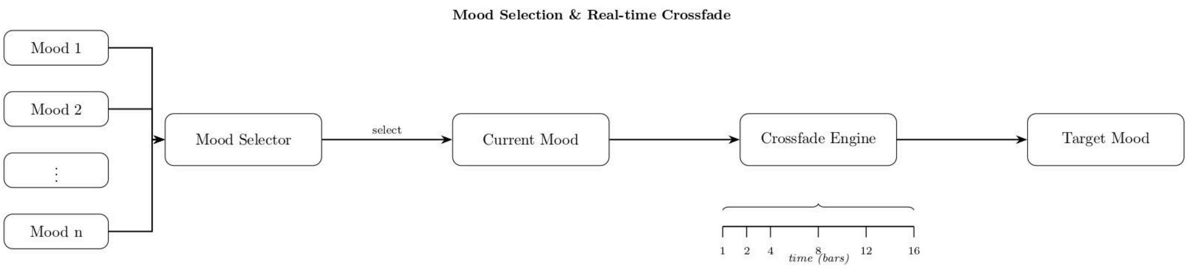


Figure 12. Schema showing the workings of the Mood Transitions and Crossfade Engine.

6.3.5 Perfect Looping System

Because the workings of the MusicBox share several characteristics with game music, such as potentially indefinite duration, interactive use, and technical constraints, the perfect looping system draws on game-audio principles as described by Sweet (2014).

The system has two main aims. Technically, it seeks to avoid audible clicks caused by discontinuities at loop boundaries. These occur when playback starts or stops at a point where the waveform is not at a zero-crossing. Therefore, the MusicBox, every loop iteration begins at a point where the amplitude is 0. Artistically, we also want some audio clips to include a pre-entry (intro) and a post-exit (tail). This not only helps to mask any remaining artefacts at the boundaries, but also allows us to incorporate elements such as risers, reverb tails, and other transitional gestures.

To support both aims, a stated, we impose a simple structural constraint on all loops in the MusicBox: each loop iteration consists of one bar of intro, followed by eight bars of the main loop, followed by one bar of tail (see Figures 13 and 14). During continuous playback, the tail of one iteration overlaps with the intro of the next, while both are underpinned by the main eight-bar loop. The only exception is the very first iteration, where we hear the single bar of intro on its own, ensuring that the MusicBox always starts with an explicit introduction before moving into seamless looping.

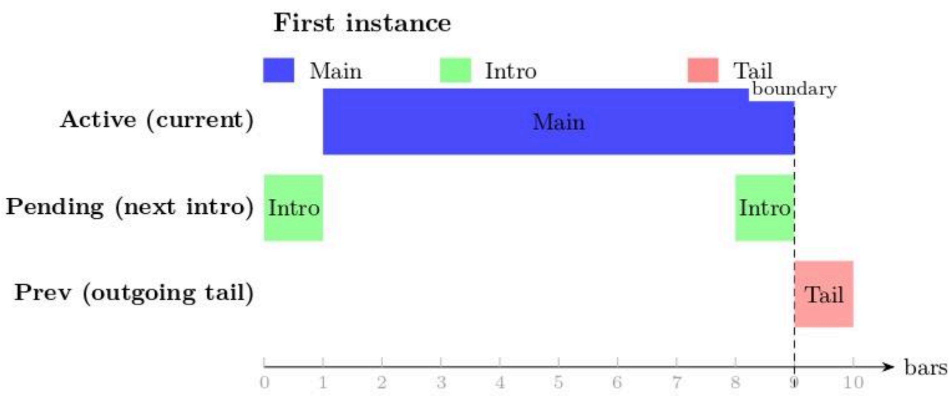


Figure 13. First instance of the loop iterations.

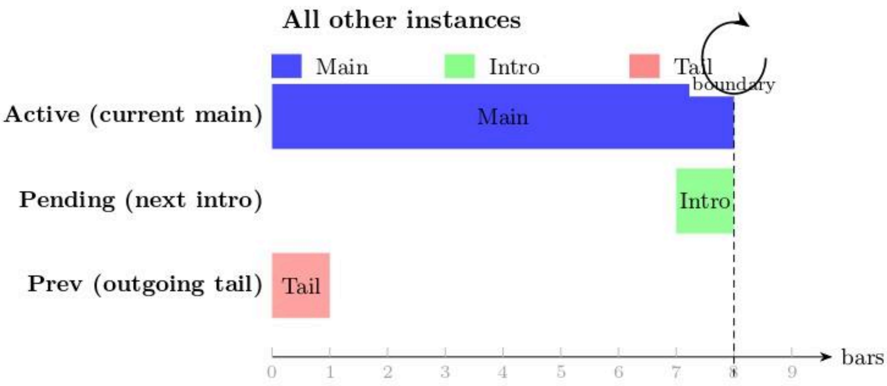


Figure 14. All other instances of loop iterations.

6.4 Compositional Guide

To support the composition of music for the MusixBox system, we have developed a compositional guide. This guide focuses primarily on the technical considerations involved in the creation of music for the system's dynamic architecture. However, it is important to acknowledge that the aesthetic and creative dimensions of crafting soundscapes for narrative audio productions are equally, if not more, important. A comprehensive exploration of these artistic aspects lies beyond the scope of this thesis; however, a short overview of some of our guidelines and observations from the industry has been included under the aesthetic considerations segment (6.4.4).

6.4.1 Compositional rules

The compositional rules, or dynamic arrangement rules, are to be met for the algorithms of the MusicBox to function properly.
Firstly, each composition must be an independent mood. This mood must represent a single abstract feeling or vibe. It must adhere to the naming rules of moods (6.4.3).
Each mood must consist of two musical segments, *A* and *B*, where *A* is the main segment, and *B* is a variation.
For each of these segments, there must exist the three modes of rhythm, pad, and chords, interpretation thereof is subjective. For each of the modes, there must exist three loop iterations, all with a slight variation in structure, not in aesthetics.
In addition to these, each mood should also have a set of themes, ranging from 8 to 10 small signal motive themes. For clarification, please see Figure 15.

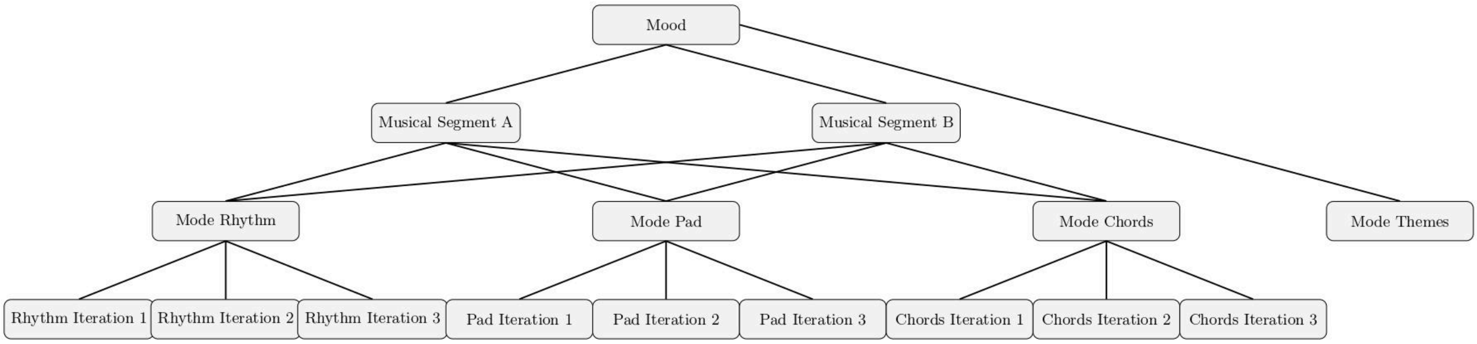


Figure 15. Structure of the Compositional Elements for the MusicBox - a larger version can be found on frederikbjorn.com/figures

In addition to these structural rules, as we must also be able to switch between moods at any given point in time, they must all be in the same musical key of either C Major or the equivalent of A Minor.

6.4.2 Technical rules

In making sure that each mood works correctly, the composer must also follow a set of technical rules. Firstly, the tempo of every single loop in a mood must be set to 105 BPM[7]. The length of each loop iteration must be 18 bars, 1 bar as fade-in, 16 bars as the main part of the loop and 1 bar as fade-out.
In addition, all audio renders must be stereo .wav files with a sample rate of 44.100 kHz, all normalised to a true peak of -4.00 dB.

6.4.3 Naming rules of moods

As for the rules of naming moods, due to past experiences with naming musical perception and feelings, we knew the importance of the names of these moods.
As the name mood implies, it is set to describe a feeling or emotion which the music can conjure for the listener. Our first intention was to focus on what these feelings should be, i.e. looking at theories of 'basic or primary emotions' (Ekman, P., & Friesen, W. V., 1971). But evidently, this approach posed three major problems.

The first being that a musical composer would need to compose quite strictly in order to meet a single specific basic emotion. This would inevitably lead to a more superficial and limited composition. Secondly, the user might rely on these emotions at face value, which would limit their serendipitous discovery by simply choosing the emotion which they feel matches their storyline. Thirdly, the perception of these moods is a matter of subjectivity, as Philip Tagg (2013, page 75) states, these moods denote a *"state of*

[7] Beats Per Minute

mind in abstracto" before further explaining, *"consider, for example, the emotion of affect behind the verbal label 'joy'. Are we talking about: [a] the joy of a small boy excitedly bubbling over as he plays with a new toy; [b] a calm, confident sense of joy slowly welling up inside someone realising that the end of the tunnel may be in sight …"*.
So in order to circumvent these issues, we consulted with a composer who suggested a purely imaginary terminology for these moodes. The thesis being that it forces the user to listen, interact and discover their own interpretation of emotions, which potentially opens up for a much broader classification of emotions.

Following this logic, we drew on research into the Bouba–Kiki effect (Gómez Milán et al., 2013). This effect demonstrates that humans tend to associate certain spoken sounds with particular visual forms and shapes, i.e. consonants with abrupt, high-energy articulations, such as the voiceless plosives (/k/, /t/), voiceless fricative (/s/, /ʃ/, /f/) and high front vowels (/i/), are typically matched with angular, spiky shapes, like Kiki, whereas consonants produced with more continuous or resonant airflow, such as voiced plosives (/b/, /g/, /d/), nasals and approximants (/m/,/n/, /l/, /r/) and rounded vowels (/u/, /o/, /a/), are more often linked to rounded, curvilinear shapes, like Bouba.

Building on this phenomenon, we developed a systematic guide for generating non-lexical labels, see Attachment 3.

6.4.4 Aesthetic Considerations

As a short section of context, we will, from a musicology perspective, provide a bit of background for the aesthetic considerations of composing background music and soundscapes for narrative audio stories. Firstly we will explain to concepts which we have found useful throughout our compositional journeys, naimly the difference of note music and sound music as well as the concept of reduced listening.

These contrasting terms of note music and sound music, follow Leigh Landy's distinction in "On the paradigmatic behaviour of sound-based music" (Landy, L., 2008). Landy labels what most would simply call music, *note music*, characterised by the primacy of rhythm, melody, harmony, and traditional motivic development, typically representable in musical notation. In contrast, *sound music* foregrounds timbre, ambience, and mood, and may be used to shape auditory environments and atmospheres (Therapontos, N, 2012).

Reduced listening or l'écoute réduite, as introduced by Pierre Schaeffer (Kane, B., 2014), is a mode of listening that sets the audio source in parentheses and focuses on the intrinsic properties of sound, i.e. pitch, loudness and timbre. Emerging from acousmatic practices, it emphasises sonic characteristics and affective associations rather than the physical origin of sounds (Moore, A., 2017).

Compositions must adhere to our definitions of soundscape (2.2.2.3) and background music (2.2.2.1), here by stating it must be subtle and mainly take the sonic placement of field (van Leuween, T., 1999). On the placement in the sonic scape, the compositions must adhere to the writings on reduced listening as well as be oriented towards sound music,

In terms of the medium of this music, in our experience, instrumentations tending towards the acousmatic, i.e. where the sound source is not readily attributable to a specific instrument, thereby reducing culturally specific associations, supporting its use as background music or soundscapes rather than a musical foreground. A reason for this is also to keep the space open for diegetic sound[8], i.e. if part of the story line happens to be set in an Irish pub, we can use the diegetically fitting instrumentation of fiddle and flageolet[9].

On a note crossing the technical and aesthetic, it is also important that the single loop iterations only imprint a minimum of any wet post-processing (reverb, delay, etc.), as these effects can be added in the MusicBox, to the preference of the narrative audio storyteller.

Our five moods, *Rora*, *Moma*, *Baupa*, *Sheeshe* and *Melin*, were all composed according to these principles.

[8] Diegetic sound, as opposed to non-diegetic sound, refers to audio elements which originate from the story world and are perceivable by the characters, such as sounds that are live in the scene, i.e. music from an orchestra or a playing from a radio, as well as environmental sounds or foley.
[9] Traditional Irish Instruments.

7. Results

This section encapsulates the findings, as derived from testing the MusicBox v0.1, as stated in the methodology section. As we have already stated how we have done the survey, this section will be focused on the results of our study.
The data is synthesised from three sources: the observations and discussions undertaken during sessions with each of the participants, the empirical data from the study questionnaire of the Creativity Support Index and their production Hindenburg PRO sessions from each of their two treatment tasks.

7.1 Quantitative Results (The Creativity Support Index)

As stated in methodology (4.5.1), we utilised the Creativity Support Index (CSI) as proposed by Cherry, E., & Latulipe, C. (2014) (see Attachment 1), with an additional question added regarding serendipity.
As stated in the study design (4.4), after each of the treatment tasks, the subjects would fill out the CSI. For each of the questions in the CSI, we have synthesised the data into two charts: a paired score showing the individual trajectories as well as a mean score ($\pm$ Standard Deviation) of the baseline vs. condition treatments. We can use these two visualisations to see the overall conjectures, the agreement of the participants, and to see if there are any outliers.

7.1.1 Results of the CSI

Metric	**Baseline (Mean)**	**Condition (Mean)**	**Improvement?**
Mental Effort	5.17	4.00 (SD $>$ 1.5)	No
Enjoyment	3.58 (SD $<$ 1)	6.50 (SD $<$ 0.5)	Yes
Exploration	3.33 (SD $>$ 1.5)	5.92	Yes
Expressiveness	2.42 (SD $<$ 1)	5.33	Yes
Immersion	3.67 (SD $>$ 1.5)	4.75 (SD $>$ 1.5)	Yes
Results Worth Effort	4.08 (SD $>$ 1.5)	4.92 (SD $>$ 1.5)	Yes
Collaboration	3.33 (SD $>$ 1.5)	5.08	Yes

Table 2. An overview of the Baseline and Condition treatments' mean score, with high and low standard deviation noted.

To interpret these findings, as we can see in Table 2, the most immediate observation is the consistent upward shift from the baseline mean to the condition mean across nearly

all of the CSI metrics. With the exception of mental effort, every measure shows an improvement when participants used the MusicBox. Several metrics, most notably enjoyment, exploration, and expressiveness, exhibited increases of more than 2.5 points on the 7-point Likert scale, indicating a substantial perceived enhancement in creative support.
The measures, such as immersion and results worth effort, showed relatively high standard deviations ($SD > 1.5$), indicating substantial variability across participants. This is to be expected given the diversity of participants' workflows and experiences.
All taken together, the quantitative results indicate a general trend toward improved perceived creative support when using the MusicBox, especially in domains closely tied to exploratory and expressive work. These findings provide a useful quantitative backdrop for our qualitative observations.
A diagram for both individual trajectories as well as mean results for each of the CSI measurements can be found under Attachment 4.

7.1.1.1 Serendipity Self-Report

As described in the methodology section (4.5.1.1), we added a self-report item on serendipity to the CSI questionnaire.
When examining these self-reports, a clear pattern emerges: serendipitous events were reported far more frequently in the MusicBox condition than in the baseline workflow. Only one participant indicated experiencing serendipity during the baseline task, whereas all six reported at least some degree of serendipitous discovery when interacting with the MusicBox (see Figure 31 in Attachment 4).

However, these findings should be interpreted cautiously, as participants were introduced to the concept of serendipity as part of the study procedure, which may have primed them to identify such moments more readily.
Nonetheless, the strong contrast between conditions suggests that the dynamic and exploratory nature of the MusicBox afforded more opportunities for unexpected yet creatively useful outcomes.

7.2 Workflow Behaviours and Cognitive Patterns

As stated in the background section, our main focus has been on the analysis of cognitive load. In particular, the two aspects of levels of operation and the divergent and convergent thinking or phases.

7.2.1 Levels of Operation

As outlined in the Problem Analysis (3.3), static music inherently forces producers into a low-level, detail-oriented editing setting. This pattern was evident across all participants in the baseline condition and was frequently described as an unavoidable and undesirable aspect of working with fixed audio files.

7.2.1.1 Levels of Operation - Baseline treatment

A representative example comes from subject 5, who explicitly searched for a music track that could be turned into a perfect loop. The subject articulated several criteria: the track needed to be rhythmically simple, have minimal reverb tail, contain a clear onset and ending, and it should "sound good" once manually looped. After identifying a promising segment, the subject imported it into Hindenburg PRO and attempted to determine the tempo by physically tapping on the table to find the tempo. When the perceived loop point approached, the subject relied on the B shortcut in Hindenburg PRO to execute a cut at precisely the right moment. However, as both the subject and observer noted, this process was fragile and error-prone. Even slight imprecision resulted in audible discontinuities, like clicks, timing shifts, or tonal mismatches or more. An example of this issue can be heard at timestamp 0:33 in "Subject 5 – Brief 1 excerpt" (see Attachment 8).

A second recurrent issue relates to knowledge limitations during exploratory listening. Because participants needed to audition large numbers of static tracks to locate a suitable candidate, most listened only to the initial 5–10 seconds before moving on. This meant that when a track was selected, participants often had incomplete knowledge of its full structure. In several cases, a later, unforeseen musical element pulled participants out of the high-level editing setting and forced into low-level corrective editing. One striking case was the royalty-free track 'heartbeat_full'. Three participants (subjects 2, 3 and 5) initially chose it as a promising piece of background music, but all encountered frustration when an unexpected shift in the music occurs, as it introduces programmed drums and a harsh string synth 20 seconds into the piece (see Attachment 7). This shift disrupted the intended mood and pulled the subject out of their high-level editing setting, and a visible sense of frustration was evident. Another example of this was the struggle of ending a musical track, especially observed for subjects 4 and 6. In addition, as static music has no way of being extended, either the subject has to do manual low-level editing to try to extend the musical passage or find another suitable track, which is longer. This created obvious frustrations for the subjects, as both of these possibilities broke their high-level focus.

7.2.1.2 Levels of Operation - Condition treatment

In contrast, these issues did not appear in the MusicBox sessions. Participants used the system's dynamic features, mode faders, mood switching, and musical segments to create musical passages of arbitrary length without encountering structural

interruptions or unwanted musical elements. This allowed participants to remain in a high-level editing setting for longer periods.
For example, subject 1 used the faders and delay settings to create a controlled musical exit: by gradually increasing delay before stopping the playback, they produced a smooth, musically coherent post-fade that could transition cleanly into other material. A demonstration of this behaviour can be seen at the 1:35 mark in Subject 1's Brief 2 project; on the green audio clip on the fourth track named 'MusicBox' (see Attachment 5).

Across participants, the MusicBox enabled workflows that relied less on precise, technical edits and more on high-level shaping of mood, pacing, and texture. While some challenges remained, such as understanding segment switching or replicating previously discovered configuration, the overall pattern suggests that the MusicBox reduced the frequency of low-level corrective tasks inherent to static audio and enabled a workflow more aligned with the high-level operational demands of narrative audio production.

7.2.2 Results in Relation to Divergent vs. Convergent Phases

As described in the Problem Analysis (3.2.2), we conceptualise these modes of thought as two complementary phases within the creative workflow. Across the study, five of the six participants explicitly recognised a shift between these phases during their sessions, even though none had formal terminology for it. Participants instead used expressions such as "exploration phase" or "ideation phase" for the divergent phase, and "putting it all together/collecting it" or "making it real" for the convergent phase.

7.2.2.1 Divergent vs. Convergent Phases - Baseline treatment

As for the baseline treatment, all subjects started out by playing the speaker track to understand the story, feeling and so on. After listening to the speaker, they would then listen to the ambience and background music. Four out of the six subjects kept the narrative speech running while listening to the ambience and background music, to "get a feel for what the different compositions do" (subject 5). The subjects has various ways of saving or marking the musical tracks, some put them into a separate muted audiotrack, some in a different container in the clip board, some wrote the name down on a piece of paper, the method varied but the result was the same, all subjects created a smaller curated source-pack, which they where going to use in the convergent phase.
However, because the baseline relied on static music, participants were required to adopt practical strategies for managing the large number of available tracks. Each subject showed their own method for marking promising candidates, placing them on a muted track, storing them in the clipboard in Hindenburg PRO, or writing down the

filenames on a piece of paper. Although these strategies varied, the purpose was consistent: to curate a smaller, manageable subset of materials for the later convergent phase. This behaviour reflects the necessity of early divergent exploration, but also highlights the organisational overhead imposed by static material.

7.2.2.2 Divergent vs. Convergent Phases - Condition treatment

A consistent pattern across participants was the use of the MusicBox for exploratory interaction early in the task. Several subjects began by placing the MusicBox as the field under the narrative track and listening while they adjusted mode faders, post-processing, or playback speed. This behaviour aligns with divergent thinking: evaluating mood, texture, and astereic fit to the story. Participants often described this mode of interaction as “fun,” “playful,” or “composing with faders,” suggesting reduced extraneous load relative to static editing.
As the condition treatment supported the divergent phase very well, the convergent phases however showed a few issues. Like the difficulty in re-creating certain states, a difficulty in understanding what will happen from a specific interaction, suggesting a learning curve and a difficulty in creating the final finished music track, suggesting a knowledge of the convergent phases but a difficulty in executing it.

7.3 Interaction Patterns With the MusicBox

To get a deeper insight into how the subjects used the MusicBox, we have here tried to synthesise the points into different categories.

7.3.1 Use of Modes and Faders

As explained in the technical description (6.1.1.2), the MusicBox contains three modes, where their volume can be adjusted via a slider in real time, see Figure 4. This gave the subjects the possibility to mix the music live, giving the subjects a sense of “playing the MusicBox, like an instrument” (subject 5).
This feature was used in many different ways, both convergently and divergently, i.e. as a way of creating a soundscape or to modulate said soundscape live underneath the speaker. The faders also gave way for the subjects to end their score, by fading out the three different modes with an offset, creating a dynamic and personal way of ending a soundscape.

7.3.2 Interaction With Mood Selection

The mood selection was, by all subjects, the first thing they interacted with (presumably given the prior instructions). Most subjects did not interact with the fade bars selection, keeping it at the default value of 2. Some subjects commented that they wanted to be able to make the switch at an instant, but when inquiring them as to how

and why, their focus was mostly on the speed at which they could listen to all moods. Subject 3 requested a way to listen to all the modes of each mood, to more easily get an overview of what each mood encapsulates.
The moods of the MusicBox were most often explored non-linearly, giving an indication of the importance of the naming of the moods. Three subjects recognised that they had chosen a mood based on the look and feel of the name, without knowledge of the baupa and kiki phenomenon, see the naming rules of moods (6.4.3) section for reference.

7.3.4 Use of the Musical Segments A and B

The musical segment selector was one of the least-used functionalities of all of the sessions. Most of the exploration was therefore done on the default musical segment, A. The reason for this might be due to either time constraints or instructional fatigue, as subject 5 noted afterwards, “I did not know that it was possible to change musical segments!”
One points was however similar across the subjects who did interact with the musical segments, was that the time frame constraints was very annoying and most often broke the Heuristics of “User Control and Freedom” and “Visibility of System Status” (Nielsen, J., 1994a), as the subjects were unsure of how to use it and thought that the switch happes immediately.

7.3.4 Interaction With Themes

The themes were also of the lesser-used features, and most often the subjects used the themes not as their intended use, but more to create soundscapes with. Most subjects did not notice the theme selector switching along with the mood selector - serving yet as another instance of issues with the heuristics, “Visibility of System Status”. Subjects 2 and 3 noted that the simple layout of a dropdown might be to blame, as it was quite cumbersome to choose a theme and then press play to hear it. The subjects 5 and 6 did, however, use the theme as their somewhat intended use of patching holes, either in prior generated music from the MusicBox or between segments of narrative speech or dialogue.

7.3.4 Interaction With the Playback Speed Dial

The playback speed was one of the most used features. As we knew how much changing the playback speed of music changes the emotional shaping, we made the button quite big, and just immediately under the play button, see Figure 3.
All participants found great use of the playback speed, to an extent we had not even imagined. Subject 2 captured a ‘drunken’-soundscape by simply playing only the chords mode and then manually tuning the playback speed - fitting perfectly with the

storyline, see Attachment 6, the effect starts on the green audio track named 'music 2', at timestamp 0:20 of the video.

7.3.5 Post-Processing (Filters, Reverb, Delay, Freeze)

The post-processing tools were also used quite extensively. For the filters, the presets helped some of the participants and gave them a mental picture of what the effect is. The wet effects of reverb and delay were utilised for a certain "wash-out" effect, as explained by subject 3. This effect is used across the industry and is simply a way to end any musical underlay at any point in time, without sounding like an error.
The freeze effect was used sporadically. When we showed the subjects how to use it, they seemed to understand the purpose and use case, but did not do it themselves. This might be due to previously stated factors like time constraints or introduction fatigue.

7.4 Difficulties and Breakdowns

7.4.1 Challenges With Dynamic Music

As all of the participants in the study were only used to static music, the concept of dynamic music as a whole seemed quite foreign. When needing to define structure for endings and transitions, we could observe that the subjects often at times relied on the visual cues of Hindenburg PRO. Like how some chose a different colour for a certain type of track, or expecting specific affordances like cutting, duplicating and repositioning, these could of course be done at a later point in time in Hidenburg PRO, but this switch in area of responsibility was a point of content.
As the MusicBox has no way of storing a particular state, subjects had a difficult time re-creating compositions. Subjects 2 and 5 explicitly mentioned that they were missing a template feature.

7.4.2 Mental Models

As described in relation to the analysed research prototypes (3.2.1), the mental models of most existing tools are modelled around musicians. The MusicBox explecitly tries to be an exception to this, but this has proven a difficult switch to make, as some of the terminology and interactions of the MusicBox still created confusion, especially for subject 4, who spent a considerable amount of time understanding the MusicBox and its functionality.

7.4.3 Technical Constraints

A number of other missing features in the MusicBox were also mentioned by subjects. Like the possibility of solo a single mode (subject 3), the possibility to revert to a previous state (subject 4), saving current state feature, creating templates and the possibility to have an immediate mood switch possibility like having fade bars at a value of 0 (subject 1), as well as a full native integration with Hindenburg PRO.

7.5 Results Summary

Across all of our empirical data, a consistent pattern emerged: the MusicBox shifted participants' workflows towards high-level editing settings, with exploratory and expressive modes of interaction, while reducing the low-level corrective tasks. The CSI scores showed clear improvements in enjoyment, exploration and expressiveness. A number of observations exemplify a case for aiding an increase in serendipitous discoveries. Baseline sessions revealed recurring breakdowns caused by the structural rigidity of static audio, such as looping difficulties, unexpected musical events, and challenges in extending or ending tracks. In contrast, the condition sessions supported manipulation of mood, texture while enabling participants to play the MusicBox as if it were an instrument.

However, several usability issues also surfaced, including limited discoverability of features such as musical segments and themes, uncertainty around transition timing, and difficulties reproducing states. These challenges highlight points where participants' existing mental models, shaped by static-music workflows and DAW usage, did not align with the dynamic logic of the MusicBox system. Collectively, the findings suggest that while dynamic music offers clear cognitive and creative advantages, its implementation into a producer's workflow requires careful attention to learnability, visibility, and additional support for the convergence phase.

8. Discussion and Reflections

To further interpret and discuss these findings, we will look at our main research question as well as each of our sub-research questions in order to zoom in on the different aspects.

8.1 Interpreting Our Main Research Question: An Evaluation of the MusicBox v0.1

In trying to answer our quite broad main research question, we will try to evaluate our proposed system, the MusicBox, v0.1 (1.6).

Our findings suggest that creating an interactive dynamic music system for narrative audio storytellers is definitely a viable solution to the current state-of-the-art. Our main points of focus are that it should reduce low-level manual tasks, support high-level goal-oriented exploration, make musical structure manipulable without using the mental models of musicians and be designed in a generative way, which can only render usable soundscapes working within the current framework of narrative audio soundscapes. The MusicBox demonstrates that such a system can increase exploration, expressiveness, and enjoyment while enabling serendipitous discovery. At the same time, our study highlights several design challenges, especially surrounding a potential high learning curve and the support for a proper convergent refinement phase. This must be addressed for such a system to integrate into professional workflows.

Evaluating a tool like the MusicBox requires considering the system as an interconnected whole. The prototype contains a relatively large set of features, and many of these only become meaningful when the different components operate together. For this reason, when assessing the MusicBox, it is less a matter of judging individual functions in isolation and more about understanding how the system supports, or fails to support, the workflow of narrative audio storytellers.

Our findings indicate that several elements of the system, like the mood system, the mode system, the playback-speed dial, and the post-processing tools, were consistently perceived as the most valuable. Although none of these features are without limitations in their current form, they collectively supported longer periods of high-level editing setting, facilitated divergent exploration, and created conditions for serendipitous discovery. In this sense, the MusicBox demonstrates clear potential: even as an early prototype, it enabled participants to remain focused on shaping mood and narrative soundscapes rather than becoming preoccupied with low-level manual tasks.

8.2 Transition Between the Workflows: Static Music and Dynamic Music

To address the transition from static to dynamic music systems, we return to our sub-research question 1 (1.6.1).

As we expected, none of the participants had prior experience with dynamic music systems. Their existing practices, shaped entirely by static audio, influence how they interpret musical structure, timing, and control. As shown in the results, a dynamic system can alleviate many of the low-level manual operations, such as finding perfect loops, extending musical passages, or creating fitting endings for soundscapes. These capabilities reduce the frequency of workflow interruptions and technical breakdowns, enabling users to remain zoomed out and focused on the higher-level, narrative goal.

However, these benefits come with new challenges. Dynamic systems introduce a different set of expectations: the absence of familiar visual cues, the difficulty of reconstructing earlier soundscapes, and a possible steeper learning curve as compared to a static music workflow. Thus, while dynamic music offers clear cognitive advantages, it risks introducing new sources of extraneous load if users are not supported in understanding how the system behaves. Clear communication about the system's logic, like how modes, musical segments, moods and their transitions interact, must therefore be a central part of the instructional design.

When examining extraneous load in static music workflows, it becomes clear why a dynamic alternative is attractive. Tasks such as manually creating perfect loops, correcting timing, or finding workarounds for abrupt endings repeatedly pull users out of high-level creative thinking and into detail-oriented problem-solving. This shift strains cognitive resources because attention must be redirected from the narrative goal to technical repair. A clear example of this occurred when subject 5 selected the royalty-free track 'heartbeat_full' (see Attachment 7), which introduced unexpected musical elements and forced them into corrective editing, see results (7.2.1.1).

By contrast, a dynamic music system reduces extraneous load by automating or softening these technical demands. Participants spent less time cutting, matching, and structurally adapting audio and more time evaluating mood and narrative fit. The MusicBox handled tasks such as extending compositions, generating seamless loops, and providing musically coherent endings (e.g., through fading modes or using the "wash-out" effect). This allowed participants to focus on creative judgement rather than technical manipulation, lowering cognitive strain and supporting more coherent aesthetic outcomes.

8.3 Cognitive Load & Serendipity

To address specifically the subject of Cognitive Load & Serendipity, we return to our sub-research question 2 (1.6.1).

As stated, we changed the model of divergent and convergent thinking into phases of operation (3.3.2). This slight shift then mapped the concept perfectly to what we observed in the study, namely that the subjects could identify two separate phases in their interaction with the MusicBox (7.2.2).

We observed divergent exploration through several forms of interaction, such as participants "playing" the mode faders like an instrument, using the filters to edit the sonic quality of the soundscapes (subjects 5 and 6) and using playback speed to create new explorations like the "drunken" sound effect (7.3.4). These behaviours illustrate divergent exploration that does not interrupt the user's high-level focus.

However, when participants attempted to finish their compositions and integrate them into their storylines, the need for stronger support in the convergent phase became clear. Several subjects noted that they lacked core functionality, such as a saving current state functionality, templates, and automation, which made it difficult to refine and finalise their work. These challenges stem from two factors: first, the absence of standard creative-software features like saving, templates, and automation; and second, the shift from working with static to dynamic music, which alters the familiar workflow for fitting soundcape compositions into a narrative storyline. Improving this aspect would require further investigation into how producers can best work with dynamic systems.

As we can see in the results of the CSI as well as our observations, a clear trend emerges: the user will spend longer periods of time in a high-level editing setting level of operation when working with a dynamic music system as compared to a static one.

8.3.1 The Non-Specific Goal Strategy

As described in the background section (2.3.1), cognitive load theory identifies means-ends analysis as a problem-solving strategy in which users attempt to reduce the difference between the current problem and the clearly defined goal. While effective for conventional tasks, this approach imposes a high extraneous cognitive load due to the need to maintain goal states, track progress, and plan the steps in the process.

When extraneous load becomes excessive, a non-specific goal strategy can be used. By removing a specific end goal, the user is encouraged in working forward from the information available at a current point in time, focusing on current actions rather than a continuous goal comparison. This forward-working process reduces extraneous cognitive load and can facilitate serendipitous discovery.

In our study, several participants worked toward affective outcomes rather than predefined goals. For example, subject 4 aimed to create a romantic soundscape but did not follow a predefined plan or roadmap in achieving such. Instead, the participant engaged in an iterative exploratory process: creating material, listening to the result, and evaluating whether it evoked the intended feeling. This approach exemplifies a non-specific goal strategy, in which the user maintains a general direction without committing to explicit steps or a fixed outcome, thereby reducing cognitive load while supporting creative exploration.

8.3.2 Serendipity as a Creative Resource

As we've stated in the background section (2.3.4), it is not inherent that serendipity reduces extraneous cognitive load, but as we've observed, it keeps the user in the high-level editing setting. This makes a strong case for using serendipitous discovery as a measurement for reducing cognitive constraints. As the goal of the MusicBox is to engage the user in creating their own unique soundscape that fits their exact idea, this was exemplified by subject 2's use of the playback speed dial to create a "dunken"-effect (7.3.4).

The reason for including Serendipitous Discovery was originally a way to measure 'happy accidents', but evidently, through our study, we can experience that the measurement fits the audio medium more broadly, especially when compared to visual media, which is usually the default medium for measuring UI and UX, like via Nielsen's (1994) heuristics. Serendipitous discovery seems like a strong tool to use as a medium to measure the usability of creating auditory streams and processes. Especially in the case of the MusicBox, where many different components, like moods, modes and musical segments, work together in creating a new composition, which enables a broader discovery space, as compared to static music, the serendipitous discovery is a vital tool in keeping the user focused on the narrative goal, rather than occupied with technical low-level operations.

8.4 A Bridge Between Musical Composers and Narrative Audio Storytellers

To focus on a point which has not directly been measured in our survey, the case for the MusicBox as a musical bridge between composers and narrative audio storytellers, we return to our sub-research question 3 (1.6.1).

While our primary focus has been on the tools through which a narrative storyteller can manipulate and select the soundscapes provided by the MusicBox, the other end of this pipeline, namely, the composers creating music for the system, is equally critical to its success. Accordingly, we propose a guide for composers working with the MusicBox (6.4), which addresses not only integrational and technical considerations but also the aesthetic dimensions of composing soundscapes and background music for narrative audio stories.

Mainly, our focus here has been on the differences between the mental models of musicians and narrative radio and podcast producers. These models are evident as addressed by Philip Tagg (2013) and are therefore the underlying reason for the terminology used in the MusicBox, as well as the simplified design of our functionalities. An example of this is the pre-assigned values to the post-processing tools, leaving only the dry-wet value accessible in the UI, enabling the user to play the MusicBox like an instrument.

As the concept of the MusicBox as a tool for aiding deliverables between musical composers and narrative storytellers has been out of scope for this study and thesis, we feel there is a strong case for utilising the MusicBox for this purpose, not only for audio-exclusive media like podcasts and radio, but also for other media such as films.

8.5 Design Implications for Future Iterations

As the goal of our development of the MusicBox has been to create a limited version of a possible commercial product, we have also gathered insights into what a more sustainable and viable product could look like.

Firstly, there are some of the more obvious functionalities missing in the MusicBox v0.1, like a state-saving functionality, which should support a 'saving of current state' as well as a template functionality. In addition to this, our findings also show a somewhat unintuitive UI and UX, showing the need for improvements to certain possible interactions, like subject 3 asking for a 'solo modes feature', or subject 1 thinking that the musical segment change should take place immediately. So a re-design with focus on the heuristics of 'Visibility of System Status', 'User Control and Freedom' and 'Recognition Rather than Recall' (Nielsen 1994a), with a thorough testing schema, would do a great improvement to the interaction between the user and the MusicBox. For exclusively UI, a more beautiful and friendly interface would also need deployment in order to meet users' expectations of what a real tool should look like.

All of this would aid in providing better affordances for the user to understand what is going on behind the scenes. As we experienced, the inner workings of the MusicBox systems are quite complex, and a great emphasis needs to be put on the explanation of how to interact with the MusicBox, while simultaneously not constraining the users' workflows.

8.5.1 Future Integration with Hindenburg PRO

A more fully integrated version of the MusicBox within Hindenburg PRO is already under discussion with the Hindenburg Systems development team. A central design direction under consideration which explicitly supports the two distinct phases of divergent and convergent thinking. One proposed approach involves providing two complementary interface configurations: a feature-full interface intended to support exploratory, divergent interaction, and a more simplified interface for the convergent phase.

In this model, users would actively switch into a convergent mode, at which point the interface would present a reduced set of controls designed to simplify refinement and integration. For example, the interface might expose only a single filter control, a global dry-wet control governing all post-processing effects enabled (e.g. reverb and delay), and a smaller, curated set of moods. Such a configuration would aim to strengthen affordances for focused refinement by reducing visual and interactional complexity during convergence.

Further integration could also leverage contextual information from the Hindenburg session itself. For instance, narrative emphasis points, similar to those proposed in the UnderScore system (Rubin et al., 2012; 2013), could be visually represented within the MusicBox, potentially through countdown indicators aligned with upcoming emphasis moments.

8.5.2 Limitations of Interpretation

As the findings of our study and research, a few limitations of interpretation are justified. Firstly, as the MusicBox v0.1 tested was a quite early version, as described, some of the features involved were tested by all subjects as their intended purpose, making their measurability difficult. The collective nature of all the elements working together also contributed to some more muddy interpretations. Measuring qualitatively in HCI and design in general is always a difficult task, and by including standardised measures like the CSI and trying to measure cognitive load, rather than some visually aligned usability matrices, we hope that these findings can become useful for future studies and auditory design explorations alike.

Measuring a participant group as diverse as narrative storytellers for podcast and radio also imposes certain challenges, as described in the limitations of the study (4.6), we only measured a sliver of a much longer process, which also varies quite extensively from team to team. Some suggestions for improvements have been provided in the different possible study design approaches segment (4.6.4).

All taken together, the discussion demonstrates that the MusicBox v0.1 offers meaningful advantages in supporting high-level creative work, exploratory interaction, and serendipitous discovery, while also revealing important challenges related to learnability and convergence. These findings inform not only the evaluation of the prototype but also broader considerations for the design of dynamic music systems for narrative audio. The following conclusion situates these insights within the wider research context and reflects on the overall contribution of this thesis.

9. Conclusion

9.1 The Research Problem

Narrative audio storytellers rely heavily on music and soundscapes to shape mood, emotion, and meaning, yet the tools and workflows available to them or proposed in academia are built around static, linear music assets and music-centric mental models. As a result, producers are repeatedly forced into low-level technical editing tasks like sourcing royalty-free music, creating loops, timing the music, ending the music, and overall trying to recreate a musical structure. These tasks are all misaligned with their actual creative goal of shaping narrative atmosphere. This dependence on static music introduces unnecessary cognitive load, disrupts exploratory and high-level thinking, and limits opportunities for serendipitous discovery during the creative process. This thesis, therefore, set out to investigate how an interactive, dynamic music system could better support narrative audio storytellers by shifting their focus from trying to emulate a musician to what's important: shaping narrative atmosphere in a way that aligns with how they think and work.

9.2 Contributions

In order to address this problem, we have made three contributions. Firstly, the conceptual contribution of analysing current workflows through the lens of cognitive load theory, with support from semiotic musicology and a bit of UX/UI theory. Our findings here include the framing of the static and dynamic workflows, an emphasis on the importance of keeping a high level of operationalism and the importance of identifying the two distinct workflow phases, the divergent phase and the convergent phase. Secondly, the design contribution, our tool, the MusicBox, is a proposed dynamic music system, with a focus on supporting the mental models of narrative audio producers, rather than musicians, and a guide on how to produce with the tools, as well as how to compose music for it. Lastly, we have made an empirical contribution, the results from our study, which shows evidence of reduced extraneous load and increased exploration, as well as providing in-depth examples and explanations of how narrative audio producers interact with dynamic audio, their preferences towards the use of soundscapes, as well as their capabilities.

9.3 Reflections on the Research Questions

In creating this thesis, we devised one main research question and three sub-research questions (3.4).
To address the main research question, this thesis demonstrates that an interactive, dynamic music system can meaningfully support narrative audio storytellers in creating soundscapes by aligning musical interaction with their cognitive and creative practices.

Through design and evaluation of the MusicBox, the study shows that dynamic music enables producers to shape mood and atmosphere without relying on music-centric metal models or low-level technical editing. Rather than treating music as a fixed asset to be adapted, the system allows for a flexible, responsive, and narrative-driven creative process.
To address the sub-research question 1, our findings indicate that dynamic music can effectively replace key aspects of static music workflows by removing the need for manual looping, editing timing, and structural correction. The MusicBox demonstrated that musical continuity, variation and duration can be handled algorithmically while remaining under user control. This shift enables producers to work with music as an evolving process, rather than a fixed object, thereby improving flexibility and reducing workflow interruptions.
To address the sub-research question 2, our study shows that a dynamic music workflow can reduce extraneous cognitive load by minimising low-level technical operations and supporting high-level engagement. Participants spent more time shaping mood and considering the narrative fit, while exploratory interaction increased opportunities for serendipitous discovery. This suggests that dynamic systems can better support divergent thinking during the early exploratory phases.
To address the sub-research question 3, this thesis suggests that a dynamic music system can function as an effective bridge between composers and narrative audio storytellers by separating musical structure from musical control. By embedding compositional decisions into the system design and exposing only high-level controls, the MusicBox enables composers to create reusable musical frameworks while allowing storytellers to shape outcomes without any musical know-how. This approach points toward a collaborative model where compositional intent and narrative practice coexist within a shared system.

9.4 Implications for HCI

This thesis contributes to HCI by illustrating how creative tools can better support users when designed around domain-specific mental models rather than technical abstractions. The work highlights the role of cognitive load in creative interaction, particularly how static representations can force unnecessary low-level operations. It further suggests that within the domain of narrative audio, dynamic systems, though often more difficult to learn, may offer increased expressive potential. This points to a great HCI challenge; the balancing of learnability while preserving users' capacity for exploratory and serendipitous work.

9.5 Limitations

As a design research project, this thesis is shaped by a set of constraints which frame the scope of its contributions. The evaluated application, the MusicBox, has been an application at an early stage, which prioritises exploration, interaction and conceptual validation over completeness, integration of usual features of audio software, as well as a long-term integration with Hindenburg Pro. The empirical study was conducted via short, laboratory-style sessions, enabling close observation and within-subject comparison but limiting insight into longer-term adoption, the integration with the other stages of narrative audio production, as well as learning curves, and possible collaborative use. Finally, the participant group was intentionally small and domain-specific, to focus on the differences of nature within narrative audio storytelling rather than aiming for broad generalisation. These constraints do not undermine the findings but define the level at which conclusions can be drawn, situating the results as design insights rather than definitive solutions.

9.6 Future Work

Future work on the MusicBox is in a practical direction. From a design perspective, a more mature implementation would require deeper integration into the existing narrative audio workflows, as is common in Hindenburg PRO. Such an integration would include an editorial context, supporting smoother transitions between exploratory and refinement phases, and reducing friction between dynamic music interaction and conventional DAW-based editing.

Building on the observed distinction between divergent and convergent phases, future iterations should explicitly support these modes through separate interface configurations: a feature-rich, exploratory interface for ideation and serendipitous discovery, and a more constrained, streamlined interface for convergence, refinement, and integration into a narrative timeline. To support professional use, standard creative software features such as saving, templates, automations or a system with equivalent functionality, and improved visual feedback and aesthetics, would be essential.

Finally, the expressive range of the system would benefit from an exposition of the musical material, following the guidance of our proposed compositional guide. A broader and more diverse set of moods, variations, and themes would allow the MusicBox to better accommodate different narrative genres and editorial styles. Together, these directions point toward the MusicBox not only as a research prototype, but as a foundation for a viable creative tool whose continued development could be evaluated through longer-term use, within professional narrative production environments.

As sound plays a central role in narrative meaning-making by shaping mood, emotion and meaning. Designing tools for such practices, therefore, requires sensitivity not only to technical capability but to the cognitive processes through which creators explore, judge, decide and refine narrative intent. This thesis argues that when creative software respects these processes, by aligning interaction with users' mental models and supporting high-level judgement over low-level correction, it can better sustain creative flow and enable more expressive freedom. In this sense, the MusicBox is less a proposal for a specific tool than a reflection on how dynamic systems might be designed to support narrative creativity without constraining the ways in which stories are imagined, shaped, and told through sound.

10. References

Adar, E., Dontcheva, M., & Laput, G. (2014). CommandSpace: modeling the relationships between tasks, descriptions and features. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology (UIST '14)* (pp. 167–176). ACM. https://doi.org/10.1145/2642918.2647395

Alpern, A., Techniques for Algorithmic Composition of Music - Hampshire College, 1995. https://laurawhitfield.wordpress.com/wp-content/uploads/2012/10/techniques-for-algorithmic-composition-of-music-alpern.pdf

Albers, M. J., & Tracy, J. P. (2006). Measuring cognitive load to test the usability of web sites. In *Usability and information design* (pp. 255–260). https://www.researchgate.net/publication/253713707

Busch, C. (2022). *Towards a theory of serendipity: A systematic review and conceptualization.* Journal of Management Studies, 61(3), 1110–1151. https://doi.org/10.1111/joms.12890

Cage, J. (1969). *Silence: Lectures and writings.* The MIT Press. (Original work published 1961)

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction, 8*(4), 293–332. https://doi.org/10.1207/s1532690xci0804_2

Cheng, S., Hsu, F., & Hsu, C.-K. (2019). Soundscape emotions categorization and readjustment based on music acoustical parameters. In *Proceedings of the 23rd International Congress on Acoustics (ICA 2019).* International Commission for Acoustics. https://doi.org/10.18154/RWTH-CONV-239406

Cherry, E., & Latulipe, C. (2014). Quantifying the creativity support of digital tools through the Creativity Support Index. ACM Transactions on Computer–Human Interaction, 21(4), Article 21.

Chion, M. (1994). *Audio-vision: Sound on screen* (C. Gorbman, Trans.). Columbia University Press.

Cooper, G. (1998). *Research into cognitive load theory and instructional design at UNSW.* University of New South Wales.

Dean, R. T., & McLean, A. (2018). *The Oxford handbook of algorithmic music*. Oxford University Press.

Design Council. (2007). *Eleven lessons: Managing design in eleven global brands — A study of the design process*. Design Council.

Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology, 17*(2), 124–129. https://doi.org/10.1037/h0030377

Fraser, C. A., Dontcheva, M., Winnemöller, H., & Klemmer, S. (2016a). DiscoverySpace: Crowdsourced suggestions onboard novices in complex software. In *CSCW '16 Companion: Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work and Social Computing Companion* (pp. 29–32). ACM. https://doi.org/10.1145/2818052.2874317

Fraser, C. A., Dontcheva, M., Winnemöller, H., Ehrlich, S., & Klemmer, S. (2016b). DiscoverySpace: Suggesting actions in complex software. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems (DIS '16)* (pp. 1221–1232). ACM. https://doi.org/10.1145/2901790.2901849

Fraser, C. A., Ngoon, T. J., Dontcheva, M., & Klemmer, S. (2019). RePlay: Contextually presenting learning videos across software applications. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)* (pp. 1–13). ACM. https://doi.org/10.1145/3290605.3300527

Fraser, C. A., Markel, J. M., Basa, N. J., Dontcheva, M., & Klemmer, S. (2020a). ReMap: Lowering the barrier to help-seeking with multimodal search. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology (UIST '20)* (pp. 979–991). ACM. https://doi.org/10.1145/3379337.3415592

Fraser, C. A. (2020b). *Contextually recommending expert help and demonstrations to improve creativity* (Doctoral dissertation, University of California San Diego).

Frich, J., Nouwens, M., Halskov, K., & Dalsgaard, P. (2021). How digital tools impact convergent and divergent thinking in design ideation. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)* (Article 431, 11 pages). ACM. https://doi.org/10.1145/3411764.3445062

Gómez Milán, E., Iborra, O., de Córdoba, M. J., Juárez-Ramos, V., Rodríguez Artacho, M. A., & Rubio, J. L. (2013). The kiki-bouba effect: A case of personification and ideaesthesia. *Journal of Consciousness Studies, 20*(1–2), 84–102.

Hart, S. G., & Staveland, L. E. (1988). *Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research.* In P. A. Hancock & N. Meshkati (Eds.), Human Mental Workload (pp. 139–183). North-Holland.

International Organization for Standardization. (2014). *ISO 12913-1:2014 Acoustics—Soundscape—Part 1: Definition and conceptual framework*. Author.

Kane, B. (2014). *Sound unseen: Acousmatic sound in theory and practice*. Oxford University Press. https://global.oup.com/academic/product/sound-unseen-9780199347841

Landy, L. (2008). On the paradigmatic behaviour of sound-based music. In *Proceedings of the EMS08: Electroacoustic Music Studies Network International Conference*. EMS Network.

van Leeuwen, T. (1999). *Speech, music, sound*. Macmillan Education UK.

Matyas, K. (n.d.). Convergent vs. divergent thinking: How to use both to think smarter. *Maestro*. https://maestrolearning.com/blogs/convergent-vs-divergent-thinking/

Merriam-Webster. (n.d.). Background music. In *Merriam-Webster.com dictionary*. Retrieved November 5, 2025, from https://www.merriam-webster.com/dictionary/background%20music

Moore, A. (2017). *Sounds in the dark: The art of the acousmatic* [Video]. Arts Enterprise, University of Sheffield. YouTube. https://www.youtube.com/watch?v=s1S7rbPhvdw

Nielsen, J. (1994a). *10 Usability Heuristics for User Interface Design*. Nielsen Norman Group. https://www.nngroup.com/articles/ten-usability-heuristics/ Nielsen Norman Group - Updated in January 2024.

Nielsen, J. (1994b). Enhancing the explanatory power of usability heuristics. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '94)* (pp. 152–158). ACM. https://doi.org/10.1145/191666.191729

Norman, D. A. (2013). *The Design of Everyday Things* (Revised and expanded edition). MIT Press.

Oviatt, S. (2006). Human-centered design meets cognitive load theory: Designing interfaces that help people think. In *Proceedings of the 14th ACM International Conference on Multimedia (MM '06)* (pp. 871–879). ACM. https://doi.org/10.1145/1180639.1180831

Oviatt, S., Arthur, A., & Cohen, J. (2006). Quiet interfaces that help students think. In *Proceedings of the 19th Annual ACM Symposium on User Interface Software and Technology (UIST '06)* (pp. 191–200). ACM. https://doi.org/10.1145/1166253.1166284

Parmerud, Å. (2009). What is "meaning" in music? https://www.parmerud.com/downloads/Meaning%20in%20music.pdf (Retrieved November 5, 2025)

Phillips, W. (2014). A Composer's Guide to Game Music - MIT Press, 2014

Rubin, S., Berthouzoz, F., Mysore, G. J., Li, W., & Agrawala, M. (2012). UnderScore: Musical underlays for audio stories. In *Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology (UIST '12)* (pp. 359–366). ACM. https://doi.org/10.1145/2380116.2380163

Rubin, S., Berthouzoz, F., Mysore, G. J., & Agrawala, M. (2013). Content-based tools for editing audio stories. In *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology (UIST '13)* (pp. 113–122). ACM. https://doi.org/10.1145/2501988.2502005

Schafer, R. M. (1994). *The soundscape: Our sonic environment and the tuning of the world*. Destiny Books. (Original work published 1977)

Shi, Z., & Mysore, G. J. (2018). LoopMaker: Automatic creation of music loops from pre-recorded music. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (pp. 1–6). ACM. https://doi.org/10.1145/3173574.3174028

Sweet, M. (2014). *Writing Interactive Music for Video Games: A Composer's Guide.* Addison-Wesley Professional.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4

Tagg, P. (2013). *Music's meanings: A modern musicology for non-musos*. The Mass Media Music Scholars' Press. https://tagg.org/mmmsp/NonMusoInfo.htm

Therapontos, N. (2012). Sound-based music curriculum: Finding the meaning and meaningfulness of sound. In *Proceedings of the Electroacoustic Music Studies Network (EMS) Conference: Meaning and Meaningfulness in Electroacoustic Music*. EMS Network.

Tingen, P. (2005, October). Brian Eno—recording Another Day on Earth. *Sound On Sound*. https://www.soundonsound.com/people/brian-eno

Truax, B. (2012). Soundscape composition. Simon Fraser University. Retrieved November 5, 2025, from https://www.sfu.ca/~truax/scomp.html

Zhou, Y., Koyama, Y., Goto, M., & Igarashi, T. (2021). Interactive exploration–exploitation balancing for generative melody composition. In *Proceedings of the 26th International Conference on Intelligent User Interfaces (IUI '21)*. ACM. https://doi.org/10.1145/3397481.3450663

11. Attachments

All attachments can be found on frederikbjorn.com/attachments

11.1 Attachment 1 - Study Questionnaire, the Creativity Support Index

The Creativity Support Index (CSI) is a quantitative framework made by E. Cherry and C. Latulipe (2014), which is a proposed standard way of measuring the effectiveness of Creative Software on seven different points: Mental Effort, Enjoyment, Exploration, Expressiveness, Immersion, Results Worth Effort, and Collaborative Enablement. In addition to this, we have added an entry regarding Serendipity.

Study Questionnaire

Creativity Support Index (CSI) measurement of MusicBox v. 0.1

Participant Name: ____________ **Date:** ____________

☐ Baseline
☐ Condition 1

Creativity Support Index (CSI) (E. Cherry and C. Latulipe, 2014)

Please rate the tool you just used for this task. 1 = Strongly Disagree, 7 = Strongly Agree.

Mental Effort (overall)

Overall mental demand of the task.

Not at all 1 2 3 4 5 6 7 Extremely

Enjoyment

I would be happy to use this system or tool on a regular basis.

Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree

I enjoyed using the system or tool.

Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree

Exploration

It was easy for me to explore many different ideas, options, designs, or outcomes, using this system or tool.

Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree

The system or tool was helpful in allowing me to track different ideas, outcomes, or possibilities.

Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree

Expressiveness

I was able to be very creative while doing the activity inside this system or tool.

Strongly Disagree 1 2 3 4 5 6 7 Strongly Agree

The system or tool allowed me to be very expressive.

Strongly Disagree	1	2	3	4	5	6	7	Strongly Agree

Immersion

My attention was fully tuned to the activity, and I forgot about the system or tool that I was using.

Strongly Disagree	1	2	3	4	5	6	7	Strongly Agree

I became so absorbed in the activity that I forgot about the system or tool that I was using.

Strongly Disagree	1	2	3	4	5	6	7	Strongly Agree

Results Worth Effort

I was satisfied with what I got out of the system or tool.

Strongly Disagree	1	2	3	4	5	6	7	Strongly Agree

What I was able to produce was worth the effort I had to exert to produce it.

Strongly Disagree	1	2	3	4	5	6	7	Strongly Agree

Collaboration*

The system or tool allowed other people to work with me easily.

Strongly Disagree	1	2	3	4	5	6	7	Strongly Agree

It was really easy to share ideas and designs with other people inside this system or tool.

Strongly Disagree	1	2	3	4	5	6	7	Strongly Agree

** see section 4.6 of E. Cherry and C. Latulipe, 2014*

Serendipity (Happy Accidents)
Did anything unexpectedly useful happen while working?

No Yes

If yes, did it change your plan? Not at all Slightly Moderately Significantly

2

Figure 17.

11.2 Attachment 2 - Our pre-development MoSCoW

Moscow for the Musicbox categorised into Functional Requirements (FR) and Non-Functional Requirements (NFR)

Must have

- The ability for the user to 'play' and or engage with the musicbox in a playful manner, NFR *- Implemented*
- Basic mixer interface on master track (gain slider and a filter), FR *- Implemented*
- Perfect looping system, FR *- Implemented*
- States, to manage emotions, FR *- Implemented*
- Give the user freedom to always change the feeling of the music instantly at any point in time, NFR *- Implemented*
- A volume slider/knob for each instrument, FR *- Implemented*
- Three instruments or modes? - chords, rhythm and pad?, FR *-Implemented*
- Separate Theme player, FR *- Implemented.*
- Playback speed, FR *- Implemented.*
- Randomisation of loops, FR *- Implemented.*
- Randomised loops categorised into segments, FR *- Implemented.*
- A design for the filters, FR *- Implemented.*

Should have

- The ability to easily interact with the post-processing capabilities, NFR.
- Post-procession functionality like
    - Reverb, FR - *Implemented*
    - Delay FR *- Implemented.*
    - Harmonic overtones? - *Not implemented.*

- A way to interact with the post-processing together in an intuitive and functional way, NFR *- Not implemented.*
- A way to modulate the freeze of the reverb, FR - *Implemented.*

Could have

- A beautiful UI, NFR - *Not implemented.*
- A smooth feel for the modes faders to help user with volume editing, NFR - *Not implemented.*
- Eight different modes?, FR - *Not implemented.*
- Individual processing for each instrument, FR - *Not implemented.*

- Templates that would work for all modes, like a way to engage serendipitous discovery. Like setting the filter in a certain place, speed a certain place, reverb etc, FR - *Not implemented.*
- Each mode has a value of energy, 1 through 3. Something like aa filter taking out top or maybe increasing the gain, FR - *Not implemented.*

Won't have

- An overall thorough design, NFR - *Not implemented.*
- Midi editing/theme functionality, FR - *Not implemented.*
- Save state system, FR - *Not implemented.*
- One shots system, FR - *Not implemented.*
- Automations, FR - *Not implemented.*

11.3 Attachment 3 - the 'Kiki' and 'Bouba' framework

According to Gómez Milán et al. (2013), Bouba-like phonemes are associated with adjectives such as round, soft, calm, and pleasant.

Phonetic Bouba set

- Voiced plosives: /b/, /d/, /ɡ/
- Nasals: /m/, /n/
- Approximants: /l/, /ɹ/ (r)
- Rounded vowels: /u/, /o/, /a/
- Example pseudowords: *moba, lumo, banu, gola*

In contrast, Kiki-like phonemes are associated with adjectives such as sharp, energetic, tense, and edgy.

Phonetic Kiki set

- Voiceless plosives: /p/, /t/, /k/
- Voiceless fricatives: /s/, /ʃ/ (sh), /f/
- High/front vowels: /i/, /e/
- Example pseudowords: *kiti, shapa, feki, tisu*

Using this schema, the composer is first asked to categorise their composition as either Bouba-oriented or Kiki-oriented. From this categorisation, they can then construct a pseudoword mode name of two to three vocal/continuant clusters, aligned with the corresponding phonetic set.

11.4 Attachment 4 - Charts for Creative Support Index (CSI)

Mental effort

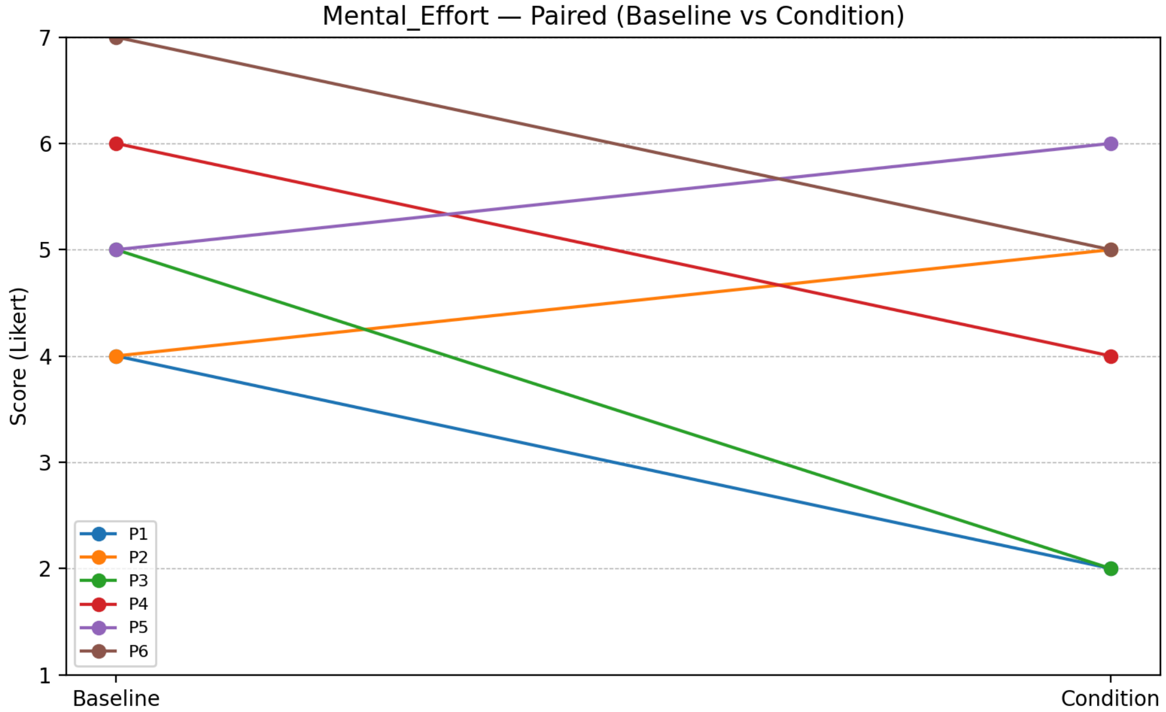


Figure 18. - Paired Mental Effort Scores / Individual Trajectories.

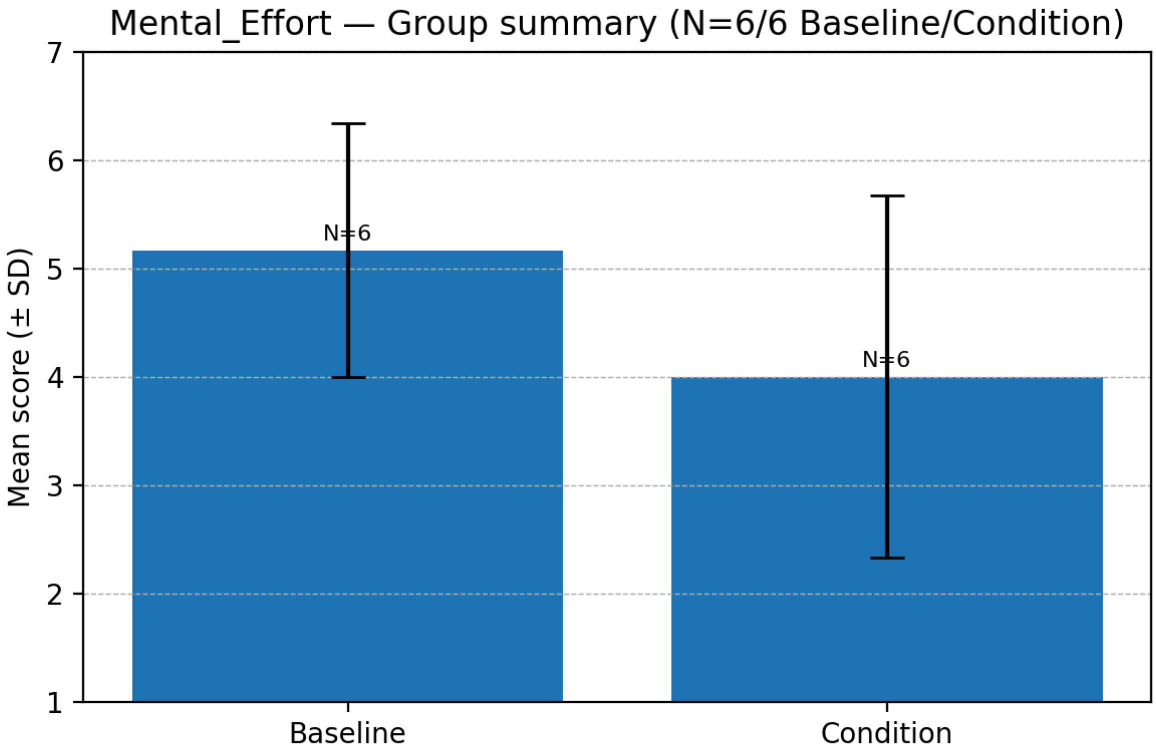


Figure 19. - Mean Mental Effort (Baseline vs. Condition)

Enjoyment - Q1 and Q2 summed

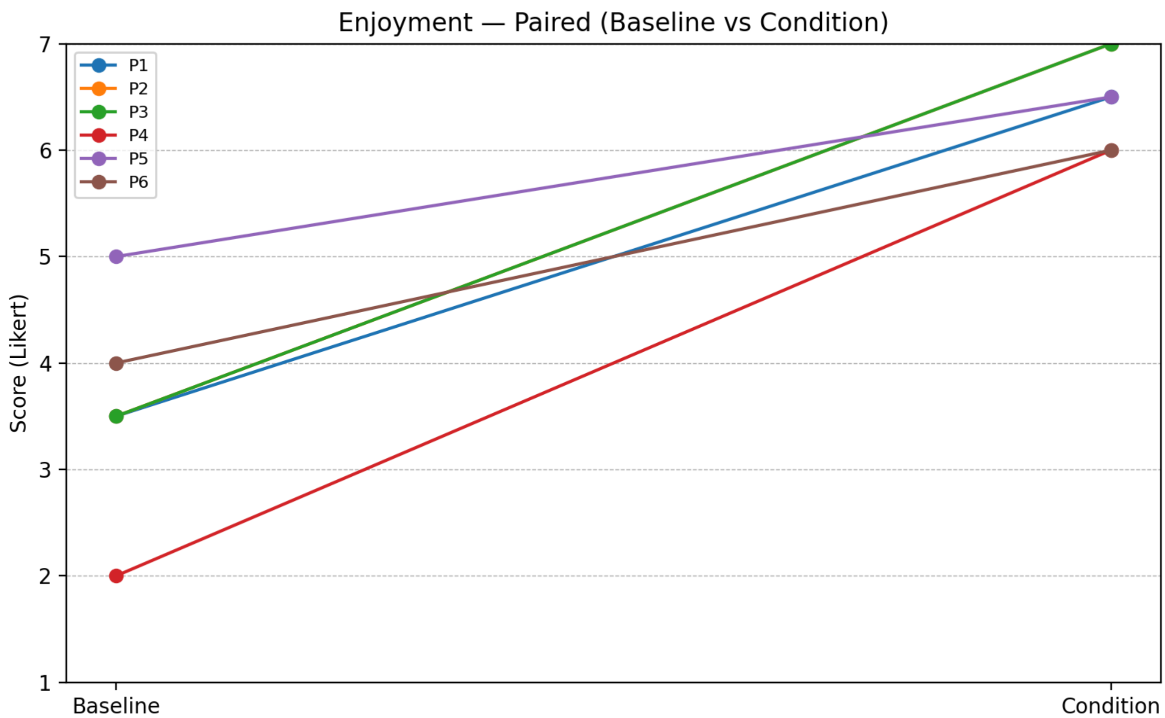


Figure 20. - Paired Enjoyment Scores / Individual Trajectories.

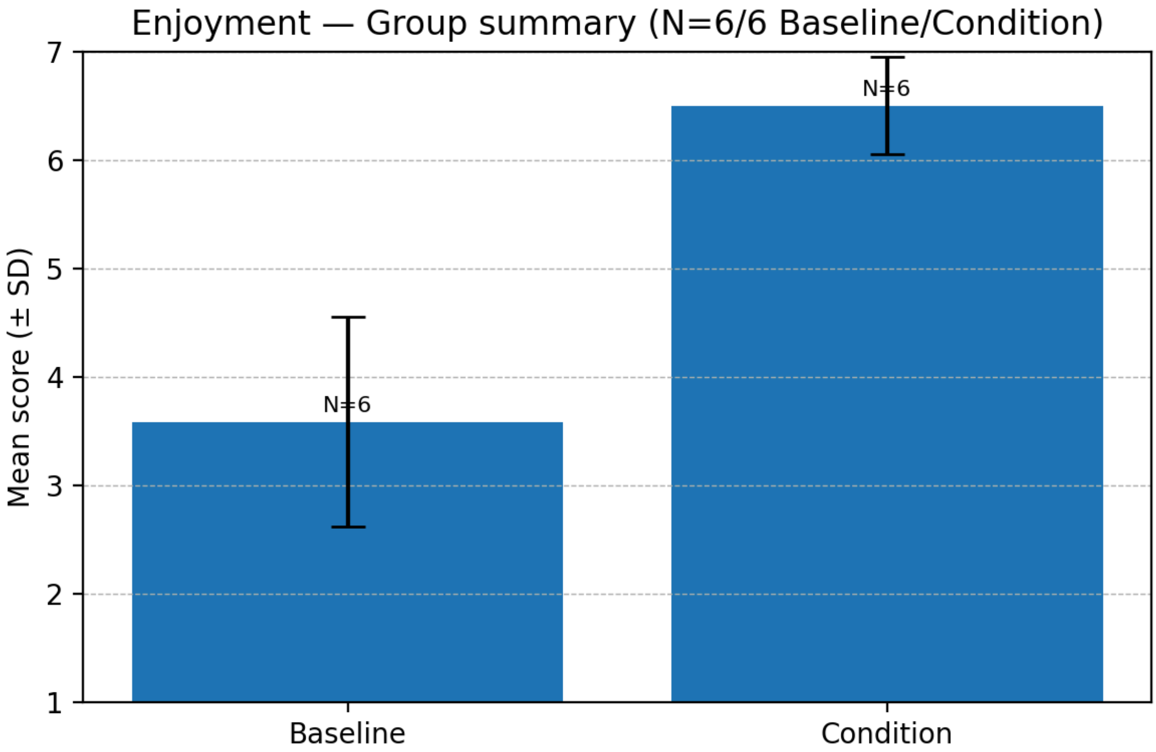


Figure 21. - Mean Enjoyment (Baseline vs. Condition).

Exploration - Q1 and Q2 summed

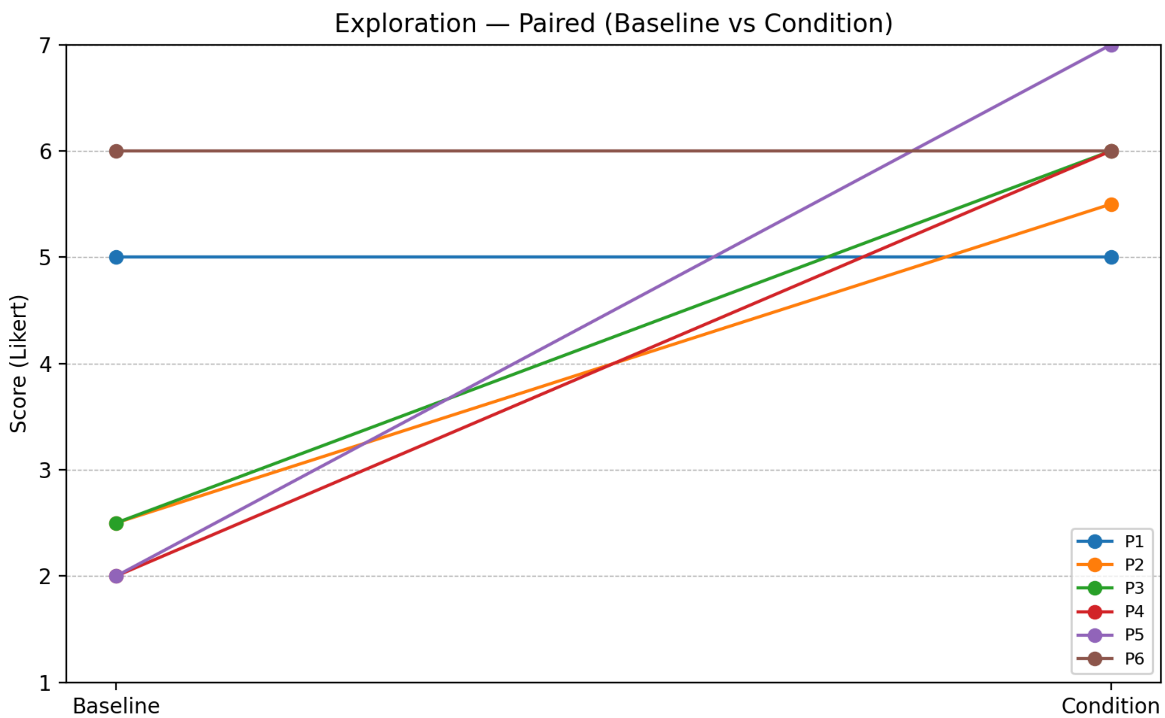


Figure 22. - Paired Exploration Scores / Individual Trajectories.

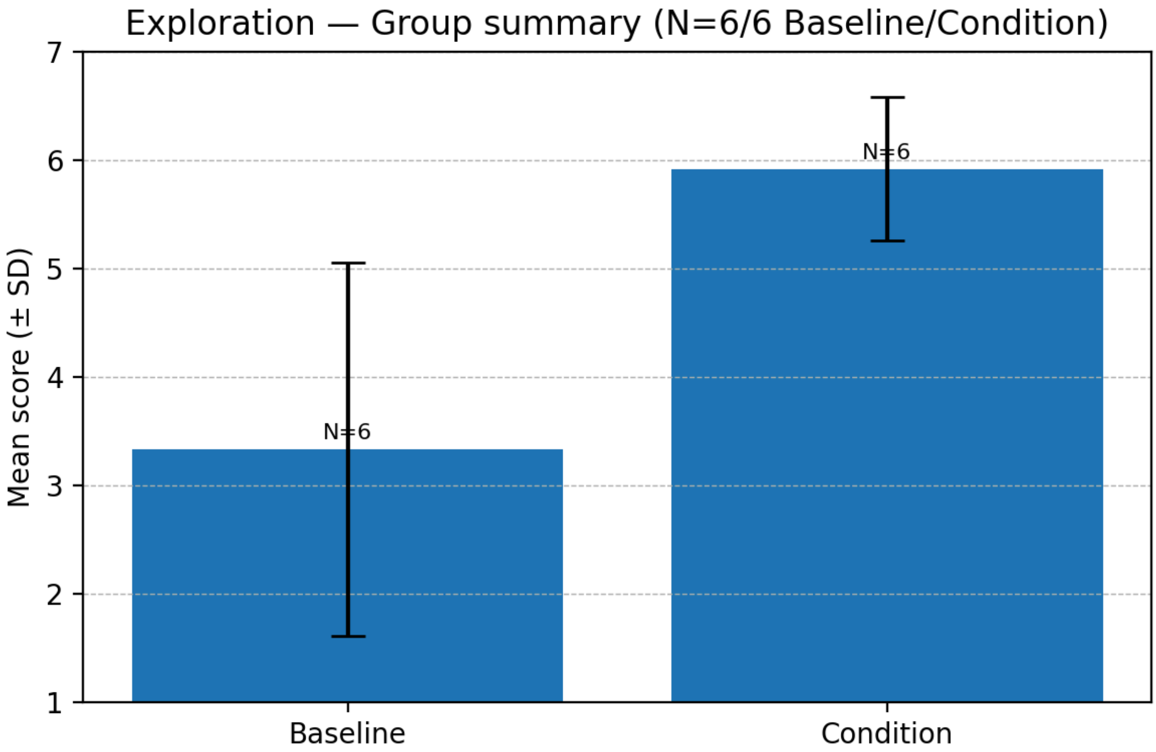


Figure 23. - Mean Exploration (Baseline vs. Condition).

Expressiveness - Q1 and Q2 summed

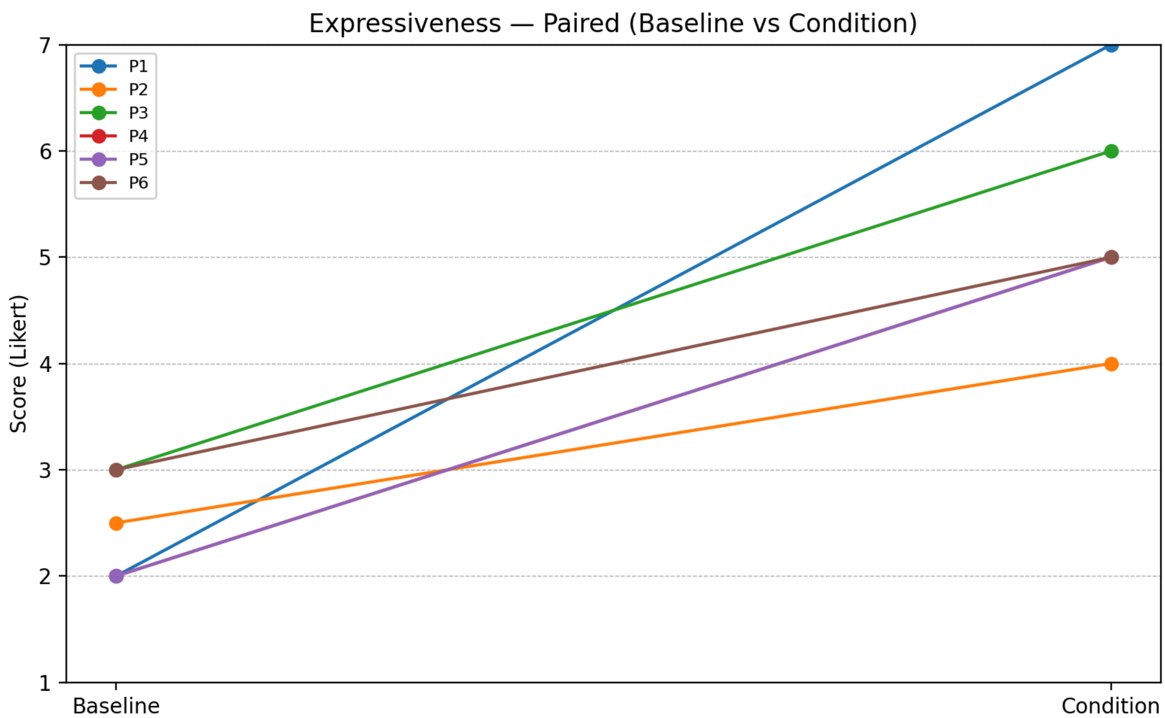


Figure 24. - Paired Expresiveness Scores / Individual Trajectories.

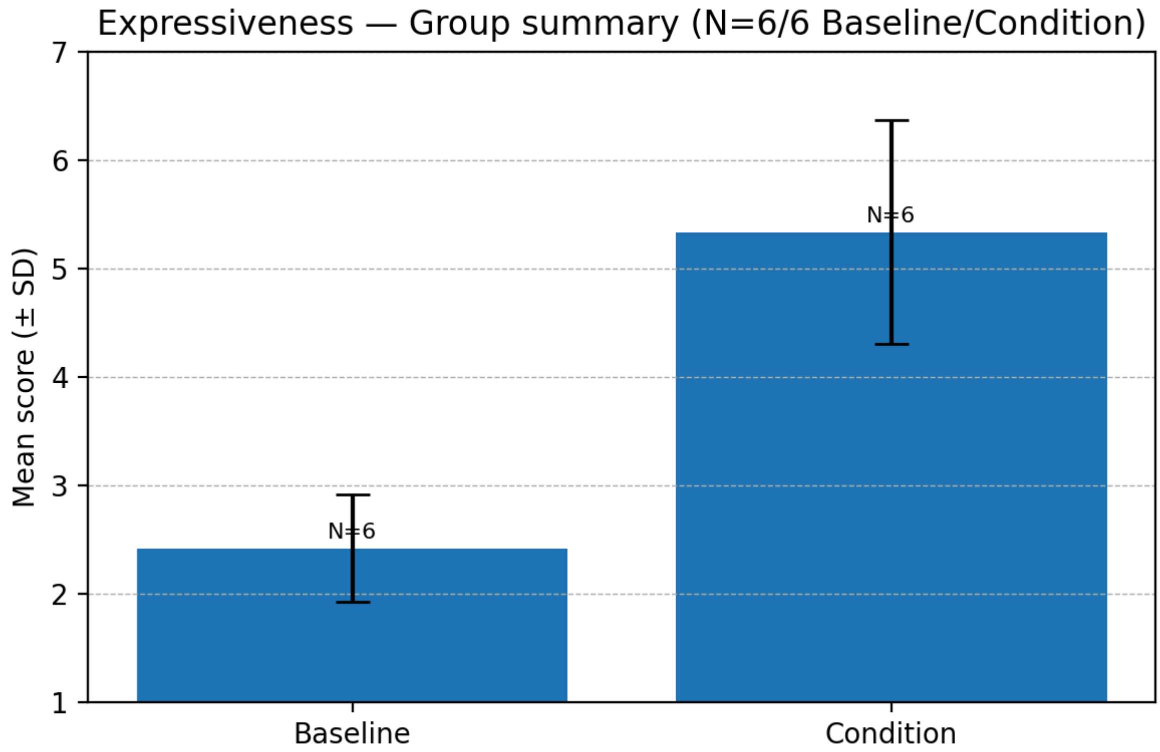


Figure 25. - Mean Expressiveness (Baseline vs. Condition).

Immersion - Q1 and Q2 summed

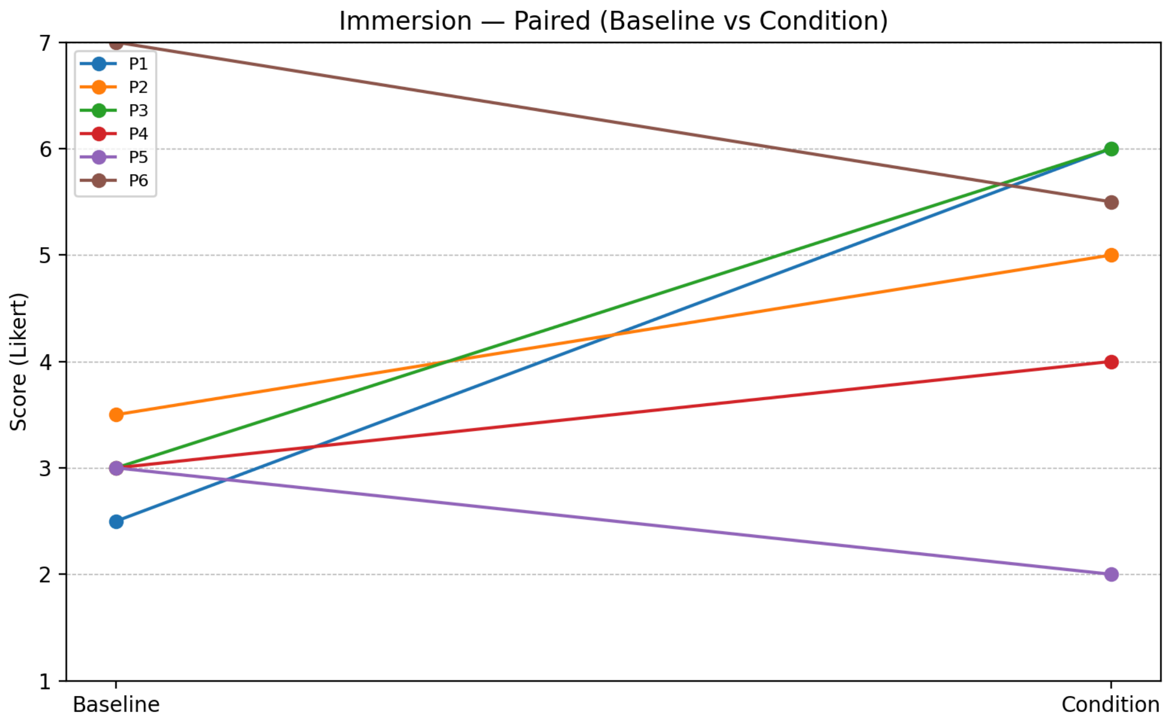


Figure 26. - Paired Immersion Scores / Individual Trajectories.

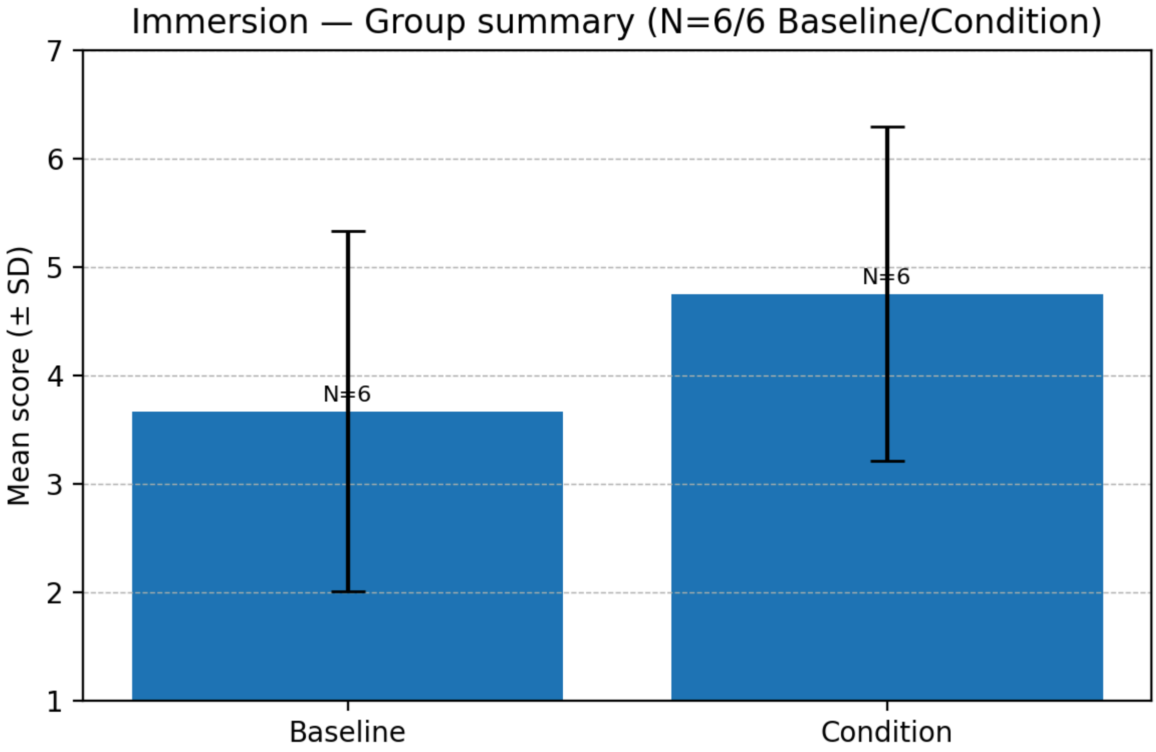


Figure 27. - Mean Immersion (Baseline vs. Condition).

Results Worth Effort - Q1 and Q2 summed

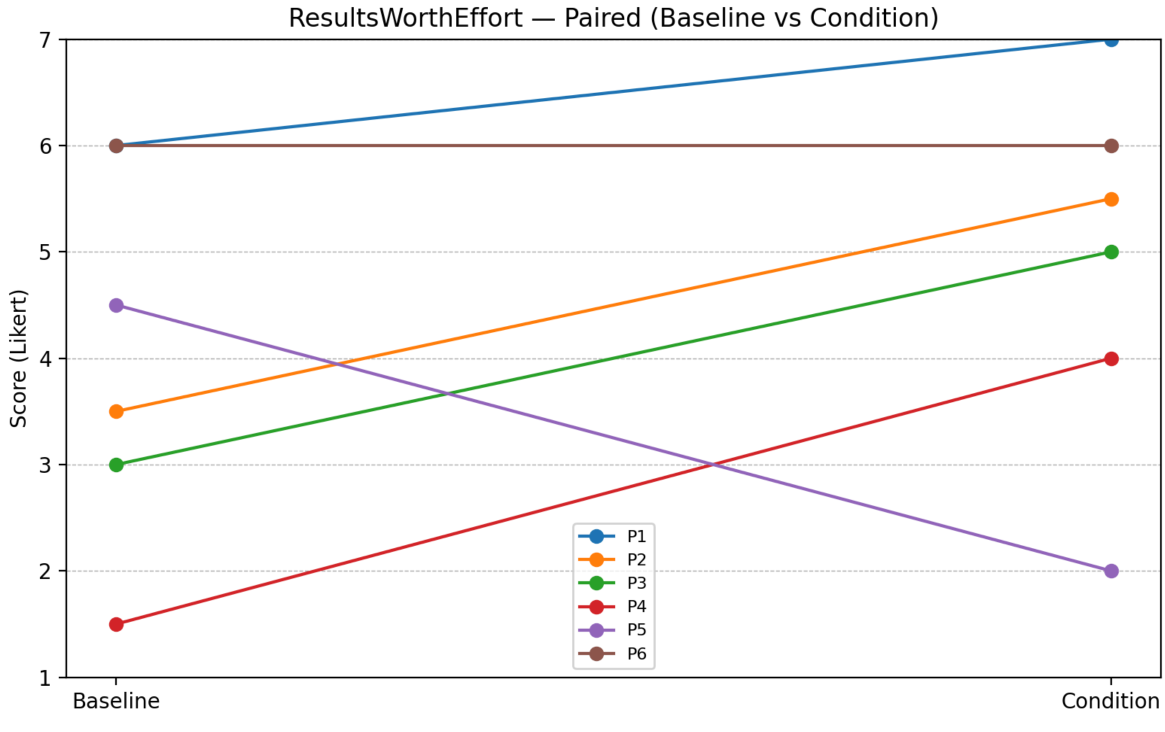


Figure 28. Paired Results Woth Effort Scores / Individual Trajectories.

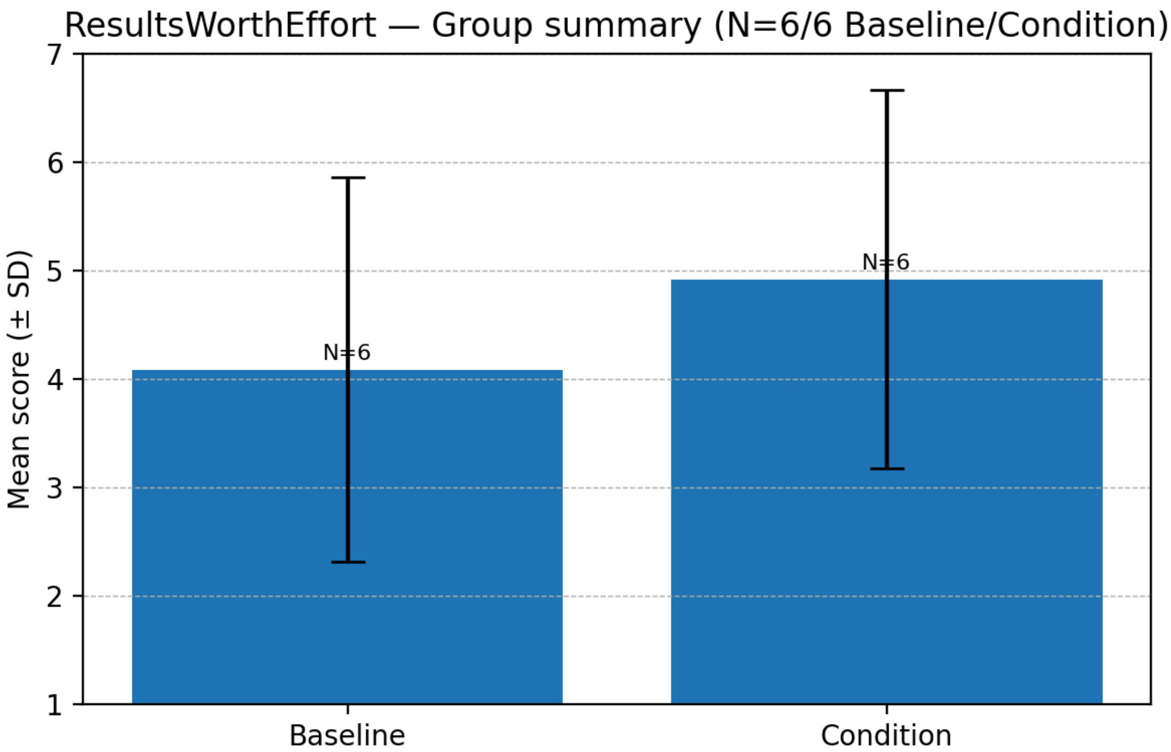


Figure 29. - Mean Results Worth Effort (Baseline vs. Condition).

Collaboration - Q1 and Q2 summed

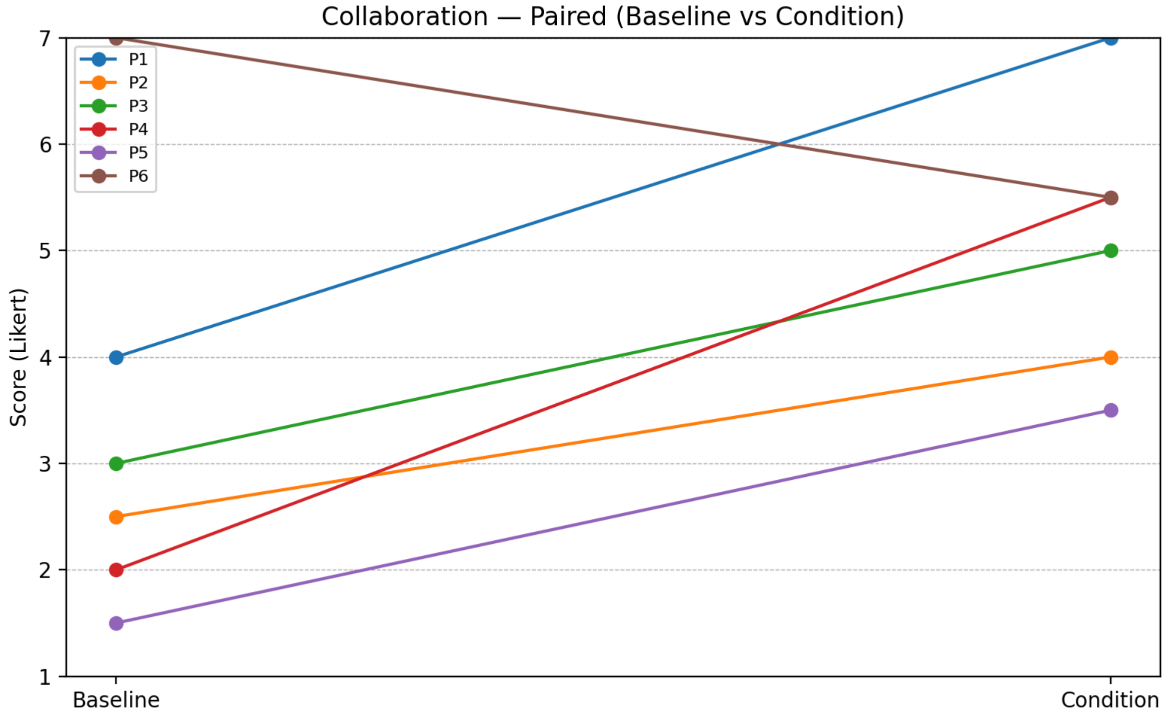


Figure 30. - Paired Results Worth Effort Scores / Individual Trajectories

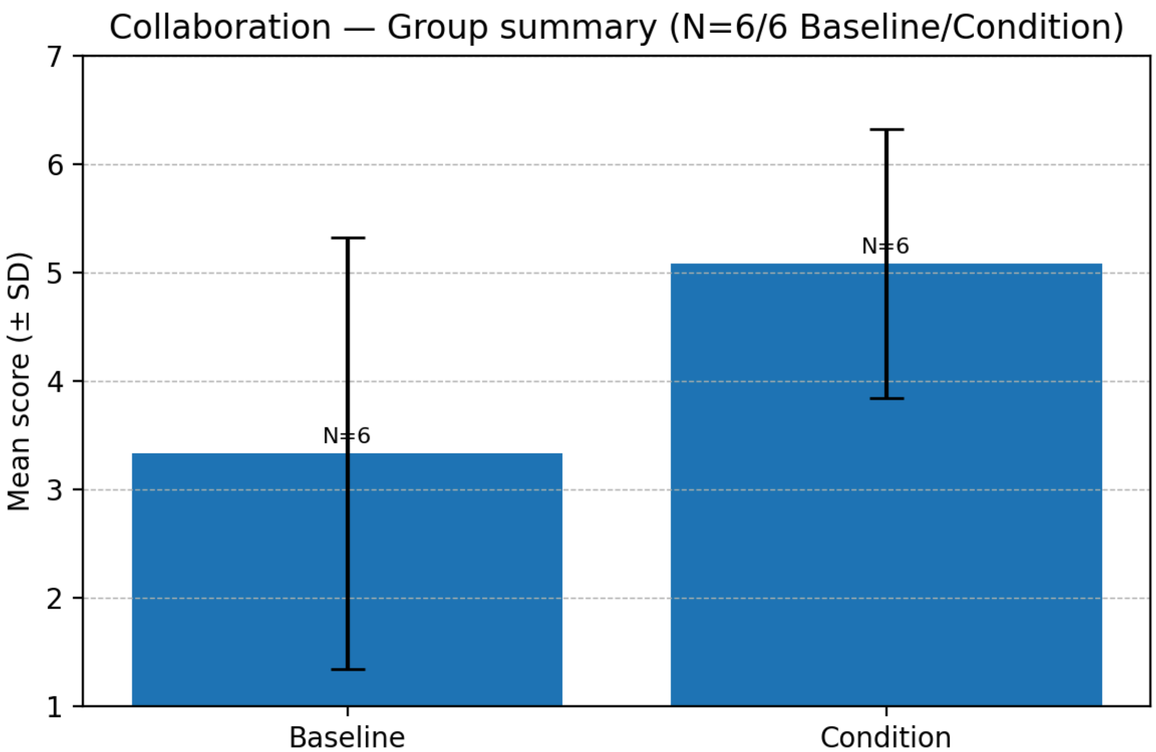


Figure 31. - Mean Results Worth Effort (Baseline vs. Condition).

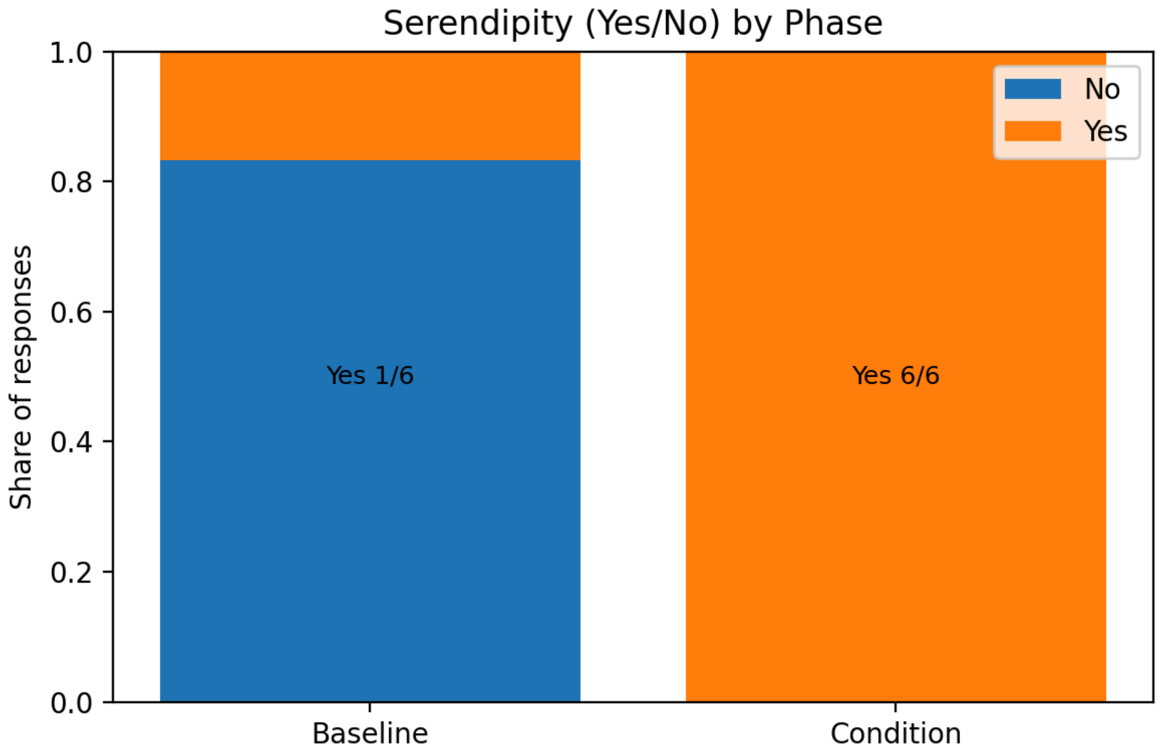


Figure 32. - Serendipity - Yes or No measurement.

11.5 Attachment 5 - Video of Subject 1's Brief 2 Hindenburg Project

Video can be found on frederikbjorn.com/attachments

Name - 'Attachment 5 - Video of Subject 1's Brief 2 Hindenburg Project'
Type - MP4 file
Size - 58 mb
Length - 6:20

11.6 Attachment 6 - Video of Subject 2's Brief 1 Hindenburg Project

Video can be found on frederikbjorn.com/attachments

Name - 'Attachment 6 - Video of Subject 2's Brief 1 Hindenburg Project'
Type - MP4 file
Size - 13.2 mb
Length - 1:24

11.7 Attachment 7 - Heartbeat_Full Music Track

Audio can be found on frederikbjorn.com/attachments

Name - 'Attachment 7 - Heartbeat_Full music track '
Type - WAV file
Size - 3.2 mb
Length - 2:17

11.8 Attachment 8 - Subject 5 - Brief 1 Excerpt

Audio can be found on frederikbjorn.com/attachments

Name - 'Attachment 8 - Subject 5 - Brief 1 excerpt'
Type - MP3 file
Size - 2.7 mb
Length - 1:41

11.9 Attachment 9 - Observer Log of Each of the Subjects

Subject 1

Baseline - Brief 1

- Starts by listening to speak.
- Royalty-free music has different gain levels, which confuses
- Uses multiple instances of the same musical track
- Applies fades to all the musical tracks.
- A lead in one of the musical tracks ruined an otherwise great composition.

MusicBox Condition - Brief 2

- No prior knowledge of dynamic or algorithmic music.
- Understood the theme of the podcast right away. This has been edited in newer versions of Brief 2.
- Uses the musicbox directly under the speaker af only 3 minutes of editing,
- Using one mood quite extensively might be due to the background as a musician and producer.
- Master vol fader used to create live fades
- Does not listen to all themes, mostly uses theme 1.
- Does not understand that themes are unique for each mood 🙁

- Uses multiple instances of the musicbox on top of one another,
  Only engages the musical segment B button after 25 minutes.
- Thinks the fades are too slow, even at the fade bar set to 2.

Subject 2

Baseline - Brief 2

- Starts by listening and taking very extensive notes. Sonically oriented.
- Ran a bit out of time, too focused on the ambience and SFX sounds ...
- Drags and drops most music untouched
- Thinks it is easy to know the music based on the name
- Tries to use the heartbeat_full music track, but is thrown off by the ending.
- However annoyed to learn that each music track is not a 'simple loop', but a more fully arranged piece which changes during the arrangement.

Condition 1 - Brief 1

- No knowledge of dynamic or algorithmic music, except listening to Brian Eno...
- Starts by listening and taking very extensive notes. Sonically oriented.
- Opens the MusicBox after 15 minutes.
- Tried to directly fade between moods, sets in the fade bar at 1.
- Thinks that the moods change from happy at the top to more mysterious at the bottom, with some more general category in the middle (e.g. Baupa).
- Surprised that the music suddenly stops, no wash automatically, just a complete cut.
- Spends more time on the music compared to the baseline - this observation might be biased, as I was sitting in the room.
- Uses a technique where the music boxes start at -30 for all modes, and then fade in the modes gradually, jumping between them.
- Serendipitous moment around 4:10 - 5:10 in the Brief 2 session. Uses the playback speed with only Melin Chords playing, creating a drunken-disoriented feeling (sort of acting like a manual LFO interaction with the knob).
- Thinks that the moods should have some sort of preview, or maybe the possibility to jump between them at 0 fade time, to more easily get an overview of what they contain.
- Did not really interact with the themes that much.

Subject 3

Baseline - Brief 1

- Starts by listening to speak
- Set makers for scenes like "st. Johns er en lille by."

- Listens to musical pieces, but only the first 10-30 seconds.
- Starts the episode-specific speech with a cinematic woosh and music.
- Get surprised at the lead synth that appears in the 'Heatbeat_full' music piece. This pulls the subject out of high-level editing into low-level task mode, as this now needs "fixing".
- Cuts both speak and music to match the emphasis points.
- Creates a new instance of music and tries to match the timing of it to the prior instancealow-levell task, tasks too long.
- Sections of the music tracks which actually fit are not long enough.
- Uses a woosh from soundly to create a wash-out effect.
- Creates a music RVB track to feed small sections of music into, helping with the wash-out effect of the woosh.

Condition 1 - Brief 2

- Does not know what is meant by dynamic music, but has played quite a few video games ...
- Listen to the music box during the interview/speak
- Misses a solo and mute button for each of the modes.
- Explores the music box a bit in solo, might take longer to understand than the subject first anticipated.
- Has a hard time ending the musical section, due to the different fundamental way of working with the music box as compared to the subject's "normal" workflow.
- Uses a theme as a jingle, tries to match it to a pre-recorded section from the musicbox.
- Stuck with having a single instance of segment A, and now wish to create a segment B piece afterwards. The subject was not thinking of re-recording and doing it "live".
- Creates a soundscape using only the pad mode from Melin Mood.
- The subject thinks the names are irrelevant.
- The subject thinks it is hard to replicate objectives after the exploratory phase.
- The subject suggests a feature to export musical loops out of the musicbox.

Subject 4

Baseline -Brief 2

- Not a Windows user
- Starts by listening to the speaker
- Uses multiple ambient recordings, layered.
- Listens quite thoroughly to all the different music tracks and the SFXs in the source pack.

- Matches speak and ambience to the music.
- Duck the music for ambience.
- Normally uses self-recorded foley/ambience.
- Does not use background music, due to the inability to play music and dissatisfaction with royalty-free music libraries.
- Only uses music if they get it from a trusted source, i.e. a friend or colleague who has made it.
- Can identify the same type of woosh used in other productions, hinting at the overuse of some sounds in the industry

Condition 1 - Brief1

- Starts by just trying and exploring the music box
- Comments on the inability to recreate a 'deafult'-state for the application.
- Tries to create a 'romantic'-underlay
- Listens with the speaker in the background, but stops this to focus more intensely on the exploration - might have underrated the ease-of-use of the musicbox.
- Seems to be searching for something specific.
- Comments that it would be nice to have the music in a separate view.
- Uses the same theme in different playback speeds to create a (very) unique theme.
- A lot of starting a stopping the musicbox, I can't really seem to understand why ..
- Creates some wooshes with the musicbox, using the fader and pad mode.
- Has many music-bites of audio in the session - maybe a bit overwhelming, but the subject is not cleaning the session in any way.
- Thinks this style of bedding and working with music is fun!
- Thinks that five moods are too few.
- Asks if it can work with existing music.
- Subject thinks that it is a bit much with the themes, but enjoys the drag-and-drop functionality.
- Subject could imagine the themes would be nice for a skiller/bumper.
- Likes that the themes have no names,
- Would like to mark a theme, i.e. to have a list of usable themes when coming out of the exploratory phase.
- The subject could, sort of, remember the moods and how they sounded.
- Subject thinks that the music might be a bit much for bedding a podcast - but also hits at versatility and states that a learning period of the application might assist in creating more tailored background music.
- Can recognise the two phases, convergent and divergent thinking.
- Need to get used to composing in this workflow.

- Thinks it is quite hard to reenact a certain soundscape - i.e. from memory

Subject 5

Baseline - Brief 1

- Starts listening to thespeakerk.
- Listens thoroughly to almost all of the music tracks.
- Settles quite early on the “Heartbeat_full_2” music track, too early as the terrible lead synth comes in, the subject realises that this music track is not very good.
- The music track also ends sort of midway, with inconvenience the flow of the subject.
- These two tthingskeeps dragging the subject out of high-level task flow and into low-level editing mode.
- Cope pastes the music track into a different section of the session (another place with dialogue between two actors), so in a way, working entirely visually - see brief for reference.
- Wants a longer fade for the music track, a wish which cannot be met.
- Shows me his normal workflow - which is to listen to the music and make the loop in real-time, feeling the rhythm, and then using the theHindenburgg shortcut ‘B’ to cut the music track at the exact point in time.
- Notes that not all music can do this, perfect looping.
- Can’t identify these types of tracks in the exploratory phase, i.e. when searching in an epidemic sounds or in another music library. Has the speaker been running in the background, or is this process?

Condition 1 - Brief 2

- No knowledge of dynamic music.
- Starts exploring the moods with the speech running underneath,
- “Learns” the musicbox, tries all the knobs, etc., quite quickly.
- Composes using the faders.
- Gets a feel for what the different compositions do
- Plays the same mood + segment for quite some time, while the session is running,
- Needs a reminder that segment B exists.
- After the exploratory phase, listens to the output and makes edits in Hidenburg.
- Uses the themes to fit “holes” in the interview - working a bit visually again.
- Enjoys the “wash-out” effect.
- Want to plan a narrative sequence of operations for the musicbox.
- Deletes all the recorded audio to try again - composing style.

- Makes more space for the music.
- Starts the recording in medias res.
- “Plays” with the rhythm fader.
- Playing the MusicBox, like an instrument ..
- Tries to recreate the whole thing again...
- Thoroughly composes a small piece, adjusting rhythm, volume, filter knobs and reverb live.
- Thinks that the themes are too long.
- Enjoys working with the musicbox - thinks it is fun.
- Thinks that he can get a lot of different music out of it.
- Good for “finding a path to walk on”.
- Wants a way to time the segment switching a bit better. Maybe also hinting at the same for the moods.
- Want to automate the music box.
- Wants to make a plan, on paper or in the musicbox.
- Suggests that the workflow is similar to having a set of guitar pedals that then have oan n/off.
- Would love to work in a preset structure - with fade times between the presets.

Subject 6

Baseline - Brief 2

- Has made this tool - i.e. Hindenburg PRO
- Is NOT a Windows user
- Listen to different passages in the musicPre-selectss the target music to use.
- Listens and utilises fade functionality.
- Struggling with ending the selected musical section.
- Only uses two musical tracks.
- Wants to attach music to a person or a sequence in the story.
- Kiera has a good starter.
- Listens specifically to the start and end of each track to ensure usability.
- Thinks that too many of the available tracks are “too much”.
- Wants to be able to automate distance and/or top frequency.
- Make the voice profiler opposite the music.
- Stretch/freeze effect would be grand for bedding.

Condition 1 - Brief 1

- Listens to the speaker with the musicbox underneath.
- Selects a single mood after approximately 7 minutes.

- Uses faders, especially rhythm and chords, along with the filters, to “play” with the musicbox.
- Does not record the first iteration.
- Chooses a mood based on prior knowledge of the moods.
- Feels trapped by the knowledge of recording.
- Feels that the convergent thought process needs a tweak.
- The exploration of the moods should be done faster.
- Suggests different stages of a bedding track. How to categorise these?